UTS UNIVERSITY OF TECHNOLOGY SYDNEY

Intelligent Data-Driven Models for Accurate Multi-factors Prediction of Carbon Credit Prices

by Najlaa Saad Alshatri

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

under the supervision of Dr. Safaa Ghannam

University of Technology Sydney
Faculty of Engineering and Information Technology

September 2025

Certificate of Authorship/Originality

I, **Najlaa Saad Alshatri**, declare that this thesis is submitted in fulfilment of the requirements for the award of Doctor of Philosophy in the Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualification at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Signature: _____

Date: _____

Acknowledgements

First and foremost, I extend my deepest gratitude to Allah for blessing me with the strength, perseverance, and opportunities to complete this journey. Without His guidance and grace, none of this would have been possible.

Success is never a solo achievement; it is built upon the foundation of shared efforts and unwavering support from those around us.

My heartfelt gratitude goes to Dr. Safaa Ghannam. Her guidance, patience, and unwavering support have been instrumental throughout this process. She has been more than a supervisor to me; she has been a mentor and a source of inspiration. I also extend my thanks to Professor Farookh Khadeer Hussain. His expertise and invaluable advice have significantly shaped my academic journey.

I would like to express my heartfelt gratitude to my parents. Your constant love, encouragement, and faith in me have been my greatest source of strength. Thank you for teaching me the value of education, for believing in me even when I doubted myself, and for your endless support throughout my journey.

To my husband, Faris, and my beloved children, Wafi, Wreef, and Wesam: your love, patience, and unwavering support have been my foundation. Your encouragement has given me the courage to persevere and pursue my dreams. I am profoundly grateful for your understanding, sacrifices, and the joy you bring to my life every day.

I would like to express my heartfelt gratitude to my brothers and sisters. Thank you for believing in me, cheering me on, and always being there when I needed you. Your kindness and solidarity have made all the difference, and I am deeply grateful to have you by my side.

I am also grateful to the University of Technology Sydney (UTS) for providing

me with an excellent academic environment and resources to pursue my research. UTS has been a supportive and inspiring place for my studies.

I also extend my sincere thanks to the University of Jeddah and the Ministry of Education, Kingdom of Saudi Arabia, for their generous support and for providing me with the opportunity to pursue my academic goals.

Lastly, I extend my thanks to everyone who has played a part in my academic journey. Whether your contribution was large or small, know that your support has been invaluable in reaching this milestone. This achievement is a testament to your belief in me and your encouragement along the way.

ABSTRACT

Intelligent Data-Driven Models for Accurate Multi-factors Prediction of Carbon Credit Prices

by

Najlaa Saad Alshatri

The aim of carbon credit trading is to reduce emissions at the lowest possible cost while meeting environmental goals. Carbon credits are digital assets in the form of immutable tokens. Knowing the accurate price of carbon credits ensures that carbon markets remain transparent, accountable, and effective. However, carbon credit prices exhibit non-linear, non-stationary, and multicollinear characteristics, which present significant challenges for accurate modeling and prediction. These challenges arise from the fact that the price of carbon credits is influenced by a wide range of external factors, including energy prices, environmental indicators, and economic conditions. These external influences can be viewed as distinct channels of information, each contributing unique statistical insights. When these are combined, they form a complex multivariate signal. Therefore, capturing and interpreting the individual components within these multivariate signals is essential to gain deeper insight and achieve accurate predictions. A systematic literature review (SLR) was conducted to identify the research gaps and limitations.

This thesis develops an advanced Carbon Credit Multi-Factor Prediction (CCMFP) model to predict short-term carbon credit prices. The model integrates key factor identification with optimised prediction algorithms. This thesis uses correlation analysis, the MIC, and the proposed Carbon Credit Multi-Factor Identification (CCMFI) model to study the importance of the 22 identified external factors. The proposed CCMFI model integrates the random forest (RF) regression algorithm

with SHAP (SHapley Additive exPlanations) values from explainable AI. It also supports data-driven feature selection by identifying the most influential input factors. The selected key factors were then used as input to the CCMFP models. The CCMFP applies feature reduction and extraction techniques, including independent component analysis (ICA), nonlinear independent component analysis (NLICA), and principal component analysis (PCA). The extracted components then serve as input variables for the support vector regression (SVR) and multilayer perceptron (MLP) neural networks.

The daily price data for Australian Carbon Credit Units (ACCUs) are used as a case study, and the data for each influencing factor were collected from reliable and authoritative sources. To verify the effectiveness of the proposed models, extensive experiments were conducted by varying the number of factors selected as input to the prediction models. This approach enabled a systematic analysis of how the number of input features impacts prediction accuracy, ultimately identifying the optimal feature set for the most accurate predictions. All the proposed models demonstrated superior prediction performance, achieving R^2 over 97% across different key factor configurations. The optimal feature set for the highest prediction accuracy includes patents on environmental technologies (PT), carbon dioxide (CO_2) emissions, adoption of renewable energy (RE), coal prices (CP), crude oil prices (CrP), European Union Allowances (EUAs), China Emission Allowances (CEAs), raw material prices (IO) and product prices (NONF). These findings enhance market confidence among carbon credit traders and investors, helping to mitigate the financial risks associated with price fluctuations. Ultimately, this thesis supports global efforts to combat climate change by promoting carbon credits as a sustainable tool for environmental practices.

List of Publications

The following is a list of research papers produced during this PhD study.

Published Papers

C-1. **Alshatri, N.**, Ismail, L., and Hussain, F.K. (2024). *A Systematic Review of the External Influence Factors in Multifactor Analysis and the Prediction of Carbon Credit Prices.* In: Barolli, L. (eds) *Complex, Intelligent and Software Intensive Systems. CISIS 2024.* Lecture Notes on Data Engineering and Communications Technologies, vol 87, pp. 1–13. Springer, Cham. https://doi.org/10.1007/978-3-031-70011-8_1

C-2. **Alshatri, N.**, Saberi, M., and Hussain, F.K. (2024). *A Comprehensive Analysis for Determining Key Influencing Factors on Carbon Credit Prices.* In: *2024 IEEE International Conference on e-Business Engineering (ICEBE)*, Shanghai, China, pp. 152–159. https://doi.org/10.1109/ICEBE62490.2024.00032

C-3. **Alshatri, N.**, Ghannam, S., and Hussain, F.K. (2025). *Multi-factor Predictive Models of Carbon Credit Prices.* In: Barolli, L. (eds) *Advanced Information Networking and Applications. AINA 2025.* Lecture Notes on Data Engineering and Communications Technologies, vol 247, pp. 200–212. Springer, Cham. https://doi.org/10.1007/978-3-031-87769-8_18

Submitted Papers

J-1. **Najlaa Alshatri**, Safa Ghannam, Leila Ismail, and Farookh Khadeer Hussain. "Factor-Driven Carbon Credit Price Time Series Analysis and Prediction Models: A Systematic Literature Review." Submitted to the Journal of IEEE Access.(Under review).

J-2. **Najlaa Alshatri**, Safa Ghannam, and Farookh Khadeer Hussain. "Intelligent Data-Driven Models for the Accurate Multi-factor Prediction of Carbon Credit Prices." Submitted to the Discover Artificial Intelligence journal. (Under review).

Glossary of Terms

CC: Carbon Credit.

CER: Clean Energy Regulator.

UNFCCC: United Nations Framework Convention on Climate Change.

ACCUs: Australian Carbon Credit Units.

EUAs: International carbon market (European Union Allowances)

CEAs: International carbon market (China Emission Allowances)

NG: Australian natural gas

CP: Australian BJ power coal spot prices

CrP: Average global crude oil

EP: Australian electricity spot prices

MA: Australian air temperature

AQI: Australian Air Quality Index

GHG: Australian greenhouse gas emissions

CO2: Australian carbon (CO2) emissions

RE: Australian renewable energy

BC: Business confidence indicator

SP: S&P/ASX 200 Index price

AUD: AUD/USD exchange rate

GDP: Australian per capita GDP growth rate

POP: Australian population growth rate

IO: Australian iron ore

NONF: Australian Basic non-ferrous metal product manufacturing

IA: Australian households internet access percentage

PT: Patents on environmental technologies

IND: Australian industry emissions

REI: Australian Renewable usage in industry

MAE: Mean Absolute Error.

MSE: Mean Squared Error.

RMSE: Root Mean Square Error.

R^2**:** R-squared, the coefficient of determination

SLR: Systematic Literature Review.

EDA: Exploratory Data Analysis.

SHAP: Shapley Additive Explanations.

XAI: Explainable Artificial Intelligence.

MIC: Maximum Information Coefficient.

VIF: Variance Inflation Factor.

CCMFI: Carbon Credit Multi-Factor Identification.

DT: Decision Trees.

RF: Random Forests.

CCMFP: Carbon Credit Multi-Factor Prediction.

PCA: Principal Component Analysis.

ICA: Independent Component Analysis.

NLICA: Non-Linear Independent Component Analysis.

SVR: Support Vector Regression.

MLP: Multilayer Feedforward Perceptron neural networks.

Kernel: A function used in SVR that transforms input data into a higher-dimensional space.

RBF: Radial Basis Function.

C: Regularisation parameter in SVR.

Epsilon: Epsilon in the SVR model defines a margin of tolerance.

Gamma: Kernel coefficient for 'RBF' in SVR.

Hidden Layer Sizes: Number of neurons in the hidden layers of the MLP.

Activation: Activation function for the hidden layers in the MLP.

ReLU: Rectified Linear Unit.

Tanh: Hyperbolic tangent.

Solver: Algorithm used for weight optimisation in the MLP.

Adam: An optimisation algorithm that computes adaptive learning rates for each parameter.

SGD: Stochastic Gradient Descent.

Alpha: L2 penalty parameter (regularisation term) in the MLP.

Table of Contents

Certificate ii

Acknowledgments iii

Abstract v

List of Publications vii

Glossary of Terms ix

List of Figures xviii

List of Tables xxvi

1 Introduction 1

1.1 Introduction 1

1.2 Research objectives and aims 4

1.3 Scope of the thesis 5

1.4 Research contribution 7

1.4.1 Scientific contribution 7

1.4.2 Social contribution 8

1.5 Thesis structure 8

2 Literature Review 13

2.1 Overview 13

2.2 Introduction 14

2.2.1 Background 14

2.2.2 The need for a carbon credit price prediction model 15
2.3 Systematic review protocol 17
2.3.1 Search the literature 18
2.3.2 Inclusion and exclusion criteria 20
2.3.3 Study selection process 21
2.3.4 Data extraction and synthesis 23
2.4 Literature review: Multi-factor analysis methods 25
2.4.1 Statistical methods for multi-factor analysis 25
2.4.2 Time series methods for multi-factor analysis 34
2.5 Literature review: Multi-factor price prediction methods 46
2.5.1 Statistical methods 46
2.5.2 Regression-based methods 49
2.5.3 Neural Network-based methods 51
2.5.4 Hybrid methods 52
2.6 Relevant studies analysis 87
2.7 Conclusion 104

3 Problem Definition and Research Questions 106

3.1 Overview 106
3.2 Key terms and concepts 106
3.3 Key gaps in the literature 110
3.4 Research questions 112
3.5 Research objectives 113
3.6 Conclusion 114

4 Research Methodology and Solution Overview 115

4.1 Overview 115
4.2 Design science research methodology 115
4.2.1 Awareness of the problem 116
4.2.2 Suggestion stage 117
4.2.3 Development stage 118
4.2.4 Evaluation stage 119
4.2.5 Conclusion stage 119
4.3 Overview of the proposed solution 120
4.4 Overview of the proposed solution to achieve objective 1 125
4.5 Overview of the proposed solution to achieve objective 2 127
4.6 Overview of the proposed solution to achieve objective 3 132
4.7 Overview of the proposed solution to achieve objective 4 135
4.8 Conclusion 136

5 Identifying the Factors Influencing Carbon Credit Prices 138

5.1 Overview 138
5.2 The importance of identifying external influencing factors 139
5.3 Approach to identifying external factors 139
5.4 Results and discussion 142
5.4.1 Similar factors: international carbon markets 147
5.4.2 Energy price factors 148
5.4.3 Environmental factors 150
5.4.4 Macroeconomic factors 152
5.4.5 Policy factors 153
5.4.6 Societal factors 155

5.4.7 Economic factors 156
5.4.8 Sustainable industry factors 157
5.4.9 Public awareness factor 159
5.5 Investigating the reasons for conflicting literature on the role of external factors 160
5.6 Conclusion 162

6 Feature Engineering: Data Collection, Data Preparation and Exploratory Data Analysis 164

6.1 Overview 164
6.2 The importance of feature engineering 164
6.3 Dataset for feature engineering 165
6.3.1 Carbon market 165
6.3.2 External factors 166
6.4 Applying data pre-processing techniques 169
6.4.1 Results and discussion 177
6.5 Checking the data characteristics and challenges 182
6.5.1 Non-linearity test 183
6.5.2 Non-stationary test 183
6.5.3 Multicollinearity test 183
6.5.4 Results and discussion 184
6.6 Conclusion 187

7 Key Factor Selection 188

7.1 Overview 188
7.2 Methods for identifying and selecting key factors 189

7.2.1 Correlation matrix analysis 189
7.2.2 Maximal information coefficient (MIC) 190
7.2.3 The proposed CCMFI model 191
7.3 Experiment analysis 197
7.3.1 Dataset used to study feature importance 197
7.3.2 The selected implementation platform 198
7.3.3 Implementation 198
7.4 Results and discussion 204
7.4.1 Results of the correlation analysis 204
7.4.2 Results of MIC 206
7.4.3 Results of the proposed CCMFI model 207
7.4.4 Comparative results 209
7.5 Conclusion 215

8 Carbon Credit Price Prediction Modelling 217

8.1 Overview 217
8.2 An Overview of SVR and MLP 219
8.2.1 An overview of the SVR model 219
8.2.2 An overview of the MLP neural network model 221
8.3 Techniques for extracting independent components of carbon credit price influences 223
8.3.1 An overview of PCA 224
8.3.2 An overview of ICA 224
8.3.3 An overview of NLICA 225
8.4 The proposed CCMFP models for predicting carbon credit prices . . . 225

8.5 Evaluation of the proposed CCMFP prediction models 227
8.5.1 Dataset used for carbon credit price prediction modelling . . . 227
8.5.2 Hyperparameter tuning for SVR and MLP models 229
8.5.3 The experimental setup and implementation 231
8.5.4 Evaluation metrics 239
8.6 Results and discussion 240
8.6.1 Evaluation of the proposed CCMFP using the ICA_SVR, NLICA_SVR and PCA_SVR models 241
8.6.2 Evaluation of the proposed CCMFP using the ICA_MLP, NLICA_MLP and PCA_MLP models 269
8.6.3 Comparative performance analysis of prediction models 296
8.7 Conclusion 307

9 Conclusion and Recommendations for Future Research 308
9.1 Overview 308
9.2 Thesis contributions 309
9.3 Limitations of the study 315
9.4 Future research directions 316

A Appendix A: Tuning of Prediction Models 318

B Appendix B: Code links 503

Bibliography 505

List of Figures

1.1 Projected market size of the Australian voluntary carbon credit market. *Source:* Adapted from Horizon Grand View Research, 2024, [58]. 4

1.2 Thesis structure. 12

2.1 The systematic literature review search and selection process. 22

2.2 Distribution of the selected articles by year on carbon credit price prediction and analysis models. 23

2.3 Categorisation of the existing literature based on their adopted techniques for multi-factor analysis and multi-factor price prediction. 24

2.4 Distribution of the selected articles across multi-factor analysis and multi-factor price prediction methods. 25

2.5 Distribution of the studied carbon markets in the selected papers. . . 96

4.1 Methodology of design science research [56]. 116

4.2 Overview of the proposed framework reflecting the research objectives. 120

4.3 The forecasting process [106]. 121

4.4 General overview of the proposed research solution. 121

4.5 Overview of the proposed research solution with the proposed approaches and methods. 122

4.6 Overview of the approach steps for identifying external factors influencing carbon credit prices. 125

4.7 Overview of data pre-processing steps. 129

4.8 The applied techniques to identify the key influencing factors. 130

4.9 The flowchart process for the proposed CCMFI model. 131

4.10 Overview of the proposed CCMFP model with the applied approach and methods. 133

5.1 Overview of the approach to identify the external factors influencing carbon credit prices. 140

5.2 Number of studies per effect of each influencing factor. 145

5.3 The taxonomy of the identified influencing factors on carbon credit prices. 146

5.4 The percentage of studies on the effect of the international carbon markets influencing factor in the similar factors category. 147

5.5 The percentage of studies on each effect of the influencing factors in the energy prices category: (a) natural gas, (b) coal, (c) crude oil, and (d) electricity. 149

5.6 The percentage of studies on each effect of the influencing factors in the environmental category: (a) air temperature, (b) air quality index (AQI), and (c) CO_2 emissions. 150

5.7 The percentage of studies on each effect of the influencing factors in the macroeconomic category: (a) industrial income, (b) stock market, and (c) exchange rate. 152

5.8 The percentage of studies on each effect of the carbon emission threshold influencing factor in the policy category. 154

5.9 The percentage of studies on each effect of the influencing factors in the societal category: (a) per capita growth rate, and (b) population growth rate. 155

5.10 The percentage of studies on each effect of the influencing factors in the economic category: (a) raw material price, (b) product price, and (c) science and technology innovation index. 156

5.11 The percentage of studies on each effect of the influencing factors in the sustainable industry category: (a) green technology maturity, (b) development degree of low-carbon industries, and (c) clean energy penetration. 158

5.12 The percentage of studies on each effect of the Web search index influencing factor in the public awareness category. 159

6.1 Overview of data pre-processing steps. 169

6.2 Overview of the dataset columns. 177

6.3 Missing value checking on the dataset. 178

6.4 Minimum and maximum closing of the carbon credit prices (a) ACCUs, (b) EUAs, and (c) CEAs. 179

6.5 Line plots for the external factors. 180

6.6 Histogram of ACCUs and all influencing factors. 181

6.7 Box-and-whisker plots of ACCUs and all influencing factors. 182

6.8 Scatter plots of ACCUs and all influencing factors, illustrating potential non-linear relationships. 185

6.9 The result of the ADF test indicates a non-stationary characteristic. . 186

7.1 The applied techniques to analyse and identify the key factors influencing carbon credit prices. 189

7.2 Illustration of the RF framework [73]. 192

7.3 Illustration of SHAP method [78]. 193

7.4 The flowchart process for the proposed CCMFI model. 195

7.5 Time series of pre-processed data of ACCUs price split into training and testing sets. 203

7.6 Spearman correlation heatmaps visualising the relationships between ACCUs and the influencing factors. 205

7.7 Pearson correlation heatmaps visualising the relationships between ACCUs and the influencing factors. 206

7.8 Feature importance using the MIC method. 207

7.9 (a) SHAP plot showing the impact of external factors on carbon credit prices with features sorted by importance. (b) SHAP bar chart quantifying the average impact of external factors using mean absolute SHAP values. 209

7.10 Comparison of feature importance for external factors on ACCUs using Spearman's rank correlation, Pearson correlation, MIC, and the proposed CCMFI model. 210

8.1 One-dimensional linear SVR [9]. 220

8.2 Structure of MLP neural network [55]. 221

8.3 Overview of the proposed CCMFP model with the applied approach and methods. 228

8.4 Code snippet for defining and applying PCA. 234

8.5 Code snippet for defining and applying ICA. 235

8.6 Code snippet for defining and applying NLICA. 237

8.7 PCA_SVR model hyperparameters tuning for different components using a different number of external factor configurations. 243

8.8 Best extracted PC components by the proposed PCA_SVR model using the top $K = 5$ factors. 244

8.9 Best extracted PC components by the proposed PCA_SVR model using the top $K = 10$ factors. 244

8.10 Best extracted PC components by the proposed PCA_SVR model using the top $K = 15$ factors. 245

8.11 Best extracted PC components by the proposed PCA_SVR model using all $K = 22$ factors. 246

8.12 Actual vs. predicted ACCUs price using the proposed PCA_SVR model with different external factor configurations. 247

8.13 ICA_SVR model hyperparameters tuning for different components using a different number of external factor configurations. 253

8.14 Best extracted IC components by the proposed ICA_SVR model using the top $K = 5$ factors. 253

8.15 Best extracted IC components by the proposed ICA_SVR model using the top $K = 10$ factors. 254

8.16 Best extracted IC components by the proposed ICA_SVR model using the top $K = 15$ factors. 255

8.17 Best extracted IC components by the proposed ICA_SVR model using all $K = 22$ factors. 256

8.18 Actual vs. predicted ACCUs price using the proposed ICA_SVR model with different external factor configurations. 257

8.19 NLICA_SVR model hyperparameter tuning for different components using a different number of external factor configurations. 263

8.20 Best extracted PC components by the proposed NLICA_SVR model using the top $K = 5$ factors. 264

8.21 Best extracted IC components by the proposed NLICA_SVR model using the top $K = 10$ factors. 265

8.22 Best extracted IC components by the proposed NLICA_SVR model using the top $K = 15$ factors. 266

8.23 Best extracted IC components by the proposed NLICA_SVR model using all $K = 22$ factors. 267

8.24 Actual vs. predicted ACCUs price using the proposed NLICA_SVR model with different external factor configurations. 268

8.25 PCA_MLP model hyperparameter tuning for different components using a different number of external factor configurations. 271

8.26 Best extracted PC components by the proposed PCA_MLP model using the top $K = 5$ factors. 271

8.27 Best extracted PC components by the proposed PCA_MLP model using the top $K = 10$ factors. 272

8.28 Best extracted PC components by the proposed PCA_MLP model using the top $K = 15$ factors. 273

8.29 Best extracted PC components by the proposed PCA_MLP model using all $K = 22$ factors. 274

8.30 Actual vs. predicted ACCUs price using the proposed PCA_MLP model with different external factor configurations. 275

8.31 ICA_MLP model hyperparameter tuning for different components using a different number of external factor configurations. 281

8.32 Best extracted IC components by the proposed ICA_MLP model using the top $K = 5$ factors. 281

8.33 Best extracted IC components by the proposed ICA_MLP model using the top $K = 10$ factors. 282

8.34 Best extracted IC components by the proposed ICA_MLP model using the top $K = 15$ factors. 283

8.35 Best extracted IC components by the proposed ICA_MLP model using all $K = 22$ factors. 284

8.36 Actual vs. predicted ACCU price using the proposed ICA_MLP model with different external factor configurations. 285

8.37 NLICA_MLP model hyperparameter tuning for different components using a different number of external factor configurations. 291

8.38 Best extracted IC components by the proposed NLICA_MLP model using the top $K = 5$ factors. 291

8.39 Best extracted IC components by the proposed NLICA_MLP model using the top $K = 10$ factors. 292

8.40 Best extracted IC components by the proposed NLICA_MLP model using the top $K = 15$ factors. 293

8.41 Best extracted IC components by the proposed NLICA_MLP model using all $K = 22$ factors. 294

8.42 Actual vs. predicted ACCU price using the proposed NLICA_MLP model with different external factor configurations. 295

8.43 Comparison of the proposed prediction models based on (a) MAE, (b) MSE, (c) RMSE, (d) R^2, (e) execution time and (f) the best-generated number of ICs, using different configurations of external factors. 299

8.44 Box-and-whisker plots comparing model performance based on (a) MAE, (b) MSE, (c) RMSE, (d) R^2, (e) execution time and (f) the best-generated number of ICs. 300

8.45 Performance evaluation of PCA-based models across different configurations of external factors. 302

8.46 Performance evaluation of ICA-based models across different configurations of external factors. 303

8.47 Performance evaluation of NLICA-based models across different configurations of external factors. 304

List of Tables

2.1 Search categories and keywords for carbon credit studies. 19

2.2 Comparison of relevant articles on statistical methods for multi-factor analysis. 61

2.3 Comparison of relevant articles on time series methods for multi-factor analysis. 67

2.4 Comparison of relevant articles on statistical methods to predict the price of carbon credit. 74

2.5 Comparison of relevant articles on regression-based methods to predict the price of carbon credit. 75

2.6 Comparison of relevant articles on neural network-based methods to predict the price of carbon credit. 76

2.7 Comparison of relevant articles on hybrid methods to predict the price of carbon credit. 77

2.8 Analysis criteria questions. 87

2.9 Evaluation of the selected articles according to the analysis criteria to identify the research gaps. 91

2.10 Feature selection approaches used in the selected studies and their limitations. 99

3.1 Research sub-questions and their motivations 112

5.1 Influencing factors and their effect on the price of carbon credits based on the shortlisted articles of the SLR. 144

6.1 Description and source of the external factors. 174

6.2 Descriptive statistics of the dataset. 178

6.3 The calculated VIF values for independent variables indicate the presence of multicollinearity. 187

7.1 The results comparison of the Spearman's rank correlation, Pearson correlation, MIC, and the proposed CCMFI model. 211

7.2 The top 10 influencing factors identified using Spearman's rank correlation, Pearson correlation, MIC, and the proposed CCMFI model. 212

8.1 Selected influencing factors on carbon credit prices, ordered by their importance from most to least significant. 229

8.2 Hyperparameters of SVR 232

8.3 Hyperparameters of MLP 232

8.4 Performance measures and their definitions. 240

8.5 Parameter grid search results for PCA_SVR model using a different number of external factor configurations. 242

8.6 ICA_SVR model hyperparameter tuning for different ICs using the top K = 5 external factors. 249

8.7 ICA_SVR model hyperparameter tuning for different ICs using the top K = 10 external factors. 250

8.8 ICA_SVR model hyperparameter tuning for different ICs using the top K = 15 external factors. 251

8.9 ICA_SVR model hyperparameter tuning for different ICs using all K = 22 external factors. 252

8.10 NLICA_SVR model hyperparameter tuning for different ICs using the top K = 5 external factors. 259

8.11 NLICA_SVR model hyperparameter tuning for different ICs using the top K = 10 external factors. 260

8.12 NLICA_SVR model hyperparameter tuning for different ICs using the top K = 15 external factors. 261

8.13 NLICA_SVR model hyperparameter tuning for different ICs using all K = 22 external factors. 262

8.14 Hyperparameter tuning results for PCA_SVR, ICA_SVR, and NLICA_SVR models, showing mean MSE and optimal ICs based on a different number of external factors. 269

8.15 Parameters grid search result for PCA_MLP model using different external factor configurations. 270

8.16 ICA_MLP model hyperparameter tuning for different ICs using the top K = 5 external factors. 277

8.17 ICA_MLP model hyperparameter tuning for different ICs using the top K = 10 external factors. 278

8.18 ICA_MLP model hyperparameter tuning for different ICs using the top K = 15 external factors. 279

8.19 ICA_MLP model hyperparameter tuning for different ICs using all K = 22 external factors. 280

8.20 NLICA_MLP model hyperparameter tuning for different ICs using the top K = 5 external factors. 287

8.21 NLICA_MLP model hyperparameter tuning for different ICs using the top K = 10 external factors. 288

8.22 NLICA_MLP model hyperparameter tuning for different ICs using the top K = 15 external factors. 289

8.23 NLICA_MLP model hyperparameter tuning for different ICs using all K = 22 external factors. 290

8.24 Hyperparameter tuning results for PCA_MLP, ICA_MLP, and NLICA_MLP models, showing mean MSE and optimal ICs based on a different number of external factors. 296

8.25 Comparison of the prediction models for different configurations of external factors (K). 298

A.1 ICA_SVR model hyperparameters tuning for different ICs using the top K = 5 external factors. 318

A.2 ICA_SVR model hyperparameters tuning for different ICs using the top K = 10 external factors. 322

A.3 ICA_SVR model hyperparameters tuning for different ICs using the top K = 15 external factors. 329

A.4 ICA_SVR model hyperparameters tuning for different ICs using all K = 22 external factors. 341

A.5 NLICA_SVR model hyperparameters tuning for different ICs using the top K = 5 external factors. 357

A.6 NLICA_SVR model hyperparameters tuning for different ICs using the top K = 10 external factors. 361

A.7 NLICA_SVR model hyperparameters tuning for different ICs using the top K = 15 external factors. 369

A.8 NLICA_SVR model hyperparameters tuning for different ICs using all K = 22 external factors. 380

A.9 ICA_MLP model hyperparameters tuning for different ICs using the top K = 5 external factors. 397

A.10 ICA_MLP model hyperparameters tuning for different ICs using the top K = 10 external factors. 402

A.11 ICA_MLP model hyperparameters tuning for different ICs using the top K = 15 external factors. 412

A.12 ICA_MLP model hyperparameters tuning for different ICs using all K = 22 external factors. 428

A.13 NLICA_MLP model hyperparameters tuning for different ICs using the top K = 5 external factors. 450

A.14 NLICA_MLP model hyperparameters tuning for different ICs using the top K = 10 external factors. 455

A.15 NLICA_MLP model hyperparameters tuning for different ICs using the top K = 15 external factors. 465

A.16 NLICA_MLP model hyperparameters tuning for different ICs using all K = 22 external factors. 480

Chapter 1

Introduction

1.1 Introduction

The aim of carbon credit trading is to reduce emissions at the lowest possible cost while meeting environmental goals [128]. A carbon credit is a digital asset in the form of an immutable token representing a reduction or removal of one tonne of greenhouse gas emissions from the atmosphere [118]. However, the dynamic nature of carbon markets, driven by multiple external factors, presents unique challenges to accurate prediction. The time series of carbon credit prices has non-stationary, non-linearity, and multicollinearity characteristics. These challenges are the result of the influence of external factors on the carbon market [94], [174], such as the price of the energy market [8], [175], environmental conditions [130], microeconomic variables [23] and financial markets [175]. Recognising and addressing these characteristics is imperative for robust data analysis and the prediction of carbon credit prices. Accurate prediction of carbon credit prices ensures that the carbon market remains transparent, accountable and effective, leading to more informed trading decisions, greater profitability, and reduced financial risk [133].

A systematic literature review (SLR) [77] was conducted to explore the advances and challenges in predicting the price of carbon credits. The primary focus of this review is to identify and evaluate relevant studies that propose prediction models that consider factors that influence the price of carbon credits. The results show that in terms of prediction accuracy, hybrid methods, which combine the strengths of different modelling techniques, are the most commonly used method to predict

the price of carbon credits because of their ability to learn complex relationships. However, there is a lack of prediction models for the price of carbon credits that incorporate a comprehensive list of influencing factors. Different studies have considered some influencing factors when predicting the price of carbon credits, but they tend to overlook or underestimate the impact of other factors. Not considering all key factors can lead to less accurate price predictions, negatively impacting the efficiency of carbon credit markets and investor confidence.

Furthermore, an examination of the current literature shows that there is a lack of studies that focus on developing a smart and effective method for selecting the key factors that influence the price of carbon credits. The feature selection approach is critical to ensure that the prediction models avoid overfitting and remain efficient by focusing only on the most relevant factors [14]. Although several factors influence carbon credit prices, not all carry the same weight. Identifying the key factors can significantly enhance the accuracy of price predictions [66], [167]. Current methods for selecting these factors are often subjective and rely on expert judgment. Therefore, there is a significant research gap in developing a smart and data-driven method to select the most important factors that affect carbon credit prices, which can help improve the accuracy and reliability of carbon credit price predictions.

Furthermore, this study addresses a significant gap in the literature by developing the first comprehensive taxonomy of external factors that influence carbon credit prices. To our knowledge, no research has systematically classified and organised these factors. Through an extensive review of the multidisciplinary literature, we identified and categorised these factors into nine groups: similar carbon markets, price of the energy markets, environmental, microeconomic, policy, social, economic, sustainable industry, and public awareness factors. These factors are used as input for the predictive models proposed in this research. It was essential to base the selection of these factors on a solid foundation from the literature rather than relying

on subjective judgment, which could result in critical variables being overlooked. By providing a structured and evidence-based framework, this taxonomy strengthens the validity of the proposed models, improves the understanding of the key drivers of carbon credit prices, and supports future research and practical decision making in the field.

Furthermore, to our knowledge, no research in the existing literature has examined how these factors contribute to prediction accuracy. Many existing methods function as 'black boxes,' offering limited insight into how predictions are made. The proposed model addresses this gap by integrating explainable artificial intelligence (XAI) to analyse the type and magnitude of the impact that external factors have on carbon credit prices. XAI improves the interpretability and transparency of the prediction model [157], [95], and empowers stakeholders with valuable information on how specific factors shape the market dynamics. This approach bridges the gap between predictive accuracy and interpretability, fostering trust and practical utility among policymakers and market participants.

This research focuses on addressing these challenges by developing advanced hybrid models tailored for short-term carbon credit price prediction, using Australian Carbon Credit Units (ACCUs) as a case study. By systematically selecting and analysing a wide range of external factors and employing advanced machine learning techniques, this research seeks to develop robust and accurate models for the short-term prediction of the price of carbon credits.

The focus on ACCUs provides a valuable case study that not only advances the understanding of the Australian carbon market but also offers insights applicable to other markets around the world. The Australian carbon market is a significant component of the global carbon trading landscape [25]. The Australian voluntary carbon credit market has experienced significant growth, with its market size pro-

jected to increase from $95.5 million in 2023 to $796.9 million in 2030, at a compound annual growth rate (CAGR) of 35.4% [58] as shown in Figure 1.1. This highlights the economic importance of the market and the need for focused research on the unique factors that influence ACCU prices. To our knowledge, there is no research in the existing literature which focuses specifically on the Australian market and the factors that influence ACCU prices. The existing literature has concentrated largely on global carbon markets, such as those in Europe and China, leaving a critical gap in understanding the drivers of carbon credit prices in the Australian context.

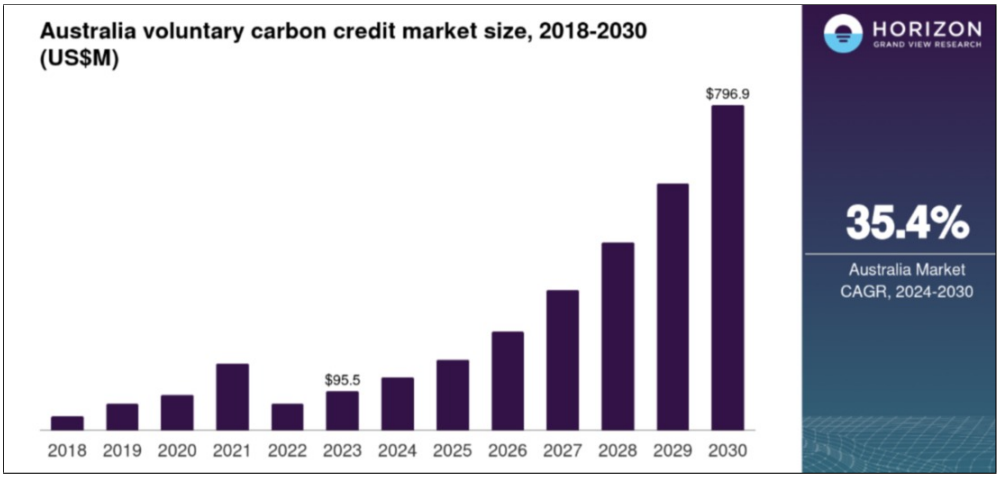


Figure 1.1 : Projected market size of the Australian voluntary carbon credit market. *Source:* Adapted from Horizon Grand View Research, 2024, [58].

This chapter is organised as follows. Section 1.2 identifies the research objectives. The scope of this study is defined in Section 1.3. Section 1.4 highlights the significance of this thesis. An overview of the thesis structure is provided in 1.5.

1.2 Research objectives and aims

The primary goal of this research is to develop advanced models for short-term carbon credit price prediction to support effective market planning and decision-making for stakeholders. In addition, it aims to intelligently identify the key factors

influencing carbon credit prices and determine the most suitable predictive model based on the characteristics of the data. The specific objectives are as follows:

- **Objective 1**: Identify and analyse a comprehensive list of external factors that affect carbon credit prices.
- **Objective 2**: Develop a model that leverages feature importance and feature engineering techniques to accurately and reliably analyse and select the most influential factors affecting carbon credit prices.
- **Objective 3**: Develop a robust and intelligent multi-factor approach to predict short-term carbon credit prices, addressing non-linearity, non-stationarity, and multicollinearity challenges.
- **Objective 4**: Evaluate and validate the methodologies developed to achieve Objectives 2 and 3.

1.3 Scope of the thesis

This research aims to accurately predict the short-term price of carbon credits by considering the key factors that affect the price of carbon credits. Several elements must be considered to determine the scope of this thesis, including the scope of the market, the key factors that influence the price of carbon credits, the prediction time frame, and the data set.

The time frame: Price prediction models could have different short- and long-term time frames. Short-term predictions typically refer to forecasts made for events or trends that are expected to occur in the near future, usually within a few days, weeks, or months [38]. Long-term predictions refer to forecasts made for events or trends that are expected to occur in the distant future, usually several years [120]. The main difference between short-term and long-term predictions is the level

of uncertainty and complexity involved. Short-term predictions tend to be more accurate and reliable as they are based on current data and trends. However, long-term predictions are based on assumptions about how various factors can evolve and interact over time, and their accuracy is less certain due to the potential influence of a wide range of unpredictable factors [97]. In general, short-term predictions are more useful for making immediate decisions, while long-term predictions are more useful for strategic planning for future trends and events. In this thesis, the predictions will be based on the short-term time, that is, the prediction for the next day. Long-term price prediction is beyond the scope of this research.

The carbon market scope: The proposed model is verified using a real-world dataset. ACCUs are used as a case study in this research. The Australian carbon market has become one of the world's largest offset trading schemes [26]. It currently operates as an economy-wide voluntary carbon offset scheme under the Emissions Reduction Fund (ERF) [26]. Each ACCU is generated through the ERF and represents one metric tonne of equivalent carbon dioxide emissions (CO2-e) that have been avoided or removed from the atmosphere through approved projects [25]. Furthermore, the data source for the influencing factors is also based on Australian datasets to reflect the real effect of these factors.

The key influencing factors: In the literature, many external factors influence the price of carbon credits. This research focuses on the 22 factors that influence carbon credit prices in seven directions: (1) the international carbon market, including the European Union Allowances (EUAs) and the China Emission Allowances (CEAs), (2) the energy market (natural gas market, the coal market, crude oil market, electricity market), (3) environmental factors (air temperature, air quality index (AQI), greenhouse gas emissions, carbon dioxide (CO2) emissions and renewable energy), (4) macroeconomic factors (industrial development, stock market and exchange rate), (5) societal factors (per capita growth rate, population

growth rate) and (6) economic factors (raw material price, product price and science and technology innovation) (7) sustainable industry factors (green technology maturity, development degree of low carbon and clean energy penetration). Each variable selected for these categories is based on relevance and availability, as evidenced by the comprehensive list of real data compiled from authoritative sources. Qualitative factors such as policy and public awareness are excluded from this study. By focusing on numerical data, we avoid the complexities of qualitative factors such as policy and public awareness, which often require different methodologies and can introduce variability and subjectivity. Excluding these factors keeps our study focused and methodologically consistent, enabling us to develop robust predictive models for carbon credit prices.

1.4 Research contribution

The research addresses several gaps that have not been fully explored and, as a result, provides a notable contribution to the field. The contribution of this study can be classified into the following scientific and social contributions.

1.4.1 Scientific contribution

1. This research is the first to conduct a systematic and comprehensive literature review on carbon credit price prediction methods, identifying key gaps and limitations in existing approaches.

2. This study is the first to conduct a systematic literature review and develop a comprehensive taxonomy that categorises the identified factors into coherent categories: similar carbon markets, energy market prices, environmental, macroeconomic, policy, societal, economic, sustainable industry, and public awareness factors.

3. This is the first research that develops a comprehensive empirical analysis of

the key factors that impact carbon credit prices.

4. This is the first research to develop an intelligent mechanism to study feature importance to select the key factors influencing the price of carbon credits.

5. This research introduces a novel intelligent mechanism for accurately predicting carbon credit prices by addressing non-linearity, non-stationarity, and multicollinearity, while incorporating a comprehensive set of external influencing factors.

6. This is the first research to provide an intelligence base to predict the carbon credit price and apply it to Australian Carbon Credit Units (ACCUs) as a case study.

1.4.2 Social contribution

1. Developing an intelligent prediction model of the price of carbon credit that considers various external influencing factors could increase the confidence of carbon credit traders and investors to make reliable trading decisions and reduce their financial risk.

2. An accurate short-term prediction model helps policymakers by providing timely information on price changes, enabling them to make informed decisions, adjust regulations, and reduce market uncertainty.

3. This research will contribute to the broader goal of mitigating the effects of global warming by promoting the use of carbon credits as a tool to encourage sustainable practices.

1.5 Thesis structure

This research focuses on the development of intelligent multi-factor prediction models for predicting carbon credit prices. Understanding the factors that influence

carbon credit prices is essential for building reliable models, as the market is highly dynamic, non-linear, and influenced by a multitude of external factors. This thesis details the development of short-term prediction models based on the key factors of external influences, including international carbon markets, macroeconomic conditions, environmental and sustainable industry factors, to name a few. This thesis is organized into nine chapters, summarized in the following section.

Chapter 2: Literature Review

- Provides a comprehensive review of the existing literature related to carbon credit prediction and analysis.
- Systematically reviews relevant articles.

Chapter 3: Problem Definition and Research Questions

- Highlights the existing gaps within the research field based on the literature review.
- Formulates clear research questions and objectives that align with the research questions to address identified gaps.

Chapter 4: Research Methodology and Solution Overview

- Presents definitions of key research terms.
- Presents the research methodology used in the study.
- Provides an overview of the proposed solutions for the identified objectives.

Chapter 5: Identifying the Factors Influencing Carbon Credit Prices

- Presents a list of external factors that influence carbon credit prices based on the literature.

- Categorisation of the selected factors and taxonomy development.
- Justification and relevance of each category to the price of carbon credits.
- Investigates the reasons for the conflicting literature on the role of external factors.

Chapter 6: Feature Engineering: Data Collection, Data Preparation and Exploratory Data Analysis

- Describes the data collection, data pre-processing, and exploratory data analysis.
- Conducts critical statistical tests to examine key data characteristics and challenges: the Augmented Dickey-Fuller (ADF) test to check for non-stationarity, scatter plot analysis to investigate non-linearity, and the variance inflation factor (VIF) test to identify multicollinearity.

Chapter 7: Key Factor Selection

- Describes the approach used to study feature importance.
- Presents a detailed description of the proposed carbon credit multi-factor identification (CCMFI) model.
- Presents the results of the key factors obtained.

Chapter 8: Carbon Credit Price Prediction Modelling

- Presents a detailed description of the proposed carbon credit multi-factor prediction models.
- Presents a brief overview of the adopted machine learning SVR and deep learning MLP models.

- Presents an explanation of the techniques used to extract independent components of the influences.
- Presents a detailed solution for the approaches used for developing the prediction model using SVR and MLP models.
- Presents the key results obtained through the development of the proposed prediction models.
- Analyses and compares all proposed prediction models comprehensively.

Chapter 9: Conclusion and Recommendations for Future Research

- Presents the conclusions and main contributions of the study.
- Discusses the limitations and suggests future research directions.

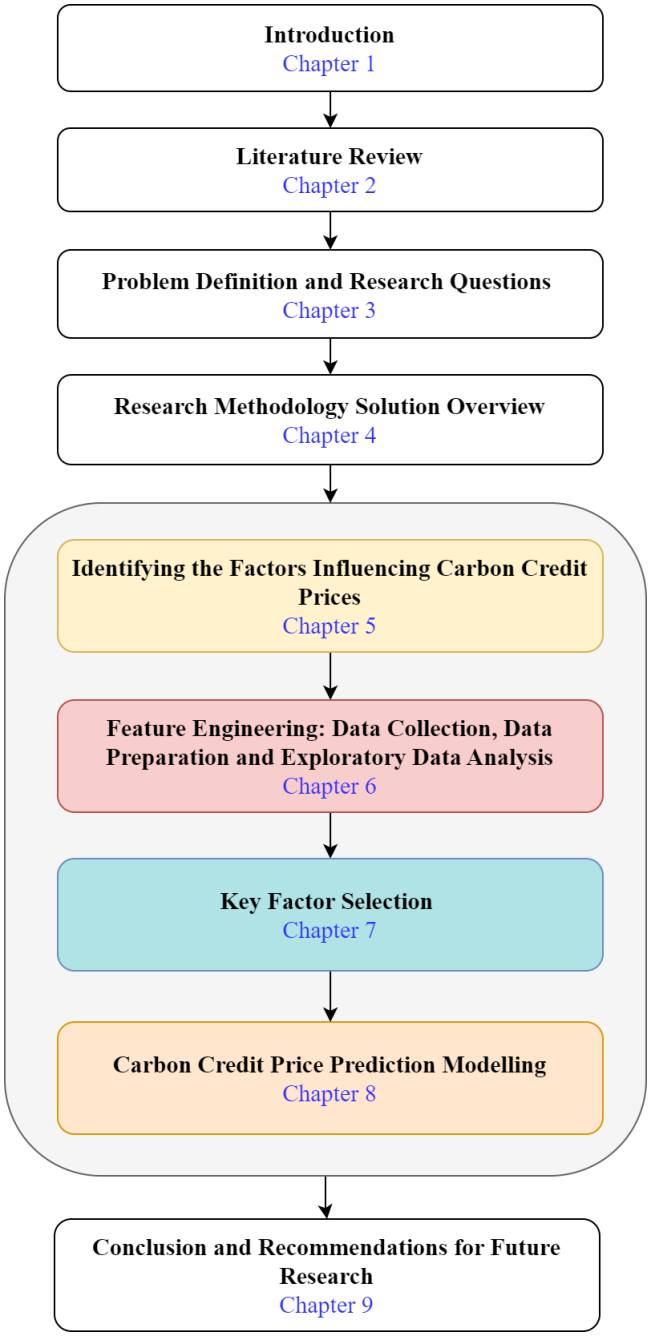


Figure 1.2 : Thesis structure.

Chapter 2

Literature Review

2.1 Overview

Carbon credit trading plays an important role in achieving global emission reduction targets at the lowest cost. The accurate prediction of carbon credit prices is essential for efficient carbon market operations and reaching these targets. However, predicting carbon credit prices is inherently complex due to their sensitivity to multiple external factors, including energy markets, economic indicators, and environmental factors. A systematic literature review (SLR) was conducted to explore the advances and challenges in predicting carbon credit prices. The primary focus of this review is to identify and evaluate relevant studies that propose prediction models, and address carbon credit price complexity, such as non-linearity, non-stationarity, and multicollinearity. The SLR was undertaken following the guidelines of Kitchenham and Charters [77] as follows:

- **Search protocol**: A predefined list of databases and search terms was used to ensure a comprehensive retrieval of relevant studies.
- **Define inclusion and exclusion criteria**: Specific criteria were established to define the scope of the review and focus on retrieving the most relevant and high-quality studies. The articles that met the inclusion criteria were selected for further analysis.
- **Study selection process**: A three-stage filtration was applied to select the most relevant articles. This involved screening titles and keywords, reviewing

abstracts, and analysing full texts.

- **Data extraction**: Key information was extracted after reviewing the selected studies.

Because it is difficult to ensure that all relevant studies are identified, backward citation is used in addition to the electronic data source search method. The same selection stages were applied to the articles found using backward citation.

This (SLR) examined 52 studies from 2005 to 2024 and categorised existing research into two areas: multi-factor analysis and multi-factor predictive models, providing a structured framework for understanding developments in the field. An in-depth analysis using 18 criteria reveals critical gaps and areas for improvement, establishing a foundation to enhance predictive accuracy in carbon credit pricing for future academic and industry applications.

This chapter is organised as follows: Section 2.2 provides an introduction including the background. Section 2.3 illustrates the detailed description of the procedure used to select relevant papers for this thesis. Section 2.4 examines articles classified under external factor analysis. Section 2.5 examines articles classified as price prediction based on external factors. Section 2.6 critically analyses the selected studies and identifies the gaps and findings. Finally, Section 2.7 concludes the chapter.

2.2 Introduction

2.2.1 Background

Climate change has become a global issue with serious impacts on both the environment and human life [71], [109]. Global warming is mainly caused by the increase in carbon dioxide emissions from human activities, such as the use of fossil fuels [3]. Carbon dioxide is released into the atmosphere when fossil fuels are burnt, acting as a blanket over the Earth, trapping heat and increasing temperatures.

Governments have redoubled their efforts to address the problem and limit carbon emissions. Consequently, several projects have emerged, such as the Kyoto Protocol [141] and the Paris Agreement [117]. These projects are the most significant global efforts to date in fighting climate change. In December 1997, in Kyoto, Japan, officials from 150 nations initiated the first global initiative to slow climate change. An agreement was reached to limit greenhouse gas emissions by 5.2%. The Paris Agreement replaced the Kyoto Protocol in 2021 and reflects the world's political determination to reduce carbon emissions and contribute to keeping global warming to below 1.5 degrees Celsius by 2050 [85].

One of the proposed solutions is carbon emission trading (CET), a market-based approach introduced in the Kyoto Protocol in 1997 [122]. CET aims to reduce greenhouse gas emissions, including CO2, by implementing a "cap and trade" mechanism [122]. Carbon credit trading is one of the primary objectives in the fight to reduce emissions at the lowest possible cost while meeting environmental goals. Carbon credits are used in carbon emission trading as tradable units that represent emission reductions, allowing entities to meet their regulatory obligations while supporting emission reduction projects [128]. These credits are digital assets, with one carbon credit equivalent to one tonne of emissions [128]. Nations are assigned specific CO2 allowances corresponding to their emission reduction targets and can trade surplus carbon credits with other countries. This concept has led to the establishment of carbon trading schemes in various regions, including Europe, North America, and Asia, with more than 40 countries and 20 sub-nations either implementing or planning such schemes [140].

2.2.2 The need for a carbon credit price prediction model

Accurate model predictions have become increasingly important, as they are essential for effective decisionmaking in various industries, including energy, finance,

and agriculture. Carbon emission credits are increasingly recognised as investment assets, similar to stocks and real estate [74]. The accurate prediction of carbon credit prices can lead to better informed trading decisions, greater profitability, and lower risk [133]. This can help investors identify emerging trends and quickly adjust their investment strategies, thereby enhancing the potential returns of their investments.

One of the biggest problems with carbon credits is that traders find it very challenging to predict their price, as the price of carbon credits over time exhibits non-stationarity, non-linearity, and multicollinearity characteristics [105]. It is difficult to predict the price of carbon credits based on historical data in a constantly changing and uncertain market [98], [155]. The concept of "non-stationarity" in the data implies that the statistical properties of carbon credit prices are not constant over time [70]. This could involve sudden spikes or drops in carbon credit prices, which do not adhere to a consistent pattern. Consequently, carbon credit prices do not follow a predictable trend, and this fluctuation can be challenging to analyse and forecast accurately. The presence of "non-linearity" within the dataset implies that the relationship between carbon credit prices and their determinants is not linear or straightforward. Instead, it can involve intricate patterns, curves, and abrupt shifts. "Multicollinearity" occurs when the independent variables in a regression model are highly correlated, leading to unreliable estimates of the regression coefficients [60]. It can cause instability in prediction models and make it difficult to determine the effect of each predictor. These challenges are the result of the influence of external factors on carbon credit prices [94], [174], such as energy market prices [8], [175], environmental conditions [130], microeconomic variables [23] and financial markets [175]. Recognising and addressing these characteristics is imperative for robust data analysis and the prediction of carbon credit prices.

A significant research gap exists due to the lack of studies that have systematically examined carbon credit price prediction models that consider the factors

that influence carbon credit prices. To our knowledge, no recent academic studies have systematically addressed this gap. As explained by Littell et al. [96], the goal of an SLR is to "find and synthesise research that relates to a particular question, using organised, clear, and repeatable procedures at each step of the process." This SLR is the first attempt to systematically explore the methods and techniques in analysing and predicting carbon credits, considering external influencing factors. Therefore, this SLR examines 52 studies from 2005 to 2024 to evaluate the main models and methodologies used to analyse and predict carbon credit prices. This review categorises the existing research into two areas: multi-factor analysis and multi-factor predictive models, providing a structured framework for understanding developments in the field. An in-depth analysis using 18 criteria reveals critical gaps and areas for improvement, establishing a foundation to enhance predictive accuracy in carbon credit pricing for future academic and industry applications. In the following section, the search protocol employed is explained and presented.

2.3 Systematic review protocol

The objective of this SLR is to systematically evaluate and analyse the existing research on the prediction and analysis of carbon credit prices based on external influencing factors. A well-focused question should be formed to identify appropriate resources and search for relevant studies. To align with the research objective, the following research question is formulated to guide the SLR: What models are used to analyse and predict the price of carbon credits considering the external influencing factors? Therefore, using the review, our aim is to address the research question, identify the research gaps, and discuss the related methods and techniques of the relevant published studies. The SLR is carried out following the structured methodology described by Kitchenham and Charters [77] which involves a four-step process to identify, evaluate, and interpret relevant studies. These steps are carried

out as follows:

- Step 1: Search the literature.
- Step 2: Select relevant articles according to the inclusion and exclusion criteria.
- Step 3: Conduct the study selection process.
- Step 4: Undertake data extraction and synthesis.

2.3.1 Search the literature

As part of this step, we identify the main data sources, search terms, and search procedures used to extract the relevant studies that address the literature research question. Two search methods were used, as follows:

Searching electronic data sources

The following electronic databases were searched to identify the relevant literature:

- IEEE Xplore Library (http://ieeexplore.ieee.org)
- SpringerLink (https://link.springer.com)
- ScienceDirect (https://www.sciencedirect.com)
- ACM Digital Library (https://dl.acm.org).

These databases were selected because they are the largest and most widely used digital databases for academic journals, proceedings, and book chapters [47]. The search was conducted in September 2024 and covers the period from 2005 to 2024, since very little consideration was given to carbon credits before this period, as the Kyoto Protocol came into force in February 2005.

The keywords selected to search for studies related to carbon credit price prediction and analysis models based on the factors influencing carbon credit prices, along with the motivations for these searches, are presented in Table 2.1. These keywords are designed to capture the most relevant research that addresses the research question. The chosen search strings were created using critical phrases from the study question, keywords from various research articles acquired by a pilot search, and a synonym list. The "OR" and "AND" operators were used to find publications connected to these phrases. The search term was used in each database to select journals and conference articles, resulting in 25743 articles being retrieved from the databases.

Table 2.1 : Search categories and keywords for carbon credit studies.

Search Category	**Keywords**	**Motivation**
Carbon credit price prediction models	("Carbon credit" OR "Carbon allowance" OR "Carbon unit" OR "Carbon offset") AND (price OR value OR cost) AND predict	To identify studies and models that focus on predicting the price of carbon credits considering external influencing factors.
Carbon credit price analysis models	("Carbon credit" OR "Carbon emission" OR "Carbon offset") AND (price OR value OR cost) AND (factor OR driver OR determinant)	To identify studies and models that focus on analysing the price of carbon credits considering external influencing factors.

Backward citation searching method

In addition to the electronic data source search method, a backward citation search method [59] was implemented as an additional strategy to ensure that important articles were not overlooked. This method allows researchers to see what

sources and theories inspired the extracted literature. This method was applied to the final list of the selected articles. The same selection steps were applied to the output of the backward method.

2.3.2 Inclusion and exclusion criteria

Selection criteria are applied to the articles recovered from the databases to determine whether particular studies should be included or excluded from the SLR. The following are the formal inclusion and exclusion criteria:

Inclusion criteria: Papers must meet the following inclusion criteria to be considered for the SLR:

- Criterion 1: Published between 2005 and 2024.
- Criterion 2: Peer-reviewed journals and conferences.
- Criterion 3: The focus area must be on predicting or analysing carbon credit prices.
- Criterion 4: All papers must consider at least one external factor.
- Criterion 5: The full text must be available and written in English.

Exclusion criteria: Papers that meet the following criteria will be excluded from the SLR:

- Criterion 1: Non-English papers.
- Criterion 2: Books or book chapters, theses, white papers, editorial comments, predatory journals, patents, preprints, surveys, and review articles.
- Criterion 3: Papers do not focus on the price of carbon credits.
- Criterion 4: Duplicate publications.

- Criterion 5: Papers with no available full text.

As a result, 177505 articles were included, and 8038 articles were excluded from this SLR. Zotero [177] was used in this work to manage the references and export details of the shortlisted articles to further filter them, as explained in the next step.

2.3.3 Study selection process

In this step, three stages of filtration were followed to decide whether to include an article in this literature review. The first stage applies to the title and the list of keywords in the article. The relevance of some of the articles could not be determined at this stage based on their titles and keywords. Therefore, the articles were not excluded and were further evaluated in the next stage. As a result of this stage, a total of 184 articles remained after thoroughly reviewing the titles and keywords of the articles. The second stage was to read the abstracts to check whether the articles met the objective of this SLR. Articles whose abstracts met the SLR objective were selected; otherwise, they were excluded. This resulted in 38 articles being moved to the next stage. The last filtration stage involved reading and analysing the full texts of the 38 articles to assess their relevance. All articles reviewed at this stage were found to be relevant and therefore included in the study.

To ensure comprehensive coverage of relevant studies, backward citation analysis is included along with electronic database searches. This approach helps identify key foundational works that may not appear in the selected database results [59], [76]. The same selection stages were applied to the articles found using backward citation, resulting in 14 additional relevant articles.

Finally, a total of 52 studies were included for detailed analysis, as explained in the next step. Figure 2.1 illustrates the entire SLR flow process, which shows the total number of studies retrieved from each digital database and each step.

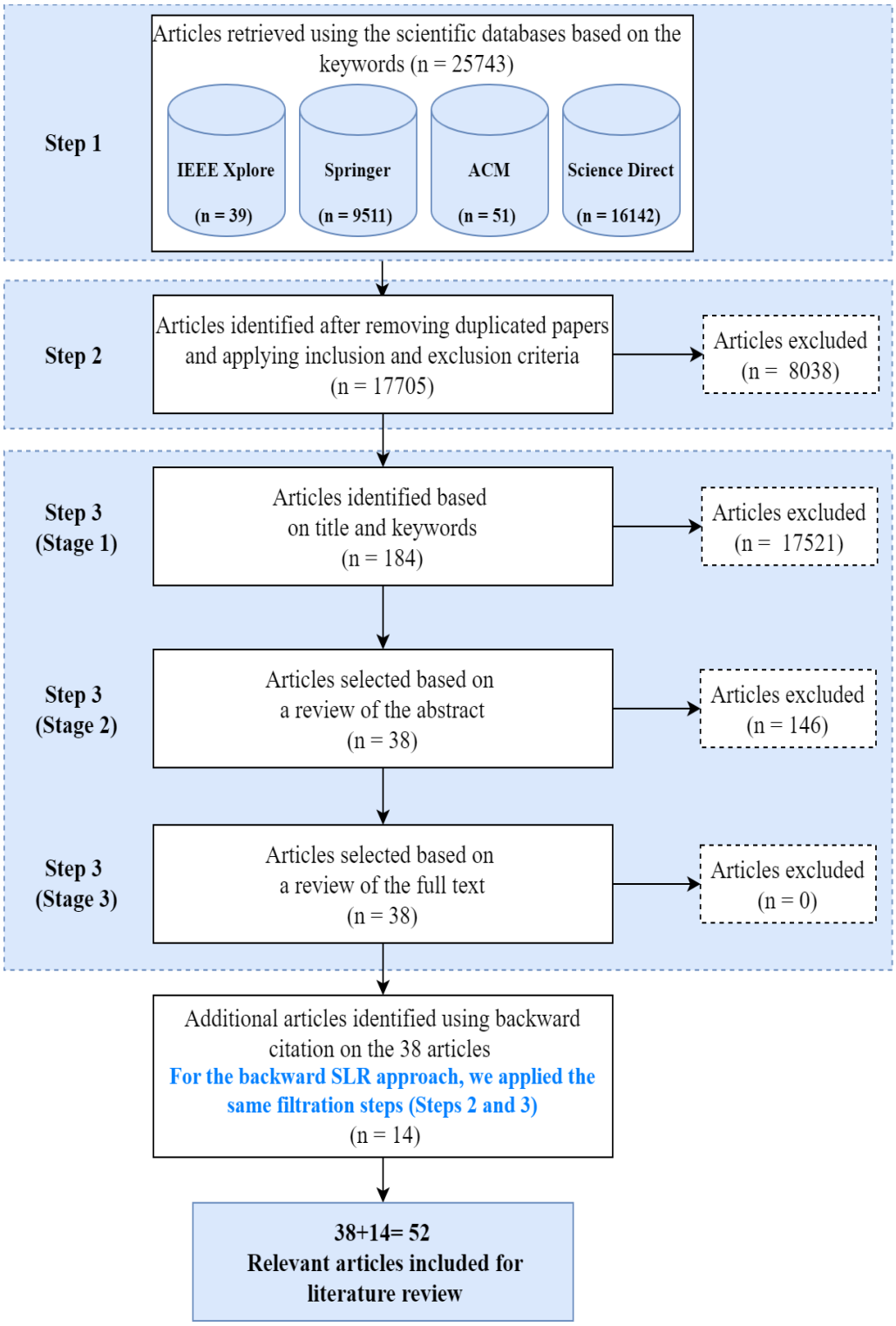


Figure 2.1 : The systematic literature review search and selection process.

2.3.4 Data extraction and synthesis

As a final step, the 52 relevant studies were carefully examined. This SLR synthesises the findings of the selected studies and organises the extracted data into spreadsheet format tables. This approach provides a clear overview of the key aspects, including the studied market, methodological approaches, and factors identified in the related literature.

As evidenced by the trend shown in Figure 2.2, the topic is gaining substantial importance and is increasingly attracting the interest of the research community, as reflected in the significant increase in the number of publications, particularly reaching its highest point in 2022.

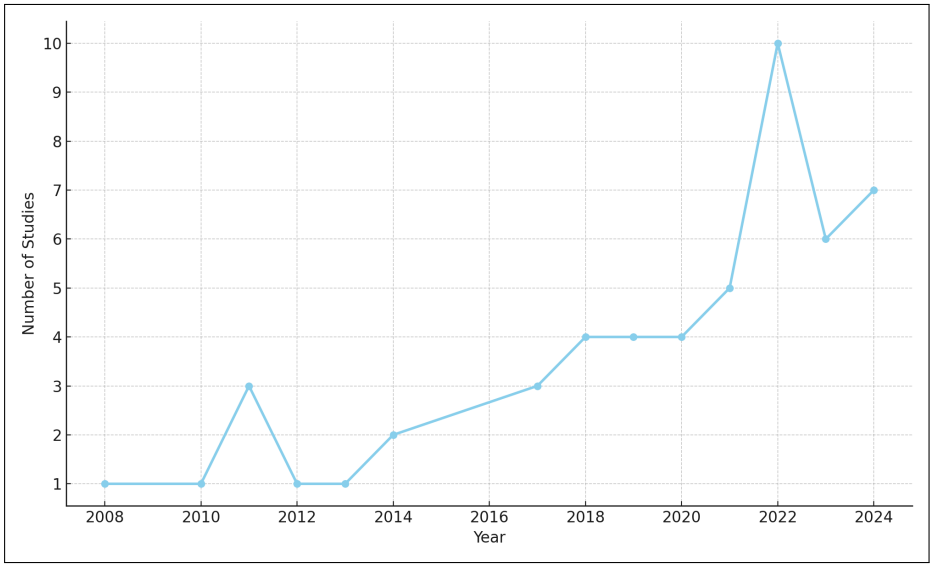


Figure 2.2 : Distribution of the selected articles by year on carbon credit price prediction and analysis models.

The chosen articles were divided into two categories, as shown in Figure 2.3: multi-factor analysis and multi-factor price prediction. The multi-factor analysis category includes the methods used to analyse the impact of external factors on carbon credit prices, focusing on statistical (AS) and time series methods (AT). The

multi-factor price prediction category consists of models that predict carbon credit prices based on external factors and is further divided into four subcategories: statistical methods (PS), regression-based methods (PR), neural network-based methods (PN), and hybrid methods (PH). Figure 2.4 shows the distribution of articles in these categories, highlighting that time series methods dominate multi-factor analysis, while hybrid methods are the most commonly used approach in multi-factor price prediction. The next section reviews the shortlisted studies to evaluate their contributions to the analysis and prediction of carbon credit prices.

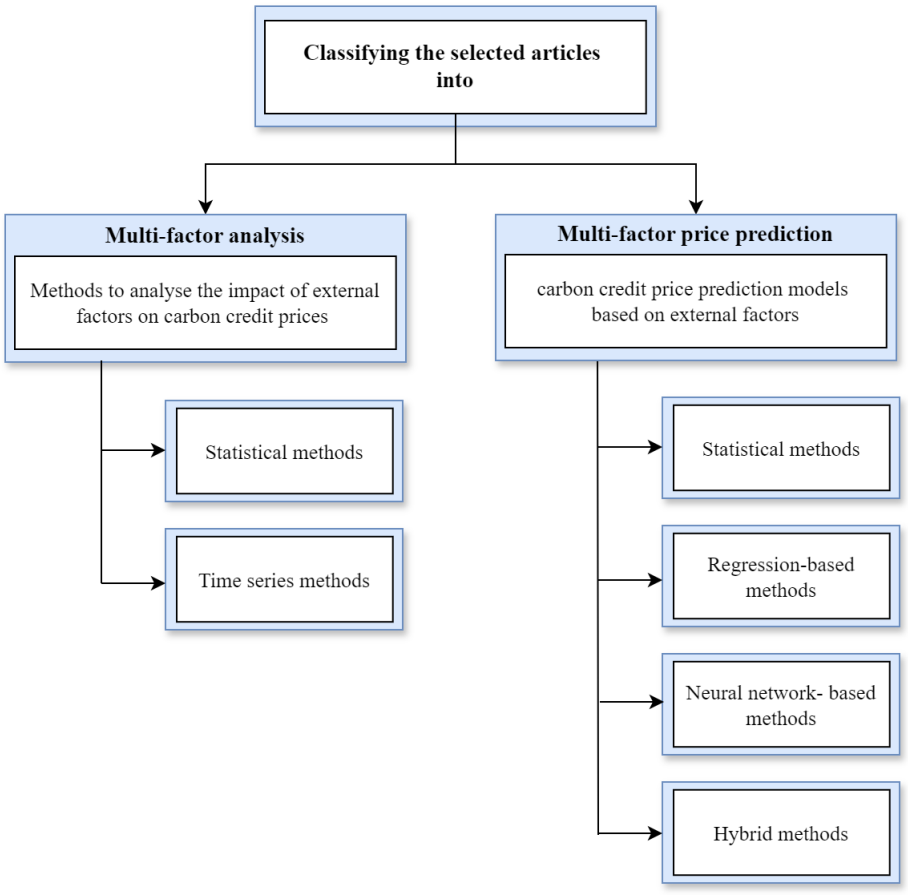


Figure 2.3 : Categorisation of the existing literature based on their adopted techniques for multi-factor analysis and multi-factor price prediction.

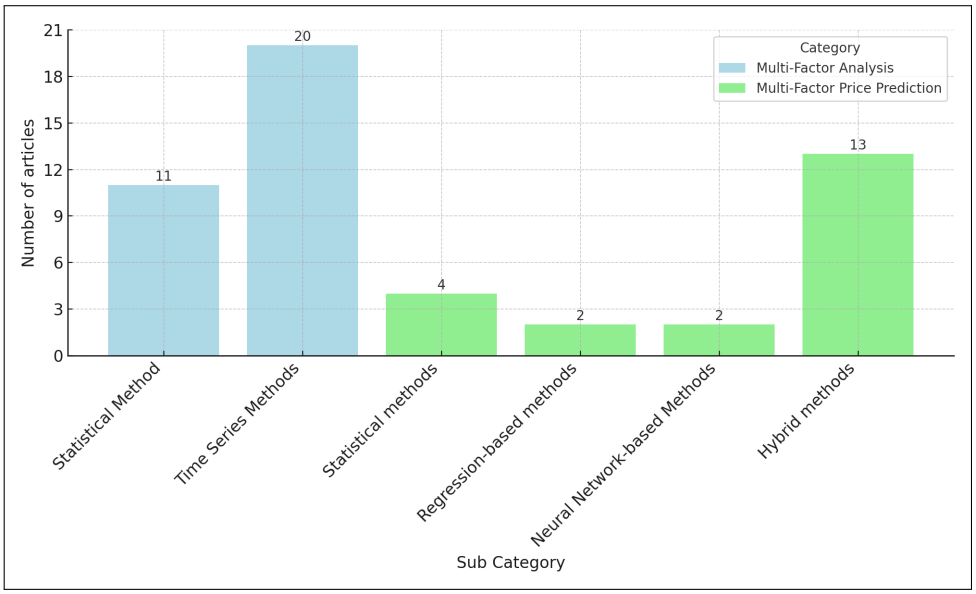


Figure 2.4 : Distribution of the selected articles across multi-factor analysis and multi-factor price prediction methods.

2.4 Literature review: Multi-factor analysis methods

This section analyses the selected studies and assesses their contributions to the prediction of carbon credit prices. In addition, we identify the factors and methods used in each study. While researchers explore ways to predict the price of carbon credits, they also examine the various factors that influence these prices. Many studies analysed the factors that affect the carbon trading market, with a particular focus on the European Union Allowances (EUA) and the China pilot carbon markets. It is important to note that many of the reviewed studies incorporate time series techniques to investigate the impact of various external factors on carbon credit prices. The following subsections provide a detailed overview of these approaches.

2.4.1 Statistical methods for multi-factor analysis

Creti et al. [31] investigated the determinants of carbon prices during the first two phases of the European Union Emission Trading Scheme (EU ETS). Their pa-

per aimed to test whether the drivers identified for Phase I of the EU ETS also applied to Phase II and to determine if the carbon price evolved toward a long-term equilibrium. The study used cointegration techniques to analyse the long-term relationship between carbon prices and their determinants, employing fully-modified OLS (FM-OLS) and dynamic OLS (DOLS) methods to estimate the cointegration relationships. The authors accounted for structural breaks, particularly the compliance break in 2006, which had a significant impact on the carbon market. The study used daily futures contracts for EUAs from June 2005 to December 2010, incorporating external factors such as energy prices (Brent crude oil, natural gas, coal), switching prices (the cost at which gas and coal-fired power plants become equally competitive), and economic indicators (Dow Jones Euro STOXX 50). A dummy variable was also used to account for the structural breakdown in 2006. The paper did not explicitly use feature selection or reduction techniques but focused on analysing the most relevant external factors influencing carbon prices. The main contribution of the paper is to demonstrate that equilibrium relationships between carbon prices and their fundamentals exist in Phases I and II of the EU ETS, with energy prices playing a more significant role in Phase II. It also provides insights into the discrepancies between observed carbon prices and predicted equilibrium values, helping to understand market inefficiencies. However, the focus of the article on a specific time frame (2005–2010) may limit its generalisability to other future phases of the EU ETS. Additionally, while the cointegration analysis handled non-stationarity, the paper did not address potential multicollinearity between the explanatory variables.

Feng et al. [44] also examined the volatility of carbon prices within EU ETS using ensemble empirical mode decomposition (EEMD) and variance ratio analysis. The paper aimed to identify the key factors that influence carbon price volatility, particularly focusing on the effects of market mechanisms, heterogeneous environments,

and temperature sensitivity. The authors proposed hypotheses about how these factors affect carbon price fluctuations. The study applied EEMD to decompose carbon price data into intrinsic mode functions (IMFs), allowing for a detailed analysis of both the high and low-frequency components of carbon price movements. The high-frequency components were associated with market mechanisms, while low-frequency components were influenced by external environments and special events such as international negotiations and financial crises. The data used included carbon futures contracts from the European Climate Exchange (ECX) from April 2005 to December 2010, with a focus on the December 10 contract. The external factors considered in the study included temperature, market mechanisms, international negotiations, and financial crises. The paper did not employ explicit feature selection or reduction techniques but used EEMD to decompose the data into relevant components. The key contribution of the work is the identification of the different frequencies at which market mechanisms and heterogeneous environments affect carbon price volatility, providing insights into both short-term and long-term price movements. However, the study limitations include the lack of attention to multicollinearity in the dataset and the focus on a specific time frame, which may reduce the generalisability of the findings to other time periods. Additionally, while the paper effectively addressed non-stationarity using EEMD, it did not compare the performance of its approach to other analytical models.

Hammoudeh et al. [53] used a quantile regression approach to investigate the impact of energy prices on the prices of CO2 emission allowances in the United States. The study specifically analysed how changes in crude oil, natural gas, coal and electricity prices affect the distribution of CO2 allowance prices, particularly during periods of extreme market conditions. The key objective was to explore the non-linear and asymmetric effects of energy prices on carbon prices at different quantiles of the carbon price distribution. The quantile regression approach differs from

traditional ordinary least squares (OLS) by estimating varying slope coefficients at different points in the carbon price distribution. The data used for the study were daily prices for CO2 emission allowances from the European Union carbon market (EEX EU allowances) and energy prices (crude oil, natural gas, coal, electricity) from July 2006 to November 2013. The study's contribution lies in demonstrating the varying effects of energy prices on CO2 allowance prices across different quantiles. The authors found that crude oil price increases lead to a substantial drop in carbon prices when carbon prices are very high, while natural gas price increases negatively affect carbon prices at low quantiles but positively affect them at higher quantiles. This quantile-based approach provides a better understanding of how energy prices influence carbon prices under different market conditions. However, the study has some limitations; although the study performed tests for non-stationarity, it did not account for non-stationarity beyond these tests. Furthermore, the study did not explicitly use feature selection or reduction techniques to address the multicollinearity between the energy price variables, which could potentially affect the accuracy of the results.

Ji et al. [64] developed a model using the autoregressive distributed lag (ARDL) approach to assess the relationships between carbon prices and key factors such as product prices, raw material prices, energy prices, financial market performance, and weather conditions in five Chinese pilot carbon markets: Shenzhen, Beijing, Guangdong, Shanghai, and Hubei. The main objective of the study was to explore the impacts of policy and economic factors on carbon prices, particularly focusing on the role of structural changes in the carbon market. The data used in this study included carbon prices from the five pilot markets between June 2013 and July 2019. External factors considered in the model were product prices (iron ore, cement, non-ferrous metals, chemical products), energy prices (crude oil, coal), financial market indicators (Hushen 300 index), and weather variables (water levels

and extreme temperatures). Although the paper did not apply specific feature selection techniques, it employed structural break tests to detect significant shifts in carbon prices due to policy changes. The ARDL model was then used to estimate both short-term and long-term relationships between carbon prices and external factors. The contribution of the paper lies in its detailed analysis of the effects of policy and market structure on carbon prices in China, particularly highlighting how structural changes, such as the allocation of carbon allowances, impact market dynamics. However, while the ARDL model accounts for non-stationarity, the study did not address potential multicollinearity between the explanatory variables.

Lin and Jia [93] explored the key factors that affect carbon prices in the context of China's ETS using a computable general equilibrium (CGE) model. The objective of the paper was to analyse how different design mechanisms within the ETS, such as industry coverage, annual decline factor, and free allowance rate, impact the carbon price and the overall market structure. The study used the CGE model, which is a widely used tool for policy analysis, especially in the areas of energy, environment, and economy. This model simulates the behaviour of households, enterprises and governments to analyse the effects of ETS design factors on the dynamics of carbon prices. The data used included sector-specific energy consumption, CO2 emissions, and economic output, which were combined into the social accounting matrix. The external factors analysed in the study include industry coverage (sectors such as cement, electricity, and steel), free allowance rates, annual decline factors, and energy consumption levels. The study did not use explicit feature selection or reduction techniques. Instead, it modelled various scenarios by adjusting key ETS design factors to observe their effects on market outcomes. The main contribution of the study is the detailed scenario analysis of how different ETS design factors affect carbon prices. The findings suggest that covering more industries in the ETS and lowering free allowances generally result in lower carbon prices, while stricter annual

decline factors lead to higher prices. However, while the CGE model addressed the non-stationarity of the data, the study did not explicitly address potential multicollinearity between the variables.

Luo et al. [104] investigated the relationship between carbon prices and energy prices in the Beijing carbon ETS, focusing on their co-movement and lead-lag dynamics in both time and frequency domains. The study used wavelet analysis to examine weekly data on oil, natural gas and carbon prices from November 2013 to April 2019, offering insights into the complex interconnections between energy and carbon markets. The wavelet analysis method was used to decompose the time series into various frequency bands, allowing the authors to capture both short-term and long-term relationships between carbon and energy prices. The study found that natural gas prices exhibited a consistent negative correlation with carbon prices, leading to changes in the carbon market over short-term periods (12-20 weeks). In contrast, oil prices showed an unstable and variable relationship with carbon prices, with leadership roles alternating between the two markets. The primary contribution of this work lies in its ability to identify the short-term dynamics between carbon and energy prices, particularly the strong influence of natural gas prices on carbon prices. However, the study's limitations include its focus on a single regional carbon market (Beijing) and a limited set of external factors (natural gas and oil prices). Furthermore, while wavelet analysis handled non-stationarity, the study did not address potential multicollinearity, though this is less of a concern given the limited number of factors.

Li et al. [91] analysed the comovement between carbon prices and various indicators of the financial and energy market using a wavelet-based approach over different time horizons. The study applied wavelet analysis to decompose the relationships between carbon prices and external factors into different frequency bands, allowing a better understanding of their short-term and long-term interactions. The focus

was on determining which factors had the most significant influence on carbon prices in varying time periods. The data used in the article included EUA price, British natural gas price, Brent crude oil, Rotterdam coal price, and stock market indices (S&P 500, FTSE 100, DAX, CAC 40). In addition, bond yields and exchange rates (EUR to RMB and EUR to USD) were incorporated into the analysis. Although no explicit feature selection or reduction techniques were mentioned, wavelet-based decomposition allowed for a focused analysis of relevant time-frequency interactions. However, the paper's limitations include the absence of feature selection techniques, and while the wavelet analysis effectively handled non-stationarity, the study did not explicitly address the multicollinearity between the external factors, which could affect the robustness of the results.

Song et al. [130] examined the key factors that influence carbon price trading in the eight pilot carbon markets. The authors used the improved grey relational analysis model (GRAM) to assess the correlation between various external factors and carbon trading prices, focusing on both spatial and temporal dimensions. The study investigates 15 external factors, including the government's free carbon quotas, the price of coal, the price of crude oil, the level of industrial development, the maturity of green technology, the penetration of clean energy, USD/CNY exchange rate, air quality index (AQI) and air temperature. After applying systematic cluster analysis to reduce the correlation between the factors, the authors retained 11 factors for further analysis. The study found that industrial development, low-carbon industry development, AQI, and green technology maturity had the strongest influence on carbon trading prices. Of these, the economic and green industry factors were particularly significant.

Tan and Wang [136] investigated the dynamic dependence between carbon prices and their key influencing factors, focusing on energy prices and macroeconomic risk factors, during the three phases of the EU ETS. The study applied quantile regres-

sion to examine the non-linear and asymmetric relationships between carbon prices and external factors, which include Brent crude oil prices, natural gas prices, coal prices, the Dow Jones Euro STOXX 50 index, the Commodity Research Bureau (CRB) index, a Treasury bill (T-Bill), and junk bond yields. The study found that the dependence between carbon prices and energy/macroeconomic factors changes across different phases of the EU ETS, reflecting the evolving influence paths of these factors. Energy prices, particularly coal and gas, showed significant effects on carbon prices, with the relationships varying between symmetric and asymmetric depending on the economic environment and the phase of the EU ETS. The paper also identified three structural breaks in the relationship, corresponding to major economic and financial crises. One of the key contributions of the paper is the identification of specific dependence paths between energy markets and carbon prices, such as the aggregated demand effect, the substitution effect, and the production restraint effect. These paths varied between the different phases of the EU ETS. The study also highlights how macroeconomic risk factors, such as T-Bill and junk bond yields, influenced carbon prices through industrial production and energy price paths. However, the study did not apply specific feature selection techniques, and, while it addressed non-stationarity through cointegration methods, it did not explicitly handle multicollinearity between the influencing factors.

Tang et al. [137] investigated the factors that influence the volatility of carbon prices in two carbon markets: the EU ET and the CDM. The study aims to explore how market mechanisms, external environments, and seasonal changes affect carbon price volatility. The paper proposes three hypotheses related to carbon price volatility: that the market mechanism influences prices with a short duration and high frequency, that the external environment affects prices with a long duration and low frequency, and that the carbon market is sensitive to seasonal temperature changes. The authors applied EEMD to decompose carbon price series into IMFs to

better understand the various components that influence volatility. In addition, they used the variance ratio test to assess the effect of seasonal changes on carbon price volatility. The data used in the analysis include daily futures contracts from the EU ETS and CDM markets (DEC12 contracts), covering the period from January 2009 to October 2012. The external factors considered in the analysis were market mechanisms (supply-demand dynamics and electricity price), external environment (international climate negotiations and policy changes), and seasonal temperature variations. The results showed that market mechanisms had a short-term but frequent impact on carbon prices, with small amplitude fluctuations. The external environment had a longer-term impact with greater price fluctuations. However, seasonal changes did not significantly affect carbon prices during the study period, contradicting earlier research. The main contribution of the study is the comparative analysis of the EU ETS and CDM markets, highlighting differences in the way market mechanisms and external environments affect carbon price volatility. However, no feature selection or reduction techniques were applied and the paper does not explicitly address the potential multicollinearity between influencing factors.

Zeng et al. [161] investigated the key factors that influence the price of carbon trading in China's carbon trading market under the background of 'double carbon'. The study covers five pilot markets: Beijing, Shenzhen, Shanghai, Hubei, and Guangdong, from January 2017 to December 2021, using daily average price data. The authors used the dynamic heterogeneous panel pooled mean group (PMG) model to analyse the relationship between the carbon trading price and influencing factors. The main factors analysed in the study include macroeconomic factors such as the Shanghai Stock Exchange Industrial Index (SSEII), energy price factors such as the Bohai-Rim Steam-Coal Price Index (BSPI) and the Liquefied Natural Gas Index (LNGI), and environmental factors such as AQI. These factors were found to have short-term and long-term impacts on the carbon market, with energy prices

and industrial development playing a critical role in determining carbon prices in the pilot markets. However, a limitation of the paper is that while it effectively addresses non-stationarity using dynamic panel models, it does not apply specific feature selection or reduction techniques, and it does not explicitly address the multicollinearity between the influencing factors, which could affect the robustness of the results.

2.4.2 Time series methods for multi-factor analysis

Aatola et al. [1] investigated the price determination of the EUA during its first five years of operation, from 2005 to 2010. The primary objective of the study was to explore how the price of EUA is influenced by market fundamentals such as energy prices and production costs. The study applied several econometric models, including OLS, instrumental variables (IV), and vector autoregression (VAR), to test the relationship between EUA prices and various market factors. The factors included in the analysis were: German electricity prices, UK gas prices, coal prices, mineral product prices, steel product prices, paper product prices, water reservoir levels, UK gas storage levels, Brent crude oil prices, and the FTSE 350 stock market index. The study found that the EUA forward price was significantly influenced by German electricity prices, as well as the price difference between gas and coal. The findings demonstrated that electricity, coal, and gas prices had a strong and stable relationship with EUA prices, indicating that energy prices play a crucial role in determining the price of emissions allowances. One limitation of the study is that while it addressed non-stationarity by applying econometric models with stationary time series, it did not explicitly handle multicollinearity among the external variables, which could affect the accuracy of the results.

In the study by Alberola et al. [8], the authors used the VAR model combined with structural break tests (Lee and Strazicich test) to explore price drivers and

structural breaks in EUA prices during the pilot phase of the EU ETS from 2005 to 2007. The main factors analysed in the study include Brent crude oil prices, natural gas prices (Zeebrugge Hub), coal prices (CIF ARA), electricity prices (Powernext), clean dark spread, clean spark spread, switch price, and temperature variables including extreme temperature deviations. The paper revealed that energy prices, particularly natural gas and coal, had a significant impact on EUA price changes. The study also identified two structural breaks in April 2006 and October 2006, corresponding to key institutional events such as the first disclosure of verified emissions data and the European Commission's stricter national allocation plans for Phase II. The findings suggested that EUA prices reacted not only to energy prices but also to unanticipated weather events, particularly during colder periods. However, while the paper effectively addressed non-stationarity by using the differencing method, it did not apply specific feature selection techniques or handle the multicollinearity between the external variables, which could impact the robustness of the results.

Chevallier [24] developed a model to examine the interactions between carbon prices, macroeconomic activity, and energy prices, focusing on EUAs in the EU ETS. The study used a Markov-switching VAR model to explore how these variables interact under different economic regimes, particularly during periods of economic expansion and contraction. The external factors analysed in the study include Brent crude oil, natural gas and coal prices and the aggregated EU industrial production index. The study found that industrial production positively influenced carbon prices during periods of economic expansion, while energy prices, particularly Brent crude oil, were critical drivers of carbon price formation. The research highlighted the dynamic interplay between macroeconomic and energy factors in determining carbon prices, revealing that regime shifts in the economy significantly impact these relationships. A key contribution of the paper is its novel application of the Markov-switching VAR model, which effectively captures regime changes in the economic

environment and how these shifts affect carbon prices. However, while the model addresses non-stationarity by accounting for regime changes using differencing, the study did not apply any feature selection or reduction techniques to manage the multicollinearity between external variables, which may influence the accuracy of the findings.

In the study by Gavard and Kirat [45], the authors examined the price dynamics of EUA and international carbon credits, specifically certified emission reductions (CER), during the second phase of the EU ETS. Their primary objective is to assess whether the flexibility in the international market for carbon credits affects the price dynamics of these permits and how their pricing mechanisms differ. The authors used autoregressive conditional heteroskedasticity (ARCH), generalized autoregressive conditional heteroskedasticity (GARCH) and GARCH-in mean (GARCH-M) models to analyse the impact of fundamental drivers on the price dynamics of both EUA and CER. The key factors considered in the analysis include natural gas and coal prices, the EURO STOXX 50 index, policy changes, and market changes. The results showed that EUA prices were predominantly driven by demand, as expected in a regulated market. In contrast, CER prices were more influenced by supply-side factors due to the flexibility in their issuance under the CDM. The study also found differences in the long-term relationships between EUA prices and CER prices, with economic activity having a more significant impact on EUA prices than on CER prices. However, while the ARCH models effectively capture volatility, the study does not explicitly address non-stationarity, and the potential multicollinearity among the fundamental drivers is left unexamined.

Guan [49] investigated the determinants of carbon emission prices using a vector error correction (VEC) model. The study specifically focused on the relationship between carbon emission futures prices, macroeconomic performance, and energy prices. By constructing the VEC model, the paper sought to explore the long-

run equilibrium and short-run dynamics between carbon prices and these external factors. The key factors considered in the study include the spot price of thermal coal in the European ARA Port, the WTI crude oil futures, the natural gas price, and the EURO STOXX 50 index. The study found that carbon emission futures prices were positively correlated with the EURO STOXX 50 index and negatively correlated with crude oil prices in the long run. Specifically, during periods of economic expansion, the demand for energy increases, leading to higher carbon emissions and a rise in carbon prices. In contrast, an increase in crude oil prices tends to reduce carbon prices as higher oil costs lower industrial output, reducing emissions and demand for carbon futures. The VEC model allowed the author to capture both long-term equilibrium relationships and short-term deviations. The study found that when short-term fluctuations deviated from the long-run equilibrium, the correction occurred at a relatively slow rate. A limitation of the study is that it did not apply specific feature selection techniques, and while the VEC model handled non-stationarity, the paper did not address the potential multicollinearity among the variables.

Jiang et al. [68] applied the vector autoregressive model (VAR) to analyse weekly data from November 25, 2013, to March 19, 2017, in seven pilot regions: Beijing, Tianjin, Shanghai, Chongqing, Hubei, Guangdong, and Shenzhen. The results show that the carbon price is primarily influenced by its own historical prices. The study found that the coal price and the stock index have a negative effect on carbon prices, while the oil price initially has a negative effect in the short term but shifts to a positive effect over time. The findings suggest that enterprises reduce coal usage when coal prices increase, leading to a decline in carbon prices. The performance of the stock market also negatively impacts carbon prices, which may be linked to macroeconomic factors. The paper concludes that more regulatory attention is needed to improve the efficiency of carbon trading markets in China. However, the limita-

tions include the absence of specific feature selection techniques, and while the VAR model handles non-stationarity using cointegration, the potential multicollinearity between the influencing factors is not explicitly addressed.

Kim and Koo [75] examined the factors that influence the volume of carbon allowance trading in the US carbon allowance market, focusing specifically on the Chicago Climate Exchange. The key external factors considered in the study include the price of coal, crude oil, and natural gas, the temperature, represented by dummy variables for hot and cold weather, and the economic crisis using a dummy variable for the 2008 financial crisis. The study used the ARDL model to analyse both the short- and long-term relationships between the carbon allowance trading volume and the identified factors. The results of the study indicate that, in the long run, coal prices have a significant negative effect on the volume of carbon allowance trading, suggesting that as coal prices rise, the demand for carbon allowances decreases. In the short term, changes in crude oil, natural gas and coal prices significantly influence the trading volume of carbon allowances. The study also found that hot and cold weather, as well as the 2008 financial crisis, did not have a significant impact on the carbon allowance market. However, the study did not apply any specific feature selection techniques, and while it handled non-stationarity, the issue of multicollinearity between the energy prices was not explicitly addressed, which may affect the robustness of the results.

The study by Kotishwar [81] examined the impact of the global recession on carbon credits, with a specific focus on India's carbon credits. The research covers the period 2008 to 2019 and analyses the relationship between economic variables and carbon credit prices during the global financial crisis. The study used bivariate correlation and OLS regression to assess the impact of the recession on carbon credits and applied the VAR model to predict future movements of carbon credits. The external factors included in the study are the pollution index, the forest

growth index, the automobile production index, the consumption of petrol, the infrastructure index, the gross domestic product (GDP) of India and the industrial production index (IIP). The study found that carbon credits have a strong positive relationship with the pollution index, automobile production, and petrol consumption, suggesting that as pollution and production increase, carbon credits increase in value. However, infrastructure development and forest growth have a negative impact on carbon credits, as they contribute to reducing pollution. GDP had a positive but non-significant relationship with carbon credits, while IIP showed a negative but non-significant impact. However, the article did not apply the feature selection technique, and while the models addressed volatility using the ARCH family models, the study did not explicitly address multicollinearity between variables.

Li and Lei [88] applied the ARCH model to analyse carbon trading data from June 2014 to December 2016. The primary objective is to explore how energy prices, industrial activity, government intervention, and weather conditions affect the carbon trading market of the Hubei Province of China. The key external factors included in the study are industrial income, the Hubei coal price index, the retail price of liquefied natural gas, daily maximum temperature, daily rainfall, government intervention, and number of carbon quota management companies. The results show that energy prices, particularly coal and natural gas, had the most significant impact on carbon prices, with their influence increasing over time. Government intervention also had a stable and significant effect on carbon prices, while the number of participating enterprises was another significant factor. Interestingly, weather variables (temperature and rainfall) did not have a significant impact on carbon prices during the study period. However, while the study addresses volatility through the ARCH model, it does not apply feature selection techniques or explicitly handle the potential multicollinearity between variables, which could affect the robustness of the results.

Li et al. [89] focused on analysing the dynamic impacts of green bonds and economic policy uncertainty on carbon prices in China, specifically in the Hubei and Guangdong carbon markets. The study employs a time-varying parameter VAR (TVP-VAR) model and a Bayesian VAR (BVAR) model to capture the time-varying relationships between these factors from April 2014 to March 2022. The study includes green bonds and external factors of economic policy uncertainty. The key findings suggest that economic policy uncertainty has a positive short-term impact on carbon prices, while green bonds tend to have a negative short-term impact. The paper highlights the heterogeneous effects across the two provinces, with Hubei being more sensitive to economic policy uncertainty compared to Guangdong. The results emphasise the importance of stable policies to mitigate price fluctuations in the carbon market. Furthermore, the study suggests that green bonds can offer risk management tools to stabilise carbon markets over the long term. Although the use of TVP-VAR and BVAR models addresses time-varying relationships, the issue of multicollinearity between external factors is not directly tackled, which could affect the robustness of the results.

Lovcha et al. [100] examined the dynamic interaction between carbon prices and their determinants within the EU ETS, using the SVAR model to explore the relationship between energy markets and the share of fossil fuels in electricity generation, economic activity and the carbon market. They applied frequency domain analysis to assess short-term and long-term fluctuations in carbon prices due to fundamental market factors and speculative shocks. The study used data from 2008 to 2018. The key external factors analysed included Brent crude oil, natural gas prices, coal prices, electricity prices, the STOXX index for the EU and the share of renewable energy. The paper found that up to 65% of carbon price fluctuations were explained by these fundamental factors, with notable variations over time. Oil and gas prices were particularly influential in long-term carbon price dynamics, while short-term

fluctuations were mainly driven by speculative shocks. The study highlights the importance of low-frequency movements in policy interventions. However, it is limited by its focus on demand-side effects, as the data did not support a detailed analysis of permit supply shocks. Furthermore, no feature selection techniques were applied and the multicollinearity between variables was not directly addressed.

Wang et al. [146] examined the relationship between energy prices and carbon prices in Guangdong's carbon market from January 2019 to March 2021. They used a mediating effect model and tested it using OLS, least squares dummy variable (LSDV), and system generalised method of moments (SYS-GMM) models for robustness. External factors considered in the study include thermal coal prices, natural gas prices, the Shanghai Stock Exchange Index, the Shanghai Industrial Index, and EU emission quota prices. The key findings suggest that energy prices have a negative impact on carbon prices, with the main transmission path being through economic fluctuations. The study also reveals that energy prices indirectly influence carbon prices through speculative demand, investment demand, and transaction demand. The masking effect of speculative demand is noted, where market price expectations play a role in reducing the direct effect of energy prices. While the models handle non-stationarity, the multicollinearity among the external factors is not explicitly addressed.

Wei, Qi et al. [150] examined the impact of supply and demand factors, macroeconomic variables, energy consumption, technological innovation, and extreme weather conditions on the volatility of the carbon price. The authors used the multiple mediating effect model to quantify the impact of these factors on carbon price changes across seven carbon pilot markets: Beijing, Shanghai, Guangdong, Shenzhen, Tianjin, Hubei, and Chongqing, covering data from 2013 to 2018. The results show that gross industrial output and coal consumption have a significant positive impact on carbon prices, indicating that higher industrial activity and energy consump-

tion drive carbon prices upward. In contrast, technological innovation negatively influences carbon prices, suggesting that advances in technology reduce emissions and, therefore, lower the demand for carbon credits. In addition, extreme weather events were found to increase carbon prices, as they increase energy consumption. To eliminate multicollinearity, the authors conducted univariate regressions of the dependent variable against each independent variable separately. Then they incrementally added these variables in a stepwise manner to the multiple regression model, monitoring changes in coefficients and significance levels to identify multicollinearity issues.

Wu et al. [152] developed an integrated system to explore the short-term and long-term links between the carbon price and influencing factors in China, considering the impact of COVID-19. They applied the noise-assisted multivariate empirical mode decomposition (N-A MEMD) method to decompose time series data into IMFs and used VAR and VEC models to simulate correlations between carbon prices and various factors. The study analysed the data before and after the pandemic and found that energy prices (including crude oil, coal, and natural gas), macroeconomic indicators (such as the Goldman Sachs Commodity Index and STOXX Europe 600 Index), the exchange rate and epidemic mobility trends were the main external factors affecting carbon prices in the short term. In the long term, the Market Stability Reserve (MSR) mechanism was identified as stabilising the carbon price. Mobility trends during COVID-19, using Apple's mobility data, were also included to reflect the impact of the pandemic on economic activities. Although non-stationarity was handled through data decomposition, the potential multicollinearity between the influencing factors was not explicitly addressed.

Yu and Mallory [158] applied the SVAR model to examine the dynamic interactions between exchange rates, energy markets, and carbon credit prices within the EU ETS. It investigates how shocks in the Euro/USD and Euro/Swiss Franc ex-

change rates influence carbon prices through energy substitution between coal and natural gas. The study found that exchange rate shocks have a statistically significant impact on carbon prices primarily through this energy substitution mechanism, whereas the influence of natural gas prices is less pronounced. The study addressed non-stationarity by using first-differenced variables, but it did not explicitly examine multicollinearity, which could have impacted the robustness of the results, especially given the potential correlations between exchange rates and energy prices.

Zeng et al. [160] employed the VECM model to analyse both the short-term and long-term relationships between China's pilot carbon markets, focusing on the Hubei carbon market and six external factors: international carbon prices (EUA and CER), the dollar exchange rate and energy prices (coal, crude oil and natural gas). The study found that in the long run, the price of coal and crude oil positively influences carbon prices, while the price of natural gas has a negative effect. The results suggest that higher coal and crude oil prices lead to increased demand for carbon credits, while natural gas promotes the shift to cleaner energy, reducing the demand for carbon credits. Furthermore, the dollar exchange rate significantly impacts carbon prices, with dollar appreciation leading to higher carbon prices in China due to higher export demand. The paper also highlights that the short-term impacts of these factors vary, with innovations in coal, crude oil, and EUAs having negative short-term effects on carbon prices. The limitations of the study include the lack of feature selection techniques and the absence of a detailed examination of the multicollinearity between external factors.

Zhang et al. [162] explored the factors that influence the price dynamics of secondary sCER under CDM. The study used the exponential GARCH (EGARCH), power transformed ARCH (PARCH), threshold GARCH (TGARCH), and VAR models. These models were applied to analyse the volatility, transmission of risk and dynamic relationships between sCER prices and several external factors from 2008

to 2016. The key external factors identified were international carbon-reduction policies, macroeconomic conditions (represented by the S&P Global 1200 Index), energy prices (coal, oil, and natural gas), electricity prices, the volume of CER issuance and the pricing power of developing countries (represented by the China volume ratio). The study found that EUA carbon futures prices, coal prices, and macroeconomic conditions significantly influence sCER price fluctuations. Although the pricing power of developing countries also positively impacts sCER prices, its effect is relatively small. The research concluded that energy prices, particularly coal, have the largest negative impact on sCER prices due to substitution effects, where increases in coal prices lead to higher demand for cleaner energy sources such as oil and natural gas. The paper also identified significant policy-driven price fluctuations, particularly after the transition to Phase III of the Kyoto Protocol. Limitations include the study's focus on supply-side factors, with minimal consideration of demand-side dynamics. In addition, the multicollinearity between energy prices and other macroeconomic factors was not explicitly addressed.

In the study by Zhang et al. [165], the authors analysed the multidimensional correlation between carbon prices in the Guangdong pilot carbon market and eight influencing factors using the autoregressive moving average-GARCH-vine copula model (ARMA-GARCH-vine copula). The key external factors considered include natural gas, coal and Brent crude oil prices, West Texas Intermediate (WTI) crude oil prices, CER and EUA futures prices, the CSI 300 index, and the China industry index. The study assessed the spillover risk of these factors in the carbon market using the copula-CoVaR approach. The study found that the crude oil market, especially Brent crude oil prices, plays the most significant role in risk spillovers to the carbon market. In addition, international carbon markets and energy prices exhibit asymmetric risk spillover effects, with higher upside risks than downside risks for China's carbon market. The results suggested that China's carbon market is still

regional and policy-driven, with a limited direct link to international energy markets. However, crude oil prices, particularly Brent, have substantial indirect impacts through their effect on industrial production. The study also identified a negative correlation between coal prices and carbon prices, driven by the substitution effect of cleaner energy sources as coal prices rise. Multicollinearity and non-stationarity were addressed by applying the ARMA-GARCH model, as it filtered out both autocorrelation and conditional heteroscedasticity.

Zhou and Li [173] applied the VAR and VEC model to investigate the dynamic relationships between carbon emission prices and key external factors, including coal prices, the Shanghai and Shenzhen 300 Index (HS300), the Shanghai Industrial Index, and the AQI. The study found a long-term equilibrium relationship between carbon emission prices and these factors. The results showed that macroeconomic indicators, particularly the Industrial Index, have the most significant impact on carbon prices, followed by the HS300. Coal prices have a smaller impact, while air quality plays a relatively minor role. Furthermore, the authors applied the GARCH family of models to explore the fluctuation characteristics of carbon prices, identifying the presence of volatility clusters and non-normal distribution in the carbon market. They also found a positive leverage effect, where bad news has a greater impact on price fluctuations than good news. The limitations of the study include the lack of focus on multicollinearity, which could affect the accuracy of the results.

Zhu et al. [175] conducted a multiscale analysis of carbon price drivers within the EU ETS from 2009 to 2016, using the multivariate empirical mode decomposition (MEMD) method. The MEMD method allowed the authors to decompose carbon prices and their influencing factors into different timescales (short, medium, and long-term) to identify the specific drivers on each scale. The study also used cointegration analysis and error correction models to capture the long-term equilibrium

relationships between carbon prices and these external factors. The key external factors analysed were the price of oil, coal, natural gas, electricity, stock market performance (represented by the STOXX Europe 600 index) and commodity prices (represented by the Goldman Sachs Commodity Index excluding energy). The results show that at the original data level, electricity prices, coal prices, and the stock index were the most important drivers of carbon prices. Electricity and the stock index had significant positive impacts, particularly on short timescales, while coal prices exhibited a negative impact. At medium timescales, the effect of coal prices became more pronounced, and at long timescales, the impacts of oil and natural gas prices were more significant, with both showing negative effects. The findings suggest that carbon prices are driven by different factors on different time scales, with electricity prices and the stock index consistently playing a significant role. The authors addressed non-stationarity using the augmented dickey-fuller test, confirming that the original data were nonstationary, but after decomposition and differencing, all variables became stationary. For multicollinearity, the variance inflation factor test did not reveal significant multicollinearity in the model. In addition, the Newey-West estimator was employed to correct for potential heteroskedasticity and autocorrelation, ensuring the robustness of the analysis.

2.5 Literature review: Multi-factor price prediction methods

2.5.1 Statistical methods

Statistical methods have demonstrated their ability to provide a short-term prediction of carbon prices [147]. However, methods under this approach rely heavily on strict statistical assumptions like linearity and smoothness. Consequently, their accuracy suffers when dealing with non-linear and non-stationarity time series.

Adekoya [4] proposed a predictive model using the feasible quasi-generalised least squares (FQGLS) model, which accounts for various statistical challenges such as

heteroskedasticity, persistence, serial correlation and endogeneity in the dataset. The research focuses on how asymmetric changes in energy prices affect the prediction of carbon allowance prices in the European Union (EU). The dataset used in this research covers futures prices of EU carbon allowances along with energy prices (crude oil, natural gas, and coal) from 18 October 2009 to 5 April 2020. The study also considers asymmetric and symmetric models, finding that accounting for asymmetry in energy prices improves prediction accuracy, particularly for oil and coal prices. Although the paper did not apply specific feature selection or reduction techniques, it used FQGLS to handle data complexities. The contribution of the work lies in its ability to demonstrate the importance of asymmetries in energy prices when predicting EU carbon allowance prices. However, its limitations include the use of only two evaluation metrics (RMSE and MAE) and the absence of a detailed benchmark comparison with other machine learning models. Additionally, while the FQGLS model addressed non-stationarity effectively, the study did not tackle multicollinearity challenges limiting the robustness of the model in complex environments.

Kim et al. [74] developed and compared two predictive models, multiple linear regression analysis (MRA) and autoregressive integrated moving averages (ARIMA), to assess their effectiveness in predicting carbon prices in the Korea Emissions Trading System (K-ETS) over short-term periods. For their analysis, they used search query volume data from the NAVER platform, reflecting market participant behaviour. The dataset included 21 search query variables related to carbon emissions, productivity, and stock performance, along with 37 weeks of Korean Allowance Unit (KAU) trading prices from July 2018 to March 2019. Feature selection was performed through correlation analysis, where only variables with a correlation coefficient greater than 0.5 were included in the models. The key contribution of this study is its demonstration that search query volume data can be an effective predic-

tor of carbon emission credit prices. The ARIMA model showed particularly strong predictive performance with low error rates, exceeding the MRA model in long-term forecasts. This finding offers practical implications for investors and policymakers in managing risks within the carbon market. However, the study has several limitations. The error rate of the MRA model increased over longer prediction periods, and ARIMA struggled to capture rapid price fluctuations. Furthermore, the study did not compare its results with more advanced machine learning models. Although the ARIMA model addressed non-stationarity using differencing, the study did not tackle multicollinearity, limiting its accuracy in more complex datasets.

Liu and Song [98] investigated the impact of uncertainty in output growth on the prediction of future carbon prices using the EZ climate model integrated with the Epstein-Zin utility function. The study aimed to address how uncertainty in population and per capita output growth rates influences the prediction of carbon prices represented by the marginal abatement cost. The data used for the model included global carbon emission data, with the key external factors being population growth rate and per capita output growth rate. However, the study lacks comparisons with other models and does not address the multicollinearity challenges. Although the use of dynamic modelling techniques, such as recursive dynamic programming, helped handle non-stationarity, this study's overall comprehensiveness in capturing key factors affecting carbon prices is limited.

Song et al. [131] developed a fuzzy stochastic model to predict carbon price volatility in China's carbon market, focusing on the influence of demand-related policies. The model simulates how different free allowance policies and corporate trading strategies impact carbon prices, particularly under varying market demand conditions. The study used historical trading data from the Shanghai Environment and Energy Exchange to analyse the effects of demand-related policies, such as free allowance allocation, corporate participation, and carbon auction mechanisms, on

the carbon price. Although no specific feature selection or preprocessing techniques were applied, the model effectively captures price volatility by incorporating a fuzzy stochastic differential equation with Poisson jump terms. The main contribution of this work is its ability to predict carbon price fluctuations driven by policy changes, making it valuable in policy-driven markets such as China's carbon market. However, the model has several limitations, including the absence of comparisons with other models and a limited applicability to mature carbon markets like the EU ETS. In addition, while the stochastic differential equations with Poisson jump processes addressed non-stationarity by handling sudden shifts in carbon prices, the study did not address multicollinearity, which may affect the accuracy of predictions in more complex datasets.

2.5.2 Regression-based methods

Unlike statistical models, regression models are not constrained by statistical assumptions and can effectively capture the intricate non-linear relationships present in the data [2]. Models using this approach can process large amounts of data from different sources and recognise complex patterns that humans may miss [30].

Alanazi et al. [7] proposed a machine learning-based solution to address the challenge of predicting carbon credit prices, which is influenced by multiple factors. The authors used the random forest (RF) model for this task, as it effectively handles complex datasets with non-linear patterns. The dataset used in their research is the RCPI-data-public dataset, where the authors focused on three columns: price, countries (with an emphasis on the United States) and GHG levels (carbon emissions). Although the paper emphasises the importance of high-quality data preprocessing, there are several limitations. No detailed preprocessing methods were mentioned, which could affect the data quality and the reliability of the model. Furthermore, no feature selection or reduction techniques were explicitly discussed. The authors used

only one evaluation metric, MAE, limiting the completeness of the model performance assessment. Furthermore, the study did not include a benchmark comparison with other models, making it difficult to evaluate the performance of the RF model relative to alternative approaches. The paper also did not address challenges such as non-stationarity and multicollinearity in carbon price data, which are crucial factors that could impact prediction accuracy in dynamic carbon markets. These limitations reduce the generalisability and robustness of the model.

Guðbrandsdóttir and Haraldsson [50] developed a multiple linear regression model to identify the key drivers of carbon credit price changes in the EU ETS, using UK energy market data and global equity indices. The dataset included various factors such as CER, energy prices, and weather conditions, with correlation analysis used for feature reduction. The study found that CERs are the only significant predictor of same-day EUA prices, while other factors such as UK power market data did not demonstrate a strong correlation. Although predictions for the same day using CER were accurate, correlations decreased significantly when the data lagged by one day, reducing the effectiveness of the model. No feature selection techniques beyond correlation analysis were used. The contribution of this work lies in its clear identification of CERs as the strongest same-day predictor of EUA prices, which provides insight into the dynamics of the carbon market. However, there are several limitations. The model only used one evaluation metric (R^2), which limits the robustness of the performance evaluation. No advanced preprocessing was applied and there was no comparison with other machine learning models, making it difficult to assess the relative performance of the approach. Moreover, the study did not handle key challenges such as non-stationarity and multicollinearity in carbon price data, which are essential considerations for improving prediction accuracy.

2.5.3 Neural Network-based methods

Cao et al. [20] developed a novel probabilistic model called the quantile temporal convolutional network (QTCN) to predict carbon price fluctuations, particularly addressing the challenges posed by uncertain factors in China's Hubei carbon market. The study used daily carbon spot prices from 2014 to 2022, with external factors including energy prices (coal, crude oil, natural gas), macroeconomic indicators (exchange rate, stock indices), international carbon markets (EU ETS and New Zealand ETS prices), environmental factors (air quality index, temperatures), uncertain factors (economic policy uncertainty, geopolitical risk), and public sentiment (Baidu search index and transaction volume). The authors showed that coal prices and EU carbon prices had the most significant influence on Hubei's carbon prices, while the contribution of the AQI was minor. They used Z-score normalisation for data preprocessing, grey relational analysis (GRA) for feature reduction, and permutation importance analysis to assess the importance of influencing factors. The QTCN model outperformed the other models in terms of prediction accuracy and profitability, offering more robust risk management and trading strategies for carbon market participants. The study addressed non-stationarity by applying differencing to the data, but did not explicitly address the challenges of multicollinearity, which could further improve the robustness of their predictions.

Xie et al. [154] proposed a text-based framework for predicting carbon prices using online news data and text mining to create a climate-related variable. The study focused on predicting carbon prices in the Hubei and Guangdong carbon markets in China using two machine learning models: long short-term memory (LSTM) and RF models. The data used included over 1.8 million Sina News headlines and 30,922 pieces from the China Meteorological Administration, combined with historical carbon prices, macroeconomic indicators, energy prices, environmental factors, and public sentiment measured by the Baidu search index for "carbon emissions

trading". The authors used the RF model for feature selection and their results showed that including the climate-related variable significantly improved prediction accuracy. Although the study made an important contribution by demonstrating how climate-related textual data can enhance carbon price predictions, it was limited to only two Chinese carbon markets, which may affect the generalisability of the results to other regions. Furthermore, the study did not account for non-stationarity and multicollinearity challenges that are often present in time-series data.

2.5.4 Hybrid methods

Hybrid methods, which combine two or more models, are gaining significant attention in the field of carbon credit price prediction. By integrating the strengths of different models, hybrid models are particularly well-suited to addressing the complexity of carbon markets, where the non-linear, non-stationarity and multicollinearity nature of prices presents unique challenges [54], [167], [171]. As research evolves, hybrid models are increasingly recognised as a promising approach to improving prediction outcomes in this domain.

Feng et al. [43] proposed a carbon price prediction model called GWO–XGBOOST–CEEMDAN, combining the grey wolf optimiser (GWO), extreme gradient boosting (XGBOOST), and complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN). This hybrid model addresses the limitations of single machine learning models in predicting carbon prices by improving accuracy and stability. The model was tested on carbon markets in Guangdong, Hubei, and Fujian to verify its effectiveness. The model selection process involved an RF model for feature selection and a partial autocorrelation function (PACF) to analyse historical carbon price data. The main external factors considered were Shanghai Stock Exchange Industrial Index, Dow Jones Industrial Average, S&P 500 Index, CSI300 Index, the price of Brent crude oil, Bohai Sea Power coal, natural gas, EUA, and

AQI. The study effectively addressed the non-stationarity nature of carbon price data through the CEEMDAN method but did not focus on handling multicollinearity in the dataset.

Hao and Tian [54] developed a hybrid framework for predicting carbon prices that addresses the challenges of nonlinearity, non-stationarity, and multifrequency in the Hubei and Shenzhen markets in China. The framework combines the improved complementary ensemble empirical mode decomposition with Adaptive Noise (ICEEMDAN) for data preprocessing to remove noise, and kernel-based extreme learning machine (KELM) optimised by a multi-objective chaotic sine cosine algorithm (MOCSCA) for price prediction. The authors included the price of coal, crude oil, gas, electricity, the US dollar exchange rate against RMB, international carbon market prices (EUA and CER) and AQI as external factors in their model. In addition, the Baidu search index was used as a proxy for public interest in carbon-related topics. The framework applies feature selection using PACF and max-relevance min-redundancy (mRMR) to select optimal input features. Despite its superior prediction accuracy, the study did not explicitly address multicollinearity, which could impact the effectiveness of the selected features. Although mRMR was used to reduce redundancy between features and partially address multicollinearity, it may not have fully eliminated the issue.

Mu et al. [108] proposed a novel predictive model called MS-IHHO-LSTM, which integrates multiple data sources, the improved Harris-Hawks optimisation (HHO) algorithm, and the LSTM model. Data sources include both technical indicators and sentiment analysis from news related to the carbon market. Technical indicators include carbon market prices from Hubei, Shanghai, and Shenzhen and emotional values derived from sentiment analysis. The study addressed the non-stationarity nature of carbon price data using the ICEEMDAN technique, which decomposes the data into IMFs to handle non-stationarity. The HHO algorithm was enhanced

by adding inertia weight and optimising LSTM hyperparameters, which allowed the model to avoid local optima and improve performance. This study highlighted the importance of incorporating emotional values to capture market sentiments. The limitations of the study include the increased computational cost due to swarm intelligence optimisation, which, while improving accuracy, also added to the model's complexity and runtime. Furthermore, the study did not explicitly address multicollinearity issues in the dataset.

Sun and Wang [134] proposed a carbon price prediction model based on a combination of EMD, least squares support vector machine (LSSVM), and improved particle swarm optimisation (IPSO). The external factors used in the model are EUA and CER futures, the price of Brent crude oil, natural gas, coal, USD/CNY and EUR/CNY exchange rates, Shanghai Industrial Index, SH300, and temperature changes in various regions of China. Factor analysis was applied to reduce the dimensionality of the input features, while PACF was used to analyse historical carbon price data. The IPSO method was used to optimise the LSSVM parameters. The model demonstrates improved robustness and prediction accuracy over alternative models, validated with data from China's Hubei, Shenzhen, and Beijing carbon markets. The framework was designed to address the non-linearity and non-stationarity in carbon price data by decomposing the data into IMFs using EMD. However, the study did not explicitly address multicollinearity in the data.

Wang et al. [144] developed a carbon price prediction model that integrates multi-source information fusion (MSIF) and a hybrid multi-scale decomposition (HMSD) method. The model was applied to China's Hubei carbon market, and various external factors namely energy prices (coal, oil, natural gas, electricity), macroeconomic indicators, international carbon market prices (EU ETS), and unstructured data such as the Baidu search index and social media sentiment were used. The use of HMSD allowed for the decomposition of complex carbon price signals into more

predictable components. Feature selection was handled using Spearman correlation analysis, and dictionary-based sentiment analysis was used for unstructured data processing. Although the model successfully demonstrated high prediction accuracy and addressed non-stationarity issues through its hybrid decomposition techniques, it did not explicitly tackle multicollinearity issues, which may affect feature interpretation.

Wang et al. [145] proposed a hybrid prediction model for carbon price prediction, integrating LightGBM, bald eagle search (BES), and bidirectional long short-term memory (BiLSTM), based on environmental impact factors. The model aimed to address the nonlinear, non-stationarity, and multi-frequency characteristics of carbon credit prices. The environmental factors used are AQI, PM2.5, PM10, sulfur dioxide (SO2), carbon monoxide (CO), ozone (O3) and nitrogen dioxide (NO2). The proposed framework begins with symplectic geometry mode decomposition (SGMD), which decomposes the carbon price data into components that capture periodic and frequency features. Then, LightGBM was used to extract the most important environmental factors affecting carbon prices. The top three selected features (AQI, PM2.5, and PM10) were then fused with the decomposed carbon price components. To improve prediction accuracy, BiLSTM was optimised with the BES algorithm. The model was validated on carbon trading data from the Shanghai Carbon Emission Exchange, covering the period from October 2016 to October 2020. Despite its effectiveness, the study notes that the complexity of integrating multiple models requires careful orchestration, which may increase computational cost and complexity. In this study, while environmental factors were selected based on their importance, it is unclear whether the multicollinearity between these factors was assessed or mitigated during the modelling process.

Xu et al. [155] proposed a novel hybrid prediction framework for carbon price prediction that integrates data decomposition, feature dimension reduction, and

advanced machine learning models. The framework addresses the complex characteristics of carbon prices, such as nonlinearity and non-stationarity using variational mode decomposition (VMD) to decompose the carbon price series into simpler and smoother components. These components are further refined through a stacked autoencoder (SAE) to remove noise and extract the key features. The model also incorporates various external influencing factors to enhance prediction accuracy. These factors are EUA, crude oil, natural gas, coke, steam coal and coking coal prices, the USD/CNY exchange rate, the China Industrial Index, AQI, PM2.5, PM10, sulfur dioxide (SO2), carbon monoxide (CO), ozone (O3), nitrogen dioxide (NO2). In addition, unstructured data from the Baidu web search index of carbon-related keywords were used. A two-stage feature dimension reduction method, combining R for feature selection and SAE for feature extraction, was used to manage the complexity and redundancy in these exogenous variables. In the prediction phase, the BiLSTM model, optimised by the cuckoo search (CS) algorithm, was applied to predict the carbon price based on the reconstructed data and the selected features. To quantify uncertainty, Gaussian process regression (GPR) with a hybrid kernel function was used for interval prediction. The model was tested on seven real carbon trading pilot datasets in China. Despite its effectiveness, the paper notes that the complex architecture increases computational costs and does not explicitly address multicollinearity in the data.

Zhao et al. [167] developed a hybrid prediction model named variable grey model with extreme learning machine (VGM-ELM), designed to predict carbon prices. The model addressed the non-linear, volatile nature of carbon price data using the Hodrick-Prescott (HP) filter for decomposition and an improved grey model (GM) for long-term trends. The extreme learning machine (ELM) was used for short-term fluctuations. The model incorporates external factors, including Brent crude oil prices, S&P Clean Energy Index, the STOXX-50 Index, the yields of the US 3-month

T-bill, FTSE100, the European ARA Ports coal price and IPE natural gas prices, all of which significantly influence the price of carbon credits. The input data underwent a two-stage feature selection process using PACF and LASSO regression to determine the key factors. The model was tested using EUA futures data and the empirical analysis demonstrated that the proposed model outperforms 10 benchmark models. However, the complexity of integrating multiple models increases the computational cost of the proposed model.

Zhou and Wang [171] proposed a hybrid carbon price prediction model based on a secondary decomposition algorithm and influencing factors. The model uses EMD and VMD to break down carbon price data into more regular components, while the sparrow search algorithm (SSA) optimises the parameters of the kernel extreme learning machine (KELM) to improve prediction performance. The model integrates both structural and non-structural influencing factors. The structural factors are EUAs, CER futures, Rotterdam coal, Brent crude oil and natural gas prices, and the USD/RMB exchange rate. Non-structural factors were derived from the Baidu search index, including keywords such as "Paris Agreement", "Low Carbon", "Kyoto Agreement", "Energy" and "Global Warming". The key factors were selected using the mRMR algorithm. The model demonstrated superior performance in predicting carbon prices in the Guangdong and Hubei markets, outperforming other benchmark models. However, the paper does not explicitly address potential multicollinearity issues among the selected factors, which may impact the stability of the predictions.

Wang et al. [147] proposed a CEEMDAN-PCA-DARNN framework for short-term carbon price prediction, focusing on China's Hubei carbon market. The proposed framework combines CEEMDAN, principal component analysis (PCA) and dual-stage attention-based recurrent neural network (DARNN). The CEEMDAN method was used to handle the non-stationarity nature of carbon price data, which breaks down the time series into more stationary components, capturing underlying

trends and reducing noise. In addition, the use of PCA for dimensionality reduction addresses multicollinearity by transforming correlated variables into principal components. DARNN was then used for predictive modelling, employing feature attention and temporal attention mechanisms for enhanced accuracy. The model includes 23 external factors, such as energy prices, macroeconomic indices, and environmental variables, which were validated through GRA to determine their relevance to the dynamics of carbon prices. This approach demonstrates improved prediction accuracy over traditional models, highlighting the effectiveness of attention mechanisms in carbon market analysis.

Wei and Ouyang [149] presented a scaled principal component analysis model (s-PCA) for carbon price prediction combined with linear regression and LSTM, s-PCA outperformed standard PCA and PLS in accuracy and efficiency, achieving higher returns and Sharpe ratios in market timing predictions. The model was tested on carbon price data from the Hubei carbon market. The external factors used in the model include technical indicators such as momentum rules, moving averages and oscillators, financial indicators such as the 3-month T-bill rate, S&P 500 index, Dow Jones Composite Index, the Shanghai Composite Index and the USD/CNY exchange rate, and commodity indicators such as ICE-Brent crude oil futures, ICE-coal Rotterdam futures, and S&P GSCI non-energy commodity indexes. The s-PCA approach prioritises these predictors according to their relevance to the target variable, effectively addressing multicollinearity. Although the model's use of s-PCA and LSTM improves accuracy, the study highlights that this combination increases computational complexity.

Xiao and Liu [153] proposed a multifrequency data fusion deep learning model for carbon price prediction (MFF-CPPM), which integrates various data frequencies and advanced deep learning techniques to improve the accuracy of carbon price forecasts. The model consists of three components: a feature-extraction frontend, a

multifrequency data fusion transformer, and a fusion regression layer. The feature extraction uses a combination of one-dimensional convolutional layers, bidirectional gated recurrent units (BGRU), and a ResNet-18 layer for precise feature identification at different data frequencies. The multifrequency transformer applies an attention mechanism to align and fuse data from different frequencies, ensuring that critical information from high- and low-frequency data is captured. The key external factors used in the model are EUA, CER, Brent crude oil, IPE Rotterdam coal and natural gas prices, the CSI 300 index, the S&P 500 index, and the USD/RMB exchange rate. The model accounts for different data frequencies (daily, weekly, monthly) in the input data. The MFF-CPPM model was tested on carbon trading data from China's carbon trading pilot markets, including Guangdong, Hubei, and Beijing, and it demonstrated superior performance in predicting carbon prices compared to baseline models. The model effectively addressed the non-stationarity and non-linearity in the data through the use of data fusion and attention mechanism techniques. However, the study acknowledges the complexity of integrating data at different frequencies, which increases the computational cost of the model.

Zheng et al. [169] proposed a multi-factor hybrid model for carbon price prediction, designed within a decomposition-integration framework to improve both the accuracy of the point and the interval prediction. The model employs successive variational mode decomposition (SVMD) to decompose carbon price data into simpler sequences, followed by categorising them using the sample entropy (SE) method into low- and high-frequency components. The low-frequency sequences were predicted using the CatBoost model for long-term trends, while the high-frequency sequences were handled using KELM, optimised by the SSA algorithm. The key external factors considered in the model are the Brent crude oil futures, EUA price, and the Hushen 300 index. These factors were selected through a combination of PACF and RF models. The model effectively addresses the non-stationarity na-

ture of carbon price data by using SVMD to decompose the data into components. The model demonstrates strong predictive performance, particularly when tested on data from the Hubei carbon market. The use of Shapley additive explanations (SHAP) increases the interpretability of the model by revealing that historical carbon prices, Brent crude oil futures, and EUA futures positively influence carbon price predictions, while the Hushen 300 index has a negative effect. While RF and PACF methods are used to select the most relevant factors, an additional technique to directly manage multicollinearity, such as PCA, could further enhance the model's robustness.

However, despite its widespread use for dimensionality reduction, PCA is unsuitable for feature selection because it does not identify the most relevant features. PCA is fundamentally a dimensionality reduction technique, not a feature selection method. It transforms the original variables into a new coordinate system, where all variables contribute to the projections in the lower-dimensional space. However, this approach has three main limitations: (1) it considers only linear relationships between variables, (2) it does not account for higher-order interactions or the multivariate structure of the data, and (3) it assumes that features with the highest variance are the most important, which may not align with the problem's objectives. Although PCA reduces feature space, it does not identify the most relevant features, making it an unsuitable choice for feature selection.

Tables 2.2, 2.3, 2.4, 2.5, 2.6, and 2.7 provide a detailed overview of the selected studies, including the study market, external factors, applied methods, and the techniques used. Overall, the tables reveal trends in the literature: some methods are more frequently applied than others, certain external factors are commonly considered across studies, and there are variations in the approaches used for prediction and analysis. These key trends will be further discussed in Section 2.6, where a detailed criteria-based analysis is presented.

Table 2.2 : Comparison of relevant articles on statistical methods for multi-factor analysis.

ID	**Article**	**Studied market**	**External factors**	**Factor analysis technique**	**Factor selection technique**	**Prediction technique**
AS1	Creti et al. [31]	EUA ETS	Brent crude oil, natural gas, coal, Dow Jones Euro STOXX 50 and a dummy variable to account for the structural break in 2006.	Cointegration analysis	NO	NO
AS2	Feng et al. [44]	EUA ETS	Market mechanisms (supply-demand disequilibrium, allowance changes, power demand), heterogeneous environment (international negotiations, political events and financial crises) and temperature.	EEMD and variance ratio analysis	NO	NO

Table 2.2 : Comparison of relevant articles on statistical methods for multi-factor analysis (continued).

ID	**Article**	**Studied market**	**External factors**	**Factor analysis technique**	**Factor selection technique**	**Prediction technique**
AS3	Hammoudeh et al. [53]	EUA ETS	Crude oil prices (West Texas Intermediate), natural gas prices (Henry Hub natural gas), coal prices (Coal Intercontinental Exchange API2) and electricity prices (South Path 15 Firm Peak electricity).	Quantile regression framework	NO	NO
AS4	Ji et al. [64]	China pilot ETS	Product prices (iron ore, cement, non-ferrous metals, chemical products), energy prices (Daqing crude oil, Qinhuangdao power coal), extreme high/low temperatures, water levels and Hushen 300 index.	ARDL	NO	NO

Table 2.2 : Comparison of relevant articles on statistical methods for multi-factor analysis (continued).

ID	**Article**	**Studied market**	**External factors**	**Factor analysis technique**	**Factor selection technique**	**Prediction technique**
AS5	Lin and Jia [93]	China pilot ETS	Industry coverage (chemicals, cement, electricity, transportation), allowance allocation, free allowance rate, penalty rate and the CDM.	CGE	NO	NO
AS6	Luo et al. [104]	China pilot ETS	Henry Hub Natural Gas Spot Price and crude oil Price.	Wavelet analysis	NO	NO
AS7	Li et al. [91]	EUA ETS	Natural gas price, crude oil price, coal price, bond market indicators (T-BILL, JUNKBOND, ESTB, ELTB), stock market indices (S&P 500, FTSE 100, DAX, CAC 40), commodity market (CRB index) and exchange rates (EUR/RMB, EUR/USD).	MSVAR	LASSO	NO

Table 2.2 : Comparison of relevant articles on statistical methods for multi-factor analysis (continued).

ID	**Article**	**Studied market**	**External factors**	**Factor analysis technique**	**Factor selection technique**	**Prediction technique**
AS8	Song et al. [130]	China pilot ETS	Policy factors (government free carbon quotas, government carbon quotas, Government carbon threshold, Government subsidies for energy saving and emission reduction), green technology maturity, development degree of low-carbon industries, clean energy penetration, coal, crude oil and natural gas prices, USD/CNY exchange rate, industrial development level, Shanghai composite index, AQI and air temperature.	GRAM	Systematic Cluster Analysis	NO

Table 2.2 : Comparison of relevant articles on statistical methods for multi-factor analysis (continued).

ID	**Article**	**Studied market**	**External factors**	**Factor analysis technique**	**Factor selection technique**	**Prediction technique**
AS9	Tan and Wang [136]	EUA ETS	Brent crude oil, natural gas (Henry Hub), coal (Australia BJ Power Coal Spot Prices Index), Dow Jones Euro STOXX 50 Index, CRB Index, Treasury bill yield (T-BILL), and Moody's junk bond yield.	Quantile regression and structural breaks analysis	NO	NO
AS10	Tang et al. [137]	EUA and sCER	Market mechanisms (supply-demand dynamics and electricity price), external environment (international climate negotiations and policy changes), and seasonal temperature variations.	EEMD and variance ratio analysis	NO	NO

Table 2.2 : Comparison of relevant articles on statistical methods for multi-factor analysis (continued).

ID	**Article**	**Studied market**	**External factors**	**Factor analysis technique**	**Factor selection technique**	**Prediction technique**
AS11	Zeng et al. [161]	China pilot ETS	Shanghai Stock Exchange Industrial Index (SSEII), Bohai-Rim Steam-Coal Price Index (BSPI), natural gas index, AQI.	Dynamic Heterogeneous Panel PMG	NO	NO

Table 2.3 : Comparison of relevant articles on time series methods for multi-factor analysis.

ID	**Article**	**Studied market**	**External factors**	**Factor analysis technique**	**Factor selection technique**	**Prediction technique**
AT1	Aatola et al. [1]	EUA ETS	German electricity prices, gas prices (UK gas), coal prices, mineral product prices, steel product prices, paper product prices, water reservoir levels, gas storage levels (UK), Brent crude oil prices, FTSE 350 stock market index.	OLS, IV and VAR	NO	NO
AT2	Alberola et al. [8]	EUA ETS	Brent crude oil prices, natural gas prices (Zeebrugge Hub), coal prices (CIF ARA), electricity prices (Powernext), clean dark spread, clean spark spread, switch price, and temperature variables (including extreme temperature deviations).	VAR	NO	NO

Table 2.3 : Comparison of relevant articles on time series methods for multi-factor analysis (continued).

ID	**Article**	**Studied market**	**External factors**	**Factor analysis technique**	**Factor selection technique**	**Prediction technique**
AT3	Chevallier [24]	EUA ETS	Brent crude oil, natural gas and coal prices and aggregated EU industrial production index.	MS-VAR	NO	NO
AT4	Gavard and Kirat [45]	EUA ETS	natural gas prices, coal prices, the Euro STOXX 50 index, policy changes and market shifts.	ARCH, GARCH and GARCH-M	NO	NO
AT5	Guan [49]	EUA ETS	Thermal coal, WTI crude oil and natural gas prices and Euro STOXX 50 index.	VEC	NO	NO
AT6	Jiang et al. [68]	China pilot ETS	Qinhuangdao coal price, oil price and Shanghai Composite Index	VAR	NO	NO
AT7	Kim and Koo [75]	US	Coal price, crude oil price (WTI), natural gas price (Henry Hub), and temperature (hot and cold weather dummies)	ARDL	NO	NO

Table 2.3 : Comparison of relevant articles on time series methods for multi-factor analysis (continued).

ID	**Article**	**Studied market**	**External factors**	**Factor analysis technique**	**Factor selection technique**	**Prediction technique**
AT8	Kotishwar [81]	India	Population growth rate and per capita output growth rate.	Bivariate correlation, OLS and VAR	NO	NO
AT9	Li and Lei [88]	China pilot ETS	Industrial income, the relative price of natural gas to coal, maximum daily temperature, daily rainfall, government intervention, and the number of carbon quota management enterprises.	ARCH	NO	NO
AT10	Li et al. [89]	China pilot ETS	economic policy uncertainty index and green bond index	TVP-VAR	NO	NO

Table 2.3 : Comparison of relevant articles on time series methods for multi-factor analysis (continued).

ID	**Article**	**Studied market**	**External factors**	**Factor analysis technique**	**Factor selection technique**	**Prediction technique**
AT11	Lovcha et al. [100]	EUA ETS	The price of Certified Emission Reductions (CERs) and 21 search query volume data from NAVER includes company stocks related to carbon emissions (e.g., FOOSUNG, HUCHEMS, HOMEDECO), productivity-related search terms, emission-related search terms (e.g., CO2 emissions, global warming).	Structural SVR	NO	NO
AT12	Wang et al. [146]	China pilot ETS	Thermal coal prices, natural gas prices, Shanghai Stock Exchange Index, Shanghai Industrial Index and EU emission prices.	OLS, LSDV, SYS-GMM and mediating effect	NO	NO

Table 2.3 : Comparison of relevant articles on time series methods for multi-factor analysis (continued).

ID	**Article**	**Studied market**	**External factors**	**Factor analysis technique**	**Factor selection technique**	**Prediction technique**
AT13	Wei, Qi et al. [150]	China pilot ETS	Gross industrial output, Coal consumption, Science and technology innovation index and Extreme weather conditions.	Multiple regression model	NO	NO
AT14	Wu et al. [152]	China pilot ETS	Crude oil, coal, natural gas, the Goldman Sachs Commodity Index, STOXX Europe 600 Index, EUR/USD exchange rate and mobility trends in the COVID-19 period.	N-A MEMD, VAR and VEC	NO	NO
AT15	Yu and Mallory [158]	EUA ETS	Natural gas and coal prices and the exchange rate (Euro/USD, Euro/Swiss Franc).	SVAR	NO	NO
AT16	Zeng et al. [160]	China pilot ETS	EUAs, CERs, the US dollar exchange, crude oil, natural gas and coal prices.	VAR and VECM	NO	NO

Table 2.3 : Comparison of relevant articles on time series methods for multi-factor analysis (continued).

ID	**Article**	**Studied market**	**External factors**	**Factor analysis technique**	**Factor selection technique**	**Prediction technique**
AT17	Zhang et al. [162]	sCER	International carbon-reduction policies, S&P Global 1200 Index, coal, oil, natural gas and electricity prices, the issuance volume of CERs, and the pricing power of developing countries.	EGARCH, PARCH, TGARCH and VAR	NO	NO
AT18	Zhang et al. [165]	China pilot ETS	Natural gas, coal and Brent crude oil prices, CER and EUA futures prices, CSI 300 index and Industry index.	ARMA-GARCH-vine copula	NO	NO
AT19	Zhou and Li [173]	China pilot ETS	Coal prices, Shanghai and Shenzhen 300 Index (HS300), Shanghai Industrial Index and AQI.	VAR and VEC and GARCH	NO	NO

Table 2.3 : Comparison of relevant articles on time series methods for multi-factor analysis (continued).

ID	**Article**	**Studied market**	**External factors**	**Factor analysis technique**	**Factor selection technique**	**Prediction technique**
AT20	Zhu et al. [175]	EUA ETS	Oil, coal, natural gas and electricity prices, the STOXX Europe 600 Index, and the Goldman Sachs Commodity Index.	MEMD, Cointegration Test, Error Correction, and Newey–West Estimator	NO	NO

Table 2.4 : Comparison of relevant articles on statistical methods to predict the price of carbon credit.

ID	**Article**	**Studied market**	**External factors**	**Factor analysis technique**	**Factor selection technique**	**Prediction technique**
PS1	Adekoya [4]	EUA ETS	Brent crude oil, natural gas and coal prices and EUA.	Quantile regression technique	NO	FQGLS estimator
PS2	Kim et al. [74]	Korea ETS	CERs, and 21 search query volume data from NAVER.	ARIMA	Correlation analysis	MRA
PS3	Liu and Song [98]	EUA ETS	Population growth rate and per capita output growth rate.	Epstein-Zin utility function	NO	IAM and EZ climate
PS4	Song et al. [131]	China pilot ETS	Demand-related policies, corporate participation, trading volume and carbon allowance allocation schemes.	Fuzzy stochastic differential	NO	Fuzzy stochastic differential

Table 2.5 : Comparison of relevant articles on regression-based methods to predict the price of carbon credit.

ID	**Article**	**Studied market**	**External factors**	**Factor analysis technique**	**Factor selection technique**	**Prediction technique**
PR1	Alanazi et al. [7]	United States	CO2 emission level, new regulations	NO	NO	RF
PR2	Guðbrandsdóttir and Haraldsson [50]	EUA ETS	EUA Spot, CER, coal, WTI and Brent crude oil, gas oil and natural gas futures, electricity, aluminum, dark spread, spark spread, switching price, weather, Nasdaq 100 Index, Standard & Poor's 500 Index, FTSE 100 Index, Deutscher Aktien Index, Amsterdam Stock Exchange, Portuguese Stock Index, Exchange rate (USD/GBP, USD/EUR, EUR/GBP).	Linear regression	Correlation analysis	Multiple linear regression

Table 2.6 : Comparison of relevant articles on neural network-based methods to predict the price of carbon credit.

ID	**Article**	**Studied market**	**External factors**	**Factor analysis technique**	**Factor selection technique**	**Prediction technique**
PN1	Cao et al. [20]	China pilot ETS	Demand-related policies, corporate participation, trading volume and carbon allowance allocation schemes	Grey relational analysis	Grey systems theory	TCN with quantile regression
PN2	Xie et al. [154]	China pilot ETS	EUA and CER futures prices, Brent and WTI crude oil and coal, domestic energy prices, air temperature, air quality index, exchange rate of (USD/CNY and EUR/CNY, SSEC), Baidu search index for "carbon emissions trading" and News headlines from Sina News and China Meteorological Administration	NO	NO	LSTM and RF

Table 2.7 : Comparison of relevant articles on hybrid methods to predict the price of carbon credit.

ID	**Article**	**Studied market**	**External factors**	**Factor analysis technique**	**Factor selection technique**	**Prediction technique**
PH1	Feng et al. [43]	China pilot ETS	Shanghai Stock Exchange Industrial Index, Dow Jones Industrial Average, S&P 500 Index, CSI300 Index, Brent crude oil, coal and natural gas prices, EUA price, and AQI.	PACF	RF	GWO–XGBOOST–CEEMDAN combines GWO for hyperparameter optimisation, CEEMDAN for decomposition and XGBOOST for prediction
PH2	Hao and Tian [54]	China pilot ETS	Coal, crude oil, natural gas and electricity prices, exchange rate of USD/RMB, EUA price, CER price, and AQI.	Multi objective chaotic sine cosine algorithm	PACF and mRMR	ICEEMDAN-MOCSCA-KELM combines ICEEMDAN for decomposition, MOCSCA for optimisation and KELM for prediction

Table 2.7 : Comparison of relevant articles on hybrid methods to predict the price of carbon credit (continued).

ID	**Article**	**Studied market**	**External factors**	**Factor analysis technique**	**Factor selection technique**	**Prediction technique**
PH3	Mu et al. [108]	China pilot ETS	Sentiment analysis based on news articles related to carbon markets.	NO	NO	MS-IHHO-LSTM combines IHHO for hyperparameter optimisation and LSTM for prediction

Table 2.7 : Comparison of relevant articles on hybrid methods to predict the price of carbon credit (continued).

ID	**Article**	**Studied market**	**External factors**	**Factor analysis technique**	**Factor selection technique**	**Prediction technique**
PH4	Sun and Wang [134]	China pilot ETS	CER price and 21 search query volume data from NAVER includes company stocks related to carbon emissions (FOOSUNG, HUCHEMS, HOMEDECO), productivity-related search terms, emission-related search terms (CO2 emissions, global warming).	Factor analysis	KMO and Bartlett statistical tests	EMD-IPSO-LSSVM combines EMD for decomposition, IPSO fo optimisation, LSSVM for prediction

Table 2.7 : Comparison of relevant articles on hybrid methods to predict the price of carbon credit (continued).

ID	**Article**	**Studied market**	**External factors**	**Factor analysis technique**	**Factor selection technique**	**Prediction technique**
PH5	Wang et al. [144]	China pilot ETS	Coal and oil prices, and unstructured data (Baidu index, social media sentiment).	BosonNLP	Spearman correlation	DWT, EMD, SSA, and VMD are used to extract the main features of MSIF, HMSD for decomposition, Holt's linear trend model, ARIMA, SVR, BPNN, and LSTM for prediction

Table 2.7 : Comparison of relevant articles on hybrid methods to predict the price of carbon credit (continued).

ID	**Article**	**Studied market**	**External factors**	**Factor analysis technique**	**Factor selection technique**	**Prediction technique**
PH6	Wang et al. [145]	China pilot ETS	AQI, PM2.5, PM10, SO2, CO, O3, NO2, and temperature.	NO	LightGBM	LightGBM-BES-BiLSTM combines SGMD for decomposition, LightGBM for feature extraction, BES for hyperparameter optimisation, and BiLSTM

Table 2.7 : Comparison of relevant articles on hybrid methods to predict the price of carbon credit (continued).

ID	**Article**	**Studied market**	**External factors**	**Factor analysis technique**	**Factor selection technique**	**Prediction technique**
PH7	Xu et al. [155]	China pilot ETS	EUA price, crude oil prices, natural gas prices, coke prices, steam coal prices, coking coal prices, the USD/CNY exchange rate, the China Industrial Index, AQI, PM2.5, PM10, SO2, CO, and NO2.	RF	RF and SAE	VMD-SAE-CS-BiLSTM-GPR combining VMD for decomposition, SAE for feature extraction and CS for optimizing the parameters of BiLSTM, and GPR
PH8	Zhao et al. [167]	EUA ETS	Brent crude oil, coal, and natural gas prices, S&P Clean Energy Index, STOXX-50 Index, T-bill yields, FTSE100.	PACF and LASSO	PACF and LASSO	VGM-ELM combine the HP filter for decomposition, VGM to predict long-term trends and ELM for short-term prediction.

Table 2.7 : Comparison of relevant articles on hybrid methods to predict the price of carbon credit (continued).

ID	**Article**	**Studied market**	**External factors**	**Factor analysis technique**	**Factor selection technique**	**Prediction technique**
PH9	Zhou and Wang [171]	China pilot ETS	EUA price, CER price, Rotterdam coal prices, Brent crude oil, natural gas and electricity prices, exchange rate of USD/RMB, AQI and Baidu search index.	mRMR	mRMR	EMD-VMD-SSA-KELM combines EMD and VMD for decomposition, SSA for hyperparameter optimisation, and KELM

Table 2.7 : Comparison of relevant articles on hybrid methods to predict the price of carbon credit (continued).

ID	**Article**	**Studied market**	**External factors**	**Factor analysis technique**	**Factor selection technique**	**Prediction technique**
PH10	Wang et al. [147]	China pilot ETS	Temperature, AQI, SSE 50 Index, Composite Index, Gold and Silver Spot Price, CSI 300 Index, USD/RMB Exchange Rate, CBOE China ETF Volatility Index, Power Index, CITIC Industry Index, Brent crude oil, coal, gas and EUA price.	GRA	NO	CEEMDAN-PCA-DARNN combining CEEMDAN for decomposition, PCA for dimensionality reduction, and DARNN for predicting

Table 2.7 : Comparison of relevant articles on hybrid methods to predict the price of carbon credit (continued).

ID	**Article**	**Studied market**	**External factors**	**Factor analysis technique**	**Factor selection technique**	**Prediction technique**
PH11	Wei and Ouyang [149]	China pilot ETS	T-bill, national bond rate, S&P 500 index, Composite Index, Bond Index Yield, Energy Global Innovation Index, Clean Energy Index, bond spreads, government bond, USD/CNY and Policy Uncertainty Index), commodity indicators (natural gas, coal, Brent crude oil and electricity prices) and EUA price.	NO	NO	s-PCA-LSTM combines s-PCA for dimensionality reduction and LSTM for prediction

Table 2.7 : Comparison of relevant articles on hybrid methods to predict the price of carbon credit (continued).

ID	**Article**	**Studied market**	**External factors**	**Factor analysis technique**	**Factor selection technique**	**Prediction technique**
PH12	Xiao and Liu [153]	China pilot ETS	EUA, CER, oil and coal prices, CSI 300 index, the S&P 500 index, and pilot carbon exchanges.	NO	BGRU and esNet-18	Mixed-frequency attention mechanism and BGRU
PH13	Zheng et al. [169]	China pilot ETS	EUA, coal, Brent crude oil and natural gas prices, USD/RMB exchange rate, Hushen 300 index, AQI.	NO	RF and PACF	SVMD-SSA-CatBoost/KELM combines SVMD for adaptive decomposition, SSA for hyperparameter optimisation, and a combination of CatBoost for long-term trend and KELM for short-term predictions

2.6 Relevant studies analysis

This section analyses the selected research papers based on the predefined analysis criteria questions presented in Table 2.8.

Table 2.8 : Analysis criteria questions.

Criteria	**Criteria Question**
C1	Which specific carbon market(s) does the paper study?
C2	Does the paper consider external factors influencing carbon credit prices?
C3	Does the paper propose a solution to predict carbon credit prices considering the influencing factors?
C4	Are the experiments and experimental setup clearly described in the paper?
C5	Does the paper provide clear findings with justifiable conclusions and results?
C6	Does the paper analyse the effect of external influencing factors on carbon credit prices?
C7	Does the paper apply feature selection techniques to the influencing factors?
C8	Does the paper take into account the non-linearity nature of the carbon data in its solution?
C9	Does the paper take into account the non-stationarity nature of the carbon data in its solution?
C10	Does the paper deal with multicollinearity issues in the dataset?
C11	Does the paper use real-world databases for the experiments?
C12	Does the paper involve multiple input variables in its analysis or model?
C13	Does the paper consider multiple periodicities in the data?
C14	Does the paper address multi-horizon prediction?
C15	Is model validation clearly described and implemented in the paper?
C16	Does the paper compare the proposed solution to benchmarks?
C17	Does the paper use three or more evaluation metrics to assess the performance of the model?
C18	Does the paper propose a trading strategy based on the prediction model?

The selection of these analysis criteria (C1 to C18) was motivated by the need to identify the research gaps in the literature. The research papers were evaluated against these selected criteria, where a score of 1 indicates that the study fulfilled a particular criterion, while a score of 0 indicates that the study did not meet that criterion. Table 2.9 summarises the analysis of the relevant articles according to the defined criteria. Each row represents a specific study, and the columns describe various aspects of the study evaluation. The meaning of each criteria is as follows:

- **C1: Studied Market:** This item identifies which carbon market(s) the paper studies as a dependent variable (for example, China, EU ETS, US).
- **C2: Considering External Influencing Factors:** a binary value which indicates whether the article considers external factors that affect the price of carbon credits.
- **C3: Predictive Solution:** The binary value which indicates whether the paper proposes a solution for predicting carbon credit prices, considering the external influencing factors. A value of 0 is assigned if the study's objectives do not explicitly focus on predicting carbon credit prices, suggesting that the model developed in the study is not designed for this purpose. Conversely, a value of 1 is assigned if the paper explicitly aims to predict carbon credit prices based on external influencing factors. For example, if a study analyses factors affecting carbon markets without directly proposing a price prediction model, a value of 0 is assigned.
- **C4: Experimental Setup:** a binary value which indicates whether the experiments and the experimental setup are clearly described in the paper.
- **C5: Clear Findings:** a binary value which indicates whether the paper provided clear findings with justifiable conclusions and results.

- **C6: Analysis of Influencing Factors:** a binary value which indicates whether the paper analyses the effect of external influencing factors. A detailed analysis of these factors is important, as it provides the basis for accurate predictions and informed trading decisions.
- **C7: Feature Selection:** a binary value which indicates whether the paper applies feature extraction techniques. Feature reduction is essential in prediction models to avoid overfitting and ensure that only the most relevant factors are included in the models.
- **C8: Non-linearity:** a binary value which indicates whether the paper considers the non-linearity nature of carbon credit prices in its analysis or models.
- **C9: Non-stationarity:** a binary value which indicates whether the paper considers the non-stationarity nature of carbon credit prices.
- **C10: Multicollinearity:** a binary value which indicates whether the paper addresses multicollinearity issues in the dataset.
- **C11: Real Database:** a binary value which indicates whether the paper uses real-world databases for its experiments.
- **C12: Multiple Input:** a binary value which indicates whether the model can handle multiple input variables.
- **C13: Multiple Periodicity:** a binary value which indicates whether the paper considers multiple periodicities in the data (e.g., daily, weekly, monthly data).
- **C14: Multi-Horizon Prediction:** a binary value which indicates whether the paper can make predictions for multiple time horizons where predictions are made for different time horizons, such as short-term and long-term predictions.

- **C15: Model Validation:** a binary value which indicates whether the paper includes a proper validation of the model. Validating a prediction model means evaluating its performance with data that were not used in training the model.

- **C16: Benchmark Comparison:** a binary value which indicates whether the paper compares its proposed solution with existing benchmarks or baseline models.

- **C17: Evaluation Metrics ≥ 3:** a binary value which indicates whether the paper uses three or more evaluation metrics to assess the performance of the model. This criterion ensures that the performance of the model is evaluated using at least three different metrics. Common metrics in our review include mean absolute percentage error (MAPE), root mean squared error (RMSE), and R-squared. These metrics provide a comprehensive assessment of the model's accuracy and reliability. By using multiple evaluation metrics, we ensure a well-rounded evaluation of the model's performance, rather than relying on a single measure.

- **C18: Trading Strategy:** a binary value which indicates whether the paper proposes a method to actively use the carbon credit price predictions to make trading decisions in carbon markets. A trading strategy typically involves rules and methods designed to assist the decision as to when to buy or sell carbon credits to maximise returns or minimise risks based on price predictions.

Table 2.9 : Evaluation of the selected articles according to the analysis criteria to identify the research gaps.

Article	**C1**	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**	**C8**	**C9**	**C10**	**C11**	**C12**	**C13**	**C14**	**C15**	**C16**	**C17**	**C18**
[1]	EUA ETS	1	0	1	1	1	0	1	0	0	1	1	0	0	0	0	0	0
[4]	EUA ETS	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1	0
[7]	United States	1	1	1	1	0	0	0	0	0	1	1	0	0	1	0	0	0
[8]	EUA ETS	1	0	1	1	1	0	1	1	0	1	1	0	0	0	0	0	0
[20]	China pilot ETS	1	1	1	1	1	1	1	1	0	1	1	0	0	1	1	1	0
[24]	EUA ETS	1	1	1	1	1	0	1	1	0	1	1	0	0	1	0	0	0
[31]	EUA ETS	1	0	1	1	1	0	0	1	0	1	1	0	0	0	0	0	0
[43]	China pilot ETS	1	1	1	1	1	1	1	1	0	1	1	0	0	1	1	1	0
[44]	EUA ETS	1	0	1	1	1	0	0	1	0	1	1	0	0	0	0	0	0
[45]	EUA ETS	1	0	1	1	1	0	1	0	0	1	1	0	0	0	0	0	0
[49]	EUA ETS	1	0	1	1	1	0	0	1	0	1	1	0	0	0	0	0	0
[50]	EUA ETS	1	1	1	1	1	1	0	0	0	1	1	1	0	1	0	0	0
[53]	EUA ETS	1	0	1	1	1	0	1	0	0	1	1	0	0	0	0	0	0
[54]	China pilot ETS	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	0

Table 2.9 : Evaluation of the selected articles according to the analysis criteria to identify the research gaps (continued).

Article	**C1**	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**	**C8**	**C9**	**C10**	**C11**	**C12**	**C13**	**C14**	**C15**	**C16**	**C17**	**C18**
[64]	China pilot ETS	1	0	1	1	1	0	0	0	0	1	1	0	0	0	0	0	0
[68]	China pilot ETS	1	0	1	1	1	0	0	1	0	1	1	0	0	0	0	0	0
[74]	KAU	1	1	1	1	1	1	0	1	1	1	1	0	1	0	1	1	0
[75]	United States	1	0	1	1	1	0	1	0	0	1	1	1	0	0	0	0	0
[81]	India	1	0	1	1	1	0	0	0	0	1	1	0	0	0	0	0	0
[88]	China pilot ETS	1	0	1	1	1	0	1	1	0	1	1	0	0	0	0	0	0
[89]	China pilot ETS	1	0	1	1	1	0	1	1	0	1	1	0	0	0	1	0	0
[93]	China pilot ETS	1	0	1	1	1	0	0	0	0	1	1	0	0	0	0	0	0
[98]	EUA ETS	1	1	1	1	1	0	1	1	0	1	1	0	1	1	0	0	0
[100]	EUA ETS	1	0	1	1	1	0	1	1	0	1	1	0	0	0	0	0	0
[104]	China pilot ETS	1	0	1	1	1	0	1	1	0	1	1	0	0	0	0	0	0
[91]	EUA ETS	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	0
[108]	China pilot ETS	1	1	1	1	0	0	1	1	0	1	1	0	0	1	1	1	0
[130]	China pilot ETS	1	0	1	1	1	1	0	0	1	1	1	0	0	0	0	0	0
[131]	China pilot ETS	1	1	1	1	1	0	1	1	0	1	1	0	0	1	0	0	0

Table 2.9 : Evaluation of the selected articles according to the analysis criteria to identify the research gaps (continued).

Article	**C1**	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**	**C8**	**C9**	**C10**	**C11**	**C12**	**C13**	**C14**	**C15**	**C16**	**C17**	**C18**
[134]	China pilot ETS	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	0
[136]	EUA ETS	1	0	1	1	1	0	1	1	0	1	1	0	0	0	0	0	0
[137]	EUA and sCER	1	0	1	1	1	0	1	1	0	1	1	0	0	0	0	0	0
[144]	China pilot ETS	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0
[145]	China pilot ETS	1	1	1	1	0	1	1	1	0	1	1	0	0	1	1	1	0
[146]	China pilot ETS	1	0	1	1	1	0	0	1	0	1	1	0	0	0	0	0	0
[150]	China pilot ETS	1	0	1	1	1	0	0	1	0	1	1	0	1	0	0	0	0
[152]	China pilot ETS	1	0	1	1	1	0	0	1	0	1	1	0	1	0	0	0	0
[154]	China pilot ETS	1	1	1	1	0	0	1	1	0	1	1	0	1	1	1	1	0
[155]	China pilot ETS	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0
[158]	EUA ETS	1	0	1	1	1	0	1	1	0	1	1	0	0	0	0	0	0
[160]	China pilot ETS	1	0	1	1	1	0	0	0	0	1	1	0	0	0	0	0	0
[161]	China pilot ETS	1	0	1	1	1	0	0	1	0	1	1	0	0	0	0	0	0
[162]	sCER	1	0	1	1	1	0	1	1	0	1	1	0	0	0	0	0	0
[165]	China pilot ETS	1	0	1	1	1	0	1	1	1	1	1	0	0	0	0	0	0

Table 2.9 : Evaluation of the selected articles according to the analysis criteria to identify the research gaps (continued).

Article	**C1**	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**	**C8**	**C9**	**C10**	**C11**	**C12**	**C13**	**C14**	**C15**	**C16**	**C17**	**C18**
[167]	EUA ETS	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	0
[171]	China pilot ETS	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	0
[173]	China pilot ETS	1	0	1	1	1	0	1	0	0	1	1	0	0	0	0	0	0
[175]	EUA ETS	1	0	1	1	1	0	1	1	1	1	1	0	1	0	0	0	0
[147]	China pilot ETS	1	1	1	1	1	0	1	1	1	1	1	0	0	1	1	1	0
[149]	China pilot ETS	1	1	1	1	0	0	1	1	1	1	1	0	1	1	1	1	0
[153]	China pilot ETS	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	0
[169]	China pilot ETS	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	1	0

This analysis enables the clear identification and comparison of differences and limitations in each study, supporting a comprehensive evaluation. It also helps to highlight areas where further research is necessary to improve the accuracy and reliability of carbon credit price predictions. Furthermore, this approach addresses the complexities of carbon credit markets and the challenges associated with price prediction while providing insights into the strengths and weaknesses of various modelling techniques, leading to the development of more effective and efficient models.

Based on the analysis, it can be observed that hybrid methods, which combine the strengths of different modelling techniques, dominate studies on carbon credit price prediction. The preference for hybrid methods is likely due to their ability to handle non-linearity, non-stationarity, and multicollinearity challenges in the carbon market. In contrast, regression-based and neural network-based methods are employed in fewer studies, suggesting that while these approaches are still valuable, they may not address these complexities in the carbon market as effectively as hybrid models.

Choosing the right modelling method depends on several factors, such as the complexity of the data, the goals of the research, and the available resources. Given the variety of techniques used, it is important to carefully consider which approach is most suitable to predict the price of carbon credits. The frequent use of hybrid models shows their potential to provide better results compared to using a single method.

Based on the defined criteria, additional findings and conclusions can be drawn from an analysis of the relevant studies:

- For the studied market criterion, Figure 2.5 shows the proportion of various carbon markets studied in the reviewed literature. The analysis shows that the

China pilot ETS is the most studied carbon market, accounting for 56% of the literature. This considerable attention is likely due to China's status as one of the world's largest carbon dioxide emitters, coupled with its aggressive actions to address climate change, as noted in [67] and [156]. Since its inception in 2013, China's carbon market, now the world's largest, incorporates seven pilot trading systems, further motivating research interest. Following China, 32% of studies focus on the EU ETS, reflecting the maturity and well-established nature of this market as the world's first and largest carbon market. The EUA ETS has served as a model for other emissions trading systems, which explains the substantial focus on it in academic research. These well-established and prominent carbon markets have received the majority of attention due to their size, maturity, and importance in global carbon reduction efforts. However, this concentration of research creates a notable gap in the exploration of other carbon markets, particularly in emerging economies. Markets such as the US, sCER, India, and the KAU have been underrepresented in the literature, with each receiving only 2-4% of the research focus.

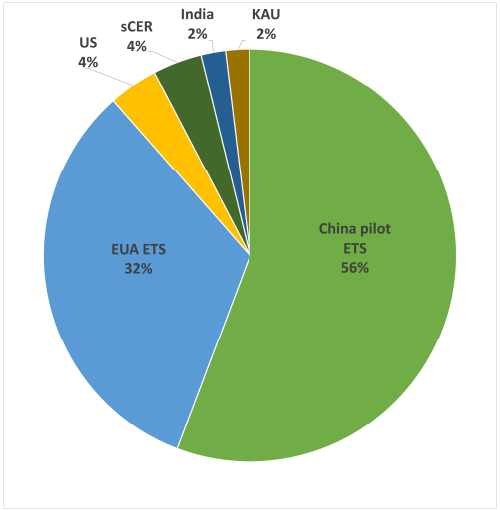


Figure 2.5 : Distribution of the studied carbon markets in the selected papers.

- All studies consider external influencing factors, recognising their essential role in predicting the price of carbon credits. In particular, the energy market was identified as the dominant external factor in most of the research. This emphasis reflects the significant impact of the energy sector on carbon pricing, given the direct link between energy consumption, emissions, and carbon credits.

- The analysis reveals that out of the reviewed studies, 31 focused on multi-factor analysis, while 21 proposed models for predicting carbon credit prices. The multi-factor analysis papers, which account for the majority, concentrate on identifying and understanding external influencing factors without explicitly developing predictive models, highlighting their contribution to factor identification rather than actionable prediction. In contrast, the multi-factor price prediction papers, which form a smaller but significant portion, present models aimed at predicting the price of carbon credits based on these external factors. This indicates that while there is a strong foundation in factor analysis, fewer studies have transitioned these insights into predictive frameworks. Although it is important to analyse the factors that impact carbon markets, studies that go further by integrating these insights into prediction models provide more actionable outcomes for stakeholders.

- The study analysis confirms that 46 of 52 studies applied methods to analyse the effect of external factors that influence the price of carbon credits. This high percentage (88%) highlights the strong focus in the literature on understanding the impact of factors such as energy prices, environmental regulations, and economic indicators on carbon markets. The thorough analysis of these external factors is important, as it serves as the foundation for developing accurate prediction models and making informed trading decisions. The results emphasise that most studies recognise the importance of investigating

these factors, reflecting the critical role they play in shaping carbon credit prices.

- The analysis shows that only 15 of 52 studies applied feature selection techniques to identify key factors influencing the price of carbon credits. This relatively low proportion (around 29%) suggests that, although many studies focus on analysing influencing factors, fewer make use of advanced techniques to reduce the number of variables in their prediction models. Feature selection is crucial to avoid overfitting, improve model efficiency, and focus on the most relevant factors [14]. This limited application highlights a gap in the literature, suggesting that more studies could benefit from incorporating feature selection methods to improve the accuracy and robustness of predictive models [66], [167]. From the literature review, prediction studies use four different approaches for feature selection, as summarised in Table 2.10.

Table 2.10 : Feature selection approaches used in the selected studies and their limitations.

Approach	**Applied techniques**	**Limitations**
No feature selection	All features directly used as inputs to prediction models.	- Introduces irrelevant or redundant features. - Increase computational complexity. - Leads to overfitting. - Reduces model interpretability and efficiency.
Statistical techniques	mRMR, KMO and Bartlett tests, SAE, PACF, LASSO, BGRU with esNet-18, Grey systems theory, Correlation Analysis	- Limited to capturing linear relationships, missing nonlinear or higher-order interactions. - Sensitive to data quality, requiring careful normalization and scaling. - Depends on arbitrary thresholds or assumptions that may not generalize across datasets. - May struggle with complex or heterogeneous data structures.

Table 2.10 : Feature selection approaches applied by the selected studies found in the literature and their limitations (continued).

Approach	**Applied techniques**	**Limitations**
Decomposition techniques	PCA	- Does not identify the most relevant features; all features contribute to the new components. - Presence of similarities and irrelevant information between the components. - Assumes linear relationships, overlooking nonlinear interactions. - Focuses on variance rather than relevance to the target variable. - Reduces interpretability by transforming features into combinations.
Machine learning-based methods	RF, LightGBM	- Computationally intensive, particularly for large datasets. - Risk of overfitting, especially if the model is not carefully tuned. - Feature selection results are heavily dependent on the careful optimization of model hyperparameters. - lack of transparency makes it challenging to derive clear and interpretable feature importance.

- The analysis shows that 69% of the reviewed studies recognise the complexity and non-linearity of carbon market dynamics. Since carbon credit prices are influenced by many unpredictable and interdependent factors, addressing non-linearity is essential for building accurate and reliable prediction models. The fact that most studies acknowledge this shows that researchers understand the need for models that can capture these complex relationships.

- The analysis reveals that 77% of studies recognise the importance of addressing non-stationarity, which is essential for accurately modelling carbon credit prices. Non-stationarity, where statistical properties such as mean and variance change over time, is a common characteristic in time-series data like carbon markets [70]. By considering non-stationarity, these studies ensure that their models are better equipped to handle the dynamic nature of carbon prices, leading to more reliable and robust predictions.

- The analysis shows that only 13 of 52 studies addressed multicollinearity issues in their datasets, representing a relatively low proportion (25%). Multicollinearity occurs when independent variables in a model are highly correlated, which can distort the interpretation of model coefficients and reduce prediction accuracy [60]. By addressing multicollinearity, models can avoid issues such as inflated standard errors and unreliable variable importance, leading to more robust and interpretable results [52]. The limited number of studies that tackle this issue suggests that a majority may overlook the potential complications arising from multicollinearity, potentially impacting model performance and interpretability. This gap in the literature highlights the need for future studies to address multicollinearity more thoroughly, ensuring that predictive models for carbon credit prices are both accurate and reliable in identifying the true influence of each factor.

- A notable aspect of the reviewed studies is that all 52 studies used real-world data for their experiments, demonstrating a strong recognition among researchers of the importance of validating models in practical, real-world data. This indicates a strong commitment across the literature to ensuring that predictive models and analyses are based on actual market conditions, enhancing the validity of the findings. Using real-world datasets also improves the reliability of the results, allowing for more accurate predictions and insights that can be directly applied to real-world carbon markets, making the models more reliable and useful for decision-making.

- The analysis shows that only a small proportion (8%) of the studies, or 4 out of 52, considered multiple periodicities, such as daily, weekly, or monthly intervals, in their data. This low adoption suggests that many studies may overlook the varying impact of external factors on carbon credit prices over different time frames. Incorporating multiple periodicities can significantly enhance prediction accuracy by capturing these time-sensitive influences. For example, Kim and Koo [75] found that the same factor, such as energy prices, may have different impacts in the short and long term: while coal prices are a major long-term determinant of carbon trading volumes, and in the short term, crude oil and natural gas prices also become influential. Furthermore, Feng et al. [44] found that within the EU ETS, the high-frequency components were associated with market mechanisms, while low-frequency components were influenced by external environments and special events such as international negotiations and financial crises. By integrating multiple periodicities, future models could better capture these dynamic effects, offering a more comprehensive and accurate view of carbon price trends that supports informed decision making in carbon markets.

- The analysis reveals that only 12 of 52 studies (23%) incorporated predictions across multiple time horizons, such as short-term and long-term forecasts. This low percentage suggests that there is limited adoption of multi-horizon prediction in carbon credit price modelling, despite its potential to improve predictive accuracy and strategic value. Multi-horizon forecasting allows models to capture both immediate fluctuations and longer-term trends, each of which can be driven by different influencing factors, enabling a more comprehensive approach to carbon market analysis. This capability is crucial for effective decision-making, as it supports short-term trading actions and long-term investment strategies

- All 21 studies proposing prediction models validated their models using a separate validation dataset, which is a good practice for model evaluation. This ensures that the model's performance is tested on data not used in training, helping to prevent overfitting and improve generalizability. The consistent use of validation across these studies demonstrates a strong commitment to building reliable models that can provide accurate predictions in real-world applications.

- The analysis shows that 19 of 52 studies compared their proposed solutions with existing baseline models. However, the fact that only about 37% of studies perform benchmark comparisons indicates that there is room for improvement in adopting this practice. Benchmarking is valuable in carbon credit price modelling as it allows researchers to evaluate how new models perform relative to established standards. By including benchmark comparisons, studies can better demonstrate the effectiveness of their models and identify areas for improvement. Expanding the use of benchmark comparisons in future research could improve model transparency and help identify the most effective

approaches for carbon credit price prediction.

- The analysis shows that only 18 of 52 studies (35%) used at least three metrics, such as MSE, RMSE, and R-squared, to assess model performance. This indicates that a minority of studies take a more comprehensive approach to evaluating model accuracy and reliability in carbon credit price prediction. Using multiple metrics provides a fuller picture of performance by capturing different aspects of predictive accuracy. Expanding the use of at least three evaluation metrics could improve the robustness and reliability of models, helping stakeholders make better-informed decisions in the carbon market.

- The study analysis confirms that none of the 52 studies extended their models to actively incorporate trading strategies based on carbon credit price predictions. This gap highlights an opportunity for future research to apply trading strategies that leverage price prediction to manage risks associated with carbon price fluctuations. Carbon emission credits are increasingly recognised as investment assets, similar to stocks and real estate [74]. An effective trading strategy could involve rules for buying or selling carbon assets based on predicted price changes, simulating real market conditions, and accounting for transaction costs. Extending models in this way would provide practical value, showing how investors and enterprises could use predictions to manage volatility and optimise carbon assets.

2.7 Conclusion

This chapter provided a comprehensive review of the relevant literature on the prediction and analysis of carbon credit prices. An SLR approach was used to select 52 studies from reputable databases and high-quality journals. These studies were analysed to identify research gaps and limitations, focussing on external influencing

factors, feature selection techniques, and methods addressing non-linearity, non-stationarity, and multicollinearity.

Based on the result of the analysis of the selected literature, the next chapter defines the research gaps, formulates the research questions, and outlines the objectives.

Chapter 3

Problem Definition and Research Questions

3.1 Overview

This chapter highlights the existing gaps within the field of carbon credit price prediction mechanisms, based on the systematic literature review (SLR) presented in the previous chapter. The chapter is structured as follows: Initially, key concepts critical to the study are defined in Section 3.2, providing the foundational understanding necessary for the discussion of the research problem. Next, Section 3.3 identifies the existing gaps within the field. By identifying these gaps, we aim to contribute to the existing body of knowledge and address areas that require further investigation. Section 3.4 presents the research questions derived from these identified gaps. Subsequently, Section 3.5 outlines the research objectives based on the research questions. The chapter concludes in Section 3.6. Through this comprehensive exploration of research gaps, research questions, and objectives, we establish a solid foundation for the subsequent chapters of our thesis.

3.2 Key terms and concepts

An explanation of the key terms and concepts used throughout the thesis is provided in this section.

- **Carbon credit:**

  A carbon credit is a digital permit in the form of an immutable token traded on the carbon market, representing a reduction or removal of one tonne of GHG emissions from the atmosphere [151].

- **Carbon market:**

  An online platform to trade carbon credits and reduce greenhouse gas emissions. In this market, members can buy and sell carbon credits [151].

- **Cap and trade:**

  A regulatory process that sets a "cap" on the level of greenhouse gas emissions that are allowed to be emitted by companies. Companies that have not exceeded their limits are allowed to "trade" (sell) any excess emissions they produce [32].

- **The Australian carbon credit units (ACCUs):**

  This is an initiative of the Australian government to eliminate or avoid greenhouse gas emissions, with an ACCU per tonne of reduction in carbon dioxide [25].

- **External Influence Factors:** In this study, an external influence factor refers to an external variable that affects carbon credit prices. These factors include market dynamics such as international carbon market prices, energy prices, environmental conditions, macroeconomic indicators, societal trends, and advancements in sustainable technology. Understanding and incorporating these factors into predictive models is essential to accurately predicting carbon credit prices and enhancing model robustness [147], [171].

- **Machine learning:**

  Machine learning is a subset of artificial intelligence (AI) that concentrates on creating algorithms and statistical models that allow computers to execute tasks autonomously, without explicit programming. These algorithms learn from patterns in sample data, called "training data," to make predictions or decisions [110].

- **Time-series:**

  Time-series refers to a sequence of observations or data values recorded over time. These data points may be collected at consistent intervals or at varying time intervals [106].

- **Prediction:**

  Prediction involves using a trained model to predict the value of a target variable based on new, unseen data. The trained model applies the relationships it has learnt from the training data to make informed guesses about the target variable for new inputs [170]. In this study, the predictions focus on estimating carbon credit prices based on historical carbon credit price data and external factors.

- **Short-term prediction:**

  Short-term prediction refers to predictions made for events or trends that are expected to occur in the near future, usually within a few days, weeks, or months [170].In this study, short-term prediction specifically refers to next-day predictions of carbon credit prices, which can help in immediate trading and investment decisions.

- **Long-term predictions:**

  Long-term predictions refer to the predictions made for events or trends that are expected to occur in the distant future, usually over several years [170].

- **Multivariate prediction model:**

  A multivariate prediction model is a predictive model that simultaneously considers multiple independent variables (external factors) to predict a target variable.

- **Hyperparameter tuning:**

  Hyperparameter tuning involves optimising the hyperparameters of a machine

learning model to improve its performance. Hyperparameters are the configuration settings used to control the learning process, such as the learning rate and the number of trees in a random forest.

- **Non-linearity:**

  Non-linearity refers to the situation where the relationship between the dependent variable (ACCU prices) and the independent variables is not a straight line [36], [70].

- **Non-stationary:**

  Non-stationary time series data implies that statistical properties, such as mean and variance, change over time [36], [70]. This can result in unpredictable trends and fluctuating patterns.

- **Multicollinearity:**

  Multicollinearity arises when the independent variables in a regression model exhibit high correlation, resulting in incorrect estimations of the regression coefficients [60]. It can cause instability in prediction models and make it difficult to determine the effect of each predictor.

- **Positive effect:**

  This implies that an increase in the factor will lead to an increase in the demand for carbon credits, increasing the price of carbon credits.

- **Negative effect:**

  Conversely, a negative effect denotes an inverse relationship between the factor and the price of carbon credit. This means that an increase in the factor either reduces the demand for carbon credits or indicates a shift towards lower-carbon alternatives, thereby decreasing the price of carbon credits.

3.3 Key gaps in the literature

In this section, the researcher identifies the literature gaps relevant to the subject of the research and highlights the shortcomings of existing studies. In doing so, the researcher aims to demonstrate the need for the study and provide a rationale for why the research question is relevant and important.

1. While many studies have explored individual external factors influencing carbon credit prices, there is still a significant gap in understanding their combined impact. The absence of such an analysis limits the development of robust predictive and forecasting models capable of capturing the complexities of carbon markets. A deeper examination of these factors is essential for policymakers aiming to design effective carbon pricing mechanisms and businesses and investors seeking data-driven strategies. Identifying and quantifying these key influences will contribute to more reliable market predictions, helping stakeholders make informed decisions and adapt to changing market conditions.

2. There is a lack of studies that focus on developing a smart and effective method for selecting the key factors influencing the price of carbon credits. Although several factors influence carbon credit prices, not all carry the same weight. Identifying the key factors can significantly enhance the accuracy of price predictions. However, current methods for selecting these factors are often subjective and rely on expert judgment. Therefore, there is a significant research gap in developing a smart and data-driven method to select the most important factors that affect carbon credit prices, which can help improve the accuracy and reliability of carbon credit price predictions. Addressing this research gap could lead to more efficient and effective carbon credit markets and provide valuable insight for investors, policymakers, and other stakeholders in

the carbon market.

3. There is a lack of prediction models for the price of carbon credits that incorporate all influencing factors. Different studies have considered some influencing factors when predicting the price of carbon credits, but they tend to overlook or underestimate the impact of other factors. Not considering all key factors can lead to less accurate price predictions, negatively impacting the efficiency of carbon credit markets and investor confidence. Therefore, there is a significant research gap in developing a comprehensive prediction model that incorporates all factors influencing the price of carbon credits for fair and accurate price predictions. Addressing this research gap could provide valuable insight for policymakers and investors, improve the efficiency of carbon credit markets, and help promote the transition to a low-carbon economy.

4. There is a lack of reliable multi-factor approaches that effectively address the inherent challenges in predicting the short-term price of carbon credits. While some studies have utilized machine learning models to predict carbon credit prices, they often fail to comprehensively tackle the issues of non-linearity, non-stationarity, and multicollinearity within the data. These challenges are essential to accurately predicting volatile carbon credit prices.

5. Although advancements have been made in recent studies on predicting carbon trading prices, a significant gap remains in the literature regarding how external factors specifically contribute to prediction accuracy. Existing models operate as black boxes, lacking interpretability and transparency, which limits their practical utility and stakeholder trust. No research has thoroughly shown the impact of external factors on carbon credit prices in a way that integrates explainable artificial intelligence (XAI). This gap highlights the need for a model that combines high predictive accuracy with clear, interpretable

insights into the impact of key external factors, helping policymakers and market participants make informed decisions.

6. There is a lack of studies that specifically analyse the factors influencing the price of ACCUs (Australian Carbon Credit Units) and predict their market trends. The existing work focuses mainly on global trading markets such as Europe, America, and China. However, no comprehensive study is available on the Australian carbon credit market. Therefore, a significant research gap exists in understanding the specific drivers of ACCU prices in Australia.

3.4 Research questions

This research will address the following main research question:

How can we accurately and intelligently predict the price of carbon credits considering the key influencing factors?

Table 3.1 : Research sub-questions and their motivations

Research Sub-Questions	**Motivations**
RQ1: What are the factors (determinants) that influence the price of carbon credits?	To identify and analyse a comprehensive list of external factors that affect the price of carbon credits.
RQ2: How can we intelligently and reliably analyse and select key influencing factors on carbon credits?	To develop a model that can accurately and reliably analyse the impact of factors that influence the price of carbon credits and identify key factors, considering the nature of the carbon credit data.

Continued on next page

Table 3.1: Research sub-questions and their motivations (continued)

Research Sub-Questions	**Motivations**
RQ3: How can we develop a reliable multi-factor approach to predict the short-term price of carbon credits that effectively addresses the challenges of non-linearity, non-stationary, and multi-collinearity?	To propose a reliable and intelligent multi-factor approach for predicting short-term carbon credit prices using machine learning algorithms that avoid the challenges of the carbon credit time-series, considering the key factors affecting carbon credit prices.
RQ4: How can we validate and verify the methodologies developed in RQ2 and RQ3?	To evaluate and validate the results of the methodologies developed in RQ2 and RQ3 to ensure their reliability and accuracy.

3.5 Research objectives

Research objectives are essential to provide a clear direction and purpose for the study. Therefore, they should be carefully designed to align with the research questions being addressed. This section establishes a framework for the research project by outlining the key objectives that will be pursued throughout the study.

The primary objective of this thesis is to develop a comprehensive, intelligent, and accurate model for predicting short-term carbon credit prices. The prediction model aims to incorporate relevant factors that influence carbon credit prices, including energy markets, macroeconomic conditions, and environmental factors. This objective will be achieved by addressing the following sub-objectives:

- **Objective 1**: Identify and analyse a comprehensive list of external factors that affect carbon credit prices. (Addressing the first research question.)

- **Objective 2**: Develop a model that leverages feature importance and feature engineering techniques to accurately and reliably analyse and select the most influential factors affecting carbon credit prices. (Addressing the second research question.)

- **Objective 3**: Develop a robust and intelligent multi-factor approach to predict short-term carbon credit prices, addressing non-linearity, non-stationary, and multicollinearity challenges. (Addressing the third research question.)

- **Objective 4**: Evaluate and test the methodologies developed in Objectives 2 and 3 to ensure their reliability and accuracy. (Addressing the fourth research question.)

3.6 Conclusion

This chapter highlighted the existing gaps and limitations within the field of carbon credit price prediction, it presented the specific research questions to be addressed in this study, and identified the research objectives to pursue. In addition, the key terms and concepts were defined to establish a common understanding. The next chapter will focus on the research methodology, providing an overview of the solutions and approaches to be used to achieve the defined objectives.

Chapter 4

Research Methodology and Solution Overview

4.1 Overview

This chapter extends the exploration of carbon credit price prediction mechanisms by outlining the methodology employed in this research. The chapter builds on the foundational work presented in the preceding chapters. The methodology is grounded in the design science research methodology (DSRM), a robust framework particularly suited to addressing problems in engineering and computer science. Section 4.2 outlines the rationale for selecting DSRM and details how it will be applied throughout this study to develop and test the proposed solutions. Each of the subsequent sections, from 4.4 to 4.7, is dedicated to proposing solutions for each research objective outlined in the previous chapter. These sections will describe the processes and techniques utilised to design, develop, and assess solutions to ensure they are robust and effective. Finally, the chapter concludes in Section 4.8.

4.2 Design science research methodology

Various research methods are commonly applied in the field of information systems, including action research, design research, laboratory experiments, and case studies [56], [113]. For this thesis, the DSRM has been selected as the most suitable approach to achieve the research objectives detailed in Chapter 3. This is because the main objective of the research is to develop a comprehensive, intelligent and accurate model to predict short-term carbon credit prices. The prediction model aims to incorporate relevant factors that influence carbon credit prices, including

energy markets, macroeconomic conditions, and environmental factors.

DSRM is defined as a process for creating and evaluating IT artefacts that are intended to solve specific problems. The methodology focuses on designing artefacts to address observed issues, contributing to research knowledge, evaluating the effectiveness of these designs, and communicating the findings to relevant audiences [56]. Following the DSRM approach, this research ensures that the methodology is scientific and rigorous. As demonstrated in Figure 4.1, the methodology for DSRM consists of five stages [56]. Based on DSRM, our proposed solution implements the following steps:

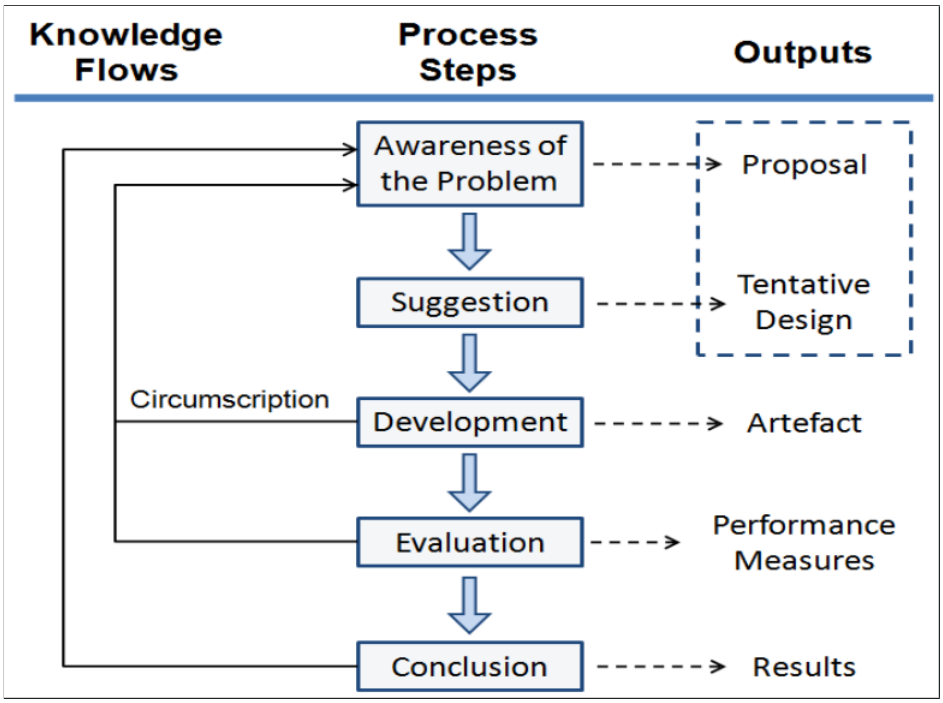


Figure 4.1 : Methodology of design science research [56].

4.2.1 Awareness of the problem

This is the first stage that involves observing the current literature to identify the limitations of the most recent state of the art related to the research topic

and identifying specific research problems. The outcome of this stage is a research proposal.

Identifying the research topic was the initial step in understanding the issue. Research topics can be chosen based on personal interests, observations, or a review of existing literature on previous theories and research in the area. Social concerns or current trends can also guide the selection process. In this case, the research topic emerged from a business concern in the carbon credit trading market.

In this thesis, we conducted an extensive study of the existing literature to identify the challenges and limitations of the current models used to predict carbon credit prices. This awareness phase is essential, as it establishes a strong foundation to create a framework that effectively addresses these challenges, ensuring that research is relevant and impactful. This step is documented in Chapters 2, 3, and 5 of this thesis.

4.2.2 Suggestion stage

In this stage, the focus is on exploring possible solutions to the identified research problem. The goal is to propose different strategies and approaches that could effectively address the challenge of predicting carbon credit prices. In this research, this involved considering several alternative methods for developing a predictive model. Different machine learning algorithms and techniques were evaluated, each offering unique strengths in handling the complexity of the data and the multiple factors that influence carbon credit prices. Various approaches were used to study feature importance to determine the key factors, since these factors will be the models' inputs. In addition, different feature extraction and reduction techniques are used to enhance the quality of input for prediction models.

The key concept suggested here is the integration of these machine learning techniques with a thorough feature process to develop a robust framework. This

framework is intended to accurately predict carbon credit prices by effectively accounting for the various external factors that influence these prices. The suggestion phase thus sets the groundwork for selecting the best approach from the alternatives and moving forward with a solution that promises to address the research problem comprehensively. This step is documented in Chapter 4, 7 and 8 of this thesis.

4.2.3 Development stage

In the development stage, the objective is to create a functional artefact based on the concepts and strategies outlined during the suggestion phase. This involves constructing the architecture of the solution, ensuring that it aligns with the design knowledge and objectives identified earlier.

This stage focuses on developing the predictive framework for carbon credit prices by implementing the models and techniques proposed previously. Specifically, two key components were developed: the CCMFI (Carbon Credit Multi-Factor Identification) model for feature selection and the CCMFP (Carbon Credit Multi-Factor Prediction) model for price prediction. During this phase, the CCMFI model is first employed to ensure that only the most impactful factors are considered. The CCMFP model is then constructed using these selected features, trained, and validated to produce accurate predictions of carbon credit prices. This ensures that subsequent predictive models focus only on the most relevant factors, enhancing their accuracy and efficiency.

The initial models were typically built iteratively as part of the development process. This iterative approach allowed for continuous refinement and improvement of the models, ensuring that each version of the framework better addressed the complexities of predicting carbon credit prices. This phase is documented in Chapters 7 and 8 of this thesis.

4.2.4 Evaluation stage

The evaluation stage is a critical component of design science research, where the developed artefact is rigorously evaluated for its effectiveness, quality, and efficiency. This phase ensures that the proposed framework is both theoretically sound and practically applicable.

In this research, evaluation involves a thorough assessment of the CCMFI model for feature selection and the CCMFP model for carbon credit price prediction. The evaluation process includes conducting experiments to determine the optimal settings for the number of external factors selected, the parameters of the machine learning models, and the number of independent components. These experiments are essential for fine-tuning the models, helping to determine whether to proceed with the current design or return to the development stage to make improvements.

Various metrics, such as mean squared error (MSE), root mean squared error (RMSE), mean absolute percentage error (MAPE), R squared (R^2), and execution time, are used to evaluate model performance.

This evaluation phase is directly related to the RQ4 research question and is essential to validate the research objectives outlined in the study. The findings of this phase are documented in the validation sections of chapters 7 and 8.

4.2.5 Conclusion stage

In the conclusion stage, the focus is on confirming that the developed framework, which includes the CCMFI and CCMFP models, provides a valuable and verifiable contribution to the prediction of carbon credit prices. This phase involves determining how the framework can be used effectively and identifying any adjustments needed for practical application. This involves submitting scholarly research publications to international peer-reviewed journals and conferences. These publications

aim to communicate the results of the proposed framework, ensuring that its findings are reached and evaluated by both researchers and professionals in the field.

4.3 Overview of the proposed solution

This section presents an overview of the proposed research solution. Once the research objectives have been clearly defined, a comprehensive framework is developed to achieve the goal of creating an intelligent and trustworthy framework to predict the price of carbon credits, capable of determining and examining the most influential factors affecting these prices. Figure 4.2 presents an overview of the proposed solution reflecting the research objectives identified in Chapter 3. The framework was built based on a thorough review of the literature and best practices in the field of establishing prediction and forecast procedures, as shown in Figure 4.3 and described by Montgomery et al. [106]. The thesis follows five methodological steps, as shown in Figure 4.4. Figure 4.5 offers a detailed overview of the proposed research solution, including the specific approaches and methods applied.

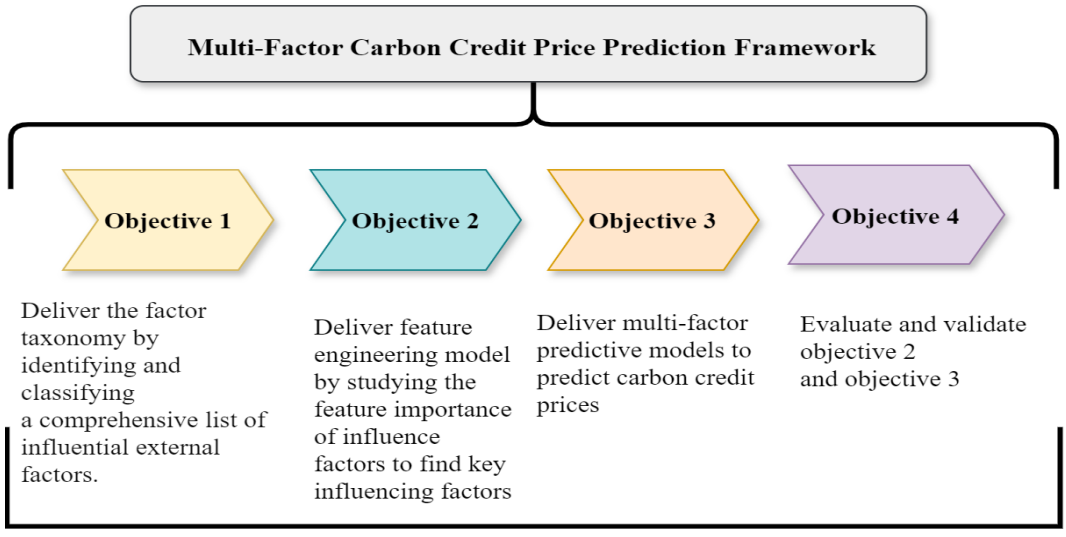


Figure 4.2 : Overview of the proposed framework reflecting the research objectives.

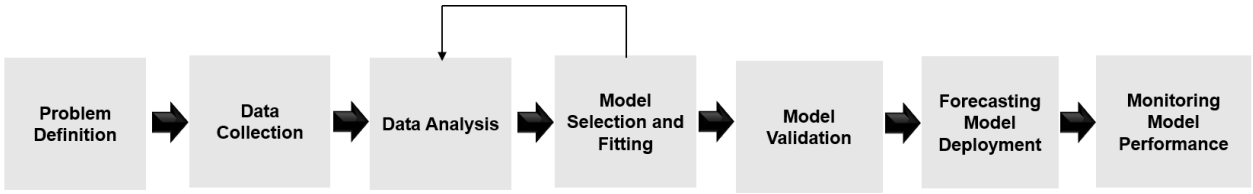


Figure 4.3 : The forecasting process [106].

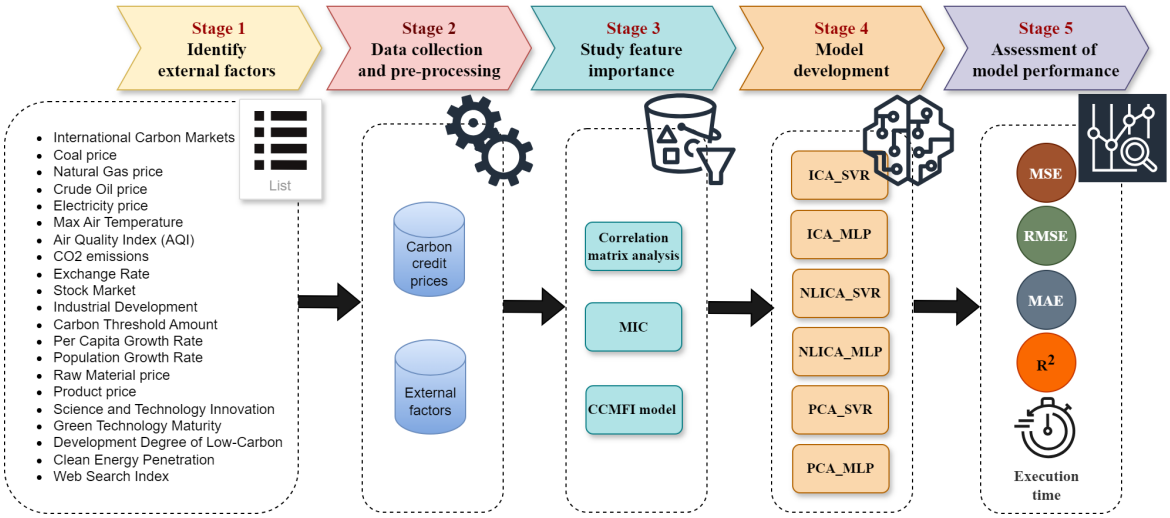


Figure 4.4 : General overview of the proposed research solution.

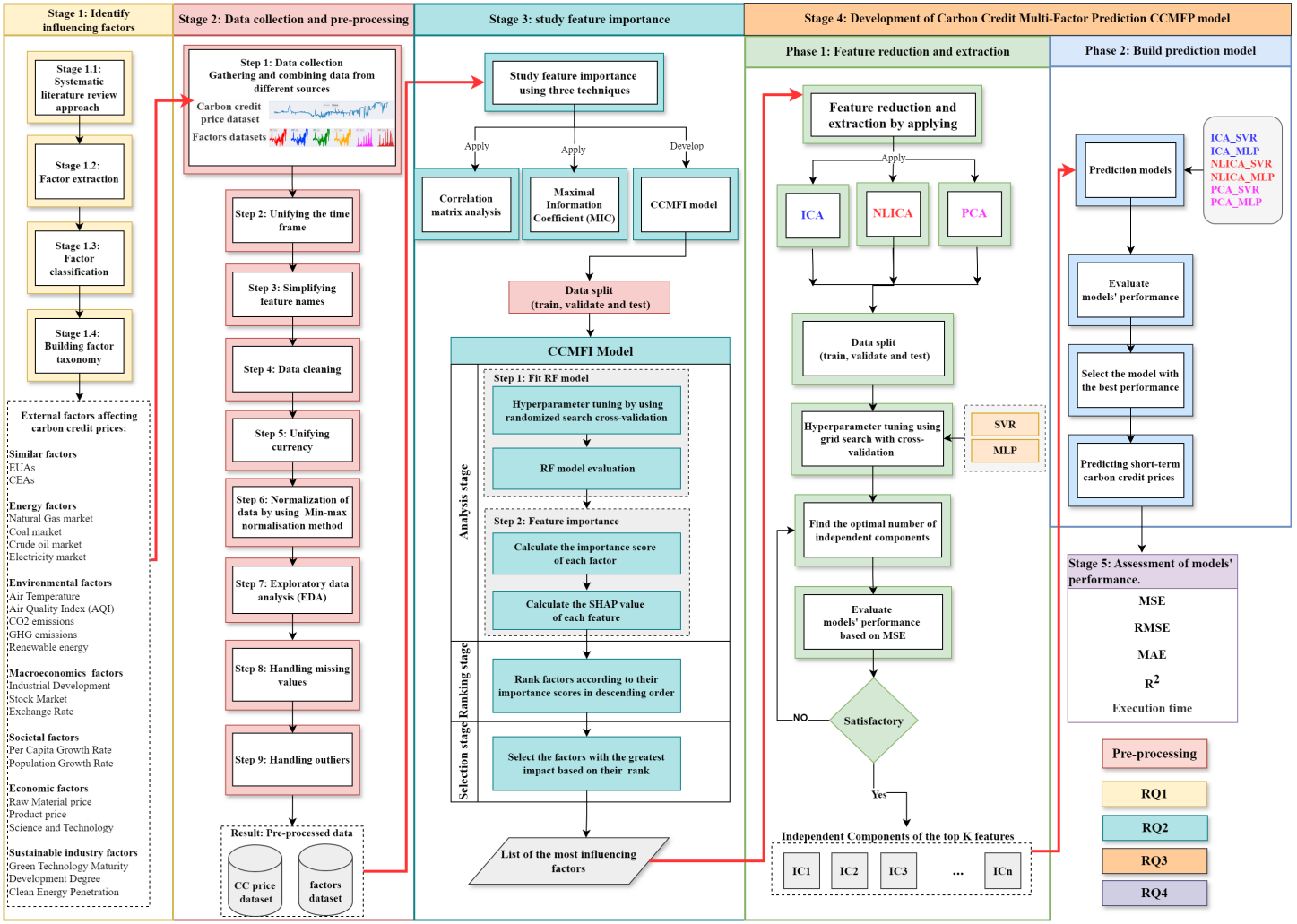


Figure 4.5 : Overview of the proposed research solution with the proposed approaches and methods.

The steps of the proposed solution are as follows:

- **Stage 1: Identify a comprehensive list of external factors that influence the price of carbon credit**

  This stage begins by searching the literature to identify and extract external factors that influence carbon credit prices. During this stage, a rigorous review of the literature is conducted to ensure that all relevant factors are taken into account. These factors are then classified and organised into a taxonomy.

- **Stage 2: Data collection and pre-processing**

  Pre-processing is a critical step in data analysis and machine learning, contributing to the robustness and validity of our research findings [16]. It involves preparing and cleaning the data to make them suitable for analysis and modelling, thus producing reliable and accurate results. First, historical data on carbon credit prices and the identified external factors are collected from reliable sources. Next, the collected data is pre-processed, which involves cleaning, normalising, and handling missing values to ensure the dataset's quality.

- **Stage 3: Select the key influencing factors**

  The third stage aims to select key influencing factors from the dataset. The selection of model inputs is a critical stage that significantly impacts models' performance [46]. The prediction accuracy of a model can be significantly enhanced by selecting appropriate input features, and a valid feature selection method can reduce the model computational time [65], [66], [167]. Eliminating redundant variables and preserving only the most relevant ones minimise the risk of overfitting. In this study, we applied three methods to study the feature importance to determine the key factors, as these factors will be the models' inputs. These methods are correlation analysis, the MIC and the proposed

CCMFI model. The proposed CCMFI model integrates RF regression and an explainable AI technique, providing a powerful tool to identify key price drivers.

- **Stage 4: Develop multi-factor predictive models to predict carbon credit prices**

  In the fourth stage, the Carbon Credit Multifactor Prediction Model (CCMFP) is developed through two key phases: the feature extractionand reduction phase, and the prediction model developing phase.

  - **Stage 4.1: Feature extraction and reduction:** This phase employs feature extraction and reduction techniques, including PCA, ICA and NLICA, to enhance the quality of input for prediction models. These techniques extract essential features and address multicollinearity among the independent variables influencing carbon credit prices. The optimal number of independent components (ICs) is determined systematically using a grid search with k-fold cross-validation, ensuring the selection of components that minimize the mean squared error (MSE) of the prediction models. By generating statistically independent components from the original variables, PCA, ICA, and NLICA effectively mitigate multicollinearity and produce inputs that are better suited for machine learning applications.
  - **Stage 4.2: Prediction model construction:** The extracted components are then utilized as inputs for support vector machines (SVR) and multilayer perceptron (MLP) algorithms to construct the prediction model.

The following sections present the solutions for each of the research objectives.

4.4 Overview of the proposed solution to achieve objective 1

This section presents the solution to achieve objective 1: **Identify and analyse a comprehensive list of external factors that affect carbon credit prices.**

Figure 4.6 outlines the detailed steps used to identify external factors that influence the price of carbon credits, including the search of the literature, the extraction, and the classification of factors. This methodological approach ensures a comprehensive and transparent review of existing research within a defined scope. The following stages are followed to achieve objective 1:

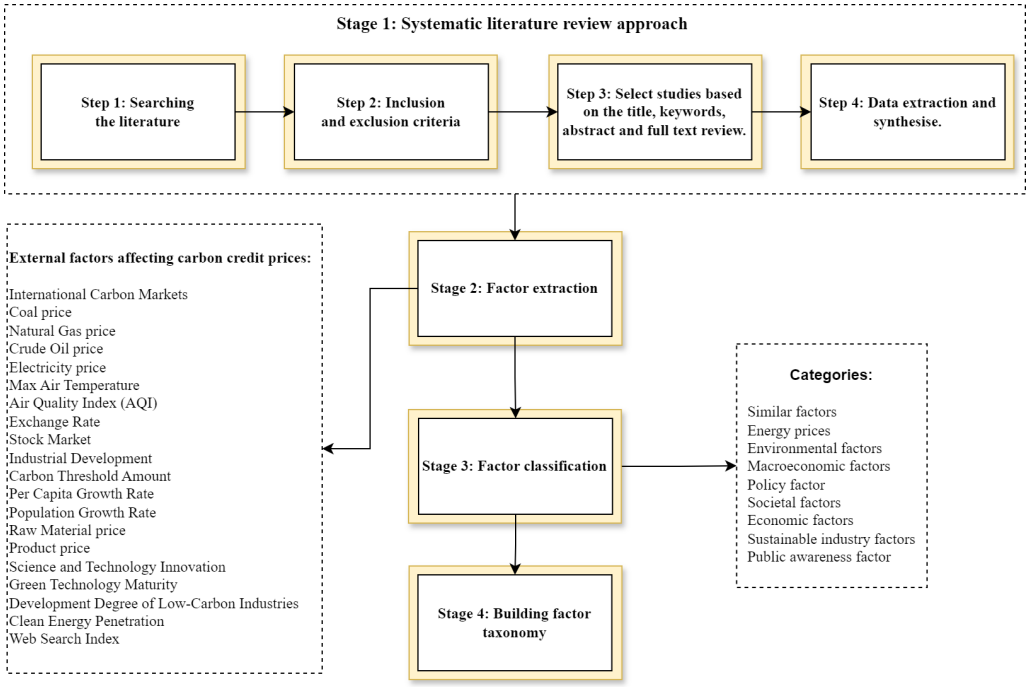


Figure 4.6 : Overview of the approach steps for identifying external factors influencing carbon credit prices.

1. The first stage involved applying the SLR methodology outlined in Chapter 2 to select articles that consider external factors that affect the prices of carbon credits. The review process involved a structured selection of studies from

major electronic databases, including IEEE Xplore Library, SpringerLink, ScienceDirect, and ACM Digital Library. The search covered articles published from 2005 to 2024, ensuring a comprehensive review of relevant literature. The selection process followed rigorous screening and refinement stages to ensure relevance. Ultimately, 52 studies were included for detailed analysis.

2. The next stage involved identifying factors that influence the price of carbon credit, as reported in the shortlisted articles. Data extraction and synthesis of the 52 qualifying articles were carried out systematically. The information was organised in a spreadsheet, capturing details such as the trading market, methodologies, models, studied factors, and the impact of each factor.

3. Next, the factor classification stage involves reviewing each identified factor to identify common themes or characteristics. Based on these themes, broad categories are created that can encompass related factors. Each identified factor is placed into the most relevant category. The categories include similar carbon markets, energy prices, environmental factors, macroeconomic factors, policy factors, societal factors, economic factors, sustainable industry factors, and public awareness factors. This step may require refining the categories to ensure that each category accurately reflects the factors. After all factors have been classified, the structure is reviewed to check whether some categories are too broad or too narrow, and adjustments are made by merging or splitting categories as needed. The final step is to confirm the categories, ensuring that each is distinct and encompasses all factors without significant overlap.

4. The final stage involved building a comprehensive taxonomy of the factors influencing carbon credit prices. This taxonomy provides a structured and organised overview of the external factors identified through the SLR, facilitating a thorough understanding of their impact and highlighting gaps in

current research.

This objective also addressed the conflicting findings that led to different results across studies. The study acknowledged the complex and multifaceted nature of the market, regional variations, and context that contribute to conflicting findings. It also noted that differences in data sources, quality, and methodological approaches impact the results. This structured approach ensures a robust and comprehensive identification of external factors that affect carbon credit prices, providing a solid foundation for accurate and reliable prediction models.

4.5 Overview of the proposed solution to achieve objective 2

This section overviews the solution to achieve objective 2: **Develop a model that leverages feature importance and feature engineering techniques to accurately and reliably analyse and select the most influential factors affecting carbon credit prices.**

After identifying the list of factors that influence the price of carbon credits, we first collect data for the price of carbon credits and each of the selected influencing factors. We use real-world datasets to validate our model. The datasets processed in this study are the daily records from January 2, 2020, to September 1, 2023, totalling 667 data samples. The source and details of the dataset are provided in Chapter 6, Section 6.3.

Next, we pre-processed the dataset by applying a series of pre-processing steps. Pre-processing is a critical stage in data analysis and machine learning, contributing to the robustness and validity of our research findings [16], [18]. It involves preparing and cleaning the data to make them suitable for analysis and modelling, thus producing reliable and accurate results [18]. In this study, pre-processing was undertaken using key steps to optimise data reliability. First, we simplify the feature

names to improve clarity and accessibility. The data frequency, quarterly, monthly, and yearly data intervals were then converted to daily frequencies, ensuring consistency between the data sets and facilitating seamless analysis and comparison. Further refinements involve unifying currency by converting the price series to AUD using the daily exchange rates obtained from https://www.foreignexchange.org.uk/.

In addition, the $fit_transform$ method in the $MinMaxScaler$ function was used to normalise the data. Transforming data can minimise noise and enhance prediction performance using machine learning techniques [164], [19]. Data normalisation is an approach to data transformation [18]. The Min-Max scaler is particularly suitable for our dataset for several reasons. Firstly, the features in our dataset exhibit varying units and scales, which could disproportionately influence the results of machine learning algorithms. Normalising these features ensures that each one contributes equally to the model, preventing any single feature from dominating due to its larger numerical range and mitigating the impact of different feature scales and units on the feature selection process. Secondly, some machine learning algorithms, such as random forests and neural networks, perform better and converge faster when the input data is normalised. By scaling the features to a uniform range, we improve the computational efficiency and stability of the model training process. Lastly, Min-Max scaling preserves the original distribution and relationships of the data, which is crucial for maintaining the integrity of the dataset's underlying patterns.

In addition, we performed exploratory data analysis (EDA) using visualisation and statistical techniques to gain insight into the distribution and relationships of the data. Subsequently, missing values are handled by removing data points containing missing information to mitigate bias. Furthermore, we conducted three critical tests to address specific data challenges: the Augmented Dickey-Fuller (ADF) test to check for non-stationary, scatter plot analysis to investigate non-linearity, and the

Variance Inflation Factor (VIF) test to identify multicollinearity. Figure 4.7 shows the pre-processing steps applied to the dataset in this study. The detailed process of the data is given in Chapter 6 of this thesis.



Figure 4.7 : Overview of data pre-processing steps.

Many researchers emphasise that selecting model input is an essential stage that significantly impacts model performance [46], [65], [66], [167]. Therefore, in this study, we identified the key factors through a comprehensive multi-factor analysis based on three techniques, as illustrated in Figure 4.8. To achieve this objective, we analysed the effect of 22 external factors on ACCU carbon credit prices using three methods: correlation analysis, the maximal information coefficient (MIC), and the proposed carbon credit multifactor identification (CCMFI) model. The proposed CCMFI model consists of three stages: analysis, ranking, and selection of

key factors, as illustrated in Figure 4.9 and explained as follows:

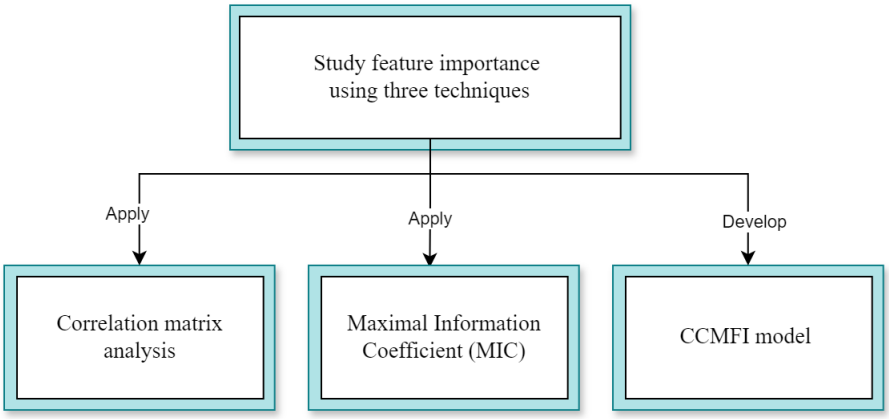


Figure 4.8 : The applied techniques to identify the key influencing factors.

- **Stage 1: Analysis Stage**: This stage involves a two-step approach: the first is to fit the RF model, and the second is to calculate the importance of features using the SHAP method. After fitting the RF model, the importance of the features is measured based on how much each feature improves the model's prediction accuracy. To determine the overall importance of each feature, we aggregated the SHAP values across the dataset by taking the mean of the absolute SHAP values for each feature across all data points.

  Although the RF model is effective for feature selection and importance scoring, incorporating SHAP values adds interpretability, captures interaction effects, and enhances the robustness of the analysis. SHAP values provide a transparent and mathematical method to interpret RF predictions by quantifying the impact of each feature. They explain how much each feature contributed to the predicted price of carbon credits compared to the average price across the dataset.

- **Stage 2: Ranking Stage**: Importance scores are sorted in descending order based on the aggregated SHAP values and then ranked according to their

scores. The higher the SHAP value, the more significant the impact of the feature on the model output, indicating a higher importance. This step helps the user of the model to see which factors impact the price of carbon credits the most.

- **Stage 3: Selection Stage**: In the final stage of the CCMFI model, the factors with the most significant impact are selected based on their rank. This step allows the user to focus on the most important factors when predicting the price of carbon credits.

The development process of the applied techniques and the proposed model is given in detail in Chapter 7 of this thesis.

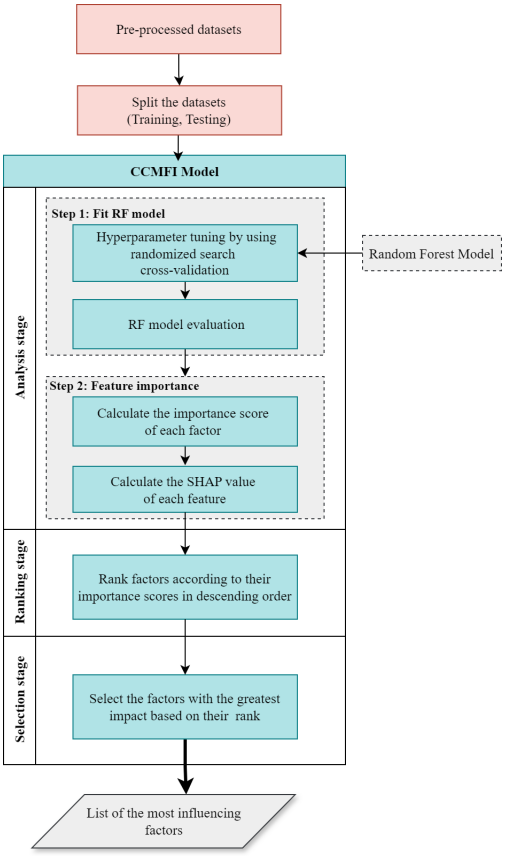


Figure 4.9 : The flowchart process for the proposed CCMFI model.

4.6 Overview of the proposed solution to achieve objective 3

In this section, we overview the solution to achieve objective 3: **Develop a robust and intelligent multi-factor approach to predict short-term carbon credit prices, addressing non-linearity, non-stationary, and multicollinearity challenges.**

To achieve objective 3, we developed the Carbon Credit Multi-Factor Prediction (CCMFP) model to predict the short-term price of carbon credits. The proposed CCMFP model employs a hybrid approach by integrating feature extraction and reduction techniques with machine learning algorithms, SVR and MLP.

The proposed methodology consists of two stages. In stage 1, we employ feature extraction and reduction techniques (PCA, ICA and NLICA) as a pre-processing step to extract essential features and mitigate the issue of multicollinearity within the set of independent variables influencing carbon credit prices [52]. Multicollinearity arises when independent factors demonstrate high correlation, which can cause instability in prediction models [60]. By identifying statistically independent components from the original variables, PCA, ICA and NLICA effectively reduce multicollinearity and generate components that are more suitable for machine learning applications [60]. Subsequently, in Stage 2, the extracted independent components are used as input for the SVR and LMP algorithms to build the prediction model. Figure 4.10 outlines the structured approach and key components of the CCMFP model. The development process of the proposed modules is given in detail in Chapter 8 of this thesis. The following is the step-by-step methodology of the proposed CCMFP approach for predicting the short-term carbon credit price using the MLP and SVR models:

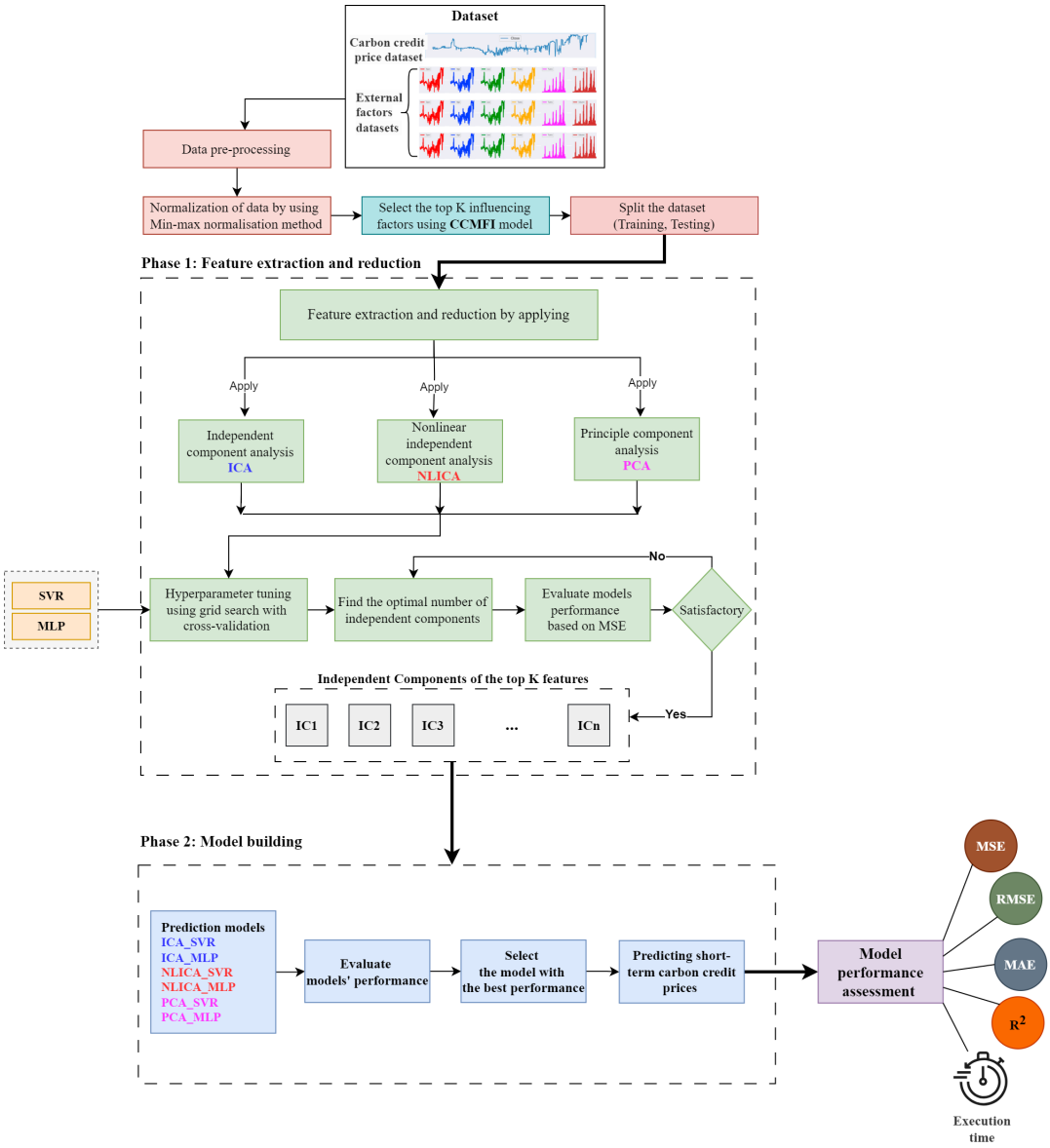


Figure 4.10 : Overview of the proposed CCMFP model with the applied approach and methods.

1. The methodology starts by pre-processing the dataset, including the data normalisation to ensure consistency across all features (details in Chapter 6).

2. Then, the top K influencing factors on carbon credit prices are selected using the CCMFI model as reported in Table 8.1. This step focuses on identifying the top key influencing factors that have a significant impact on carbon credit

prices. For a detailed explanation of the CCMFI model, refer to Chapter 7.

3. After selecting the K influencing factors, we split the pre-processed dataset into training and testing sets. This ensures that the model is trained on a portion of the data and validated on unseen data to assess its performance. In this study, an 80/20 split is used, where 80% of the data is used for training and 20% for testing.

4. Proceed with feature extraction and reduction by applying PCA, ICA, and NLICA. Each of these techniques transforms the data into a reduced set of independent components, which are expected to capture the most essential information from the original variables while minimising redundancy and noise. Therefore, the important question arises: How do we determine the optimal number of ICs that will lead to the highest accuracy? In our study, the number of components is determined using two approaches: a variance threshold, as used in PCA, or a systematic process of experimentation, where models are evaluated across a range of component numbers, as employed in both ICA and NLICA.

5. Then, we used the generated ICs obtained from the feature extraction and reduction techniques as input to train the SVR and MLP models. Perform hyperparameter tuning using grid search with cross-validation to optimise the performance of these models. The best model configuration is selected based on the cv grid search method's performance, measured by the lowest MSE.

6. Finally, six prediction models are developed: ICA_SVR, NLICA_SVR, PCA_SVR, ICA_MLP, NLICA_MLP and PCA_MLP. These models are then used to predict the price of carbon credit (ACCUs) on the test dataset. The performance of the models is assessed using five different metrics, MSE, RMSE, MAE, R^2, and execution time.

4.7 Overview of the proposed solution to achieve objective 4

In this section, we overview the solution to achieve objective 4: **Evaluate and test the methodologies developed in objectives 2 and 3.**

Assessing the performance of predictive models is essential for accurately predicting the price of carbon credit. However, models may not always be practical due to factors such as complexity, training time, and data requirements. Therefore, to select the most suitable model, a rigorous evaluation process is necessary, relying on well-established performance metrics. In this study, we apply multiple evaluation metrics to assess the effectiveness of the developed approaches. These metrics provide a quantitative basis for measuring accuracy and reliability, helping to identify the best-performing model.

where: C_t : the actual carbon credit prices

$\hat{C}_t$: the predicted carbon credit prices

N : the number of elements in the data

1 Mean squared error (MSE) 4.1

This value is defined as the mean squared of the difference between the actual data and the predicted data. [5], [84], [143]. The lower the MSE, the more accurate the proposed model [10].

$$MSE = \frac{1}{N}\sum_{t=1}^{n}(C_t - \widehat{C_t})^2 \quad \text{for } t = 1, 2, \ldots, N \tag{4.1}$$

2 Root mean squared error (RMSE) 4.2

The only difference between RMSE and MSE is the inclusion of a square root symbol [84]. The lower the RMSE, the more accurate the proposed model [10].

$$RMSE = \sqrt{\sum_{t=1}^{N}\frac{\left(\widehat{C_t} - C_t\right)^2}{N}} \quad \text{for } t = 1, 2, \ldots, N \tag{4.2}$$

3 Mean absolute error (MAE) 4.3

The MAE totals the residual error absolute values, which implies that the sign of the error (whether positive or negative) is not significant [84].

$$MAE = \frac{\sum_{t=1}^{N} \left| \widehat{C}_t - C_t \right|}{N} \quad \text{for } t = 1, 2, \ldots, N \tag{4.3}$$

4 R-Squared (R^2) 4.4

R^2 ranges from 0 to 1 and measures how well the model results are able to track actual data trends [5]. The accuracy of the model increases as the value approaches 1. In this thesis, R^2 measures how predicted values of the model close to real values.

$$R^2 = \frac{\sum_{t=1}^{N} \left(\widehat{C}_t - C_t \right)^2}{\sum_{t=1}^{N} \left(C_t - \overline{C_t} \right)^2} \quad \text{for } t = 1, 2, \ldots, N \tag{4.4}$$

5 Execution time: The execution time required for prediction is considered an additional metric when evaluating prediction models.

These five metrics were chosen because they are commonly used in the literature focused on predicting carbon credit prices [107], [119], [172] and offer different perspectives on evaluating the performance of prediction models. Further details on the evaluation and validation process for objective 2 and objective 3 are presented in chapters 7 and 8, respectively.

4.8 Conclusion

This chapter provided an overview of the solutions proposed to address research objectives 1 to 4. The design science research method was used as the backbone of our approach to address the objectives of this study in the context of predicting the price of carbon credits. The framework architecture of the overall solution was presented. In addition, this chapter describes the proposed solutions for each of the

established objectives, which outline the applied techniques and methodology. The next chapter identifies and analyses the external factors that influence carbon credit prices, which is the first objective of this research.

Chapter 5

Identifying the Factors Influencing Carbon Credit Prices

5.1 Overview

This chapter focuses on identifying external factors that influence carbon credit prices and constructing a detailed taxonomy of these factors. Given the volatility and complexity of carbon markets, this groundwork is essential to develop a more accurate and reliable prediction model for carbon credit prices. The chapter is structured as follows: Section 5.2 discusses the importance of identifying external factors that influence the price of carbon credits. Section 5.3 describes the methodology used to identify these factors, primarily through the SLR conducted in Chapter 2. This ensures a comprehensive and unbiased identification of the factors that impact carbon credit prices. Section 5.4 presents the results and discussions with an overview of the influencing factors identified from the articles reviewed in the SLR. Section 5.5 explores the reasons for conflict findings to better understand the market dynamics. Finally, the chapter concludes in Section 5.6. Several sections of this chapter have been published in the International Conference on Complex, Intelligent, and Software Intensive Systems, Springer Nature Switzerland, 2024, under the article titled "A Systematic Review of the External Influence Factors in Multifactor Analysis and the Prediction of Carbon Credit Prices".

5.2 The importance of identifying external influencing factors

Research indicates that many external factors influence the carbon market [54], [155], [162]. Factors such as energy prices, product prices, weather conditions, international carbon markets, and environmental factors significantly impact the price of carbon credits. These factors play a significant role in determining the volatility and dynamics of carbon credit prices [162], [171]. Understanding these diverse factors is essential for accurate carbon price prediction and risk management in carbon trading markets. The integration of external influences into prediction models improves the accuracy and robustness of the prediction of the price of carbon credits [155]. Furthermore, understanding the impact of external factors on carbon credit prices is essential for companies and governments to make informed decisions about managing their carbon emissions. Therefore, this chapter aims to construct a comprehensive taxonomy of these influencing factors, laying the groundwork for a more accurate and reliable prediction model for carbon credit prices.

5.3 Approach to identifying external factors

To investigate the factors influencing carbon credit prices, an SLR was conducted following Kitchenham's guidelines [77]. This methodological approach ensured a comprehensive and transparent review of the existing research within a defined scope [96]. This section outlines the approach used to identify the external factors that impact the price of carbon credits, including the literature search and the extraction and classification of factors, as shown in Figure 5.1 and presented in the following subsections.

Stage 1: Systematic literature review

This stage is based on the SLR methodology outlined in Chapter 2 to identify and analyse the external factors influencing carbon credit prices. A structured search was

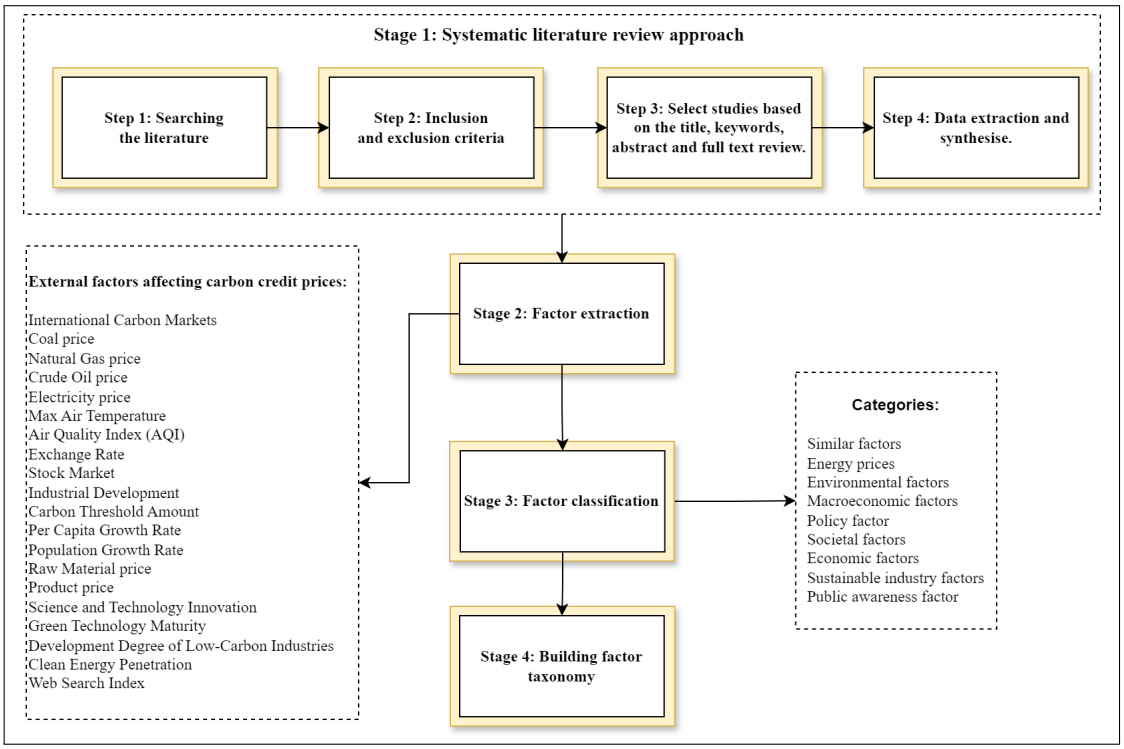


Figure 5.1 : Overview of the approach to identify the external factors influencing carbon credit prices.

performed on five major electronic databases: IEEE Xplore Library, SpringerLink, ScienceDirect, and ACM Digital Library, covering publications from 2005 to 2023 to align with the enforcement of the Kyoto Protocol. The selection process involved applying inclusion and exclusion criteria to ensure the relevance and quality of the studies. The titles, keywords and abstracts were screened, followed by a full-text review to refine the dataset. To enhance completeness, backward citation tracking was employed, leading to a final dataset of 52 studies for in-depth analysis. Zotero was used for reference management and article selection. The findings of this review form the foundation for understanding the key factors that affect carbon credit prices.

Stage 2: Factor extraction

This stage involved identifying factors that influence the price of carbon credits, as reported in the shortlisted articles. Data extraction and the synthesis of the 52 qualifying articles were carried out systematically. The information was organised in a spreadsheet, capturing details such as the trading market, methodologies, models, studied factors, and the impact of each factor.

Stage 3: Factor classification

This stage involved reviewing each identified factor to identify common themes or characteristics. Based on these themes, broad categories were created to encompass related factors. Each identified factor is placed into the most relevant category. This step may require refining the categories to ensure that they accurately reflect the factors. After all factors have been classified, the structure was reviewed to check whether some categories are too broad or too narrow, and adjustments were made by merging or splitting categories as needed. The final step was to confirm the categories, ensuring that each is distinct and covers all factors without significant overlap. This careful classification aids in the analysis and application of the data.

Stage 4: Building factor taxonomy

The final stage involved building a comprehensive taxonomy of the factors influencing carbon credit prices. This taxonomy provides a structured and organised overview of the external factors identified through the SLR, facilitating a thorough understanding of their impact and highlighting gaps in current research. This structured approach lays the foundation for future research in this critical field of environmental economics.

5.4 Results and discussion

The data analysis and validation process of the articles selected in the SLR not only facilitated a detailed understanding of the external factors that influence the price of carbon credits but also highlighted the gaps in current research, laying the groundwork for future investigations in this critical field. Using the applied process, we identified 21 external influence factors on the price of carbon credits. The factors extracted from empirical and conceptual studies are listed in Table 5.1. The table lists each external factor and its effects, categorised as positive, negative, no effect, or not specified based on the literature. Before presenting the description of the selected factors in each category, we begin by defining what negative, positive, no effect, and not specified mean in the context of the effect of each factor on the price of carbon credits as follows:

- **Positive effect:** This implies that an increase in the factor will lead to an increase in the demand for carbon credits, which increases the price of carbon credits. The number in this column represents the number of studies in the literature that suggest that this factor has a positive effect on the price of carbon credits.
- **Negative effect:** A negative effect denotes an inverse relationship between the factor and the price of carbon credits. This means that an increase in the factor either reduces the demand for carbon credits or indicates a shift towards lower-carbon alternatives, thereby decreasing the price of carbon credits. The number in this column represents the number of studies in the literature that indicate that this factor has a negative effect on the price of carbon credits.
- **No effect:** The number in this column represents the number of studies in the literature that found that this factor had no significant effect on the price of carbon credits.

- **Not specified:** The number in this column represents the number of studies in the literature that did not explicitly specify the effect of this factor on the price of carbon credits. This indicates that the available literature does not provide clear information on the impact of the factor.

Figure 5.2 visualises the data in the table showing the effect of each factor on the price of carbon credits, where blue indicates a positive effect, red indicates a negative effect, green indicates no effect, and purple indicates that the effect is not specified. It highlights the diverse effects of each factor on carbon credit prices, with some factors showing more consistent and significant impacts, while others exhibit mixed or uncertain effects according to the available literature.

Figure 5.3 classifies the factors identified through the SLR into nine categories: similar carbon markets, energy market prices, environmental influences, microeconomic conditions, policy, societal dynamics, economic factors, sustainable industry, and public awareness.

Table 5.1 : Influencing factors and their effect on the price of carbon credits based on the shortlisted articles of the SLR.

Factor Classif.	Influencing Factor	Positive Effect	Negative Effect	No Effect	Not Specified
Similar factors	International Carbon Markets	[43] [162]	[167]		[20] [54] [134] [154] [155] [160] [165] [171] [147] [149] [153] [169]
Energy prices	Natural Gas Market	[1] [8] [31] [43] [100] [160] [162]	[45] [75] [104] [91] [130] [136] [146] [152] [165] [167] [175]	[49]	[20] [24] [50] [53] [54] [88] [134] [144] [154] [155] [158] [171] [173] [147] [149] [153]
	Coal Market	[4] [43] [136] [150] [160] [161]	[1] [8] [45] [64] [68] [75] [100] [91] [130] [131] [144] [146] [152] [162] [165] [167] [173] [175]	[49]	[20] [24] [31] [50] [53] [54] [88] [134] [154] [155] [158] [171] [147] [149] [153]
	Crude Oil Market	[4] [8] [31] [64] [68] [100] [152] [161] [162] [165]	[49] [75] [91] [130] [136] [144] [160] [167] [175]		[20] [24] [50] [53] [54] [104] [134] [154] [155] [171] [173] [147] [149] [153] [169]
	Electricity Market	[1] [4] [8] [100] [162] [175]	[53]		[50] [54] [144] [149]
Environmental	Air Temperature	[8] [64] [75] [130] [150] [173]		[88] [137]	[20] [44] [50] [134] [154] [155] [147]
	Air Quality Index (AQI)	[81] [130] [145] [161]			[20] [54] [154] [155] [147]
	CO_2 emissions				[7]
Macroeconomic	Industrial Income	[161]	[81] [93] [130]	[165]	[88]
	Stock Market	[43] [81] [100] [91] [136] [152]	[31] [64] [68] [167] [175]		[20] [74] [134] [146] [154] [149] [153] [169]
	Exchange Rate	[49] [160] [162] [173]	[45] [152] [158]		[50] [54] [134] [146] [154] [155] [171]
Policy	Carbon Emission Threshold	[64] [89] [93] [130] [137] [162]	[81]		[7] [20] [44] [88] [131] [144] [153]
Societal	PerCapita Growth Rate	[81] [98]			
	Population Growth Rate	[98]			
Economic	Raw Material Price	[1]	[64]		
	Product Price	[64]			
	Science and Technology Innovation Index		[64] [150]		
Sustainable industry	Green Technology Maturity		[130]		
	Development Degree of Low-carbon Industries		[130]		
	Clean Energy Penetration	[81]	[130] [167]		
Public awareness	Web Search Index	[108] [144]			[20] [54] [74] [154] [155] [171]

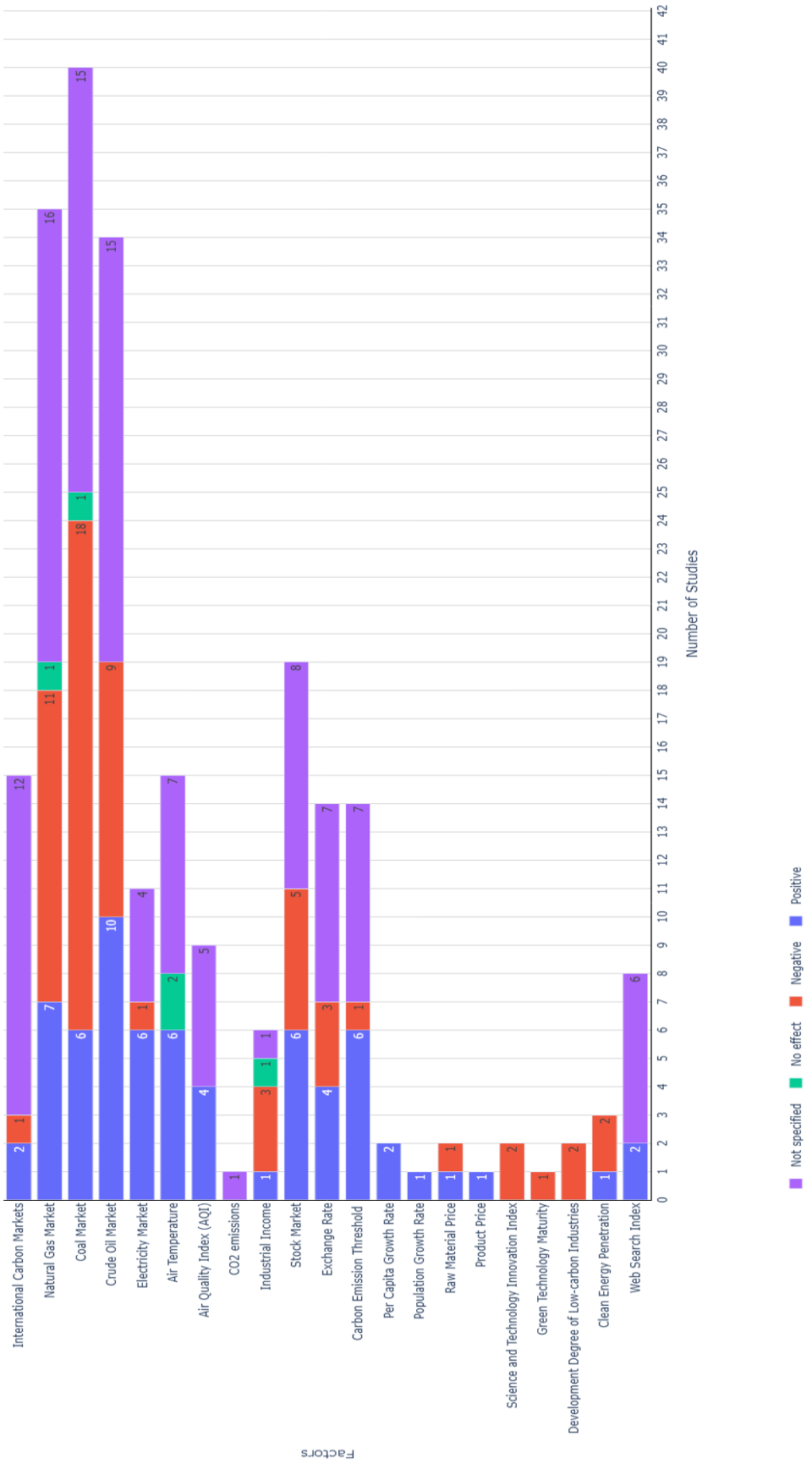


Figure 5.2 : Number of studies per effect of each influencing factor.

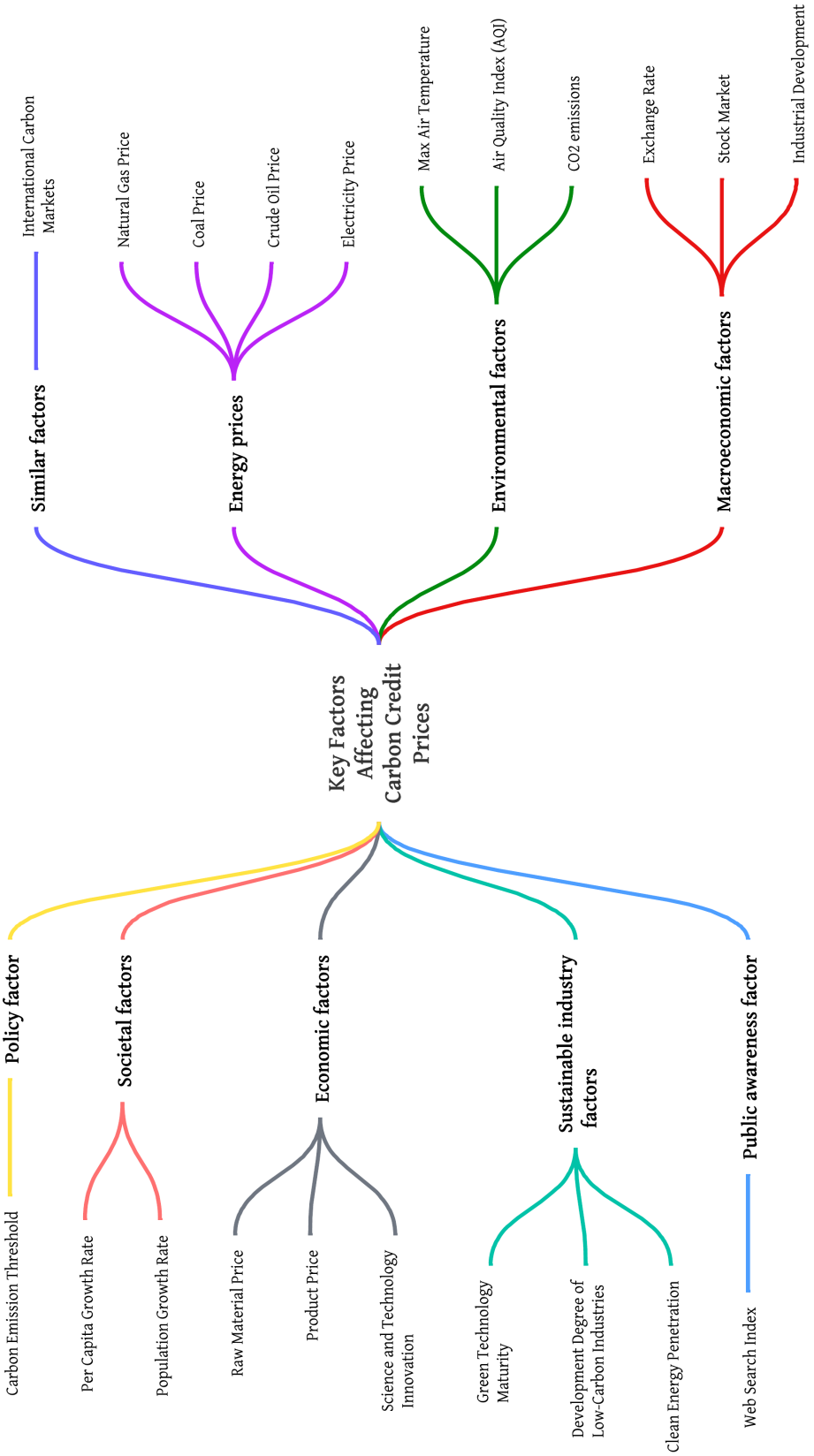


Figure 5.3 : The taxonomy of the identified influencing factors on carbon credit prices.

The following subsections examine each classification in detail, analysing the key factors and their influence on carbon credit prices based on the SLR.

5.4.1 Similar factors: international carbon markets

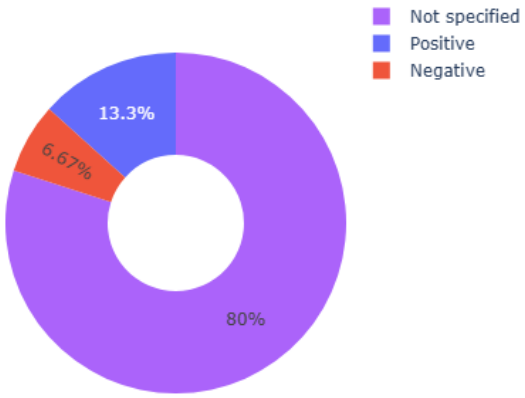


Figure 5.4 : The percentage of studies on the effect of the international carbon markets influencing factor in the similar factors category.

International carbon markets, especially the European Union Emission Trading System (EU ETS), play a significant role in shaping the carbon market. Established in January 2005 and now in its fourth phase, the EU ETS is the world's largest carbon market, providing a rich dataset and valuable insights into market trends and regulatory effects [37], [160]. Its scale and maturity make it a useful benchmark for understanding the dynamics that can also influence the price of carbon credits. In the EU ETS, two main products are traded: European Union Allowances (EUA), which are the rights to emit one tonne of CO_2, and Certified Emission Reductions (CER), issued for emission reduction projects in developing countries. In this study, EUAs are considered a similar external factor.

Table 5.1 shows that 15 studies used the EUA price as a benchmark for the global

carbon trading market due to its significant impact on carbon emission allowances. Of the 15 studies, 2 reported a positive effect, 1 a negative effect, and the remaining 12 left the effect unspecified. Figure 5.4 illustrates that 80% of these studies did not specify the final effect of this factor on the price of carbon credits. Instead, they incorporated EUA prices into prediction models without explicitly defining their impact.

5.4.2 Energy price factors

The energy markets substantially impact the price of carbon credits, as energy production and consumption are the primary sources of CO_2 emissions [91], [162], [163]. Most studies extensively explored the relationship between energy demand, carbon emissions, and carbon credit prices in the literature review. These studies investigated the impact of four key energy factors: natural gas, coal, crude oil, and electricity markets. Generally, when demand for fossil fuel-based energy sources like coal or natural gas rises, carbon emissions increase, thereby raising the demand and price of carbon credits [100], [136], [160]. The literature reveals diverse, sometimes conflicting, findings on the impact of these energy markets on the carbon market. Figure 5.5 presents the impact percentages for each of the four energy factors.

For the natural gas factor, 35 studies considered this factor, and around 45% of them incorporated it into their prediction models without explicitly analysing and reporting its effect on credit prices. However, 31% of the studies reported that the natural gas market had a negative correlation with the price of carbon credits.

For natural gas, 35 studies considered this factor, with approximately 45% incorporating it into prediction models without specifying its effect on carbon credit prices. Of the studies that reported effects, 31% found a negative impact, while 20% reported a positive correlation.

The coal market presented different findings among the 40 studies examined.



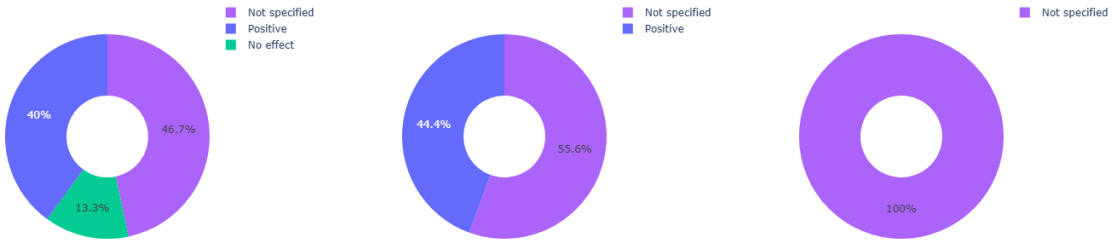
Figure 5.5 : The percentage of studies on each effect of the influencing factors in the energy prices category: (a) natural gas, (b) coal, (c) crude oil, and (d) electricity.

About 15% of the studies found a positive effect, while 45% indicated a negative effect, and 37% did not specify its impact on carbon credit prices.

For crude oil, 34 studies investigated its role of which 29% of the studies reported a positive effect on carbon credit prices, while 26% found a negative impact, and 44% left the effect unspecified.

Lastly, electricity generation contributes significantly to carbon emissions [1], [8], [100], [162]. An increase in the electricity demand from carbon-intensive sources is associated with a higher demand for carbon credits, leading to price increases [1], [8], [162]. In contrast, a shift towards cleaner energy sources, such as renewables, can help reduce carbon emissions [53]. This transition could decrease the demand for carbon credits, potentially lowering their prices. Of the 11 studies that examined this factor, 54% reported a positive effect, 9% indicated a negative impact, and 36% did not specify a clear correlation with the price of carbon credits.

5.4.3 Environmental factors



(a) (b) (c)

Figure 5.6 : The percentage of studies on each effect of the influencing factors in the environmental category: (a) air temperature, (b) air quality index (AQI), and (c) CO_2 emissions.

In the literature, we found three key environmental factors: air temperature,

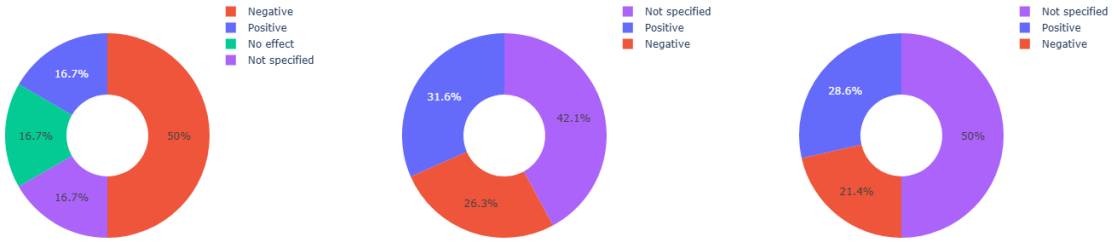
air quality index (AQI), and CO_2 emissions. The impact of air temperature and AQI on electricity production indirectly influences the price of carbon credits [8], [72]. Severe weather events, mild winters, and droughts can cause sudden fluctuations in fossil fuel energy production, leading to changes in demand for carbon allowances. For example, hotter summers can increase the demand for electricity, air conditioning, and coal, while colder winters can increase the demand for heating fuel and natural gas, resulting in a corresponding increase in carbon prices. Thus, the maximum temperature and the air quality index are selected to represent environmental factors. Figure 5.6 presents the impact percentages for each of the three environmental factors.

In the SLR, 15 studies considered the impact of air temperature on the price of carbon credits. Of these, 40% agreed that during extreme weather conditions, unanticipated temperature changes have a more positive impact on CO_2 prices than actual temperature variations, as hot and cold conditions cause abnormal and positive changes in demand. Two studies reported that carbon credit prices are not affected by air temperature [88], [137]. The remaining seven studies did not specify the type of impact, as they only included this factor to predict carbon credit prices without analysing its direct effect. However, some research suggests that the carbon market may be less sensitive to temperature fluctuations over time [88], [137].

Furthermore, AQI serves as a crucial indicator, quantifying levels of air pollution and concentrations of CO_2 and other pollutants. Changes in AQI have a direct impact on entities subject to carbon quotas; higher AQI values may increase their demand for carbon emission allowances, causing fluctuations in carbon trading prices [130], [161]. Elevated AQI levels, indicating degraded air quality and increased emissions, can increase the demand for carbon rights, thus influencing a rise in trading prices. In the SLR, nine studies considered AQI, with 55% not specifying the effect on carbon prices, while the remaining found that it had a positive impact.

Of the studies reviewed, Alanazi et al. [7] used CO_2 emissions as a variable within its prediction model for carbon trading prices but did not specify their direct impact. This approach suggests that CO_2 emissions could be a potentially influential factor in the modelling, although its exact effect on price volatility has not yet been explored.

5.4.4 Macroeconomic factors



(a) (b) (c)

Figure 5.7 : The percentage of studies on each effect of the influencing factors in the macroeconomic category: (a) industrial income, (b) stock market, and (c) exchange rate.

This category includes three macroeconomic factors: industrial income, stock market performance, and exchange rates. These factors can indirectly impact carbon credit prices by influencing production costs, industrial competitiveness, and investor sentiment. In the SLR, 39 of 52 studies examined these macroeconomic factors. Figure 5.7 presents the impact percentages for each of the three factors.

Exchange rates affect carbon credit prices by altering production costs and the competitiveness of different industries [8], [72], [105]. For example, a strong currency can make imports cheaper and exports more expensive, potentially lowering domestic demand for carbon credits. In this regard, 21% of the studies reported a negative

impact of exchange rates on carbon credit prices, while 29% observed a positive effect. However, 50% of the studies did not specify any impact.

The stock market influences the price of carbon credits through investor sentiment and economic growth [160]. A strong stock market may boost investor confidence and demand for carbon credits, whereas a weaker market may reduce it. Of the studies, 31% (6 studies) found a positive impact of the stock market, 26% (5 studies) reported a negative impact, and 42% (8 studies) did not specify any effect. In the EU ETS, studies show mixed results, some finding a positive correlation [91], [100], [136] and others an inverse relationship [31], [175], [167]. In China, similar mixed results were observed; Wu et al. [152] found a positive effect and Ji et al. [64] found a negative effect.

Industrial income is another key factor, as industrial development contributes to carbon emissions, impacting the demand and price of carbon credits. This can also affect the operating costs of companies by forcing them to adjust their operations to reduce emissions and remain competitive [88], [130], [161], [165]. Of the studies reviewed, 50% found a negative impact of industrial income on carbon credit price volatility [81], [93], [130]. However, Zeng et al. [161] observed a positive effect, while Zhang et al. [165] reported no effect, and Li and Lei [88] did not specify an impact.

5.4.5 Policy factors

Policy and regulations have a significant impact on the price of carbon credits, as they can create demand for carbon credits by setting emission limits or incentives to carbon reduction efforts [8], [23], [79]. Adopting policies that encourage carbon reduction can reduce demand and lower prices, while the absence of such policies can increase demand and prices. For example, a cap-and-trade system can create demand as companies with excess emission allowances can sell them as carbon credits to those exceeding their allocation [8].

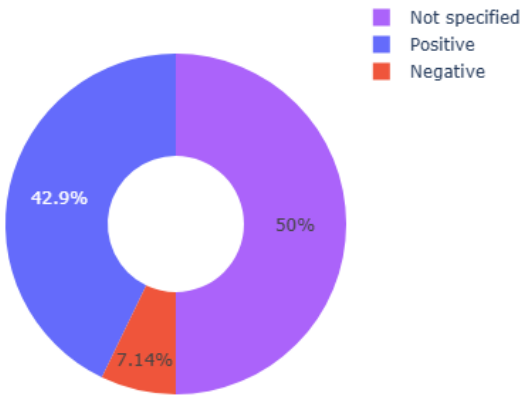


Figure 5.8 : The percentage of studies on each effect of the carbon emission threshold influencing factor in the policy category.

Policies that enforce carbon reduction can lower demand and prices, whereas the absence or reduction of such policies can have the opposite effect. For example, cap-and-trade systems create demand, allowing companies with excess emission allowances to sell carbon credits to those that exceed their allocation [8].

In the SLR, as shown in Figure 5.8, 14 studies examined the impact of policies on the price of carbon credits. Of these, 43% found a positive impact, with policies such as market mechanisms in the EU ETS contributing to price volatility [137]. Studies by Li et al. [89] and Ji et al. [64] found that green bonds, economic policy uncertainty, and the inclusion of new sectors in carbon markets can raise carbon prices, especially since companies actively trade allowances before compliance deadlines.

However, policies that allocate excess allowances or enable low-cost auction prices can reduce carbon credit prices. For example, Song et al. [130] observed that the issue of free carbon quotas and the establishment of carbon thresholds can positively influence trading prices. In the remaining 50% of studies, the effect was not explicitly

specified. For example, Song et al. [131] examined demand-related policies, Cao et al. [20] focused on uncertainty in economic policy, and Feng et al. [44] studied the role of market mechanisms. In addition, Li and Lei [88] evaluated government intervention and Wang et al. [144] analysed carbon emission allowances.

5.4.6 Societal factors



(a) (b)

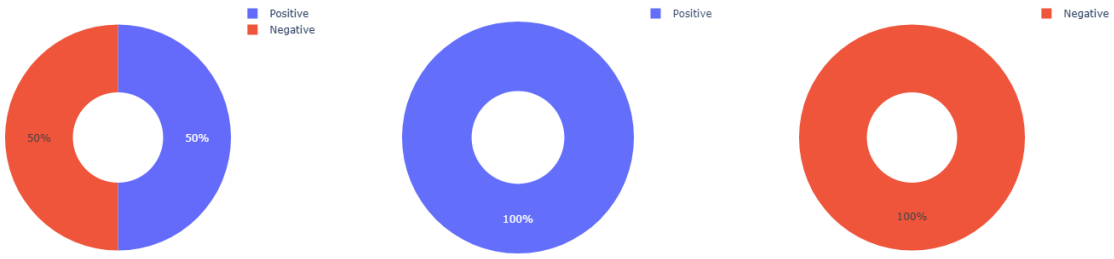
Figure 5.9 : The percentage of studies on each effect of the influencing factors in the societal category: (a) per capita growth rate, and (b) population growth rate.

Societal factors, such as per capita output growth and population growth rate, impact carbon emissions, which in turn affect the price of carbon credits [98]. Generally, a growing population and an expanding economy lead to higher carbon emissions, increasing the demand for carbon credits to offset these emissions, consequently increasing the price of carbon credits.

In the SLR, two studies examined these factors. Liu and Song [98] studied both per capita output growth and population growth rate, finding a positive influence of each on the price of carbon credits, as shown in Figure 5.9. Furthermore, Kotishwar

[81] included the growth in per capita output in their analysis and similarly reported a positive effect. These findings suggest that a growing population and an expanding economy typically result in higher carbon emissions, which in turn raises the demand for carbon credits to counterbalance these emissions.

5.4.7 Economic factors



(a) (b) (c)

Figure 5.10 : The percentage of studies on each effect of the influencing factors in the economic category: (a) raw material price, (b) product price, and (c) science and technology innovation index.

According to the existing literature, several key economic factors affect carbon credit prices, including raw material prices [1], [64], product prices [64], and the science and technology innovation index [64], [150]. In the SLR, 3 of 52 studies (approximately 6%) considered these factors. Figure 5.10 presents the impact percentages for each of the three factors.

Raw material prices, such as those of iron, cement, and chemicals, directly impact carbon credit prices because they affect production costs for industries that rely on carbon-intensive materials. When raw material prices increase, production costs increase. However, studies have reported contradictory effects on carbon credit prices. Ji et al. [64] found that raw material prices have a negative impact on

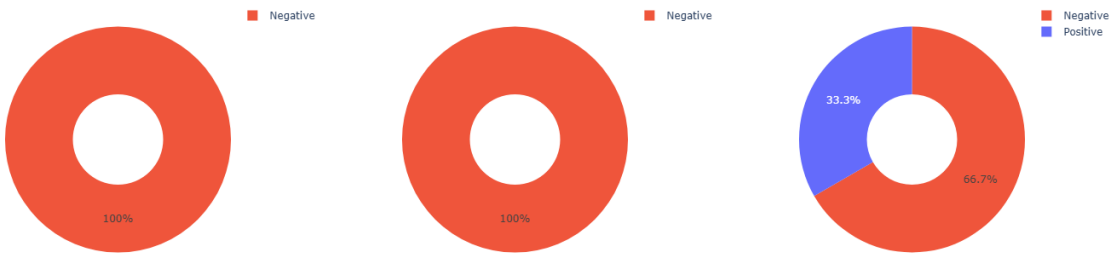
carbon credit prices, suggesting that higher costs could reduce the production and demand for carbon credits. In contrast, Aatola et al. [1] reported a positive impact, indicating that increased raw material costs can lead to higher carbon credit prices as expenses are passed on.

Product prices, including those of paper and minerals, also influence carbon credit prices. Industries producing carbon-intensive products can transfer the cost of carbon credits to consumers, leading to higher product prices. Ji et al. [64] found that product prices have a positive impact on carbon credit prices, which implies that as product prices increase, so do carbon credit prices.

The Science and Technology Innovation Index also affects the price of carbon credits [64], [150]. This factor reflects an industry's investment in innovative technologies and practices to reduce carbon emissions. Industries that invest in innovative technologies and practices to reduce carbon emissions are likely to have lower carbon credit prices compared to industries that do not prioritise sustainable practices. Adopting sustainable technologies and practices can help these industries reduce their carbon footprint, ultimately resulting in more affordable carbon credits. Both Ji et al. [64] and Wei, Qi et al. [150] found that this index has a negative impact on carbon credit prices, suggesting that greater innovation leads to reduced emissions and lower demand for carbon credits.

5.4.8 Sustainable industry factors

Under the sustainable industry category, the literature identifies three factors: green technology maturity, the development degree of low-carbon industries, and clean energy penetration. These factors play a significant role in influencing carbon emissions and carbon trading prices. In the SLR, only 3 out of 52 studies considered these factors. 5.11 presents the impact percentages for each of the three factors.



(a) (b) (c)

Figure 5.11 : The percentage of studies on each effect of the influencing factors in the sustainable industry category: (a) green technology maturity, (b) development degree of low-carbon industries, and (c) clean energy penetration.

Song et al. [130] examined the three factors and reported a negative impact for each, suggesting that as these sustainable industry factors improve, carbon credit prices tend to decrease. Similarly, Zhao et al. [167] found that clean energy penetration has a negative impact on carbon credit prices, indicating that greater adoption of clean energy reduces emissions and demand for carbon credits. In contrast, Kotishwar [81] found a positive impact of clean energy penetration, while reporting a negative impact for the development of low-carbon industries.

Green technology maturity serves as a metric to assess the level of advancement in emission-reducing technologies. As green technologies mature, their effectiveness in curbing emissions increases, leading to lower carbon trading prices [130].

The development of low-carbon industries also plays a key role in reducing carbon emissions. As industries expand, they generally contribute to lower emissions and thus decrease the demand and price for carbon credits [130].

Clean energy penetration measures the extent of clean energy use in the energy sector. Increasing the use of clean energy sources is correlated with a reduction in

carbon emissions and demand for carbon credits, which can lead to lower trading prices [130], [167].

5.4.9 Public awareness factor

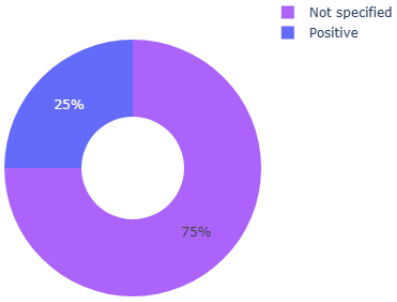


Figure 5.12 : The percentage of studies on each effect of the Web search index influencing factor in the public awareness category.

Public awareness, as reflected through factors such as the Web search index, can influence carbon credit prices by indicating societal interest in climate change and carbon reduction efforts. In the SLR, 8 of 52 studies (approximately 15%) investigated this factor. Of these, 25% (2 studies) reported a positive impact of public awareness on carbon credit prices [108], [144], while the remaining 75% (6 studies) included this factor without specifying its direct effect [20], [54], [74], [154], [155], [171].

The Web search index, in particular, serves as an indicator of public interest, indirectly affecting carbon credit prices by potentially influencing policy changes or regulatory decisions. Unstructured data from search indices provide insights into public sentiment and engagement on topics related to carbon credits, carbon markets, climate change, and renewable energy. Analysing keyword trends associated with these themes can reveal shifts in public opinion that can shape government

policies, thereby affecting demand for carbon credits and potentially driving price fluctuations.

Of the studies reviewed, most found a correlation between keyword query volumes related to carbon credits and price movements in the carbon market. However, only Wang et al. [144] specified a positive influence, indicating that increased public interest, as shown by search index data, can increase carbon credit prices.

Although this chapter identifies and classifies a wide range of quantitative and qualitative factors from the literature, qualitative factors such as policy and public awareness are not included in the final predictive models. These factors can significantly influence carbon credit prices, and their exclusion may limit the comprehensiveness of the analysis. Future studies could incorporate these qualitative elements using appropriate methodologies, providing a more comprehensive understanding of the factors influencing the price of carbon credit.

5.5 Investigating the reasons for conflicting literature on the role of external factors

This subsection aims to clarify the potential reasons behind the conflicting findings in the literature regarding the impact of external factors on the price of carbon credits. By examining these reasons, we can gain a deeper understanding of the underlying complexities that influence these diverse relationships.

- **Complexity of the carbon trading market**

  The carbon trading market is a complex and multifaceted system that involves interactions between various entities, policies, and market participants. This complexity can lead to different results in different studies, as a combination of dynamic and interconnected variables can influence the impact of each factor. For example, in 2018, the EU ETS faced a surplus of allowances, leading to

a decrease in carbon prices. However, with the implementation of the Market Stability Reserve (MSR) to address this surplus, prices began to increase again [17]. This example illustrates how a combination of policy decisions, market supply, and demand can interact in complex ways to influence carbon prices.

- **Regional variations and context**

  The carbon market operates within different geographical regions, each with its own set of environmental, economic and policy contexts. As a result, the impact of factors on carbon credit prices can vary from region to region, leading to conflicting findings in studies conducted in different regions [92]. For example, the differences in carbon credit prices between the California Cap-and-Trade Programme and the EU ETS can be attributed to regional variations. California's programme has generally experienced higher and more stable prices than the EU ETS, due to the fact that California has a controlled market.

- **Time period considerations**

  The carbon market continuously changes over time due to evolving policies, technological advances, and economic conditions. Studies conducted during different time periods can capture various market dynamics, contributing to discrepancies in the findings. For example, the Clean Development Mechanism (CDM), a global carbon market under the Kyoto Protocol, experienced significant fluctuations in the price of carbon credits over time. Prices were relatively high in the mid-1900s, but decreased drastically after 2012 due to uncertainties about the future of the Kyoto Protocol and the oversupply of credits [35]. This reflects how changes in global climate policy and market dynamics over different time periods can lead to varying impacts on carbon credit prices.

- **Data and methodological differences**

  Variations in data sources, data quality, and methodological approaches used in different studies can lead to inconsistent results. Differences in model specifications, statistical methods, and the selection of variables can influence the observed relationships between factors and the price of carbon credits. For example, in light of this SLR, it can be seen that hybrid approaches often provide more accurate and robust results compared to traditional statistical methods [54],[167] .

- **Policy, regulations, and market maturity**

  Variations in carbon trading policies, regulations, and government interventions between regions can influence the resulting impact on carbon credit prices. Varied policy environments may lead to conflicting results in studies. A comparison between the EU ETS and China's National Emissions Trading System (ETS) illustrates the impact of policy and regulatory differences on carbon credit prices. The EU ETS has undergone several phases and reforms to address issues such as over allocation of allowances and carbon leakage. These policy adjustments have gradually led to a more robust system and higher carbon prices [17]. In contrast, China's National ETS, launched in 2021, is still in its infancy and covers only the energy sector [87]. Its trading rules and allocation plans are relatively less stringent than those of the EU ETS. Consequently, carbon prices in China's ETS have been considerably lower and less volatile in its initial trading period.

5.6 Conclusion

In conclusion, this chapter has systematically identified and analysed the external factors that influence the price of carbon credits. Through a detailed taxonomy of these factors, we have established a comprehensive framework that enhances our

understanding of market dynamics. The methodology, grounded in SLR, has ensured a robust and unbiased identification of these external factors. The discussion highlights the complexities and variability within the market, addressing conflicting findings. This foundational work sets the stage for further research and application in the development of accurate and reliable prediction models for carbon credit prices. The next chapter will delve deeper into the practical application of the proposed data pre-processing and feature engineering solutions.

Chapter 6

Feature Engineering: Data Collection, Data Preparation and Exploratory Data Analysis

6.1 Overview

In Chapter 1, we presented an introduction to the thesis, followed by an extensive literature review in Chapter 2. Chapter 3 explained the problem definition, followed by an overview of the proposed solutions in Chapter 4. Chapter 5 identified the external influence factors on the carbon credit market. The primary objective of this chapter is to establish a solid foundation for building robust predictive machine learning models by emphasising the critical roles of feature engineering and data preprocessing. Pre-processing is a critical step in data analysis and machine learning, contributing to the robustness and validity of our research findings. It involves preparing and cleaning the data to make them suitable for analysis and modelling, thus producing reliable and accurate results. The structure of this chapter is as follows: The importance of using feature engineering techniques is highlighted in Section 6.2. Section 6.3 explains the dataset used for feature engineering. Data preprocessing and data analysis are discussed in Section 6.4. Section 6.5 examines the data challenges. Finally, Section 6.6 concludes the chapter.

6.2 The importance of feature engineering

Feature engineering involves creating informative features that capture relevant information and transforming the data to uncover hidden patterns. It is an essential stage in constructing robust and effective predictive machine learning models, allow-

ing them to make more accurate and reliable predictions [16]. In addition, feature engineering is critical to address data quality issues such as missing values, outliers, and skewed distributions. Without proper data pre-processing, messy data can lead to unreliable and inaccurate model performance. In this chapter, techniques such as imputation, outlier detection, and data transformation are applied, ensuring that the model learns from clean and representative data. In general, the role of feature engineering in the machine learning pipeline cannot be overstated; it bridges the gap between raw data and actionable insights derived from machine learning models, making it a fundamental component of data science and machine learning workflows [82], [168].

6.3 Dataset for feature engineering

Building on the external factors influencing carbon credit prices identified in the previous chapter, this section outlines the data sources for each factor. The datasets used in this study are real daily records from January 2, 2020, to September 1, 2023, totalling 667 data samples. At the time of writing, this dataset represents the most current and extensive collection of historical data available. Table 6.1 provides a clear reference for the data used in this study; it offers descriptions and specific sources for each factor.

6.3.1 Carbon market

This thesis aims to predict the price of carbon credits. The case study of this thesis is Australian Carbon Credit Units (ACCUs). One tonne of carbon dioxide is represented by one ACCU, issued by the Australian authorities [25]. Therefore, the target variable to be predicted is the price of ACCU. A real-world carbon market dataset was used to validate the proposed model. Daily ACCU price data were collected from the JARDEN platform [63], a recognised platform that provides in-

dicative prices for ACCUs from January 2, 2020, to September 1, 2023, consisting of 667 data samples.

6.3.2 External factors

Carbon credit prices are influenced by various external factors. Based on a comprehensive review of the literature in Chapter 5, 22 influencing factors were identified and real data for each factor were collected from reliable and authoritative sources. Table 6.1 outlines the selected variables and their data sources. These factors include mixed-frequency data, such as monthly, weekly and daily records. The selection of variables was guided by relevance and data availability. The factors are categorised into seven groups: similar factors, energy factors, environmental factors, macroeconomic factors, societal factors, economic factors and sustainable industry factors. Details of the data collection process for these factors are provided in the following.

- **Similar factors:** The international carbon market plays a major role in determining carbon credit prices. Two key indicators selected in this category are the EUAs and CEAs, both sourced from the World Bank. The EUA data, representing the largest cap-and-trade system globally, reflects the impact of the mature EU carbon market on global prices. The currency of the daily closing price of EUA is EUR. Similarly, CEA data provides insights into the Chinese carbon market, which is still developing and has significant implications for global carbon trading dynamics. Considering the recent establishment of China's national carbon market in 2022, historical data from January 1, 2017, to March 1, 2022, were compiled to calculate the average carbon prices in seven provinces (Beijing, Chongqing, Fujian, Guangdong, Hubei, Shanghai, and Shenzhen), which had established regional carbon markets before the national scheme. Post-establishment data, from December 22, 2022,

to September 6, 2023, reflect prices from CEAs. This provides a comprehensive view of carbon pricing in China, from regional variations to the national market.

- **Energy factors:** Energy markets significantly influence carbon credit prices. This study considers four key energy-related factors: natural gas, coal, crude oil, and electricity prices. The natural gas price in Australia (AU), sourced from IndexMundi, captures shifts in the natural gas market. Similarly, Australia's BJ power coal spot prices and global average crude oil prices, both obtained from the World Bank, reflect the impact of the coal and crude oil markets, respectively. Lastly, Australia's electricity spot prices, sourced from the Australian Energy Regulator (AER), highlight the role of the electricity market in shaping carbon credit prices.

- **Environmental factors:** Environmental conditions significantly affect carbon credit prices. This study considers five environmental factors: air temperature, AQI, CO_2 emissions, GHG emissions, and renewable energy. Data on Australia's air temperature is sourced from the Bureau of Meteorology (BOM), providing insights into how temperature variations impact carbon pricing. AQI in Australia, with data sourced from BOM, is another critical factor. Australia's GHG emissions data, sourced from Macrotrends, and Australia's CO_2 emissions data, also from Macrotrends, are critical for understanding carbon pricing dynamics. Finally, data on renewable energy in Australia, sourced from Macrotrends, indicates the influence of renewable energy adoption on carbon credit prices and is calculated as the percentage of electricity generated from renewables.

- **Macroeconomic factors:** Macroeconomic conditions are vital in influencing the price of carbon credits. This study considers three macroeconomic factors:

industrial development, stock market and exchange rate. The Business Confidence Indicator, sourced from OECD, reflects the impact of industrial development. The S&P/ASX 200 Index price, sourced from the Reserve Bank of Australia, represents stock market conditions and their influence on carbon pricing. The AUD/USD exchange rate data, sourced from the Wall Street Journal Markets (WSJ), indicates the impact of currency exchange rates on carbon credit prices.

- **Societal factors:** Societal factors such as economic growth and population changes influence carbon credit prices. This study considers Australia's per capita GDP growth rate and population growth rate, both sourced from the World Bank, to assess the impact of economic and demographic changes on carbon pricing.

- **Economic factors:** In this study, we consider three economic factors: raw materials, product prices, and science and technological innovation. Australia's iron ore prices, sourced from IndexMundi, provide insights into how raw material markets affect carbon credit prices. In addition, the prices of non-ferrous metal products in Australia, obtained from the Australian Bureau of Statistics (ABS), highlight the manufacturing sector's influence on carbon pricing. Finally, the percentage of households with internet access in Australia, sourced from OECD reflects the impact of science and technological innovation on carbon credit prices.

- **Sustainable industry factors:** This study considers three sustainable industry factors: green technology maturity in industries, low carbon industries and clean energy penetration in industrial activities. The number of patents on environmental technologies, sourced from the World Bank, indicates the maturity of green technology. Data on Australia's industry emissions, sourced from Statista, reflects the extent of low carbon industry development and

its impact on carbon pricing. Finally, the use of renewable energy in industry, with data sourced from OECD, indicates the penetration of clean energy in industrial activities and its influence on carbon credit prices.

6.4 Applying data pre-processing techniques

Pre-processing is a critical step in data analysis and machine learning, contributing to the robustness and validity of our research findings [16]. It involves preparing and cleaning the data to make them suitable for analysis and modelling, thus producing reliable and accurate results. The specific pre-processing steps often vary depending on the dataset and how it will be used in an application. Data pre-processing is achieved by applying the following steps to the dataset, as shown in Figure 6.1.



Figure 6.1 : Overview of data pre-processing steps.

Step 1: Data collection

The price of carbon credits from three markets (Australia, Europe, and China), along with datasets for the external factors, as identified in the previous section, were obtained from January 2, 2020, to September 1, 2023. All datasets were merged into a single CSV file, with the 'Date' column set as the index.

Step 2: Unifying the time frame

To ensure consistency across all datasets and enable accurate time series analysis, we converted variables with different frequencies (quarterly, monthly, and yearly) into a daily format. This transformation was performed using the forward fill (ffill) method, which carries the most recent available value forward until a new observation is recorded. This approach assumes that values remain constant between observations, reflecting the most recent known market state, and was preferred over interpolation methods, which could introduce unrealistic trends.

Forward fill is widely used in financial time series analysis to handle irregular or non-daily data without introducing artificial trends or structural breaks [139]. Forward fill was preferred over linear interpolation because economic and market-driven factors, such as carbon credit prices, often react suddenly to market events rather than change gradually; linear interpolation could introduce unrealistic trends between observations [86]. On the contrary, forward fill preserves the most recent actual market state, making it particularly appropriate for economic and financial factors that react suddenly to market shocks rather than evolve smoothly, as observed in carbon credit markets [11].

By applying forward fill, we ensure that all features are aligned daily while maintaining the statistical properties of the original dataset. Moreover, since economic and market-driven factors influencing carbon markets operate on different timescales, aligning them with daily data using forward fill ensures temporal con-

sistency without introducing artificial volatility [57].

Step 3: Simplifying feature names

For clarity and ease of reference, the target name was updated as ACCUs, and the feature names were updated as follows:

- International carbon market (European Union Allowances) - EUAs
- International carbon market (China Emission Allowances) - CEAs
- Australia natural gas - NG
- Australia BJ power coal spot prices - CP
- Average global crude oil - CrP
- Australia electricity spot prices - EP
- Australia air temperature - MA
- Australia Air Quality Index - AQI
- Australia greenhouse gas emissions - GHG
- Australia carbon (CO2) emissions - CO2
- Australia renewable energy - RE
- Business confidence indicator - BC
- S&P/ASX 200 Index price - SP
- AUD/USD exchange rate - AUD
- Australia per capita GDP growth rate - GDP
- Australia population growth rate - POP

- AU iron ore - IO
- AU Basic non-ferrous metal product manufacturing - NONF
- Australia households internet access percentage - IA
- Patents on environmental technologies - PT
- Australia's industry emissions - IND
- Renewable usage in industry - REI

Step 4: Data cleaning

This step involves fixing incomplete or irrelevant parts of the data. Irrelevant details that could clutter and confuse the analysis were eliminated. Symbols such as "$" and "," were removed from the numerical values to allow proper mathematical operations and analysis.

Step 5: Unifying currency

The price series was converted to AUD$ using the daily exchange rate provided by https://www.foreignexchange.org.uk/ to ensure that all factors and carbon credit price series are traded in the same currency.

Step 6: Data normalization

In this study, we applied the Min-Max scaler to normalise the dataset. The Min-Max scaler is a commonly used normalization technique that rescales feature values to a specified range while maintaining the relative differences between the original data points. Normalising the dataset is essential as it ensures that all features contribute equally to the model, preventing any single feature from dominating the learning process due to its scale [19], [164]. The formula for Min-Max scaling is as

follows 6.1:

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{6.1}$$

where X' is the scaled value, X is the original value, $X_{\min}$ is the minimum value of the feature and $X_{\max}$ is the maximum value of the feature.

The Min-Max scaler is particularly suitable for our dataset for several reasons. Firstly, the features in our dataset exhibit varying units and scales, which could disproportionately influence the results of machine learning algorithms. Normalizing these features ensures that each one contributes equally to the model, preventing any single feature from dominating due to its larger numerical range and mitigating the impact of different feature scales and units on the feature selection process. Secondly, some machine learning algorithms, such as RF, SVR and neural networks, perform better and converge faster when the input data are normalized [16], [22]. Lastly, Min-Max scaling preserves the original distribution and relationships of the data, which is important for maintaining the integrity of the dataset's underlying patterns. In this study, we used the range [0, 1] to scale the feature.

Step 7: Exploratory data analysis (EDA)

EDA helps to gain insight into the nature of the data, which can be helpful for further steps in the data analysis process [12]. EDA was used to create visualisations to explore data distributions, relationships between variables, and potential patterns. EDA helps uncover data issues such as missing values and outliers.

Step 8: Handling missing values

After performing the EDA, it was confirmed that there are no missing values in the dataset, ensuring the integrity of the data for further analysis.

Step 9: Handling outliers

This study identified outliers in the data set using a visualisation method based on interquartile ranges. This study used the box-and-whisker plot approach (box plot) to detect outliers in each feature. Following common practice in financial research [6], [28], we chose not to remove these outliers from the dataset, including carbon credit prices and external factors, because they reflect the real-world volatility of the carbon market. These extreme values often capture important events, such as sudden weather changes or economic shifts, that can heavily influence prices. For example, in the air quality index (AQI) data for Australia, we found extremely high values caused by bushfire events that occur frequently in the Australian climate. Removing such values might mean losing critical insights. In addition, the selected machine learning models are robust and designed to handle these variations, ensuring that our analysis remains realistic and comprehensive.

Table 6.1 : Description and source of the external factors.

Category	**Factors**	**Description**	**Notation**	**Data Unit**	**Data Source**	**Frequency**	**Citation**
Similar factor	International carbon market	European Union Allowance (EUAs)	EUAs	AUD/ton of CO2	World Bank	daily	[91, 163]
		China Emission Allowances (CEAs)	CEAs	AUD/ton of CO2	World Bank	daily	[160]
Energy factors	Natural gas market	Australia natural gas	NG	AUD/Therm	IndexMundi	monthly	[91, 146, 148, 163]
	Coal market	Australia BJ power coal spot prices	CP	AUD/tonne	World Bank	monthly	[91, 146, 148, 163]
	Crude oil market	Average global crude oil	CrP	AUD/BBL	World Bank	daily	[91, 148, 159, 163]

Continued on next page

Table 6.1 – continued from previous page

Category	Factors	Description	Notation	Data Unit	Data Source	Frequency	Citation
	Electricity market	Australia electricity spot prices	EP	AUD/MWh	Australia energy regulator	weekly	[163]
Environmental factors	Air temperature	Australia air temperature	MA	Celsius (°C)	Australian government bureau of Meteorology	daily	[64, 88, 130, 150]
	Air Quality Index (AQI)	Australia Air Quality Index (AQI)	AQI	Index	Australian government bureau of Meteorology	hourly	[130, 137, 161]
	Carbon (CO2) emissions	Australia carbon (CO2) Emissions	CO2	Metric tons/capita	Macrotrends	yearly	[7]
	Greenhouse gas emissions	Australia greenhouse gas GHG emissions	GHG	Metric tons/capita	Macrotrends	yearly	newly introduced in this study
	Renewable energy	Australia renewable energy	RE	% of Electricity from Renewables	Macrotrends	yearly	newly introduced in this study
Macroeconomic factors	Industrial development	Business confidence indicator	BC	Index	OECD Data	monthly	[88, 130, 161, 165]
	Stock market	S&P/ASX 200 Index price	SP	Points	Reserve bank of Australia	daily	[91, 148]
	Exchange rate	AUD/USD	AUD	AUD Currency	WSJ Markets	daily	[91, 146, 158]
Societal factors	Per capita growth rate	Australia per capita GDP growth rate	GDP	Percentage %	World Bank	quarterly	[98]
	Population growth rate	Australia population growth rate	POP	Percentage %	World Bank	yearly	[98]
Economic factors	Raw material price	AU iron ore	IO	AUD/Dry Metric Ton	IndexMundi	monthly	[64]

Continued on next page

Table 6.1 – continued from previous page

Category	Factors	Description	Notation	Data Unit	Data Source	Frequency	Citation
	Product price	AU Basic non-ferrous metal product manufacturing	NONF	Index Number	Australian Bureau of Statistics	quarterly	[64]
	Science and technology innovation	Australia households internet access percentage	IA	Percentage %	OECD Data	monthly	[64, 150]
Sustainable industry factors	Green technology maturity	Patents on environmental technologies	PT	Number	World Bank	yearly	[130]
	Development degree of low carbon industries	Australia's industry emissions	IND	Million kilograms	Statista	yearly	[130]
	Clean energy penetration	Renewable usage in industry	REI	Percentage %	OECD Data	yearly	[130]

6.4.1 Results and discussion

This section presents the results of the preprocessed data on carbon credit prices and the identified factors.

Figure 6.2 shows that using Python's $info()$ function, the dataset contains 23 columns: the dependent variable ACCUs and the 22 independent variables.

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 667 entries, 2020-02-01 to 2023-01-09
Data columns (total 23 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   ACCUs   667 non-null    float64
 1   EUAs    667 non-null    float64
 2   CEAs    667 non-null    float64
 3   NG      667 non-null    float64
 4   CP      667 non-null    float64
 5   CrP     667 non-null    float64
 6   EP      667 non-null    float64
 7   MA      667 non-null    float64
 8   AQI     667 non-null    float64
 9   GHG     667 non-null    float64
 10  CO2     667 non-null    float64
 11  RE      667 non-null    float64
 12  BC      667 non-null    float64
 13  SP      667 non-null    float64
 14  AUD     667 non-null    float64
 15  GDP     667 non-null    float64
 16  POP     667 non-null    float64
 17  IO      667 non-null    float64
 18  NONF    667 non-null    float64
 19  IA      667 non-null    float64
 20  PT      667 non-null    float64
 21  IND     667 non-null    float64
 22  REI     667 non-null    float64
dtypes: float64(23)
```

Figure 6.2 : Overview of the dataset columns.

Descriptive statistics form the basis for data analysis. Table 6.2 lists the statistical summary of the dataset by applying the *described*() function in Python. Skewness refers to the asymmetry of a distribution compared to a normal distribution. Values outside the range of -1 to 1 are considered significantly skewed [62]. In this analysis, EP (1.29), AQI (15.81), SP (-1.07), GDP (-1.51), POP (1.51), IND (-1.46) and REI (-1.45) exhibit substantial skewness, indicating that their distribu-

tions are skewed. The remaining constructs fall within the acceptable threshold.

Table 6.2 : Descriptive statistics of the dataset.

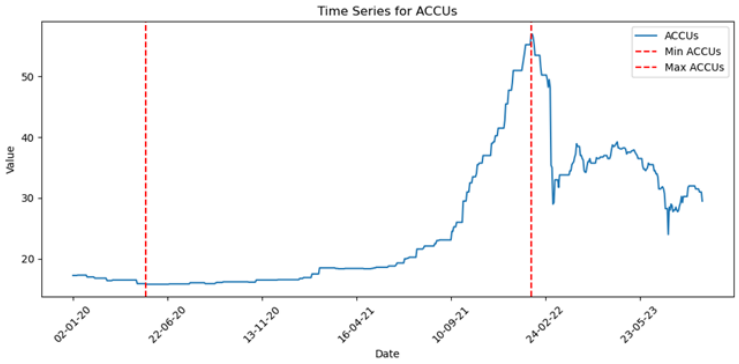
Variable	Count	Mean	Std	Min	25%	50%	75%	Max	Skew.
ACCUs	667	25.43	10.94	15.8	16.5	18.6	34.5	57	1.02
EUAs	667	105.56	53.70	29.62	52.02	101.48	163.72	203.59	0.30
CEAs	667	8.47	2.62	3.19	6.47	7.89	10.57	15.87	0.58
NG	667	4.09	1.50	2.35	3.16	3.39	5.23	8.15	0.98
CP	667	129.50	70.65	50.14	66.74	129.97	169.65	379.23	0.98
CrP	667	61.60	19.25	11.25	42.95	66.37	75.65	123.7	-0.26
EP	667	68.52	35.99	22.8	42.2	56.8	87.4	208.8	1.29
MA	667	23.62	4.67	10.1	20.2	23.2	26.8	42	0.41
AQI	667	48.68	144.86	12.5	27.35	32.37	40.32	2912.96	15.81
GHG	667	-6.81	0.06	-6.9	-6.87	-6.83	-6.79	-6.7	0.72
CO2	667	14.59	0.14	14.33	14.57	14.63	14.7	14.77	-0.91
RE	667	12.17	0.11	12	12.09	12.17	12.3	12.3	-0.30
BC	667	100.33	1.13	97.86	99.81	100.34	101.39	101.91	-0.47
SP	667	6811.21	653.90	45	6393.6	7064.7	7312.65	7628.9	-1.07
AUD	667	0.69	0.03	0.57	0.66	0.69	0.71	0.75	-0.35
GDP	667	0.93	2.84	-6.9	0.2	0.5	3.4	4.2	-1.51
POP	667	1.03	0.07	0.98	0.99	1	1.06	1.23	1.51
IO	667	132.48	34.91	84.73	110.2	122.91	155.43	214.43	0.94
NONF	667	136.82	23.35	104.6	114.2	131	160.3	173	0.11
IA	667	92.85	1.94	90.4	90.4	93.15	95	95	-0.22
PT	667	2592.89	44.31	2507.63	2572.34	2599.59	2630.25	2657.5	-0.51
IND	667	36.70	0.68	34.81	36.35	37.12	37.12	37.12	-1.46
REI	667	8.21	0.23	7.56	8.09	8.35	8.36	8.36	-1.45

Figure 6.3 confirms that there are no missing values in the dataset, ensuring the integrity and completeness of the data for subsequent analyses.

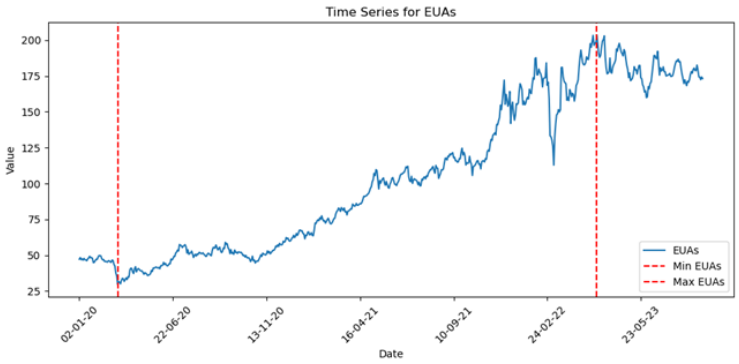
```
Old data frame length (number of observations): 667
New data frame length (number of observations after dropping NAs): 667
Total number of NA values: 0
```

Figure 6.3 : Missing value checking on the dataset.

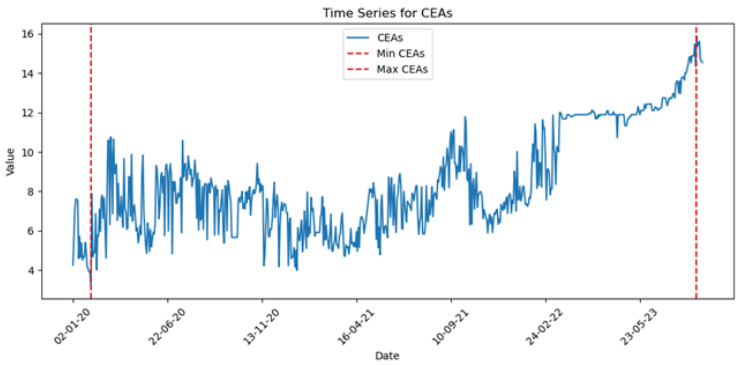
Figure 6.4 presents the lowest and highest recorded closing prices of ACCUs, EUAs, and CEAs. They illustrate the market's volatility, providing insights into the potential risks and rewards in carbon credit markets.



(a)



(b)



(c)

Figure 6.4 : Minimum and maximum closing of the carbon credit prices (a) ACCUs, (b) EUAs, and (c) CEAs.

The line graphs of the dataset are presented in Figure 6.5. These figures show how each feature has changed over time.

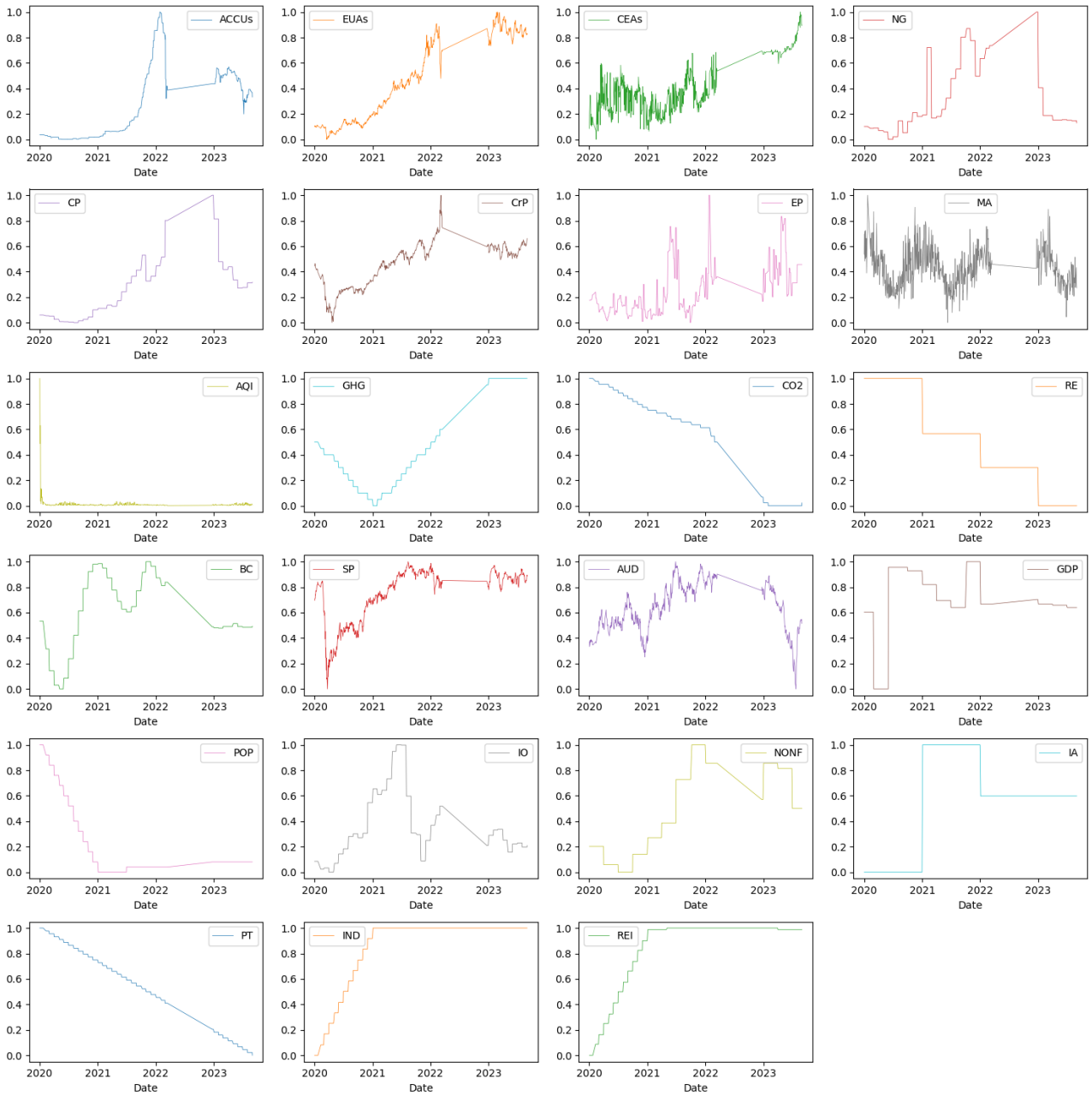


Figure 6.5 : Line plots for the external factors.

Figure 6.6 represents the histograms of each feature in the data set, providing us with an understanding of the frequency and distribution of the values for each attribute. We can quickly identify trends, patterns, and variations in the data by viewing these histograms.

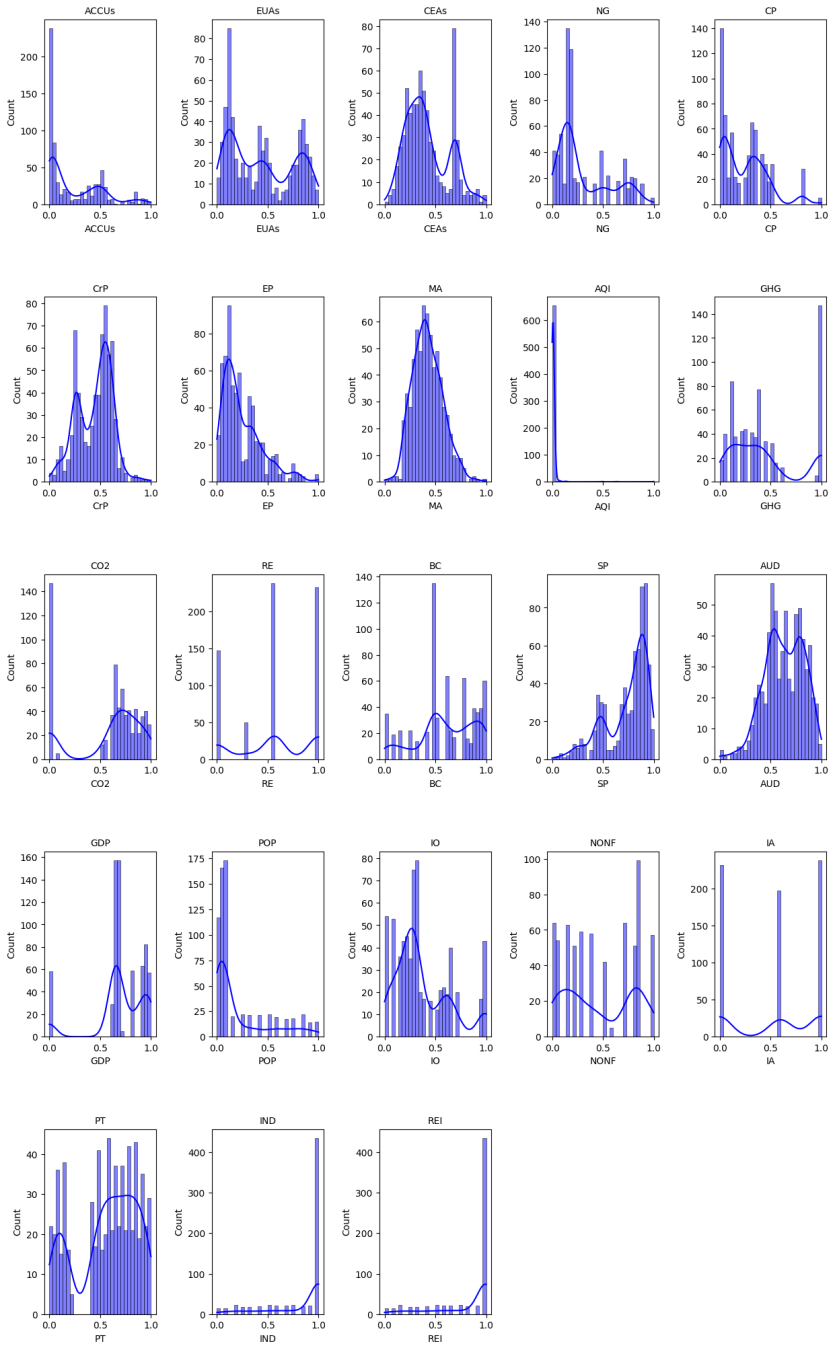


Figure 6.6 : Histogram of ACCUs and all influencing factors.

Figure 6.7 shows the box-and-whisker plots for all features, highlighting their distribution, variability, and outliers. Outliers were retained because they reflect real-world events that can significantly influence market behaviour. To limit their disproportionate impact while preserving their value, MinMax normalization was applied to scale all features to a range [0, 1]. The machine learning models used are robust to such anomalies and can effectively capture complex, non-linear relationships that include the effects of outliers.

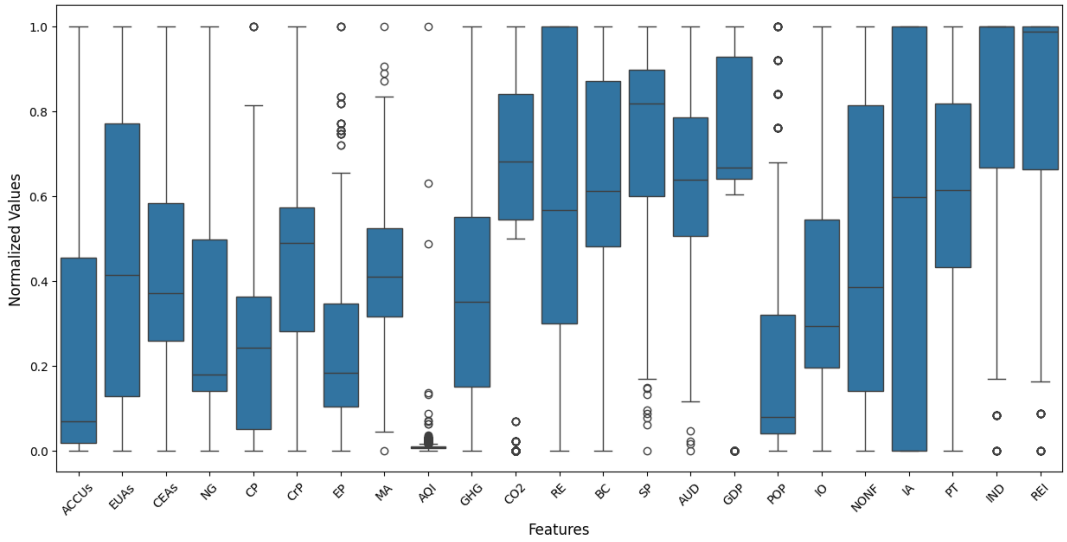


Figure 6.7 : Box-and-whisker plots of ACCUs and all influencing factors.

6.5 Checking the data characteristics and challenges

The biggest challenge with carbon credits is that traders find it very difficult to predict carbon prices accurately, as the price of carbon credits over time exhibits non-linearity, non-stationary, and multicollinearity characteristics [105]. To confirm the complexity of our dataset, we conducted three specific tests on the dataset as follows:

6.5.1 Non-linearity test

Non-linearity refers to the situation where the relationship between the dependent variable (ACCUs prices) and independent variables is not a straight line [70]. This means that changes in independent variables do not result in proportional changes in the dependent variable.

To assess non-linearity, we generated scatter plots for each independent variable against ACCUs prices using the `matplotlib` library in Python. These visualisations help identify non-linear relationships by revealing patterns such as curves and abrupt shifts [48].

6.5.2 Non-stationary test

Non-stationary in time series data implies that the statistical properties, such as mean and variance, change over time [70]. This can result in unpredictable trends and fluctuating patterns.

To test for non-stationary, it is essential to examine the presence of a unit root. There are various methods to check for the presence of a unit root in a time series, with one of the most widely used being the augmented Dickey-Fuller (ADF) test [34]. Following the suggestion in [68], we applied the ADF test using the `adfuller` function from the `statsmodels` library in Python. The ADF test helps identify the presence of unit roots in the time series data, indicating that the mean and variance are not constant over time. By rejecting the null hypothesis of the ADF test, we can conclude that the time series is stationary.

6.5.3 Multicollinearity test

Multicollinearity occurs when independent variables in a regression model are highly correlated, leading to unreliable estimates of regression coefficients [60]. It can cause instability in prediction models and make it difficult to determine the

effect of each predictor [115].

The most common way to detect multicollinearity is using the variance inflation factor (VIF) [123]. In this study, we calculated the VIF for each independent variable using the `variance_inflation_factor` function from the `statsmodels` library in Python. The VIF quantifies how much the variance of an estimated regression coefficient increases due to collinearity, making it easier to detect when independent variables are highly correlated. According to statistical guidelines, a VIF greater than 10 generally indicates severe multicollinearity that needs to be addressed [115].

6.5.4 Results and discussion

The analysis of carbon credit prices reveals several data characteristics that must be addressed to ensure accurate modelling and prediction. This section presents the findings of the non-linearity, non-stationary, and multicollinearity tests, highlighting the challenges and implications for developing robust prediction models.

We created scatter plots to investigate the potential non-linear relationships between various influencing factors and ACCUs prices. By plotting each factor against ACCUs, we can identify patterns suggesting non-linearity as shown in Figure 6.8. If the points form a straight line, it indicates a linear relationship; if they form a curve or another pattern, it suggests a non-linear relationship. The results reveal that factors such as crude oil prices (CrP), coal prices (CP), and NONF display curved patterns, indicative of non-linear relationships.

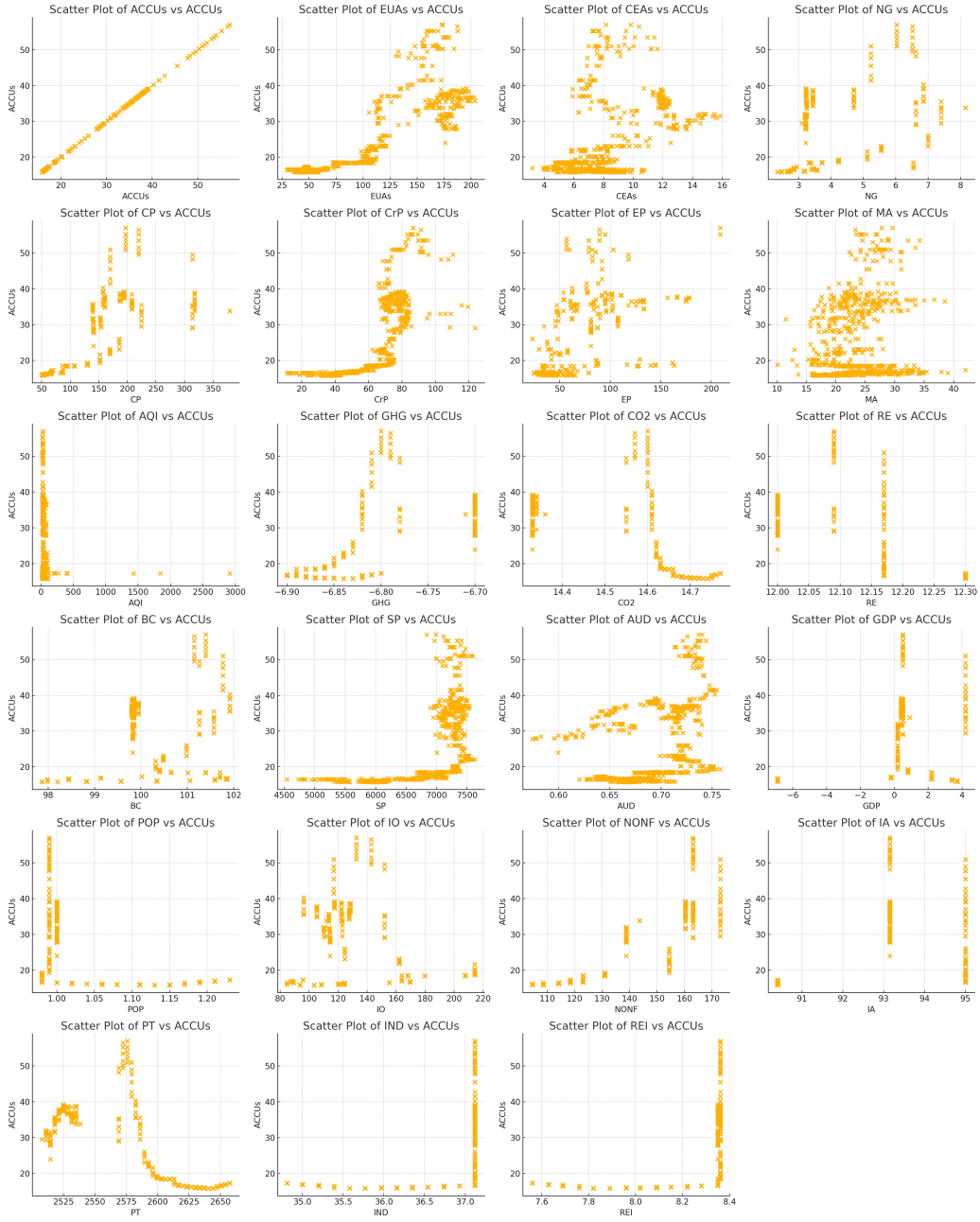


Figure 6.8 : Scatter plots of ACCUs and all influencing factors, illustrating potential non-linear relationships.

To test for non-stationary, we applied the ADF test and the results of the ADF test are presented in Figure 6.9. The value of the t-test -1.478707 was obtained against critical values at different confidence levels (1%, 5%, and 10%), and the p-value of 0.544017 suggests that we fail to reject the null hypothesis. The null

hypothesis suggests the presence of a unit root, indicating that ACCUs are non-stationary. This means that there is evidence to conclude that the time series within our study exhibits characteristics of non-stationarity.

```
Results of Dickey-Fuller Test
Test Statistic                  -1.478707
p-value                          0.544017
#Lags Used                      11.000000
Number of Observations Used    655.000000
Critical Value (1%)             -3.440373
Critical Value (5%)             -2.865963
Critical Value (10%)            -2.569125
dtype: float64
The data is non-stationary.
```

Figure 6.9 : The result of the ADF test indicates a non-stationary characteristic.

To test for multicollinearity, we calculated the VIF. Table 6.3 highlights the presence of multicollinearity in most of the features in our dataset. Factors such as REI, POP, CO2, and GHG show very high VIF values, indicating significant multicollinearity. This suggests that these factors are highly correlated with the independent variable, ACCUs. Multicollinearity can lead to overfitting and unstable coefficients, reducing the ability of the model to perform well on new unseen data.

The results obtained from the applied tests revealed that the daily carbon credit price dataset is nonlinear, non-stationary, and multicollinear. These characteristics of the dataset emphasise the complex dynamics of the carbon credit market and the need for advanced models capable of handling such complexities.

Table 6.3 : The calculated VIF values for independent variables indicate the presence of multicollinearity.

Feature	**VIF**	**Feature**	**VIF**	**Feature**	**VIF**
const	2.06e+07	EUAs	45.55	CEAs	5.69
NG	7.80	CP	19.11	CrP	22.35
EP	2.33	MA	1.89	AQI	1.27
NONF	24.68	PT	5.43	REI	13442.78
GHG	864.29	CO2	263.72	RE	58.42
BC	17.85	SP	10.37	AUD	534.58
GDP	4.25	POP	2327.03	IO	113.80
IA	129.97	IND	17.19		

6.6 Conclusion

This chapter details the methodology for collecting and pre-processing the data required to analyse the price of carbon credits and the influencing factors. Starting with collecting data for carbon credit prices and the 22 external factors identified in the previous chapter, we applied a series of pre-processing steps to ensure the data were clean and ready for analysis. In addition, we also performed EDA using visualisation and statistical techniques to gain insight into the distribution and relationships of the data. Furthermore, we performed three key tests to investigate data challenges: the ADF test to assess stationarity, scatter plot analysis to evaluate non-linearity, and the VIF test to detect multicollinearity. This detailed pre-processing and exploration phase was essential for understanding and preparing the dataset for subsequent analytical methods. The next chapter focuses on identifying the key external factors, employing various methods to identify the most significant factors affecting carbon credit prices.

Chapter 7

Key Factor Selection

7.1 Overview

The previous chapter outlined the methodology for collecting data on the carbon credit prices and 22 influencing factors. Following this, we pre-processed the dataset by applying various steps as well as exploring the data through visualisation and statistical analysis to gain insights into its distribution and relationships. This chapter identifies the most important factors that affect carbon credit prices. The selection of model inputs is a critical stage that significantly impacts the models' performance [46]. The prediction accuracy of a model can be significantly enhanced by selecting the appropriate input features, and a valid feature selection method can reduce the model's computational time and minimize the risk of overfitting [65], [66], [167]. In this thesis, we applied three methods to study feature importance to determine the key factors. These methods are correlation analysis, the MIC and the proposed CCMFI model. The CCMFI model integrates RF regression and an explainable AI technique, providing a powerful tool to identify key price drivers. The structure of this chapter is as follows: Section 7.2 outlines the techniques and methodologies used in this chapter. Section 7.3 explains the empirical study. The results and a discussion are presented in Section 7.4. Finally, Section 7.5 concludes the chapter. The contents of this chapter have been published in the IEEE International Conference on e-Business Engineering (ICEBE), under the article titled "A Comprehensive Analysis for Determining Key Influencing Factors on Carbon Credit Prices".

7.2 Methods for identifying and selecting key factors

Many researchers emphasise that selecting the input for the model is an essential step that greatly affects its performance, making feature selection a critical stage in machine learning tasks such as classification and prediction [46], [65], [66], [167]. Feature selection is the process of identifying and selecting the most relevant features in a dataset that contribute significantly to the predictive model [51]. This process enhances the effectiveness and performance of the models by using a reduced set of features [51]. Furthermore, it can reduce the overfitting possibility by removing redundant variables and retaining meaningful variables. In this thesis, we calculate the feature importance based on three techniques to analyse and identify the key factors that influence carbon credit prices, as illustrated in Figure 7.1. The correlation matrix analysis, MIC, and the proposed CCMFI model are explained in the following subsections.

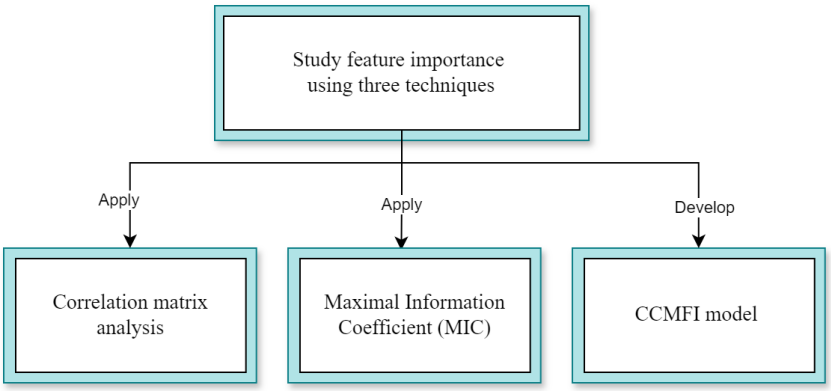


Figure 7.1 : The applied techniques to analyse and identify the key factors influencing carbon credit prices.

7.2.1 Correlation matrix analysis

A correlation matrix analysis is a fundamental tool for examining the relationships between variables in dataset [40]. This method offers insights into the strength

and direction of their relationships. This analysis helps summarise large datasets and visualise patterns inherent in the data. In this study, we applied two correlation techniques, Pearson's and Spearman's, to investigate the relationships between the price of ACCUs and the external factors in our dataset.

The Spearman rank correlation coefficient [40] clearly represents the monotonic relationship between different factors and the price of carbon credits [144]. This technique is particularly advantageous in our analysis because of its capability to capture non-linear, monotonic relationships between variables. Unlike Pearson's correlation coefficient [40], which assumes linearity and normal distribution, Spearman's correlation is more robust in handling datasets with non-linear associations. Therefore, it is suitable for our dataset, which may contain non-linear relationships and not follow a normal distribution.

7.2.2 Maximal information coefficient (MIC)

MIC is an essential statistical tool that quantifies the strength of the associations between two variables [121]. MIC has been widely used to measure the relationship between independent and dependent variables, as it helps identify significant relationships and serves as a reliable method for feature selection [116], [132], [166]. MIC is especially adept at capturing non-linear relationships that conventional correlation measures might overlook [121].

In this study, we applied MIC to investigate the associations between the price of carbon credits and external factors in our dataset. MIC is a normalisation correction for mutual information values using an unequal interval optimisation method, and its range is [0,1]. A score of 1 indicates that two variables are closely related; a score of 0 indicates that two variables are independent. Using MIC, we can gain a clearer insight into our data, ensuring a thorough understanding of the underlying relationships essential for our subsequent analyses and model development.

7.2.3 The proposed CCMFI model

This subsection presents the proposed solution, the CCMFI model, to identify the most influential factors on the price of carbon credits. Before explaining the proposed model, we explain the SHAP method and the baseline machine learning model adopted in our approach, the random forest model.

Random forests (RF)

RF is an ensemble learning method that combines multiple prediction trees during training and averages their predictions for regression tasks [14]. The trees are trained on a random subset of training data and input characteristics, which reduces overfitting and improves generalisation [14]. Figure 7.2 illustrates the operation of the RF classifier and how it produces a final class based on the results of all the decision trees. In RF, the importance of the features is usually determined by averaging the decrease in impurity or the increase in accuracy caused by each feature in all trees in the forest. The features with a higher decrease in average impurities are considered more important [14]. The RF model is a commonly used algorithm for feature selection and importance scoring in non-linear data while offering a method to evaluate variable importance [41], [143]. The use of RF to study the importance of variables provides insight into the contribution of features, which enhances the understanding of the internal mechanisms of the model and quantifies the impact of variables on the prediction of carbon credit prices [155].

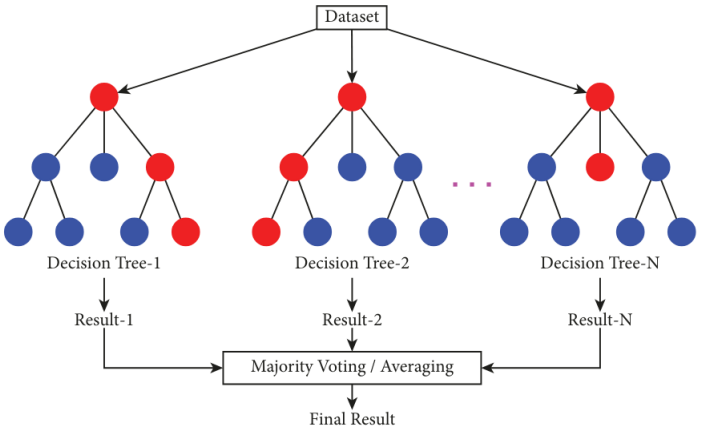


Figure 7.2 : Illustration of the RF framework [73].

SHAP method

SHAP values [103] are a powerful model interpretability method that provides a detailed explanation for each prediction made by a machine learning model as shown in Figure 7.3. SHAP leverages the principle of local explainability to construct simplified models for complex black-box models [102]. The concept behind SHAP values originates from the Shapley values in game theory [126]. Shapley values offer a fair distribution of a total gain to each player by considering their contribution across all possible coalitions. Similarly, SHAP values aim to reasonably attribute the contribution of each feature to the final prediction in a machine learning model. By making slight adjustments to input features and observing the changes in the predictions, SHAP can determine the significance of each feature for specific data points. If altering a feature's value does not significantly change the prediction, it suggests that the feature may not be an essential predictor for that instance. This ability to provide detailed, data-point-specific insights makes SHAP an invaluable tool in the interpretability toolkit for machine learning practitioners.

Given a dataset with a set of features, the challenge lies in computing the exact Shapley values because of the exponential number of possible feature combinations.

To address this, SHAP uses an approximate method that involves a special weighted linear regression, which applies to any model [103]. For ensemble tree models, SHAP employs various assumptions about feature dependence to achieve this approximation.

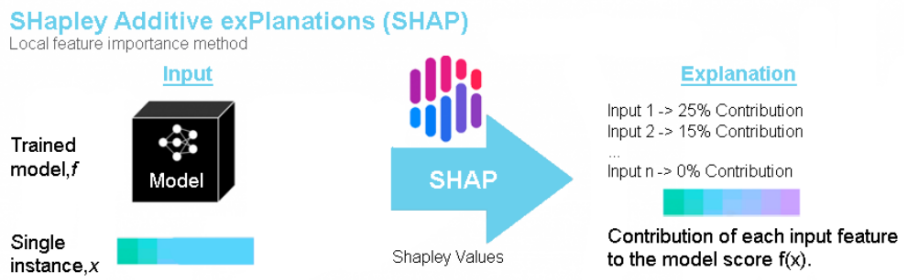


Figure 7.3 : Illustration of SHAP method [78].

The proposed CCMFI model consists of three stages: analysis, ranking, and the selection of key factors, as illustrated in Figure 7.4. The detailed steps of this methodology are described in Algorithm 1. The main stages of the proposed model are explained as follows:

- **Stage 1: Analysis Stage**: This stage involves a two-step approach: the first is to fit the RF model, and the second is to calculate the importance of features using the SHAP method. After fitting the RF model, the importance of the features is measured based on how much each feature improves the model's prediction accuracy. To determine the overall importance of each feature, we aggregated the SHAP values across the dataset by taking the mean of the absolute SHAP values for each feature across all data points.

  Although the RF model is effective for feature selection and importance scoring, incorporating SHAP values adds interpretability, captures the interaction effects, and enhances the robustness of the analysis. SHAP values provide a

transparent and mathematical method to interpret RF predictions by quantifying the impact of each feature. They explain how much each feature contributed to the predicted price of carbon credits compared to the average price across the dataset.

- **Stage 2: Ranking Stage**: Importance scores are sorted in descending order based on the aggregated SHAP values and then ranked according to their scores. The higher the SHAP value, the more significant the impact of the feature on the model output, indicating a higher importance. This step helps the user of the model to see which factors impact the price of carbon credits the most.

- **Stage 3: Selection Stage**: In the final stage of the CCMFI model, the factors with the most significant impact are selected based on their rank. This step allows the user to focus on the most important factors when predicting the price of carbon credits.

The number of key factors K that result from the CCMFI model is determined based on the specific context at hand. Experts with a deep understanding of the factors that affect carbon credit prices in that particular situation decide on the best value for K. They balance the need to include enough factors to capture the significant influences on carbon credit prices while avoiding too many factors that could lead to inaccurate predictions. Furthermore, our proposed model can serve as a preliminary step for advanced prediction, allowing adjustment of the number of selected factors based on prediction accuracy. The model establishes a solid foundation for more refined predictive analyses by identifying and selecting the most influential factors. This iterative process ensures that the selected factors are optimal, improving the accuracy and reliability of carbon credit price predictions while preventing overfitting.

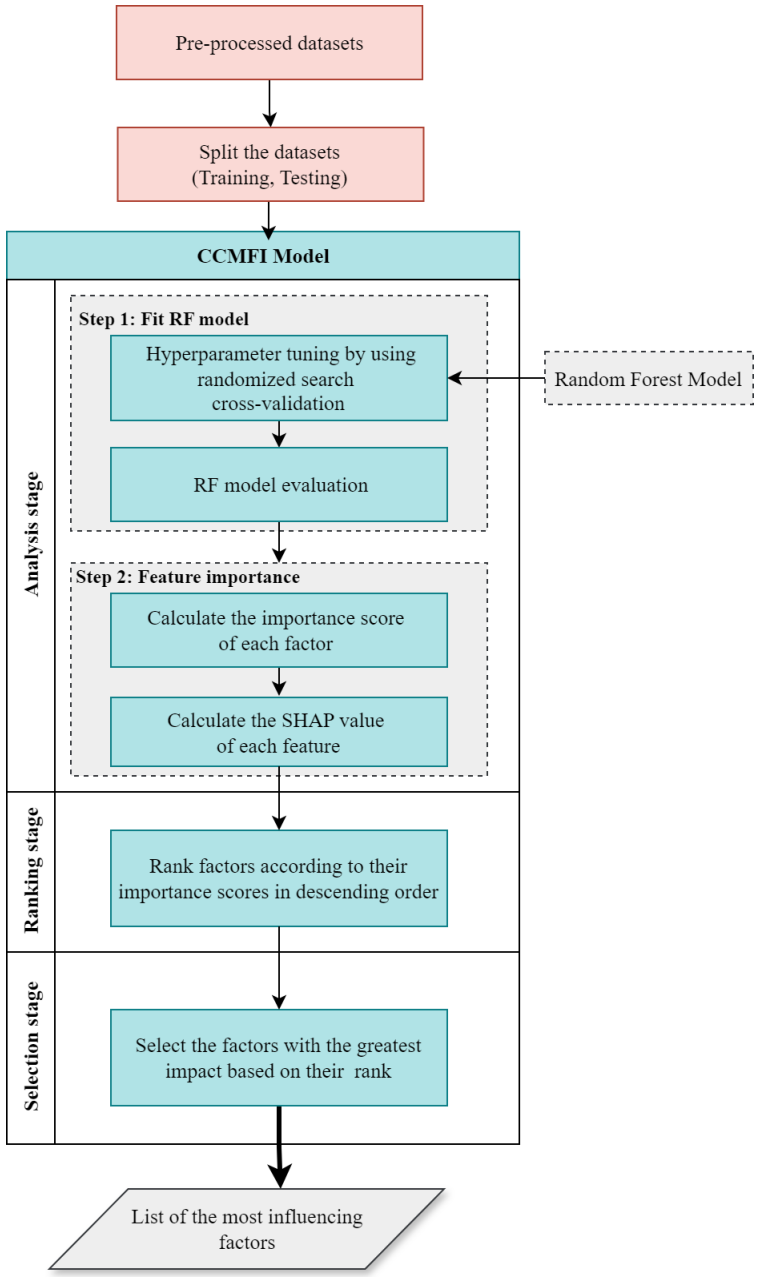


Figure 7.4 : The flowchart process for the proposed CCMFI model.

Algorithm 1 CCMFI model - Carbon Credit Multi-Factor Identification

Input:
  X_{train}: Training set of independent variables
  y_{train}: Training set of target variable (carbon credit price)
Output:
  $K[\,]$: Subset of X_i of the top selected factors
Initialisation:
  Initialise an empty list $Score[S]$ to store importance scores
  Initialise an empty list $K[\,]$ to store the top selected factors
1. Analysis stage:
  Perform *RandomizedSearchCV* to tune the model parameters using $(X_{\text{train}}, y_{\text{train}})$
  Fit the *Random Forest* model on the training data using $(X_{\text{train}}, y_{\text{train}})$
  Calculate SHAP values for all features using the trained *Random Forest* model
  Aggregate SHAP values across all samples to compute the overall importance of each feature
  Score[i] $\leftarrow$ Mean of the absolute SHAP values of feature i across all samples
2. Ranking stage:
  RankIndex $\leftarrow$ Sort the features based on their mean absolute SHAP values in *Score* in descending order
3. Selection stage:
  Select the top K factors based on their rank indices:
for each index i in the *RankIndex* indices **do**
  Append the corresponding factor X_i to the list K
end for
Output:
  Return the list K containing the selected key factors.

7.3 Experiment analysis

This section presents the experiment procedures used in this study. It is structured as follows: Subsection 7.3.1 provides an overview of the pre-processed data used to study feature importance in this study. Subsection 7.3.2 gives an overview of the selected implementation platform. Subsection 7.3.3 outlines the implementation process of the correlations, MIC, and the proposed CCMFI model.

7.3.1 Dataset used to study feature importance

In this study, a real-world carbon market dataset are used to validate the proposed model. Dataset collection and preprocessing were performed following the key steps detailed in Chapter 6. Focusing on Australian Carbon Credit Units (ACCUs) as a case study, daily price ACCU data was collected from the JARDEN platform [63]. Furthermore, we identified 22 influencing factors from the literature and then used real data to reflect each factor. The identified influencing factors include European Union Allowances (EUAs), China Emission Allowances (CEAs), Australian natural gas (NG), Australian BJ power coal spot prices (CP), Average global crude oil (CrP), Australian electricity spot prices (EP), Australia air temperature (MA), Australia Air Quality Index (AQI), Australia greenhouse gas emissions (GHG), Australia carbon (CO2) emissions, Australia renewable energy (RE), Business confidence indicator (BC), S&P/ASX 200 Index price (SP), AUD/USD exchange rate (AUD), Australia per capita GDP growth rate (GDP), Australia population growth rate (POP), AU iron ore (IO), AU Basic non-ferrous metal product manufacturing (NONF), Australia households internet access percentage (IA), Patents on environmental technologies (PT), Australia's industry emissions (IND), and renewable usage in industry (REI). Datasets were obtained from January 2, 2020, to September 1, 2023, each with 667 data samples. Table 6.1, in the previous chapter, provides a reference for the external factors considered in this study.

7.3.2 The selected implementation platform

We used Google Colab, a cloud-based Jupyter notebook environment, for our implementation. This platform enables writing and executing Python code directly in the browser without requiring local installations. Its accessibility, integration with Google Drive for seamless data management, and free access to GPU / TPU resources significantly improve machine learning tasks [13]. Additionally, Colab includes a wide range of pre-installed libraries, such as sklearn, facilitating the efficient implementation of machine learning and deep learning experiments.

7.3.3 Implementation

We implemented correlation analysis, MIC, and CCMFI methods using Python and its libraries. Matplotlib was extensively used for visualisations, including correlation matrices, MIC score plots, and graphs that illustrate data distributions, relationships, and model performance in the CCMFI model. The implementations of each approach are explained in detail in the following sections.

Implementation of correlation matrix analysis approach

Correlation analysis is used to investigate relationships among input variables, particularly focusing on the dependencies between carbon credit prices and various external factors. This analysis aims to identify the significance of each factor and to eliminate redundant variables. Two techniques, Pearson's correlation and Spearman's rank correlation, are employed to assess the associations between carbon credit prices and these factors. Heatmaps are also created to visually depict the strength and direction of the relationships between variables, providing a comprehensive view of their interdependencies. The process involved the following steps:

1. The pre-processed datasets were loaded using Pandas.

2. The Pearson correlation method was used to compute the correlation matrix. This method measures the linear relationship between pairs of features. The correlation matrix was then saved to an Excel file for further analysis.

3. A heatmap of the Pearson correlation matrix was generated using Seaborn and Matplotlib to visualise the linear relationships between features. The heatmap highlighted the strength and direction of the correlations, providing insight into potential linear dependencies.

4. The Spearman correlation method was used to compute the correlation matrix. This method measures the rank-based relationship between pairs of features, making it suitable for identifying monotonic relationships. The correlation matrix was saved to an Excel file for further analysis.

5. A heatmap of the Spearman correlation matrix was generated to visualise the rank-based relationships between features. This visualisation provided insights into potential non-linear dependencies and was particularly useful for identifying monotonic relationships.

6. The Pearson and Spearman correlation matrices were displayed to provide a comprehensive view of the relationships between features. These matrices were essential to understand the dependencies and interactions between factors influencing carbon credit prices.

Implementation of the MIC analysis approach

The MIC analysis is conducted to explore the relationships between carbon credit prices and various external factors. This method is chosen for its ability to capture both linear and non-linear dependencies, providing a more versatile measure of association compared to traditional correlation methods. By computing the MIC values for each feature relative to the target variable (ACCUs), we aim to identify the

most influential factors. The results are visualised using bar graphs to highlight the strength of the relationships, thus providing valuable information on the key drivers of carbon credit price fluctuations. The process involved the following steps:

1. The pre-processed datasets were loaded using `pandas`, ensuring that all numeric features were properly formatted for analysis.

2. MIC was computed to measure the strength of the relationship between the target variable (ACCUs) and each numeric feature. MIC captures both linear and nonlinear dependencies, making it a robust metric for identifying influential features.

3. `MinMaxScaler` from `sklearn.preprocessing` was applied to normalize all numeric features, ensuring that differences in magnitude did not affect the MIC computation.

4. MIC values were calculated for each feature relative to the target variable by iterating over the features and applying the `compute_mic` function to each pair of variables. The `compute_mic()` function was implemented using the MINE class from the `minepy` library. It computes the MIC score between two variables by following these steps:

    - The function initialises an instance of `MINE()`.
    - It then applies the `compute_score(x, y)` method, where `x` and `y` represent the two numerical variables being compared.
    - The `mic()` method is called to retrieve the computed MIC value, which quantifies the relationship strength between the two variables.
    - The resulting MIC score is returned, providing a numerical measure of dependency between the variables.

5. The computed MIC values are compiled into a `DataFrame` containing feature names and their corresponding `MIC` values with the target variable (`ACCUs`).

6. The `DataFrame` is sorted in descending order based on MIC values to highlight the most influential features affecting ACCUs.

7. A bar plot is generated using `seaborn` to visualize the MIC values. The plot provides a clear representation of feature importance, with each bar annotated with its `MIC` value for better interpretability.

Implementation of the proposed model (CCMFI)

The proposed CCMFI model was also developed using various Python libraries. Scikit-learn was used for data analysis and machine learning tasks, including model building and evaluation. SHAP was used to interpret the feature importances and understand the factors influencing carbon credit prices. The implementation process involved the following steps:

1. The pre-processed data were divided into training and testing sets using the `train_test_split` function from Scikit-learn as shown in Figure 7.5. Feature selection is an integral part of the model-building process and was performed exclusively on the training set to prevent data leakage. The test set remained completely separate and unseen during feature selection.

2. The RF regression model was implemented using the `RandomForestRegressor` class from the `sklearn.ensemble` module. This model was chosen for its ability to capture non-linear relationships and its robustness to overfitting.

3. Hyperparameter tuning was performed using *randomized search cross-validation (CV)* to find the optimal hyperparameters for the model. This process was applied solely to the training data. A range of parameter combinations was

evaluated, and the best model configuration was selected based on the CV using the lowest mean squared error (MSE).

The best parameters identified were:

- `n_estimators=500`: This is the number of trees in the forest. Increasing the number of trees generally improves the model performance, but also increases the computational cost.
- `min_samples_split=2`: This specifies the minimum number of samples required to split an internal node. Smaller values allow the model to learn more complex patterns.
- `min_samples_leaf=1`: This is the minimum number of samples required to be at a leaf node. Setting this to 1 allows the model to capture more fine-grained patterns.
- `max_features='log2'`: This indicates the number of features to consider when looking for the best split. Using 'log2' limits the number of features, which can help prevent overfitting and improve computational efficiency.
- `max_depth=30`: This defines the maximum depth of each tree in the forest. Deeper trees can model more complex relationships, but setting it too high might lead to overfitting.
- `bootstrap=False`: If set to False, the model uses the entire dataset to build each tree (without bootstrapping).

4. The RF model was trained on the training data using the `fit` method. This step involved learning the relationships between the input features and the target variable (ACCUs) from the training data.

5. The `shap.Explainer` class from the `shap` Python library was used to create an explainer for the trained random forest model, ensuring compatibility with tree-based models.

6. SHAP values were calculated for all instances in the training set, quantifying each feature's contribution to the model's predictions.

7. The absolute mean of the SHAP values was calculated across all data points, providing a measure of overall feature importance.

8. The features were ranked in descending order based on their mean SHAP values, identifying the most influential factors affecting carbon credit prices (ACCUs).

9. SHAP visualisations were generated to interpret feature importance: SHAP summary plot to display the distribution of SHAP values for all features, highlighting their impact on predictions, and SHAP bar chart to show the average absolute SHAP values, emphasising the most influential features.

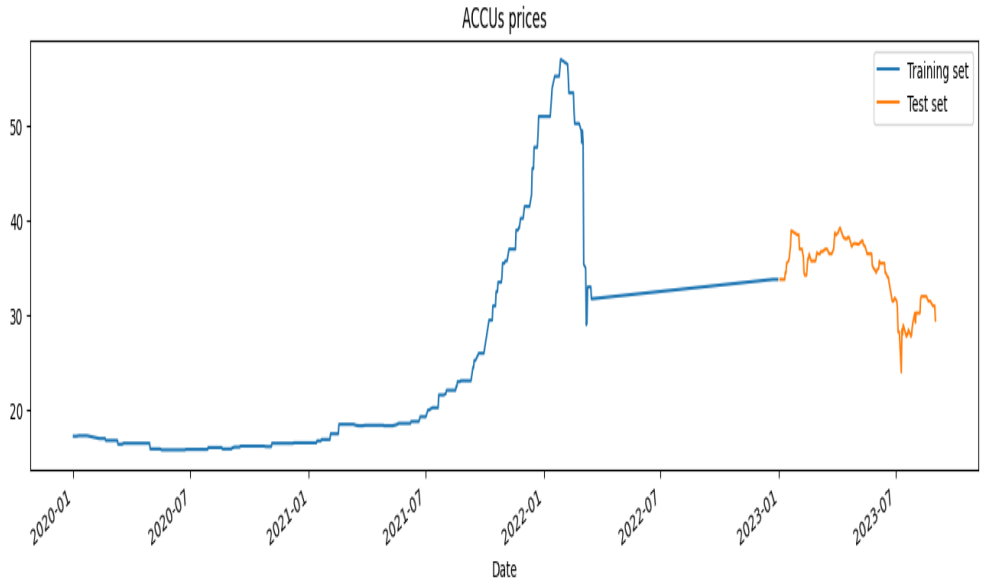


Figure 7.5 : Time series of pre-processed data of ACCUs price split into training and testing sets.

7.4 Results and discussion

This section outlines the results of our analysis, highlighting the top 10 factors identified using each method. The structure of this section is as follows: Subsection 7.4.1 details the results of the Spearman's rank and Pearson correlation. Subsection 7.4.2 covers the findings from the MIC analysis, while the results of the proposed CCMFI model are discussed in Subsection 7.4.3. Following this, we present the comparative analysis in Subsection 7.4.4.

7.4.1 Results of the correlation analysis

When considering correlation coefficients, strong positive and negative correlations are important. The strength of a correlation is determined by its absolute value, regardless of whether it is positive or negative. In this study, heat maps were used to visually show the strength of the relationships between variables through correlations. Based on the correlation analysis between ACCUs and the 22 external factors, several factors stand out for their strong relationships with ACCUs.

Using Spearman's rank correlation, the most influential factors on ACCUs carbon credit prices were identified, as shown in Figure 7.6. The NONF factor showed the highest positive correlation with a value of 0.9502, indicating a significant impact on product prices in the non-ferrous metals sector. The CP and CrP factors also exhibited strong positive correlations of 0.9281 and 0.8974, respectively, highlighting their substantial influence on the demand for carbon credits. Other significant factors included EUA with a correlation of 0.8920. CO2 emissions, PT and RE also exhibited high negative correlations of - 0.8726, -0.8615 and -0.8517, respectively, suggesting their notable impact. Furthermore, SP, REI and IND factors showed positive correlations of 0.8062, 0.7800 and 0.7788, respectively.

The Pearson correlation method highlighted similar influential factors, as shown

in Figure 7.7. NONF factor exhibited a strong positive correlation of 0.8444, indicating its significant impact on product prices in the non-ferrous metals sector. EUAs also showed a substantial positive correlation of 0.8436. CP and CrP demonstrated significant positive correlations of 0.7540 and 0.7474, respectively. RE and PT also exhibited high negative correlations of - 0.7118, -0.8615, and -0.7108, respectively, suggesting their notable impact. GHG and CO2 emissions also showed notable correlations of 0.6360 and -0.6255, respectively, highlighting their influence on ACCU prices. SP and REI exhibited positive correlations of 0.4950 and 0.4943, respectively, suggesting their notable impact.

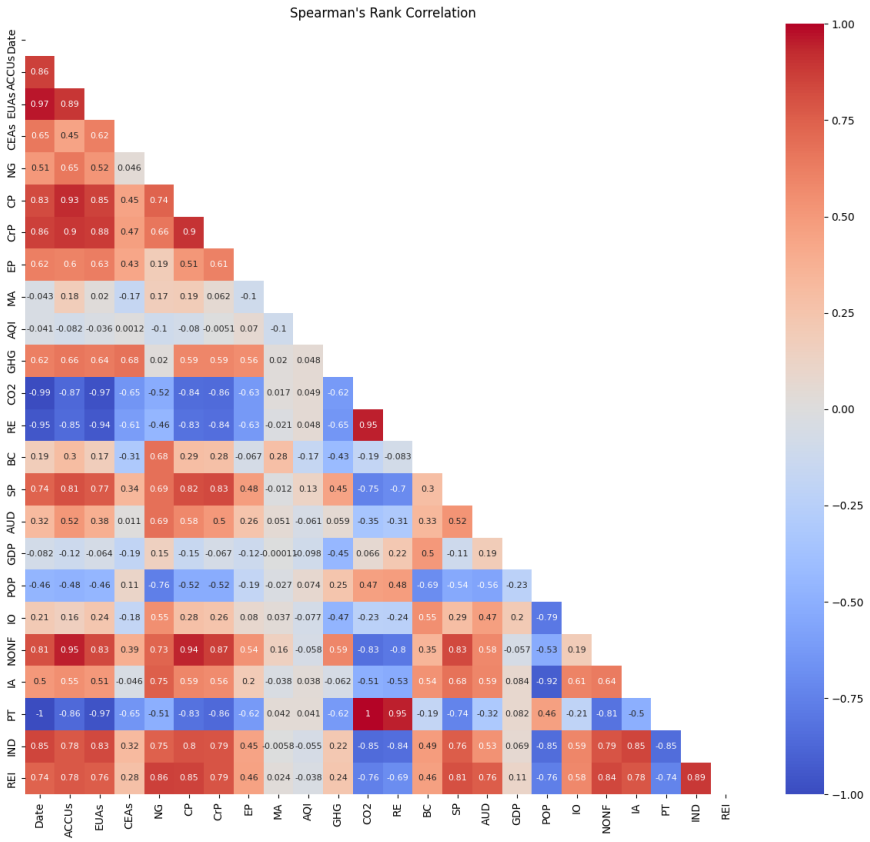


Figure 7.6 : Spearman correlation heatmaps visualising the relationships between ACCUs and the influencing factors.

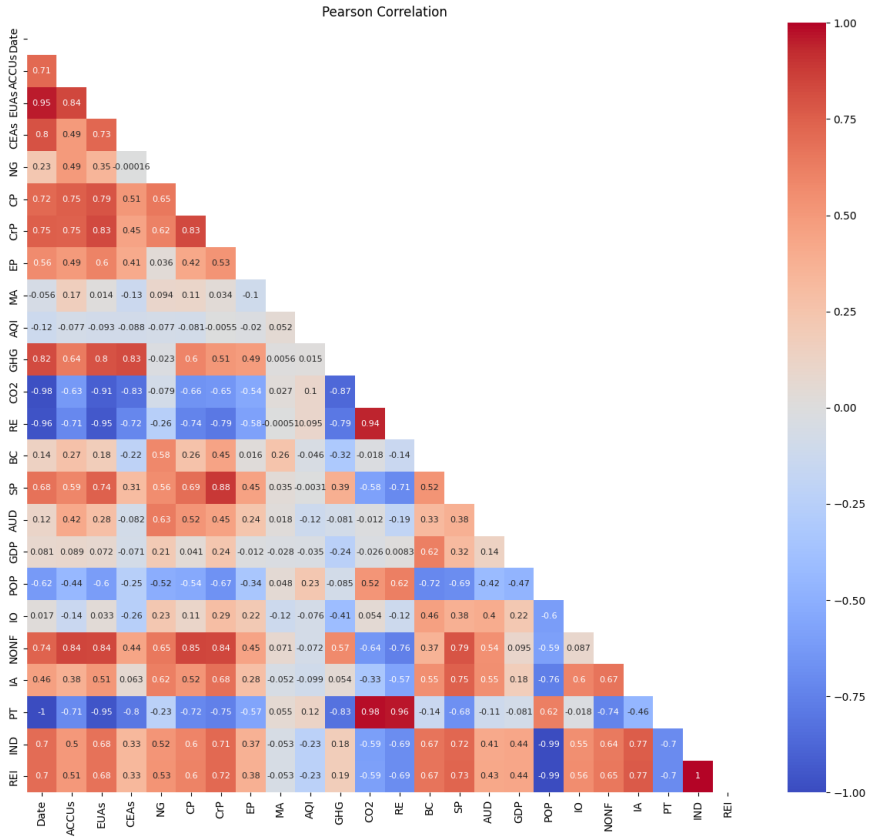


Figure 7.7 : Pearson correlation heatmaps visualising the relationships between ACCUs and the influencing factors.

7.4.2 Results of MIC

MIC is less sensitive to the distribution of data and can identify meaningful relationships even in the presence of outliers or skewed data distributions. This makes it particularly useful for selecting the main factors that influence carbon credit prices. MIC assigns a value between 0 and 1 to each pair of variables, where values closer to 1 indicate stronger associations. This quantitative measure allows for easy comparison and ranking of factors based on their influence on the price of carbon credits.

Figure 7.8 identifies the strength of the relationships between the identified factors and ACCU using the MIC method. Using the MIC method, the most influential

factors on ACCU carbon credit prices are as follows: CrP had the highest association with ACCU prices, with a MIC value of 0.9959. CO2 emissions, PT and CP showed high associations, each with a MIC value of 0.9909. NONF and EUA also showed strong associations with MIC values of 0.9887 and 0.9859, respectively. GDP was significantly associated with a MIC value of 0.9766. IO, BC and NG exhibited significant associations, each with MIC values of 0.9728.

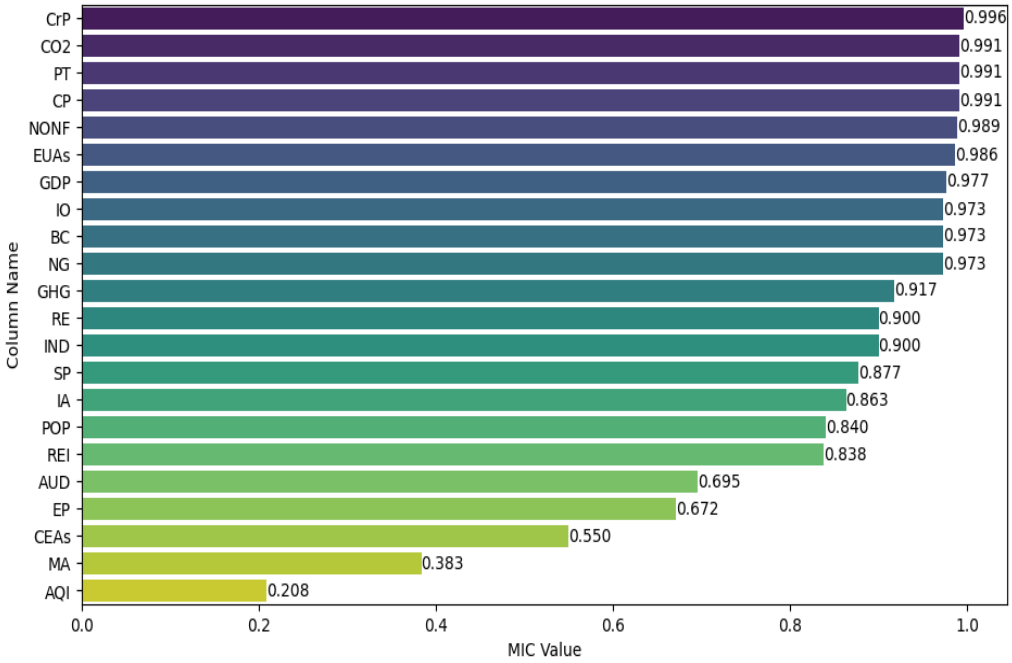


Figure 7.8 : Feature importance using the MIC method.

7.4.3 Results of the proposed CCMFI model

The proposed CCMFI model effectively identifies and ranks the factors that influence the price of carbon credits, achieving an MSE of 1.443. The three-stage approach ensures a thorough examination of potential influencing factors, allowing a detailed and comprehensive understanding of the drivers of carbon credit prices. SHAP values provide a clear interpretation of how each factor affects carbon credit prices, making the model's outcomes more understandable and actionable.

The SHAP bar shown in Figure 7.9b quantifies the average impact of these external factors using the mean absolute SHAP value, each bar representing the significance of the individual features. The features are sorted by importance, with the most influential at the top, highlighting the key factors affecting carbon credit prices. The importance order of the input features is as follows: PT > CO2 > NONF > CP > EUAs > CrP > RE > GHG > IO > CEAs > BC > GDP > AUD > IND > IA > NG > REI > SP > EP > POP > AQI > MA.

Furthermore, Figure 7.9a interprets how each factor affects carbon credit prices, making the model's outcomes more understandable and actionable. The SHAP plot uses colour coding to show feature values: red dots represent high values, while blue dots denote low values. The x-axis shows SHAP values, indicating the impact of each feature on model predictions. Positive SHAP values mean that the feature increases the price of carbon credits, while negative values imply a decrease. Each dot corresponds to a specific feature and instance, with the spread of dots showing the feature's contribution range.

The results indicate that the CCMFI model effectively captures the complex and non-linear relationships among the external factors in the carbon market, which confirms its suitability in analysing and identifying the main factors that affect the price of carbon credits. The proposed CCMFI model highlights the substantial impact of traditional energy markets, environmental factors, and economic indicators on ACCUs prices. The model identifies the PT factor as the most influential factor, with a feature importance value of 2.0154. CO2 emissions follow closely, with a value of 1.8473. NONF and CP also showed a significant influence, with values of 1.4464 and 1.1497, respectively. EUA demonstrated a notable influence with a feature importance value of 1.0316. CrP also had a considerable impact with a value of 0.7198. RE and GHG emissions showed moderate influence, with values of 0.4753 and 0.4337, respectively. IO and CEA complete the list of the main influencers, with

importance values of features of 0.2916 and 0.2211, respectively.

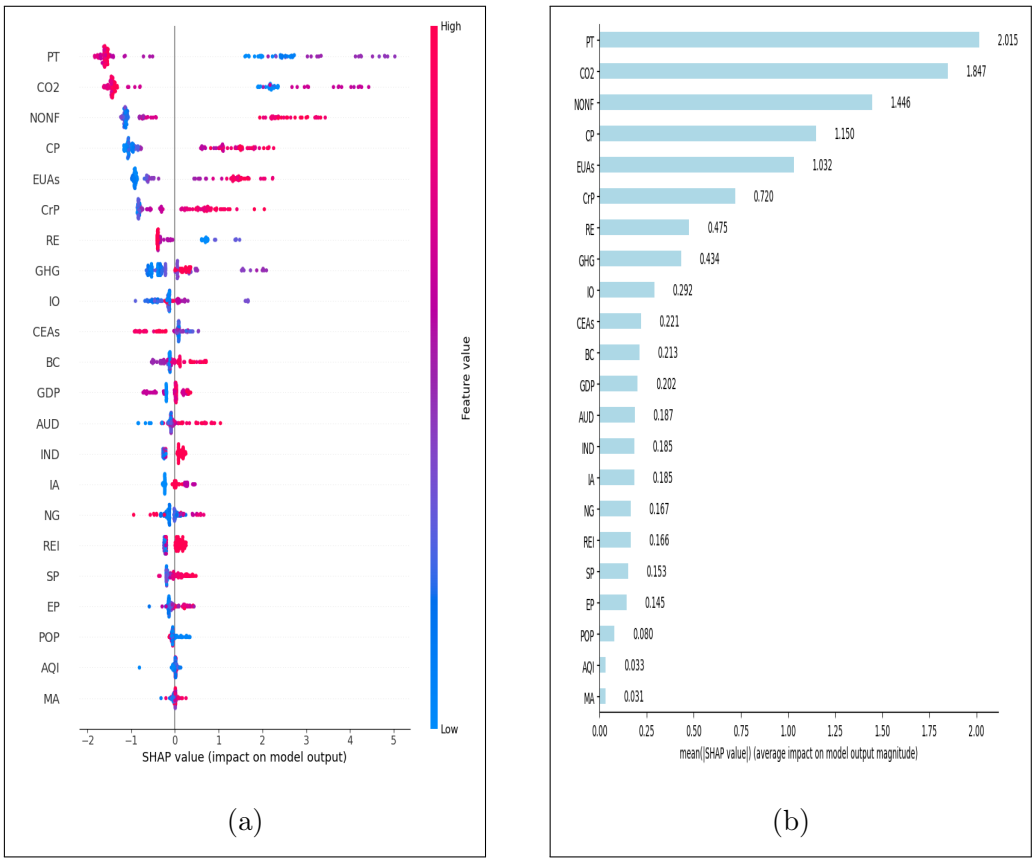


Figure 7.9 : (a) SHAP plot showing the impact of external factors on carbon credit prices with features sorted by importance. (b) SHAP bar chart quantifying the average impact of external factors using mean absolute SHAP values.

7.4.4 Comparative results

This subsection presents the comparative results in an integrated format. As shown in Table 7.1 and Figure 7.10, the results obtained through applied techniques are compared to identify the most significant factors affecting the price of carbon credits.

Table 7.2 presents the top 10 influencing factors ranked by their respective correlation and importance scores.

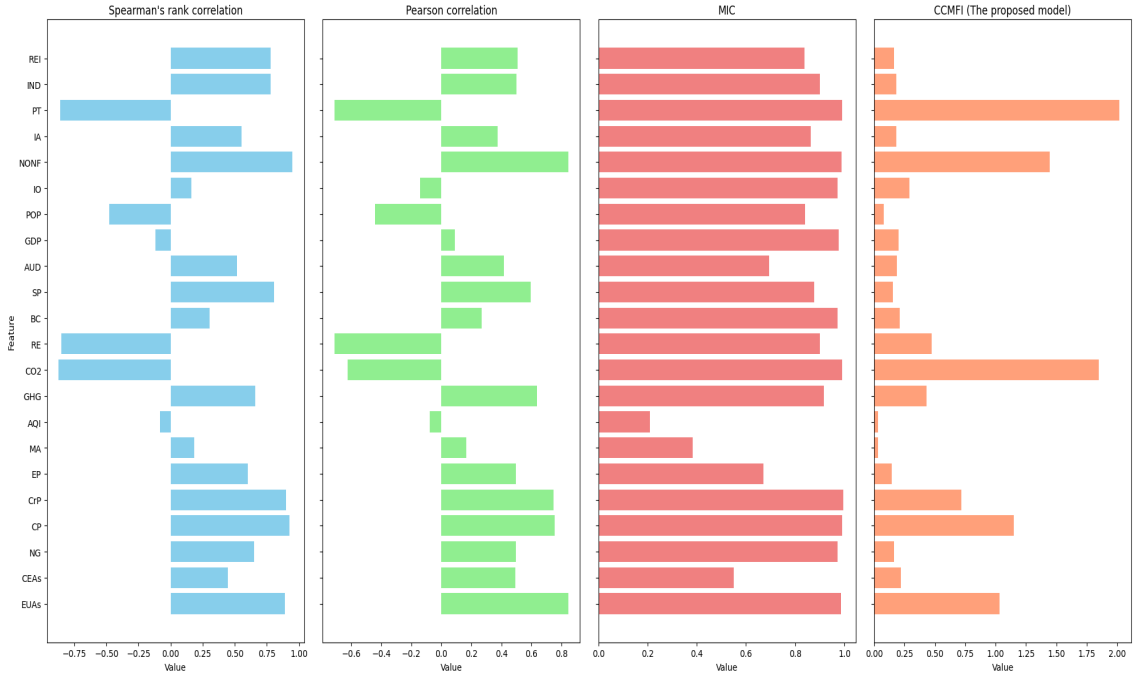


Figure 7.10 : Comparison of feature importance for external factors on ACCUs using Spearman's rank correlation, Pearson correlation, MIC, and the proposed CCMFI model.

The comparative analysis of the three methods and the proposed model revealed the key factors that appear to be consistently highly influential on the price of carbon credits, as summarised in the following.

- **PT:** In the CCMFI model, PT, which represents the green technology maturity factor, emerged as a strong influencing factor with a weight of 2.015, and the MIC analysis further supports its significance with a score of 0.991. Furthermore, the negative correlations observed in both the Spearman and Pearson correlation coefficients of -0.8615 and -0.7108, respectively, align with the findings of Song et al. [130], who claimed there was a negative correlation between the maturity of green technology and the price of carbon credits. Song et al. [130] claims that as green technology matures and becomes more effec-

Table 7.1 : The results comparison of the Spearman's rank correlation, Pearson correlation, MIC, and the proposed CCMFI model.

Feature	**Spearman's rank**	**Pearson**	**MIC**	**CCMFI**
EUAs	0.8920	0.8436	0.9859	1.0316
CEAs	0.4482	0.4932	0.5497	0.2211
NG	0.6528	0.4950	0.9728	0.1665
CP	0.9281	0.7540	0.9909	1.1497
CrP	0.8974	0.7474	0.9959	0.7198
EP	0.6025	0.4943	0.6715	0.1448
MA	0.1848	0.1677	0.3835	0.0311
AQI	-0.0821	-0.0774	0.2084	0.0331
GHG	0.6573	0.6360	0.9174	0.4337
CO2	-0.8726	-0.6255	0.9909	1.8473
RE	-0.8517	-0.7118	0.9000	0.4753
BC	0.3030	0.2665	0.9728	0.2132
SP	0.8062	0.5937	0.8768	0.1528
AUD	0.5152	0.4179	0.6952	0.1867
GDP	-0.1194	0.0887	0.9766	0.2023
POP	-0.4783	-0.4413	0.8401	0.0796
IO	0.1604	-0.1407	0.9728	0.2916
NONF	0.9502	0.8444	0.9887	1.4464
IA	0.5535	0.3754	0.8632	0.1848
PT	-0.8615	-0.7108	0.9909	2.0154
IND	0.7788	0.5004	0.9000	0.1853
REI	0.7800	0.5074	0.8381	0.1663

tive in reducing emissions, the demand for carbon credits decreases, leading to lower trading prices. This relationship is evident in the SHAP values, providing robust empirical evidence that advances in environmental technologies are inversely related to carbon credit prices.

- **CO2 emissions:** The CO2 factor is one of the most significant contributors to GHG emissions. It consistently emerges as a critical determinant of carbon credit prices, reflecting the direct relationship between GHG emissions and the

Table 7.2 : The top 10 influencing factors identified using Spearman's rank correlation, Pearson correlation, MIC, and the proposed CCMFI model.

Rank	Spearman	Pearson	MIC	CCMFI model
1	NONF (0.9502)	NONF (0.8444)	CrP (0.996)	PT (2.015)
2	CP (0.9281)	EUAs (0.8436)	CO2 (0.991)	CO2 (1.847)
3	CrP (0.8974)	CP (0.7540)	PT (0.991)	NONF (1.446)
4	EUAs (0.8920)	CrP (0.7474)	CP (0.991)	CP (1.150)
5	CO2 (-0.8726)	RE (-0.7118)	NONF (0.989)	EUAs (1.032)
6	PT (-0.8615)	PT (-0.7108)	EUAs (0.986)	CrP (0.720)
7	RE (-0.8517)	GHG (0.6360)	GDP (0.977)	RE (0.475)
8	SP (0.8062)	CO2 (-0.6255)	IO (0.973)	GHG (0.434)
9	REI (0.7800)	SP (0.5937)	BC (0.973)	IO (0.292)
10	IND (0.7788)	REI (0.5074)	NG (0.973)	CEAs (0.221)

demand for carbon credits. In the CCMFI model, CO2 was identified as the second most influential factor, with an importance value of 1.847 and a MIC analysis score of 0.991. Spearman (-0.8726) and Pearson (-0.6255) correlations. An increase in CO2 emissions could signal higher industrial emissions, causing market regulators to allocate more carbon credits to accommodate these emissions [114]. This increased supply of credits, relative to demand, may lead to a drop in the price of carbon credits. Furthermore, higher CO2 emissions could trigger increased investment in renewable energy projects as a response to mitigate increasing emissions. Over time, this could reduce the demand for carbon credits, as renewables reduce the reliance on carbon offsets, leading to a decline in prices. This relationship is supported by SHAP values, which show a negative impact of CO2 emissions on carbon credit prices,

further highlighting the inverse relationship.

- **NONF:** Reflecting a product-related factor within the economic category, NONF has been identified as a highly influential factor across all methods. With a CCMFI importance value of 1.446, an MIC score of 0.989, and strong positive correlations in both Spearman (0.9502) and Pearson (0.8444) analyses, this factor shows a significant positive impact on carbon credit prices within the non-ferrous metals sector. The SHAP values further confirm NONF's influence, consistently indicating its strong positive contribution to carbon credit price predictions. This finding is supported by Ji et al. [64], who identified a positive relationship between product prices, particularly in carbon-intensive industries, and carbon credit prices. As industries in this sector often pass on the cost of carbon credits to consumers, higher product prices lead to higher carbon credit prices.

- **CP and CrP:** The CP and CrP markets were consistently identified as key influences. The prices of coal and crude oil serve as indicators of the overall energy market, with a strong positive relationship observed between these factors and the price of carbon credits. In the CCMFI model, CP had an importance value of 1.150, while CrP showed a value of 0.720, supported by their MIC scores of 0.991 and 0.996, respectively. Correlation analyses also highlight their significance, with Spearman (0.9281 for CP, 0.8974 for CrP) and Pearson (0.7540 for CP, 0.7474 for CrP) values showing strong positive relationships. The SHAP values further emphasise their importance, indicating significant positive contributions to carbon credit price predictions. Li et al. [91], Wang and Zhao [148], and Zhang et al. [163] support this finding, highlighting that higher energy prices increase production costs for industries, which often leads to higher demand for carbon credits as companies seek to

offset their emissions.

- **EUAs:** The EUAs factor is identified as another critical factor in all methods, underscoring the strong influence of the international carbon market. In the CCMFI model, EUAs had an importance value of 1.032, with an MIC score of 0.986. Correlation analyses further highlight its significance, with Spearman (0.8920) and Pearson (0.8436) values reflecting a strong positive relationship with carbon credit prices, as confirmed by SHAP values. As the largest cap-and-trade system worldwide, changes in the supply and demand of EUAs directly affect carbon credit prices worldwide [91], [163].

- **RE:** The RE factor reflects the influence of the adoption of renewable energy on the price of carbon credits. In the analysis, the CCMFI model assigned RE an importance value of 0.475, supported by negative correlations in the Spearman (-0.8517) and Pearson (-0.7118) tests. The SHAP values further confirm their impact, showing a negative contribution to carbon credit price predictions. As more renewable energy sources (for example, wind and solar) are integrated into energy systems, carbon emissions decrease significantly because renewables replace fossil fuels, the main sources of CO2 emissions [39], [61], [127]. This reduction in emissions can reduce the demand for carbon credits, as fewer companies need credits to offset their emissions, potentially leading to reduced carbon credit prices. It should be noted that this thesis represents the first comprehensive study to investigate the effect of renewable energy adoption on the price of carbon credits. To the best of the researcher's knowledge, no prior research has systematically examined the influence of this factor within the context of the carbon market, making this an original contribution to the field.

- **IO:** The IO factor, in the economic category, represents the prices of raw

materials that directly impact the price of carbon credits by influencing the production costs of industries relying on carbon-intensive materials [1]. In the CCMFI model, IO factor had an importance value of 0.292, supported by a MIC score of 0.973. Correlation analyses further underscore its significance, with Spearman (0.8062) and Pearson (0.5937) values showing positive relationships. The SHAP values highlight IO's contribution, indicating its role in increasing carbon credit prices. When raw material prices rise, production costs for carbon-intensive industries increase, which can lead to higher demand for carbon credits and, consequently, higher carbon credit prices.

The results of this comprehensive analysis indicate that the key features may vary depending on the method applied. Therefore, the CCMFI model was selected for feature selection due to its ability to address the limitations of traditional methods, correlation analysis (Pearson and Spearman), and MIC. Correlation analysis, while effective for identifying linear or monotonic relationships, lacks the capacity to capture complex, nonlinear interactions between variables. MIC improves this by detecting broader associations but does not provide interpretability or detailed information on the impact of individual factors. In contrast, CCMFI leverages the power of the RF model to handle nonlinear, dynamic relationships and integrates SHAP values to offer interpretable, data-driven rankings of feature importance. As a result, in this thesis, the CCMFI method is used to identify the main external influences on carbon prices.

7.5 Conclusion

This chapter identified the key external factors influencing carbon credit prices through a comprehensive multi-factor analysis. To achieve this, we analysed the effect of the 22 external factors on ACCU carbon credit prices using correlation analysis, MIC, and the proposed CCMFI model. The CCMFI model, which inte-

grates RF regression, and SHAP values, demonstrated superior capability in capturing complex, non-linear dependencies and providing interpretable insights. The model effectively identified and ranked the key factors, achieving an MSE of 1.443. The CCMFI model was ultimately selected as the most effective technique due to its ability to address the limitations of traditional correlation-based and statistical methods. The findings highlight that energy markets (crude oil and coal), environmental factors (CO_2 and GHG emissions), economic indicators (NONF and PT) and the international carbon market, EUAs, are among the most significant drivers of carbon credit prices. These findings lay the foundation for the next chapter, which builds on the key factors identified to develop predictive models for carbon credit prices, addressing the third research question of this thesis.

Chapter 8

Carbon Credit Price Prediction Modelling

8.1 Overview

In recent years, machine learning has increased significantly and is one of the most promising technologies in various fields, such as the stock market. By providing insights into market trends and predicting price fluctuations, machine learning models can enhance the accuracy of financial forecasting and provide a valuable competitive advantage to investors and financial firms [135]. The accurate prediction of carbon credit prices can lead to more informed trading decisions, higher profitability, and reduced risk [133]. Moreover, machine learning models can process large amounts of data from different sources and recognise complex patterns that humans may miss [30], which can help investors identify emerging trends and quickly adjust their investment strategies, thus enhancing the potential returns of their investments. However, predicting carbon credit prices using machine learning models comes with its own set of challenges. Carbon credit prices have complex dynamics that exhibit non-stationary, non-linearity, and multiclonality features, as statistically proven in Chapter 6 section 6.5. Research indicates that many external factors influence the carbon market [54], [155], [167]. Factors such as energy prices, product prices, weather conditions, international carbon prices, and environmental factors play a significant role in determining the volatility and dynamics of carbon credit prices [171], [162]. This complexity presents significant challenges in accurately predicting the price of carbon credits [174].

This chapter provides a comprehensive solution to the third research question

and discusses the proposed prediction models in further detail. The focus of this chapter is to develop an advanced multi-factor carbon credit prediction (CCMFP) model to predict the short-term price of carbon credits. The proposed CCMFP model employs a hybrid approach consisting of two stages. In stage 1, we employ feature extraction and reduction techniques (PCA, ICA and NLICA) to improve the quality of the prediction models' input by extracting essential features and mitigating the issue of multicollinearity within the set of independent variables influencing carbon credit prices [52]. Multicollinearity arises when independent factors demonstrate high correlation, which can cause instability in prediction models [60]. The identification of the optimal number of IC components is systematically performed through a grid search with k-fold cross-validation, ensuring the selection of the optimal number of components that minimises the MSE of prediction models. Subsequently, in Stage 2, the extracted components are used as input for the SVR and LMP algorithms to build the prediction model. By identifying statistically independent components from the original variables, PCA, ICA, and NLICA effectively reduce multicollinearity and generate components that are more suitable for machine learning applications [60].

This chapter is structured as follows: Section 8.2 gives a brief overview of the adopted machine learning SVR and deep learning MLP models. The techniques used to extract independent components of the influences are explained in Section 8.3. Section 8.4 introduces the proposed CCMFP models for predicting short-term carbon credit prices. Section 8.5 provides a detailed explanation of the experimental procedures performed to obtain the results presented in this chapter, including model hyperparameter tuning. The results and a discussion are presented in Section 8.6. Finally, Section 8.7 concludes this chapter.

8.2 An Overview of SVR and MLP

This section provides an overview of the support vector regression (SVR) model and multilayer feedforward perceptron (MLP) neural networks, and investigates the use of these algorithms to accurately predict the price of carbon credits.

8.2.1 An overview of the SVR model

Rooted in the statistical learning VC-theory developed by Vapnik and Chervonenkis [142], support vector machines (SVMs) are a powerful machine learning model that is used to make predictions and generalise data [29]. SVMs are primarily designed to solve binary classification problems. Essentially, a SVM constructs a hyperplane with margins and attempts to maximise the width of this margin while correctly classifying the majority of data points. Ideally, it strives to maintain separation between the margin and the hyperplane, denoted by ϵ, as large as possible, ensuring that no data points fall within the margin.

SVR extends the principles of SVM to effectively estimate continuous values [129]. In contrast, SVR approaches the problem from an opposite angle. SVR aims to minimise the width of the hyperplane's tube while ensuring that most of the data points fall within this tube. The objective is to adjust the hyperplane and margin to accommodate most data points, thereby minimising prediction error. Minimisation implies that the distance between the data points and the hyperplane should be less than ϵ. Errors within this threshold are not considered, but any error exceeding ϵ is not tolerated. Figure 8.1 illustrates that SVR is formulated using a soft margin approach, where the slack variables ξ_1 and ξ_2 allow some data points to deviate slightly from the margin while still considered acceptable. These slack variables allow the model to include nearby points that have a minimal deviation from the margin.

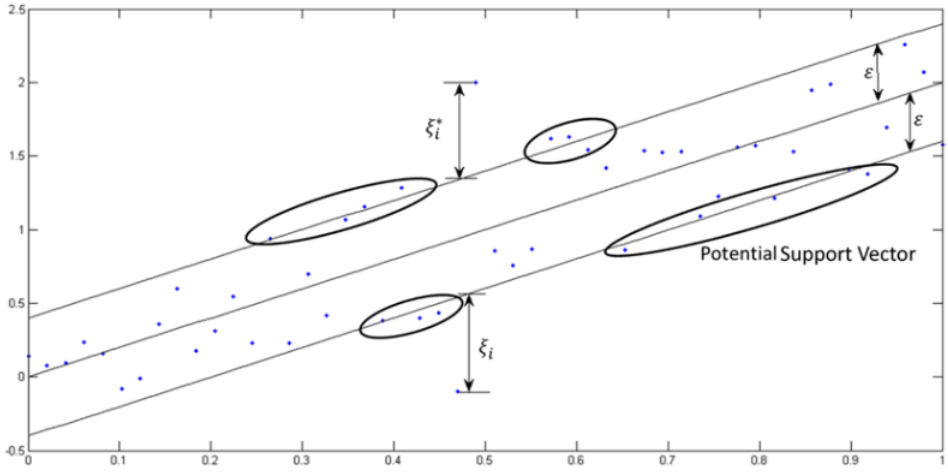


Figure 8.1 : One-dimensional linear SVR [9].

In cases where data are not linearly separable, the SVR decision boundary may fail. To address this, kernels transform the data into a higher-dimensional space, enabling SVR to handle non-linear data distributions. In SVR, the challenge lies in selecting the appropriate kernel and loss function parameters. In this thesis, our model employs the Gaussian radial basis function (RBF) kernel because of the inherent non-linear characteristics of our data. The RBF kernel function is mathematically represented as:

$$K(x_i, x_j) = \exp\left(-\frac{||x_i - x_j||^2}{2\sigma^2}\right) \tag{8.1}$$

Equation 8.1 effectively transforms the data into a higher-dimensional space, allowing SVR to manage complex, non-linear relationships within the data. In financial markets, sudden fluctuations can occur due to external factors such as energy markets and environmental factors. SVR dynamically adjusts its margins to accommodate these shifts within a certain limit, thus improving its predictive accuracy in such a volatile market. Using the RBF kernel, the SVR model adeptly captures and models these non-linear patterns, providing a robust model for predicting carbon

credit prices with higher accuracy.

8.2.2 An overview of the MLP neural network model

A multilayer perceptron (MLP) is a type of feedforward artificial neural network (ANN) used in various machine learning applications [80]. The MLP neural network processes inputs by passing them through a series of layers, including an input layer, one or more hidden layers, and an output layer. Figure 8.2 illustrates an MLP network with two hidden layers, an input and an output layer. MLPs effectively model complex relationships between inputs and outputs, making them suitable for tasks like regression and classification.

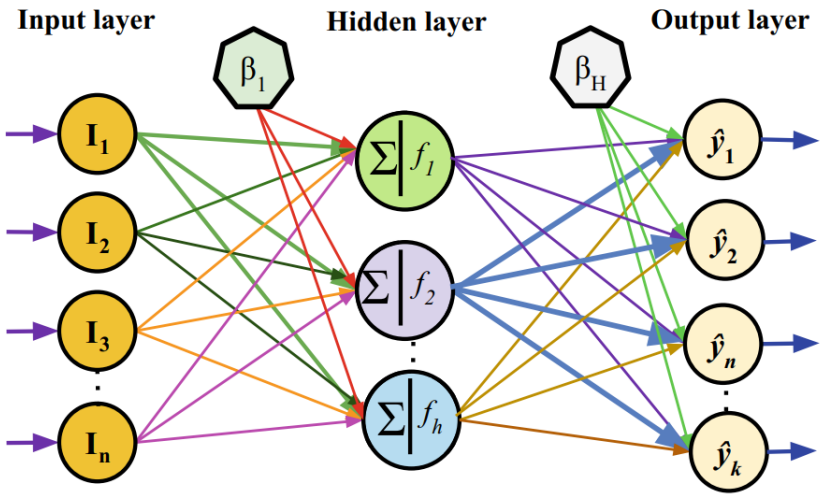


Figure 8.2 : Structure of MLP neural network [55].

The training process for an MLP uses backpropagation, which adjusts the network's weights to minimise the error between predicted and actual outputs, allowing the model to learn and improve its accuracy. For each neuron j in a layer, the input values I_i are multiplied by their corresponding weights ω_{ij}, and these products are summed together along with a bias term β_j to compute the summation output S_j, as shown in Equation 8.2.

$$S_j = \sum_{i=1}^{n} \omega_{ij} I_i + \beta_j \tag{8.2}$$

The summation output S_j is then passed through an activation function $f_j(x)$ to introduce non-linearity into the network.

Finally, the output of neuron j, denoted as y_i, is obtained by applying the activation function to the summation output, as illustrated in Equation 8.3.

$$y_i = f_j \left(\sum_{i=1}^{n} \omega_{ij} I_i + \beta_j \right) \tag{8.3}$$

This process enables the MLP to model complex, non-linear relationships in data by learning the appropriate weights and biases during the training phase. MLP has high learning potential, robustness against noise, non-linearity, parallelity, fault tolerance, and high generalisation capabilities [42], [55]. Therefore, the MLP model is appropriate to predict carbon credit prices, given the complexity of the data and the challenges associated with the carbon market. First, MLPs can effectively model the complex and non-linear relationships present in the dynamic and chaotic nature of carbon markets. Second, they do not require specific assumptions about the data distribution, making them flexible for various types of financial datasets. Third, the MLP's ability to effectively learn from limited data while capturing intricate patterns makes it ideal for predicting carbon credit prices. Fourth, once trained, MLPs can efficiently predict unseen data without the need to reprocess the training data. However, determining the optimal MLP architecture, including the selection of learning algorithms, the number of layers, the number of neurons in each layer, and the appropriate activation functions, is a complex task that requires careful consideration and experimentation.

8.3 Techniques for extracting independent components of carbon credit price influences

The motivation for this research is to address the inherent complexities in predicting carbon credit prices influenced by multiple external factors. Therefore, feature extraction and reduction techniques are essential to transform complex datasets into more informative components that reveal the underlying information and relationships in carbon markets [176]. Adopting the feature extraction and reduction technique enables one to identify critical features that contribute to the underlying patterns. This technique can improve classification and prediction models, particularly in scenarios where the data is highly correlated or noisy. For example, in finance, it can help identify and analyse hidden factors that influence asset prices and market trends [83].

By transforming correlated variables into uncorrelated components that capture most of the information, these techniques improve the quality of input to prediction models [69], [101]. Furthermore, they help eliminate noise and irrelevant data by focusing on the most significant components, thus simplifying the prediction model and reducing computational complexity without significant loss of information [60].

In the context of predicting carbon credit prices, adopting this technique helps uncover hidden factors that influence price movements, enabling machine learning models to make accurate short-term predictions. Therefore, in this study, we applied three techniques, PCA, ICA, and NLICA, to extract components that represent the underlying or hidden information in the original data to serve as input to our prediction model. The following subsections provide an overview of each of the applied techniques.

8.3.1 An overview of PCA

PCA is a powerful technique that is often used to simplify complex datasets by reducing their dimensionality while preserving as much of their original variability as possible [69], [125]. The goal is to transform the data into a new set of uncorrelated variables called principal components. These components capture the most significant patterns in the data, making it easier to understand and work with.

The process starts by normalising the data to ensure that each variable contributes equally to the analysis. This step prevents features with larger scales from dominating the results. Once the data are normalised, the mean of each variable is calculated to establish a baseline for comparison. Next, PCA involves calculating the covariance matrix, which helps us to understand how the variables interact and vary. This matrix forms the backbone of the PCA and highlights the relationships between data points. Using the covariance matrix, PCA calculates eigenvalues and eigenvectors. The eigenvalues inform us how much variance each principal component captures, while the eigenvectors point to the direction of that variance. In practice, eigenvalues are sorted from largest to smallest, identifying the principal components that capture the most variance in the dataset. These top components are then used to transform the original data, effectively reducing its complexity without losing critical information.

8.3.2 An overview of ICA

ICA is a statistical method proposed by [27], designed to describe multivariate data through linear combinations of independent components. This approach assumes that the multivariate data are composed of a linear mixture of several independent factors [138]. Generally, the number of these factors is assumed to be equal to the number of variables. ICA separates mixed signals into independent sources; it is similar to being at a party with multiple conversations happening simultaneously

and being able to focus on one conversation. By identifying and separating independent components within a dataset, each component can be analysed and understood individually, leading to a clearer understanding of the underlying phenomena.

8.3.3 An overview of NLICA

NLICA extends the traditional ICA model by relaxing the linearity assumption. While linear ICA aims to find a de-mixing matrix that linearly separates the observed data into statistically independent components, NLICA seeks to find a non-linear transformation that achieves the same goal [101]. This model is advantageous in capturing the complex, non-linear interactions between variables, which are often present in financial time-series data.

8.4 The proposed CCMFP models for predicting carbon credit prices

The CCMFP model leverages SVR and MLP as base models, ensuring reliable predictions while addressing the challenges of small dataset sizes. By selecting models suited to the data constraints, this approach improves the accuracy of prediction while maintaining computational efficiency. Deep learning models, such as long- and short-term memory (LSTM) networks and convolutional neural networks (CNNs), are highly effective for large-scale datasets but often struggle when applied to small datasets due to their high complexity and data dependency. Deep neural networks require large amounts of labelled data to achieve meaningful generalisation [112]. This challenge is further emphasised by [124], which highlights that deep learning models are inherently data-hungry and prone to overfitting when trained on small datasets. In such cases, reducing model complexity becomes crucial to prevent overfitting, as high-parameter models like deep neural networks do not perform optimally under these conditions [99]. Moreover, learning from small datasets remains a sig-

nificant challenge, as deep learning architectures often fail to generalise effectively [15].

Given that the dataset used in this study is relatively small, applying deep learning models would likely result in overfitting and unreliable predictions. Instead, SVR and MLP were chosen as they are better suited for small datasets, as they do not require extensive data to generalise effectively, making them a more appropriate choice for short-term carbon credit price prediction.

The proposed methodology consists of two stages. In stage 1, we employ feature extraction and reduction techniques (PCA, ICA and NLICA) as a pre-processing step to extract essential features and mitigate the issue of multicollinearity within the set of independent variables influencing carbon credit prices [52]. By identifying statistically independent components from the original variables, PCA, ICA, and NLICA effectively reduce multicollinearity and generate components that are more suitable for machine learning applications [60]. Subsequently, in Stage 2, the extracted components are used as input for the SVR and LMP algorithms to build the prediction model. Finally, six prediction models were developed using the proposed CCMFP model to predict short-term prices of carbon credits. Note that the SVR models that use PCA, ICA and NLICA as feature extraction and reduction are called PCA_SVR, ICA_SVR and NLICA_SVR models, respectively. Similarly, the MLP models that use PCA, ICA and NLICA as feature extraction are called PCA_MLP, ICA_MLP and NLICA_MLP models, respectively. Figure 8.3 outlines the structured approach and key components of the CCMFP model. The comprehensive approach is explained through algorithms 2 to 5, which provide detailed steps for developing and evaluating the CCMFP model.

8.5 Evaluation of the proposed CCMFP prediction models

This section provides an overview of the experimental procedures carried out to obtain the results presented in this study. This section is organised as follows: Subsection 8.5.1 provides an overview of the preprocessed data used to validate the proposed carbon credit price prediction models. The hyperparameter tuning process for SVR and MLP models is detailed in Subsection 8.5.2. Subsection 8.5.3 details the implementation steps to develop the prediction models using the SVR and MLP models. Subsection 8.5.4 outlines the metrics used to evaluate the performance of the models in predicting the short-term prices of carbon credits.

8.5.1 Dataset used for carbon credit price prediction modelling

We used a dataset from the real-world carbon market to validate the proposed model, focussing on the Australian carbon market as a case study. A total of 22 influencing factors were identified from the literature and then real data for each factor was collected from reliable and authoritative sources. Data collection and pre-processing were performed as detailed in Chapter 6. The pre-processed datasets were merged into a CSV file, using the 'Date' column as the index. The target variable (ACCUs) was then separated from the independent variables. Table 8.1 lists the external factors considered in this study, ordered by feature importance values derived from the proposed CCMFI model (further details are provided in Chapter 7).

Furthermore, we introduced lag features for the previous day to address the non-stationarity commonly observed in carbon credit price data. This approach captures temporal dependencies and is widely adopted in financial time series forecasting [90]. By incorporating lagged values, the model can better adapt to trends and patterns over time.

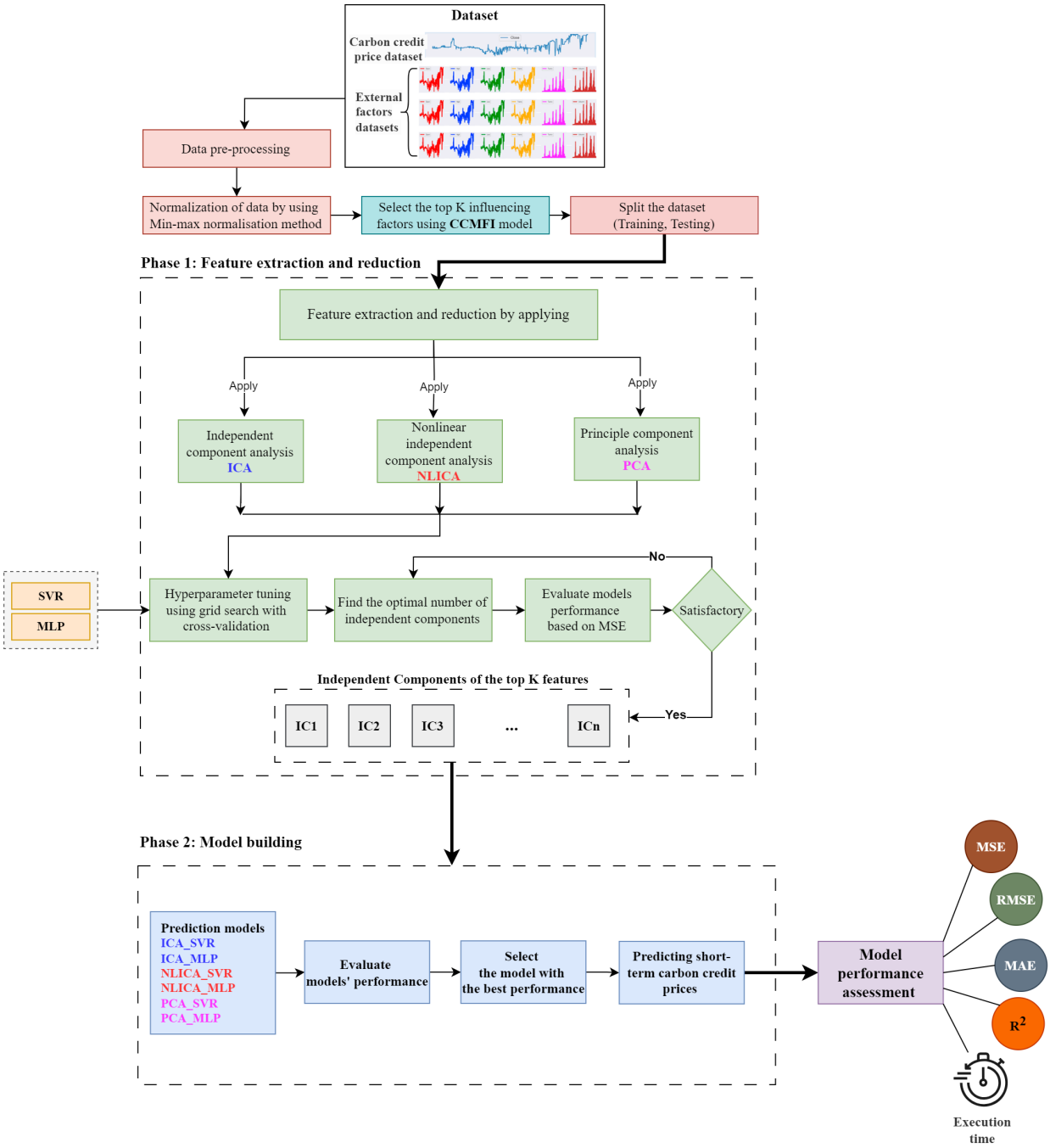


Figure 8.3 : Overview of the proposed CCMFP model with the applied approach and methods.

Table 8.1 : Selected influencing factors on carbon credit prices, ordered by their importance from most to least significant.

Rank	Factor	Description
1	PT	Green technology maturity, represented by patents on environmental technologies
2	CO2	Carbon emissions, represented by Australia's carbon CO2 emissions
3	NONF	Product price, represented by AU basic non-ferrous metal product manufacturing
4	CP	Coal market, represented by Australia's BJ power coal spot prices
5	EUAs	International carbon market, represented by European Union Allowances (EUAs)
6	CrP	Crude oil market, represented by average global crude oil
7	RE	Renewable energy, represented by Australia's renewable energy
8	GHG	GHG emissions, represented by Australia's greenhouse gas emissions
9	IO	Raw material price, represented by Australia's iron ore
10	CEAs	International carbon market, represented by China Emission Allowances (CEAs)
11	BC	Industrial development, represented by the business confidence indicator
12	GDP	Per capita growth rate, represented by Australia's per capita GDP growth rate
13	AUD	Exchange rate, represented by AUD/USD
14	IND	Development degree of low carbon, represented by Australia's industry emissions
15	IA	Science and technology innovation, represented by the percentage of Australian household internet access
16	NG	Natural gas market, represented by Australia's natural gas
17	REI	Clean energy penetration, represented by renewable usage in industry
18	SP	Stock market, represented by S&P/ASX 200 Index price
19	EP	Electricity market, represented by Australia's electricity spot prices
20	POP	Population growth rate, represented by Australia's population growth rate
21	AQI	Air quality index, represented by Australia's Air Quality Index (AQI)
22	MA	Air temperature, represented by Australia's air temperature

8.5.2 Hyperparameter tuning for SVR and MLP models

In this study, we applied the 10-fold cross-validation grid search to determine the optimal hyperparameters for both the SVR and MLP models. To do this, we used the `GridSearchCV` technique from the `sklearn.model_selection` module.

- For the SVR model, the key hyperparameters are `C`, `epsilon`, and `gamma`, which must be carefully adjusted to achieve the best generalisation perfor-

mance. The `C` parameter controls the trade-off between achieving a low training error and a large margin; higher values of `C` enforce stricter penalties on errors, leading to a more complex model. The `epsilon` parameter defines the width of the margin within which no penalty is given to training errors, influencing the number of support vectors used to build the regression model. Finally, the `gamma` parameter for the RBF kernel determines how far the influence of a single training example reaches, with larger values making the model more sensitive to individual data points, potentially leading to overfitting. When using the grid search, the parameters `C`, `epsilon`, and `gamma` were tuned by systematic hyperparameter optimisation based on the ranges specified in Table 8.2.

- For the MLP model, the key hyperparameters are `hidden_layer_sizes`, `activation function`, `optimizer`, and `alpha`. The `hidden_layer_sizes` parameter defines the neural network's architecture by specifying the number and size of the hidden layers, allowing the model to capture varying levels of complexity in the data.

  The `activation` function determines the type of non-linearity applied at each neuron, significantly influencing the model's ability to learn complex patterns. This study employs `tanh` and `ReLU`, two widely used activation functions:

  tanh (Hyperbolic Tangent) activation function: Defined as:

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

  The `tanh` function maps inputs to a range between $[-1, 1]$, ensuring zero-centred activation functions that accelerate convergence. It is particularly useful for capturing fine-grained nonlinear relationships in the data [33].

ReLU (Rectified Linear Unit) activation function: Defined as:

$$f(x) = \max(0, x)$$

ReLU is computationally efficient and mitigates the problem of vanishing gradients, making it well suited for deeper networks [111]. It allows the MLP to learn sparse, high-dimensional representations by setting negative values to zero, thereby improving gradient flow and training stability.

The `optimizer` parameter controls how the model optimises its weights during training, with `adam` and `sgd` common choices for adaptive and gradient-based optimisation. The `alpha` parameter regulates the strength of regularisation to mitigate overfitting by penalising large weights.

The parameters `hidden_layer_sizes`, `activation function`, `optimizer`, and `alpha` were tuned by systematic hyperparameter optimisation based on the ranges specified in Table 8.3.

The grid search was used to explore all possible combinations of hyperparameters within the predefined ranges, with the optimal values selected based on the lowest MSE during 10-fold cross-validation. MSE served as the evaluation metric for both hyperparameter tuning and determining the optimal number of components. A lower MSE indicates a better-performing model, with the optimal number of components corresponding to the lowest MSE. This method ensured that both models were configured for accurate short-term carbon credit price prediction.

8.5.3 The experimental setup and implementation

After exploring the feature importance, six prediction models were developed using the proposed CCMFP model to predict the short-term price of carbon credits. The CCMFP model consists of two key phases: the feature extraction and reduction

Table 8.2 : Hyperparameters of SVR

Hyperparameters	**Value**
kernel	{RBF}
C	{0.1, 1, 10}
epsilon	{0.01, 0.1, 0.5}
gamma	{scale, auto}

Table 8.3 : Hyperparameters of MLP

Hyperparameters	**Value**
hidden_layer_sizes	[(50,), (100,), (50, 50)]
activation function	{tanh, relu}
optimizer	{adam, sgd}
alpha	{0.0001, 0.001}

phase and the prediction model construction phase. The comprehensive approach is explained through algorithms 2 to 5, which provide detailed steps for the development and evaluation of the prediction models. Algorithms 2, 3 and 4 illustrate how PCA, ICA and NLICA, respectively, are applied to SVR and MLP to search for the optimal IC number based on MSE. Algorithm 5 illustrates the procedure of the proposed CCMFP model. The following is the step-by-step methodology of the proposed CCMFP approach for predicting the short-term carbon credit price using the MLP and SVR models:

1. The methodology starts by uploading the pre-processed dataset. Then, we split the pre-processed dataset into training and testing sets. This ensures that the model is trained on a portion of the data and validated on unseen

data to assess its performance. In this study, an 80/20 split is used, where 80% of the data is used for training and 20% for testing.

2. Then, the key factors influencing carbon credit prices are selected using the CCMFI model as reported in Table 8.1. This step focuses on identifying the key influencing factors that have a significant impact on carbon credit prices. For a detailed explanation of the CCMFI model, refer to Chapter 7.

3. After selecting the key influencing factors, we proceed with feature extraction and reduction by applying PCA, ICA, and NLICA. Each technique transforms the data into a reduced set of independent components expected to capture the essential information from the original variables while minimising redundancy and noise. Therefore, the important question arises: How do we determine the optimal number of ICs that leads to the highest accuracy? In our study, the number of components is determined using two approaches: a variance threshold, as used in PCA, or a systematic process of experimentation, where models are evaluated across a range of component numbers, as employed in both ICA and NLICA. The following explains the implementation of each applied reduction technique:

   (a) PCA is applied to identify the principal components (PCs) that capture the maximum variance in the dataset. The number of components to retain is determined by the `explained_variance_ratio` parameter, which specifies the proportion of total variance to be preserved. In this study, we set the `explained_variance_ratio` to 95%, ensuring that the selected principal components account for 95% of the variance in the data. The `PCA` model is initialised with this ratio, and the `fit_transform` method is used to compute the PCs and transform the data into the reduced-dimensional space. The final number of components, `n_components`, is

automatically set by the PCA algorithm based on the explained variance threshold. Figure 8.4 presents a code snippet to define and apply PCA using Python. Algorithm 2 outlines the steps for applying the PCA technique.

```
# Determine the optimal number of components based on explained variance ratio
pca = PCA(n_components=explained_variance_ratio, random_state=42)
X_transformed = pca.fit_transform(X)
n_components = pca.n_components_
```

Figure 8.4 : Code snippet for defining and applying PCA.

(b) In the case of ICA, FastICA is applied to extract independent components. Unlike PCA, which uses explained variance to determine the number of components, ICA focuses on maximising statistical independence between components. In this study, the number of ICs is not predefined; instead, we begin by defining a range from 1 to the total number of features in the dataset and systematically loop through this range. For each possible number of components, FastICA is applied, and the resulting model is evaluated based on its MSE. The optimal number of ICs is selected based on the configuration that gives the lowest MSE. The FastICA model is initialised with the `n_components` parameter, which specifies the number of ICs to extract. The `random_state=42` parameter is set to ensure the reproducibility of the results, while `max_iter=1000` limits the number of iterations to achieve convergence. The convergence tolerance is controlled by `tol=0.01`, ensuring that the algorithm stops once a sufficiently accurate solution is reached. During each iteration, the `fit_transform` function is applied to the dataset `X`, returning the

transformed data, `X_transformed`, where the original data are decomposed into the specified number of independent components. Figure 8.5 presents a code snippet to define and apply ICA using Python. Algorithm 3 outlines the steps for applying the ICA technique.

```
for n_components in n_components_range:
    ica = FastICA(n_components=n_components, random_state=42, max_iter=1000, tol=0.01)
    X_transformed = ica.fit_transform(X)
```

Figure 8.5 : Code snippet for defining and applying ICA.

(c) In the case of NLICA, an autoencoder is used to compress the data into a lower-dimensional representation, capturing non-linear structures that linear methods PCA or ICA might miss. Similar to ICA, the number of independent components is not predefined. Instead, we define a range from 1 to the total number of features in the dataset and loop through this range. For each value, the autoencoder is configured to reduce the data to that number of components. The transformed data are then used to evaluate the model's performance based on MSE, and the number of components that result in the lowest MSE is selected as the optimal configuration. The function `create_autoencoder` is used to build the autoencoder architecture, where `input_dim` represents the number of input features, and `encoding_dim` defines the number of dimensions to which the input data will be compressed (that is, the number of independent components). The model begins with an input layer, where the input shape is set to `input_dim`. This input is then passed to an encoding layer, which is a fully connected (dense) layer with `encoding_dim` neurons. The ReLU activation function is applied to introduce non-linearity, allowing the model to capture complex patterns in the data. To reduce

overfitting, a dropout layer with a dropout rate of 0.2 is added after the encoder. The decoder reconstructs the input from the compressed encoding by passing it through a dense layer with the same number of neurons as the input (`input_dim`). A sigmoid activation function is applied to map the encoded data back to its original range. The autoencoder is compiled using the Adam optimiser, which was selected for its ability to converge during training quickly. The loss function is set to MSE, making it suitable for minimising the reconstruction error between the original and reconstructed data. In the `for` loop that follows, the code iterates over different values for `n_components` (the number of independent components). For each iteration, a new autoencoder is created with the corresponding number of components. The model is trained on the dataset with 100 epochs, a batch size of 32, and a validation split of 0.2. The shuffle option is set to `True`, meaning the data is shuffled before each epoch to improve generalisation performance. After training, the encoder model is used to transform the input data into its lower-dimensional representation using the `predict` function. Figure 8.6 presents a code snippet to define and apply NLICA using Python. Algorithm 4 outlines the steps for applying the NLICA technique.

```python
# Define an autoencoder model for nonlinear ICA
def create_autoencoder(input_dim, encoding_dim):
    input_layer = Input(shape=(input_dim,))
    encoder = Dense(encoding_dim, activation="relu")(input_layer)
    encoder = Dropout(0.2)(encoder)  # Add dropout to prevent overfitting
    decoder = Dense(input_dim, activation="sigmoid")(encoder)
    autoencoder = Model(inputs=input_layer, outputs=decoder)
    encoder_model = Model(inputs=input_layer, outputs=encoder)
    autoencoder.compile(optimizer='adam', loss='mse')
    return autoencoder, encoder_model
```

```python
for n_components in n_components_range:
    autoencoder, encoder_model = create_autoencoder(X.shape[1], n_components)
    autoencoder.fit(X, X, epochs=100, batch_size=32, shuffle=True, validation_split=0.2, verbose=0)
    X_transformed = encoder_model.predict(X)
```

Figure 8.6 : Code snippet for defining and applying NLICA.

4. Next, we used the generated ICs obtained from the feature extraction and reduction techniques as input to train the SVR and MLP models. Perform hyperparameter tuning using grid search with cross-validation to optimise the performance of these models. The best model configuration is selected based on the cv grid search method's performance, measured by the lowest MSE.

5. Finally, six prediction models were developed: ICA_SVR, NLICA_SVR, PCA_SVR, ICA_MLP, NLICA_MLP and PCA_MLP. These models were then used to predict the price of carbon credits (ACCUs) on the test dataset. The performance of the models was assessed using five different metrics, MSE, RMSE, MAE, R^2, and execution time.

Algorithm 2 Function: OptimalModel_PCA

Input:
  X_{train}: Training set of independent variables
  y_{train}: Training set of target variable
  explained_variance_ratio: Variance ratio for PCA method (default = 0.95)
Output:
  PC: Optimal number of PCA components
  $bestModel$: The best model with the lowest MSE
function OPTIMALMODEL_PCA($X_{\text{train}}, y_{\text{train}}, explained_variance_ratio$)
  Construct a pipeline using PCA with *explained_variance_ratio*
  PC = pca.n_components_
  Perform *GridSearchCV* on training data to find PC that minimizes MSE
  return $bestModel$, PC
end function

Algorithm 3 Function: OptimalModel_ICA

Input:
  X_{train}: Training set of independent variables
  y_{train}: Training set of target variable
  n_components_range: Range of components for the ICA method
Output:
  IC: Optimal number of ICA components
  $bestModel$: The best model with the lowest MSE
function OPTIMALMODEL_ICA($X_{\text{train}}, y_{\text{train}}, n_components_range$)
  for each IC in *n_components_range* **do**
    Construct a pipeline using $FastICA$ with IC components
    Perform *GridSearchCV* on training data to find the number of IC that minimises MSE
  end for
  return $bestModel$, IC
end function

Algorithm 4 Function: OptimalModel_NLICA

Input:
  X_{train}: Training set of independent variables
  y_{train}: Training set of target variable
  n_components_range: Range of components for the NLICA method
Output:
  IC: Optimal number of NLICA components
  $bestModel$: The best model with the lowest MSE
function OPTIMALMODEL_NLICA($X_{\text{train}}, y_{\text{train}}, n_components_range$)
  for each IC in *n_components_range* **do**
    Construct a pipeline using an *autoencoder* with IC components
    Perform *GridSearchCV* on training data to find IC that minimizes MSE
  end for
  return $bestModel$, IC
end function

Algorithm 5 CCMFP: Carbon Credit Multi-Factor Prediction Model

Input:
  X_i: Matrix of independent variables, $i = 1, 2, \ldots, S$ (where S is the number of external factors)
  y: Vector of the target variable (carbon credit price)
  technique_list: {PCA, ICA, NLICA} // List of feature reduction techniques

Output:
  Prediction results: MSE, RMSE, MAE, R^2, execution time

Initialisation:
  Load and preprocess the dataset
  Extract independent variables X and target variable y
  Split the data into training and testing sets: (X_{train}, X_{test}, y_{train}, y_{test})
  Set *n_components_range* = range(1, S) // Range of components for ICA and NLICA
  Set *explained_variance_ratio* = 0.95 // Variance ratio for PCA

Step 1: Selection of Best Model with Optimal Components
if *technique_list* == PCA **then**
  bestModel, *PC* = *OptimalModel_PCA*(X_{train}, y_{train}, *explained_variance_ratio*) //Algorithm 2
else if *technique_list* == ICA **then**
  bestModel, *IC* = *OptimalModel_ICA*(X_{train}, y_{train}, *n_components_range*) //Algorithm 3
else if *technique_list* == NLICA **then**
  bestModel, *IC* = *OptimalModel_NLICA*(X_{train}, y_{train}, *n_components_range*) //Algorithm 4
end if

Step 2: Model Prediction
  Use the *bestModel* to make predictions on the testing dataset

Model Evaluation
  Evaluate model performance using MSE, RMSE, MAE, R^2, and execution time

8.5.4 Evaluation metrics

The predictive performance of the model is assessed using MSE, RMSE, MAE and R^2. Lower MSE, RMSE, and MAE values indicate higher accuracy, while higher R^2 signifies a better fit to the model. These metrics provide insights into the accuracy and reliability of model predictions in the testing data. Table 8.4 provides the definitions of the evaluation metrics. In addition, the execution time, required to make the predictions, is used as an additional metric to evaluate the prediction models.

Table 8.4 : Performance measures and their definitions.

Metric	**Calculation**
Mean squared error	$MSE = \frac{1}{N}\sum_{t=1}^{n}(C_t - \hat{C}_t)^2$
Root mean squared error	$RMSE = \sqrt{\sum_{t=1}^{N}\frac{\left(\hat{C}_t - C_t\right)^2}{N}}$
Mean absolute error	$MAE = \frac{\sum_{t=1}^{N}\lvert\hat{C}_t - C_t\rvert}{N}$
R-squared	$R^2 = \frac{\sum_{t=1}^{N}(\hat{C}_t - C_t)^2}{\sum_{t=1}^{N}(C_t - \overline{C_t})^2}$

Note that C_t and $\hat{C}_t$ are the actual and predicted carbon credit prices, respectively. N is the total number of data points.

8.6 Results and discussion

This section provides a comprehensive analysis of the results generated from the proposed CCMFP models, highlighting the depth and rigour of our experimental evaluation. Three advanced feature extraction and reduction techniques, PCA, ICA, and NLICA, were applied to the SVR and MLP models, resulting in the development of six prediction models: `PCA_SVR`, `ICA_SVR`, `NLICA_SVR`, `PCA_MLP`, `ICA_MLP`, and `NLICA_MLP`. To further assess the performance of these models, we conducted extensive experiments by varying the number of top-selected factors ($K = 5, 10, 15, 22$) as input to the prediction models. This variation enabled us to systematically analyse how the number of input features impacts prediction accuracy and to identify the optimal set of features to achieve the most accurate predictions.

Furthermore, hyperparameter tuning was performed using grid search with cross-validation to ensure the best possible performance for each model. This process systematically explored a wide range of parameter combinations, ensuring that each prediction model was optimised for its respective configuration.

This section is organised into three subsections as follows: Subsection 8.6.1

presents the performance of the SVR_based models, Subsection 8.6.2 presents the performance of the MLP_based models, and Subsection 8.6.3 provides a comparative analysis of the results.

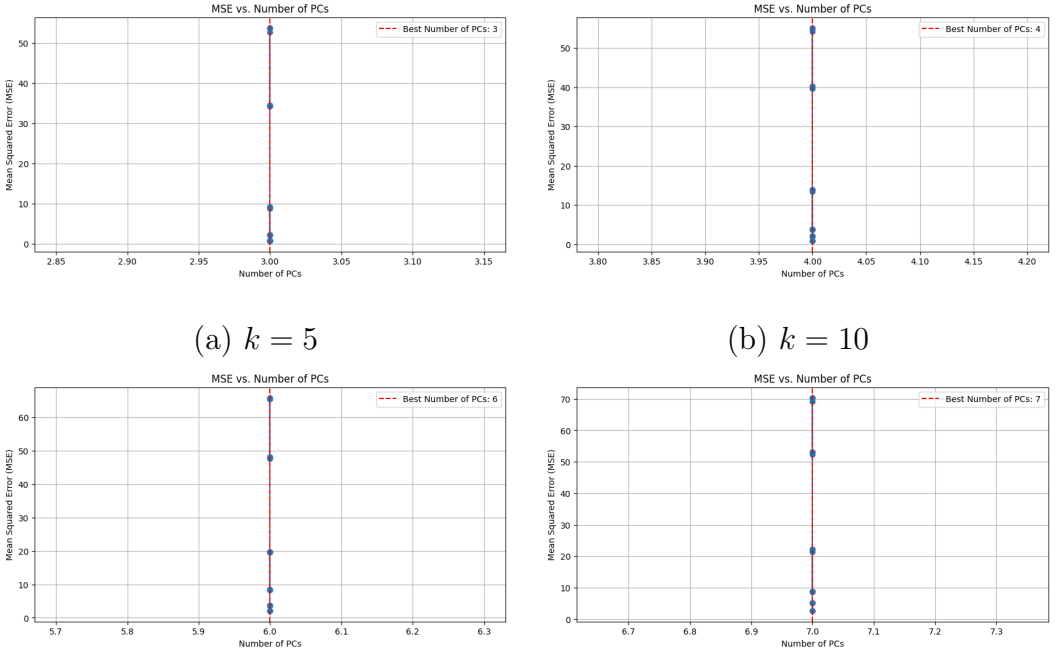
8.6.1 Evaluation of the proposed CCMFP using the ICA_SVR, NLICA_SVR and PCA_SVR models

We conducted a series of experiments to evaluate the ICA_SVR, NLICA_SVR, and PCA_SVR models, varying both the number of selected external factors K and the hyperparameters of the SVR_based models.

In the PCA_SVR model, the PCA technique and external factors are used to generate PCs, which represent the features of the input data and are used as new input variables to the SVR model. The results of the PCA_SVR model, including both the optimal hyperparameters and the corresponding optimal number of PCs, are summarised in Table 8.5. The optimal hyperparameters for the SVR model were found to be `C` = 10, `epsilon` = 0.1, and `gamma` = `scale`. The minimum mean MSE values were 0.7561, 0.8777, 2.0387 and 2.6060 when the number of selected external factors K was 5, 10, 15, and 22, respectively. Figure 8.7 provides additional insight into the tuning process conducted during the cross-validation grid search by illustrating the relationship between the MSE and the number of PCs for each configuration of K. The red dashed line in each figure indicates the optimal number of PCs that minimises the MSE. The optimal number of PCs for each configuration was 3, 4, 6, and 7.

Table 8.5 : Parameter grid search results for PCA_SVR model using a different number of external factor configurations.

C	Epsilon	Gamma	Mean MSE				Optimal number of PCs			
			K=5	K=10	K=15	K=22	K=5	K=10	K=15	K=22
0.1	0.01	scale	34.3112	40.2010	48.0089	53.2014	3	4	6	7
0.1	0.01	auto	53.6041	55.0750	65.5131	70.2878	3	4	6	7
0.1	0.1	scale	34.3908	40.2298	48.1325	53.1134	3	4	6	7
0.1	0.1	auto	53.7262	54.8322	65.5298	70.2884	3	4	6	7
0.1	0.5	scale	34.5430	39.5808	47.6781	52.3506	3	4	6	7
0.1	0.5	auto	52.6664	54.1538	65.8592	69.1483	3	4	6	7
1.0	0.01	scale	2.2729	3.8097	8.4333	8.8399	3	4	6	7
1.0	0.01	auto	9.2305	13.9441	19.7355	22.1195	3	4	6	7
1.0	0.1	scale	2.2513	3.7697	8.3797	8.8297	3	4	6	7
1.0	0.1	auto	9.0975	13.8624	19.7375	22.1501	3	4	6	7
1.0	0.5	scale	2.0976	3.8001	8.2744	8.6350	3	4	6	7
1.0	0.5	auto	8.7666	13.4679	19.6019	21.4296	3	4	6	7
10.0	0.01	scale	0.7563	0.8895	2.0440	2.6313	3	4	6	7
10.0	0.01	auto	0.8496	2.1214	3.6698	5.2331	3	4	6	7
10.0	0.1	scale	0.7561	0.8777	2.0387	2.6060	3	4	6	7
10.0	0.1	auto	0.8421	2.0929	3.6580	5.2276	3	4	6	7
10.0	0.5	scale	0.7855	0.9081	2.0743	2.6268	3	4	6	7
10.0	0.5	auto	0.8712	2.0485	3.5617	5.1985	3	4	6	7



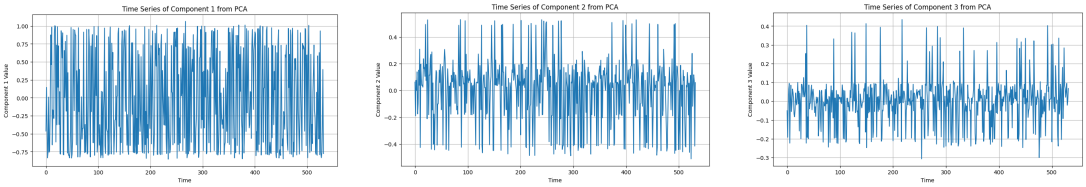
(a) $k = 5$ (b) $k = 10$

(c) $k = 15$ (d) $k = 22$ (All factors)

Figure 8.7 : PCA_SVR model hyperparameters tuning for different components using a different number of external factor configurations.

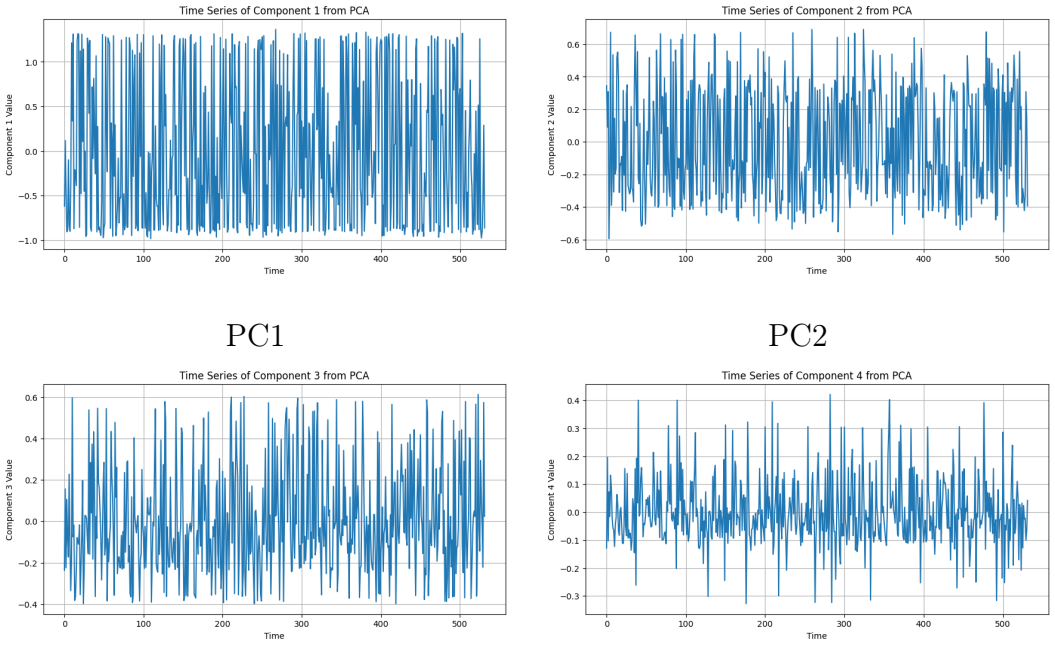
Figures 8.8 to 8.11 show the best-extracted PCs generated by the proposed PCA_SVR model from a different number of external factors ($K = 5$, $K = 10$, $K = 15$, and $K = 22$). Each PC represents a transformed version of the original input data, capturing the most significant variance within the dataset. These PCs are then used as new input variables for the SVR model.

Figure 8.12 compares the actual and predicted results for ACCUs prices using the proposed PCA_SVR model using a different number of external factors ($K = 5$, $K = 10$, $K = 15$, and $K = 22$).



PC1 PC2 PC3

Figure 8.8 : Best extracted PC components by the proposed PCA_SVR model using the top $K = 5$ factors.



PC1 PC2

PC3 PC4

Figure 8.9 : Best extracted PC components by the proposed PCA_SVR model using the top $K = 10$ factors.

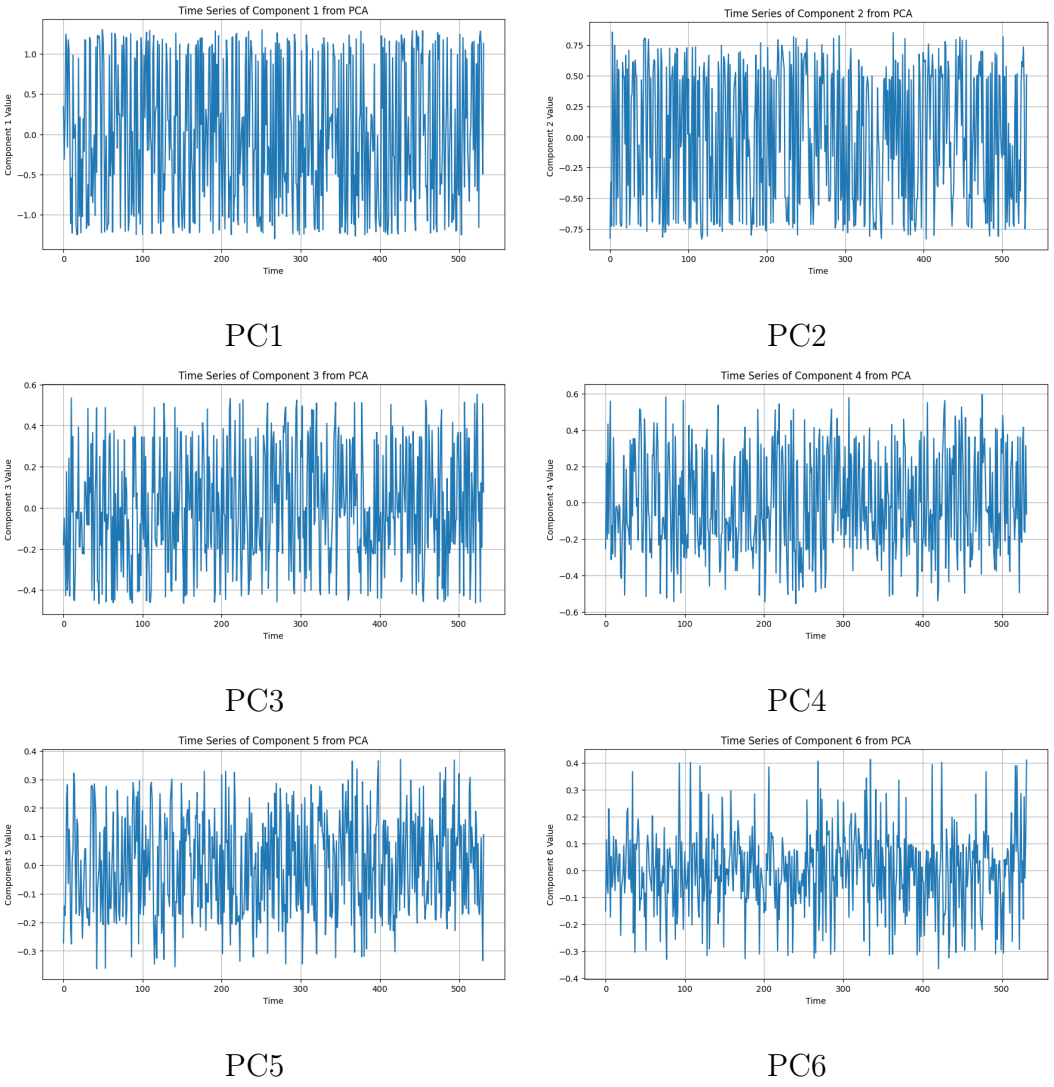
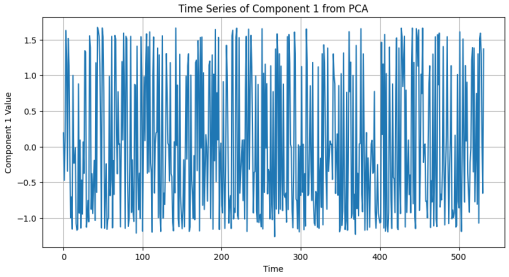
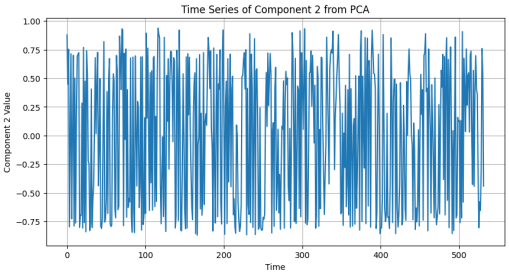


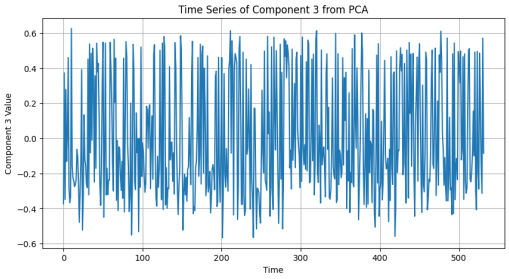
Figure 8.10 : Best extracted PC components by the proposed PCA_SVR model using the top $K = 15$ factors.



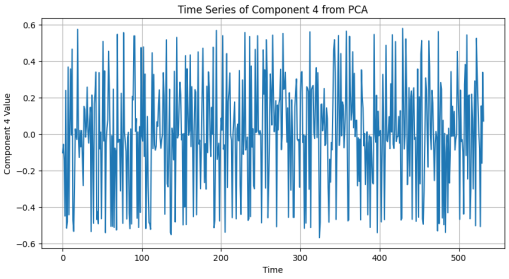
PC1



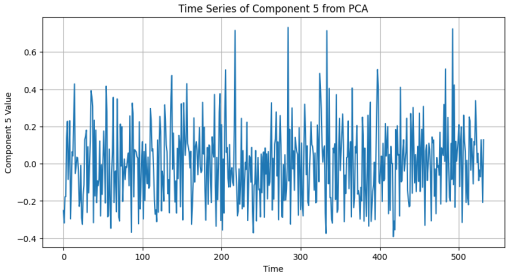
PC2



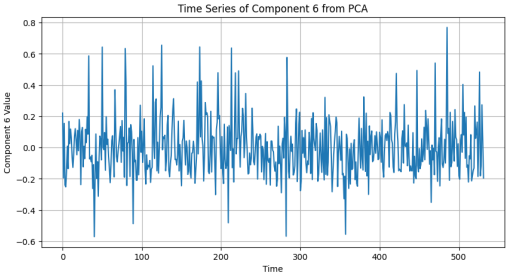
PC3



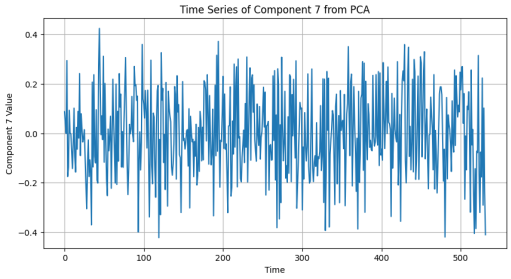
PC4



PC5

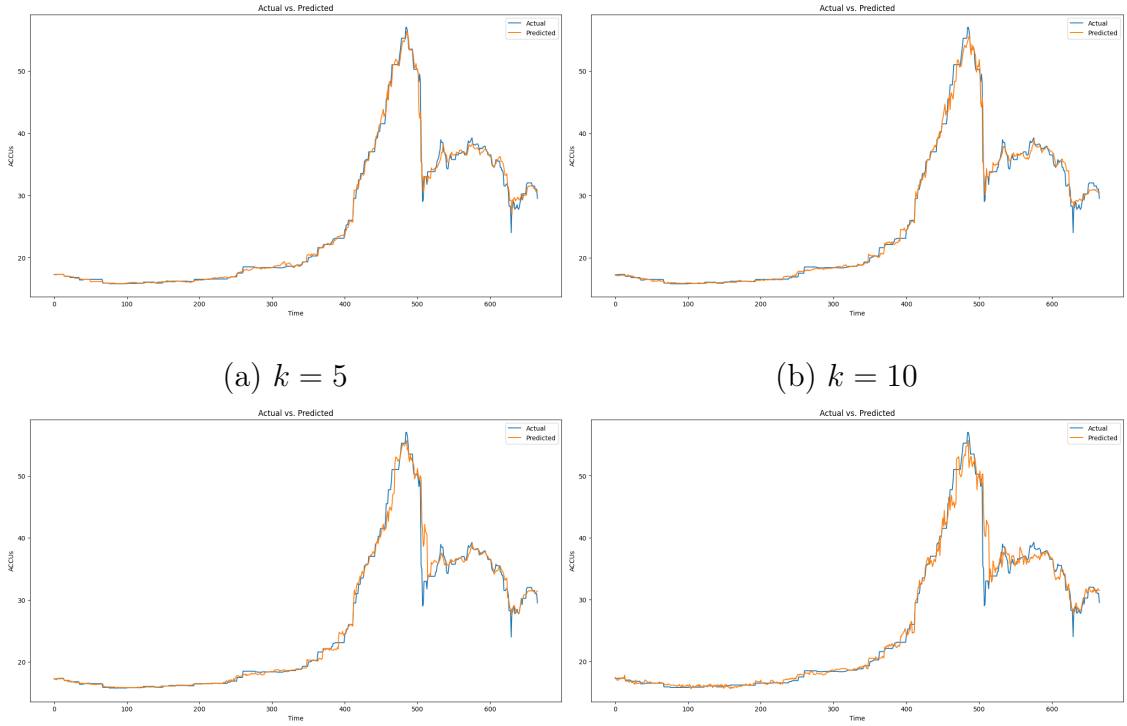


PC6



PC7

Figure 8.11 : Best extracted PC components by the proposed PCA_SVR model using all $K = 22$ factors.



(a) $k = 5$ (b) $k = 10$

(c) $k = 15$ (d) $k = 22$ (All factors)

Figure 8.12 : Actual vs. predicted ACCUs price using the proposed PCA_SVR model with different external factor configurations.

In the case of the ICA_SVR model, ICA was applied using the FastICA algorithm to extract ICs from the dataset. Unlike PCA, which selects components based on explained variance, ICA maximises statistical independence between the components. The number of ICs was not predefined; instead, a range was defined, from 1 to the total number of features in the dataset (22 external factors). For each possible number of components, the grid search was performed using cross-validation to optimise the SVR hyperparameters (`C`, `epsilon`, and `gamma`) and the model was evaluated based on its MSE performance.

Tables 8.6 to 8.9 present the results of the hyperparameter tuning for the ICA_SVR model, using a different number of external factors K. For the sake of space, the tables have been summarised to show only the key configurations and the best-

performing models with the optimal number of ICs. By systematically looping through different values for `n_components`, the lowest MSE was achieved with 3, 9, 13, 9 ICs when the number of external factors is ($K = 5$, $K = 10$, $K = 15$, and $K = 22$) respectively, as shown in Figure 8.13.

Figures 8.14 to 8.17 show the best IC components extracted by the proposed ICA_SVR model for a different number of external factors K. In addition, Figure 8.18 compares the actual and predicted results for ACCUs prices using the proposed ICA_SVR model using different numbers of external factors ($K = 5$, $K = 10$, $K = 15$, and $K = 22$).

Table 8.6 : ICA_SVR model hyperparameter tuning for different ICs using the top K = 5 external factors.

C	Epsilon	Gamma	Mean MSE	n_components
0.1	0.01	scale	35.2729	1
0.1	0.01	auto	35.2612	1
0.1	0.1	scale	35.3073	1
0.1	0.1	auto	35.3118	1
0.1	0.5	scale	35.2679	1
0.1	0.5	auto	35.2707	1
...	...	...	...	...
1	0.1	auto	22.8738	1
1	0.5	scale	22.8526	1
1	0.5	auto	22.8426	1
...	...	...	...	...
10	0.1	scale	2.7210	2
10	0.1	auto	2.7286	2
...	...	...	...	...
1	0.1	auto	1.8402	3
1	0.5	scale	2.0023	3
10	0.01	auto	0.7441	3
10	0.1	scale	0.7419	3
10	0.1	auto	0.7395	3
10	0.5	scale	0.7736	3
10	0.5	auto	0.7720	3
...	...	...	...	...
10	0.01	scale	1.2483	4
10	0.01	auto	1.2484	4
10	0.1	scale	1.2348	4
10	0.1	auto	1.2344	4
10	0.5	scale	1.2387	4
10	0.5	auto	1.2394	4
...	...	...	...	...
1	0.01	scale	3.0052	5
1	0.01	auto	3.0024	5
1	0.1	scale	2.9976	5
1	0.1	auto	2.9954	5
...	...	...	...	...
10	0.1	auto	1.1289	5
10	0.5	scale	1.1291	5
10	0.5	auto	1.1285	5

Table 8.7 : ICA_SVR model hyperparameter tuning for different ICs using the top K = 10 external factors.

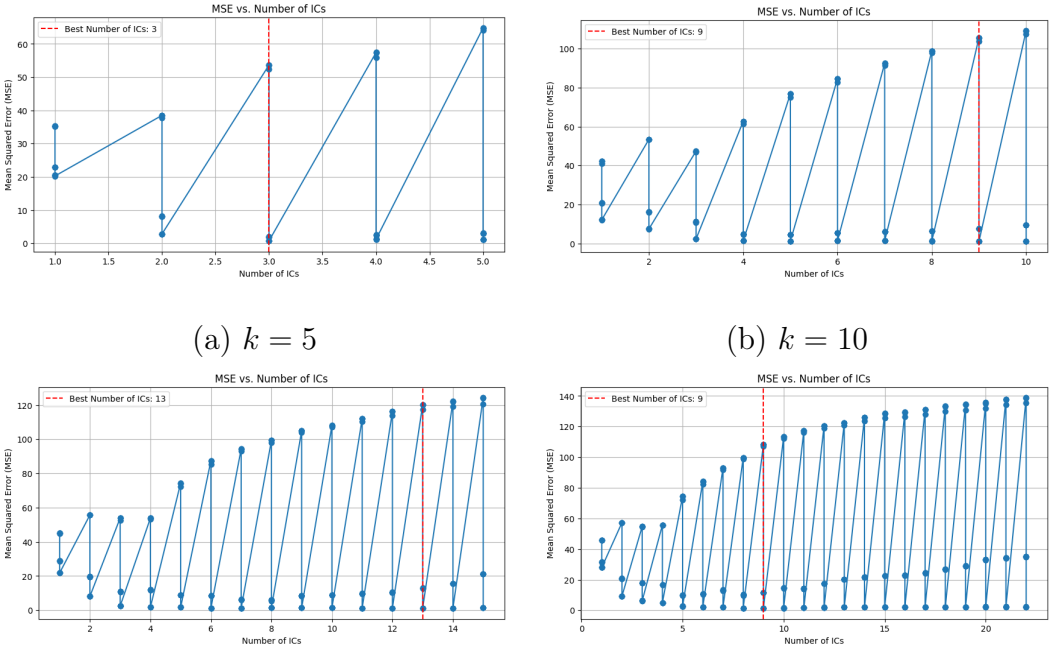
C	Epsilon	Gamma	Mean MSE	n_components
0.1	0.01	scale	42.2408	1
0.1	0.01	auto	42.2597	1
0.1	0.1	scale	42.1282	1
...	...	...	...	...
1	0.1	auto	20.8774	1
1	0.5	scale	20.6642	1
1	0.5	auto	20.6344	1
...	...	...	...	...
10	0.1	auto	7.5536	2
10	0.5	scale	7.4403	2
...	...	...	...	...
10	0.01	scale	2.2232	3
10	0.01	auto	2.2197	3
...	...	...	...	...
10	0.1	auto	1.1058	5
10	0.5	scale	1.1903	5
10	0.5	auto	1.1881	5
...	...	...	...	...
10	0.01	scale	1.3491	6
10	0.01	auto	1.3483	6
10	0.5	auto	1.3342	6
...	...	...	...	...
10	0.01	scale	1.3337	7
10	0.01	auto	1.3353	7
...	...	...	...	...
10	0.5	scale	1.2707	8
10	0.5	auto	1.2691	8
...	...	...	...	...
10	0.01	scale	1.0449	9
10	0.01	auto	1.0433	9
10	0.1	scale	1.0405	9
10	0.1	auto	1.0393	9
10	0.5	scale	1.0881	9
10	0.5	auto	1.0864	9
0.1	0.01	scale	109.2439	10
...	...	...	...	...
10	0.5	scale	1.0892	10
10	0.5	auto	1.0881	10

Table 8.8 : ICA_SVR model hyperparameter tuning for different ICs using the top K = 15 external factors.

C	Epsilon	Gamma	Mean MSE	n_components
0.1	0.01	scale	45.1689	1
0.1	0.01	auto	45.1706	1
...	...	...	...	...
10	0.1	scale	8.0298	2
10	0.1	auto	8.0258	2
10	0.5	scale	8.0575	2
10	0.5	auto	8.0635	2
...	...	...	...	...
1	0.5	scale	8.7329	10
1	0.5	auto	8.7260	10
10	0.01	scale	1.3337	10
10	0.01	auto	1.3331	10
10	0.1	scale	1.3200	10
10	0.1	auto	1.3199	10
10	0.5	scale	1.3337	10
10	0.5	auto	1.3336	10
0.1	0.01	scale	112.0237	11
0.1	0.01	auto	112.0229	11
0.1	0.1	scale	111.9270	11
...	...	...	...	...
10	0.01	auto	1.2671	12
10	0.1	scale	1.2504	12
10	0.1	auto	1.2512	12
10	0.5	scale	1.2923	12
10	0.5	auto	1.2926	12
...	...	...	...	...
10	0.01	scale	1.0668	13
10	0.01	auto	1.0655	13
10	0.1	scale	1.0454	13
10	0.1	auto	1.0439	13
10	0.5	scale	1.0963	13
10	0.5	auto	1.0950	13
...	...	...	...	...
10	0.1	auto	1.3112	15
10	0.5	scale	1.3390	15
10	0.5	auto	1.3343	15

Table 8.9 : ICA_SVR model hyperparameter tuning for different ICs using all K = 22 external factors.

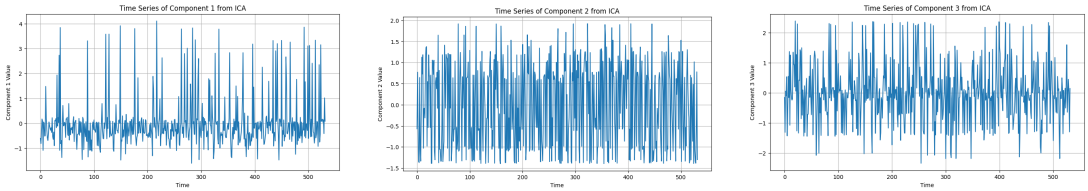
C	Epsilon	Gamma	Mean MSE	n_components
0.1	0.01	scale	45.8405	1
0.1	0.01	auto	45.8413	1
0.1	0.1	scale	45.8741	1
...	...	...	...	...
1	0.1	auto	9.8790	5
1	0.5	scale	9.6428	5
1	0.5	auto	9.6351	5
10	0.01	scale	2.6153	5
10	0.01	auto	2.6124	5
...	...	...	...	...
1	0.1	scale	11.6487	9
1	0.1	auto	11.6516	9
1	0.5	scale	11.4071	9
1	0.5	auto	11.4105	9
10	0.01	scale	1.2632	9
10	0.01	auto	1.2377	9
10	0.1	scale	1.2627	9
10	0.1	auto	1.2375	9
10	0.5	scale	1.3131	9
10	0.5	auto	1.3130	9
...	...	...	...	...
10	0.01	auto	2.1693	13
10	0.1	scale	2.1429	13
10	0.1	auto	2.1404	13
10	0.5	scale	2.1113	13
10	0.5	auto	2.1106	13
...	...	...	...	...
10	0.1	auto	2.1883	17
10	0.5	scale	2.2180	17
10	0.5	auto	2.2046	17
...	...	...	...	...
10	0.01	scale	2.1653	22
10	0.01	auto	2.1472	22
10	0.1	scale	2.1513	22
10	0.1	auto	2.1315	22
10	0.5	scale	2.3311	22
10	0.5	auto	2.3065	22



(a) $k = 5$ (b) $k = 10$

(c) $k = 15$ (d) $k = 22$ (All factors)

Figure 8.13 : ICA_SVR model hyperparameters tuning for different components using a different number of external factor configurations.



IC1 IC2 IC3

Figure 8.14 : Best extracted IC components by the proposed ICA_SVR model using the top $K = 5$ factors.

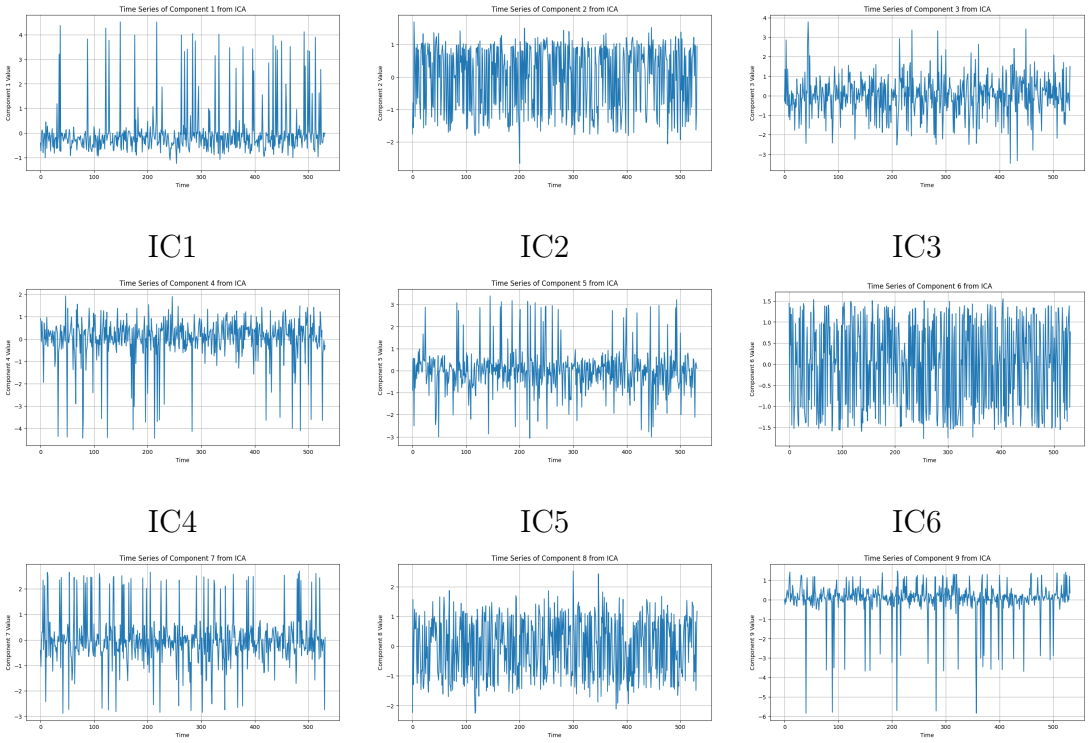


Figure 8.15 : Best extracted IC components by the proposed ICA_SVR model using the top $K = 10$ factors.

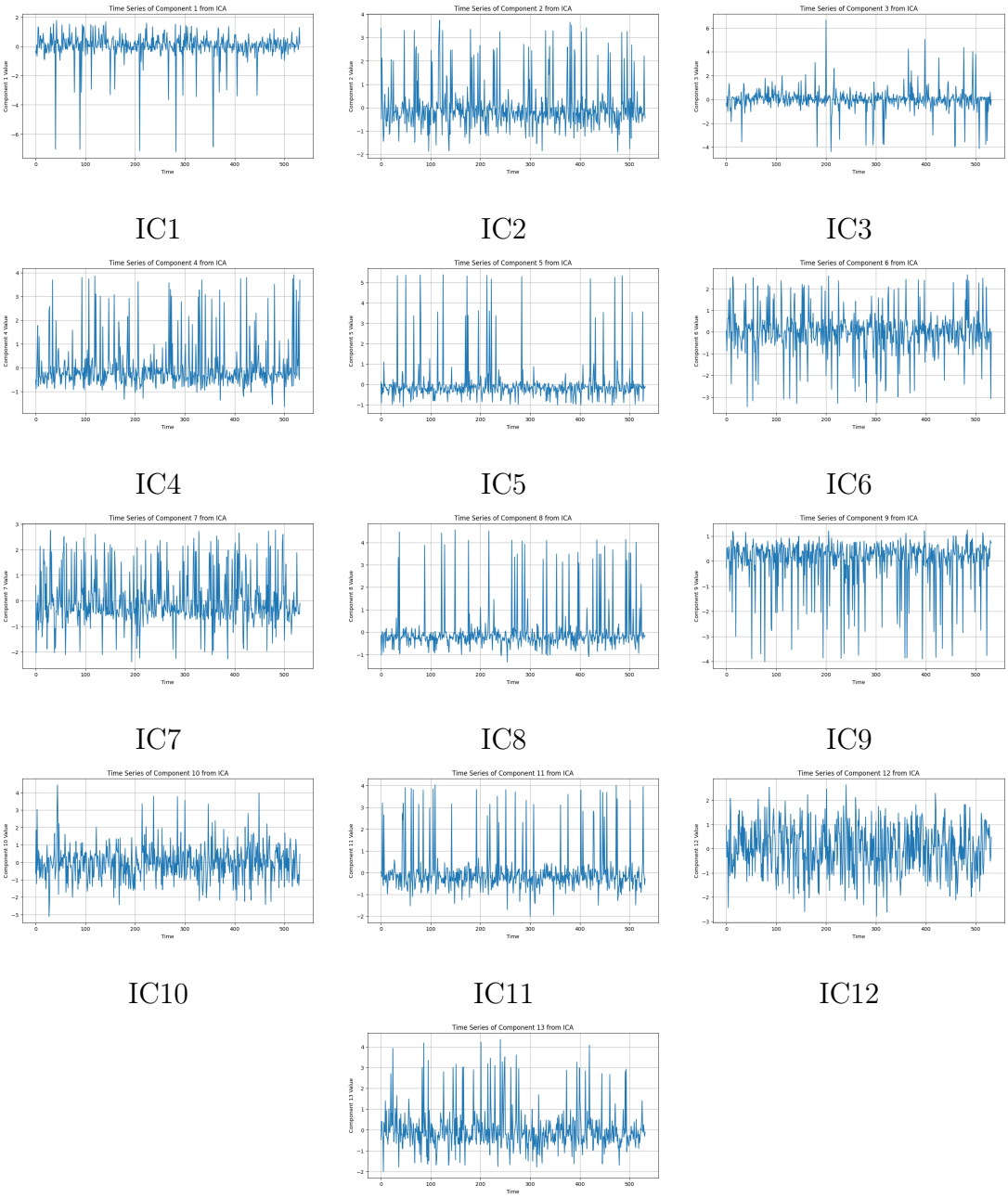


Figure 8.16 : Best extracted IC components by the proposed ICA_SVR model using the top $K = 15$ factors.

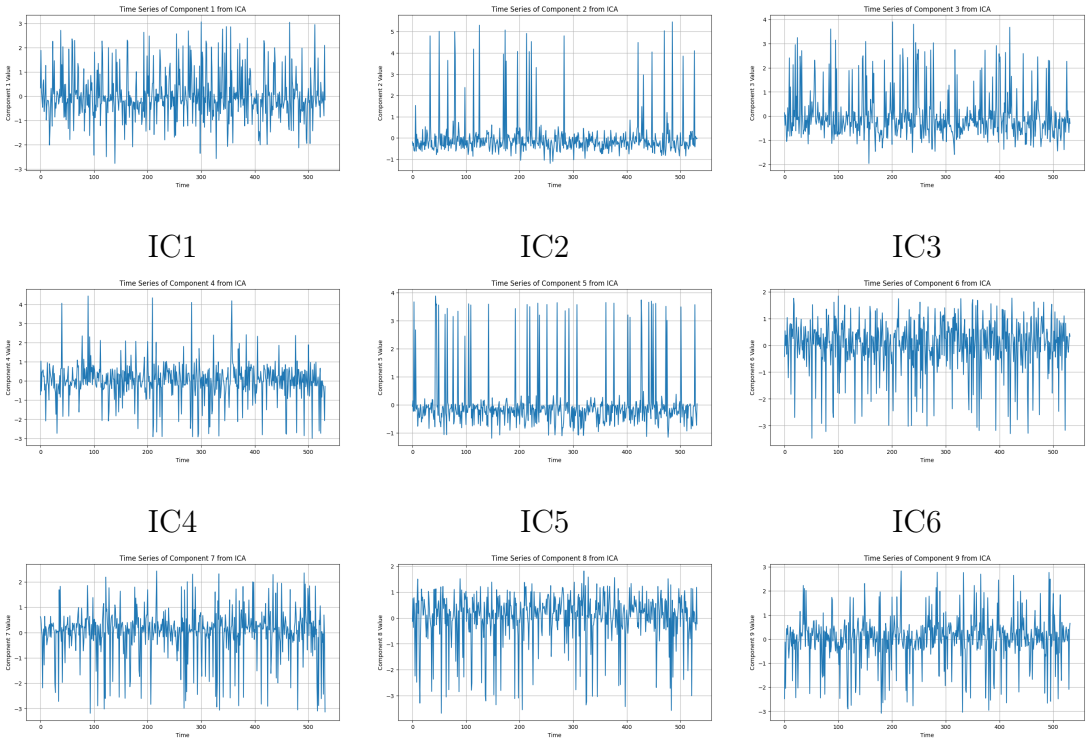
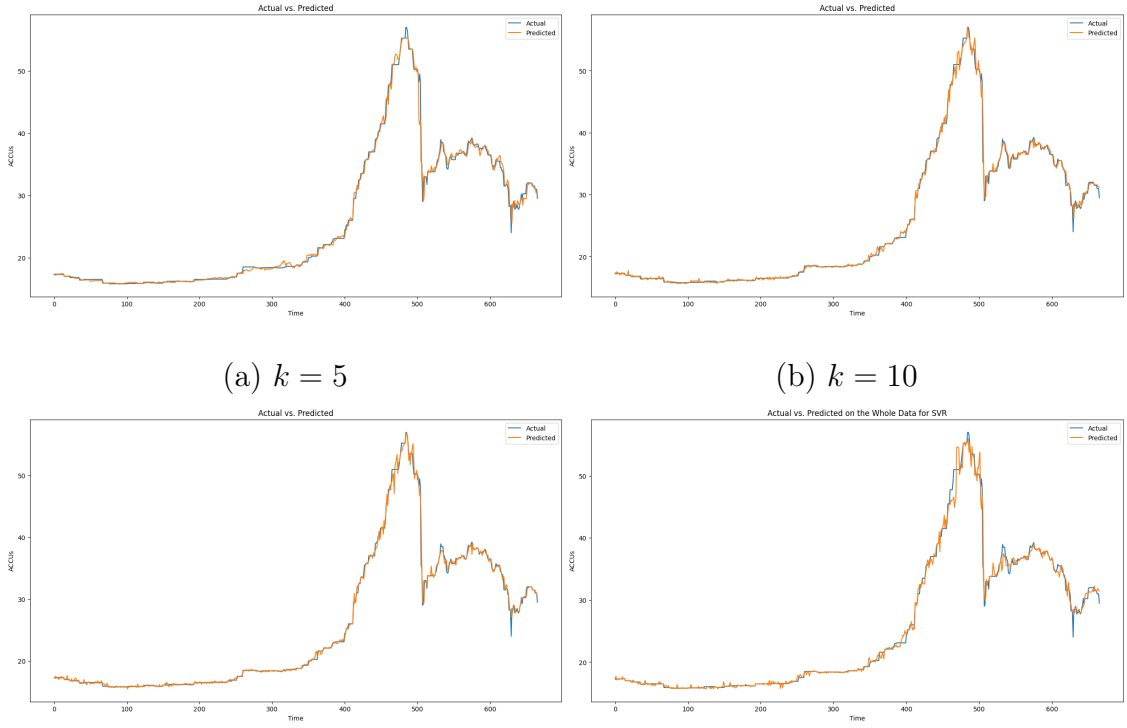


Figure 8.17 : Best extracted IC components by the proposed ICA_SVR model using all $K = 22$ factors.



(a) $k = 5$ (b) $k = 10$

(c) $k = 15$ (d) $k = 22$ (All factors)

Figure 8.18 : Actual vs. predicted ACCUs price using the proposed ICA_SVR model with different external factor configurations.

For the NLICA_SVR model, an autoencoder is used to compress the data into a lower-dimensional representation, capturing non-linear structures that linear methods PCA or ICA might miss. Similar to ICA, the number of independent components is not predefined. Instead, we define a range from 1 to the total number of features (22 factors) in the dataset and loop through this range. For each value, the autoencoder is configured to reduce the data to that number of components. The transformed data were then used to evaluate the model's performance based on MSE, and the number of components that result in the lowest MSE is selected as the optimal configuration.

For each possible number of components, the grid search was performed using cross-validation to optimise the SVR hyperparameters, and the model was evaluated

based on its MSE performance. Tables 8.10 to 8.13 present the results of the hyper-parameter tuning for the NLICA_SVR model, using a different number of external factors K. For the sake of space, the tables have been summarised to show only the key configurations and the best-performing models with the optimal number of ICs. By systematically looping through different values for `n_components`, the lowest MSE was achieved with 5, 5, 15, 17 ICs when the number of external factors is ($K = 5$, $K = 10$, $K = 15$, and $K = 22$) respectively, as shown in Figure 8.19.

Figures 8.20 to 8.23 show the best IC components extracted by the proposed NLICA_SVR model for a different number of external factors K. In addition, Figure 8.24 compares the actual and predicted results for ACCUs prices using the proposed NLICA_SVR model using a different number of external factors ($K = 5$, $K = 10$, $K = 15$, and $K = 22$).

Table 8.10 : NLICA_SVR model hyperparameter tuning for different ICs using the top K = 5 external factors.

C	Epsilon	Gamma	Mean MSE	n_components
0.1	0.01	scale	161.4290	1
0.1	0.01	auto	161.4290	1
0.1	0.1	scale	161.5702	1
0.1	0.1	auto	161.5702	1
...	...	...	...	...
1	0.1	auto	22.3064	2
1	0.5	scale	22.0157	2
1	0.5	auto	22.3199	2
10	0.01	scale	20.9489	2
10	0.01	auto	21.0390	2
10	0.1	scale	21.0092	2
10	0.1	auto	21.1134	2
10	0.5	scale	21.0938	2
10	0.5	auto	21.2113	2
0.1	0.01	scale	38.8999	3
0.1	0.01	auto	40.2692	3
0.1	0.1	scale	38.6973	3
0.1	0.1	auto	40.2062	3
...	...	...	...	...
0.1	0.1	scale	40.9790	4
0.1	0.1	auto	43.0692	4
0.1	0.5	scale	40.0686	4
0.1	0.5	auto	41.8215	4
1	0.01	scale	17.8543	4
1	0.01	auto	22.2450	4
1	0.1	scale	17.8521	4
...	...	...	...	...
1	0.1	scale	9.6702	5
1	0.1	auto	18.4134	5
1	0.5	scale	9.9734	5
1	0.5	auto	18.3406	5
10	0.01	scale	2.3257	5
10	0.01	auto	4.6360	5
10	0.1	scale	2.2097	5
10	0.1	auto	4.6173	5
10	0.5	scale	2.3118	5
10	0.5	auto	4.5230	5

Table 8.11 : NLICA_SVR model hyperparameter tuning for different ICs using the top K = 10 external factors.

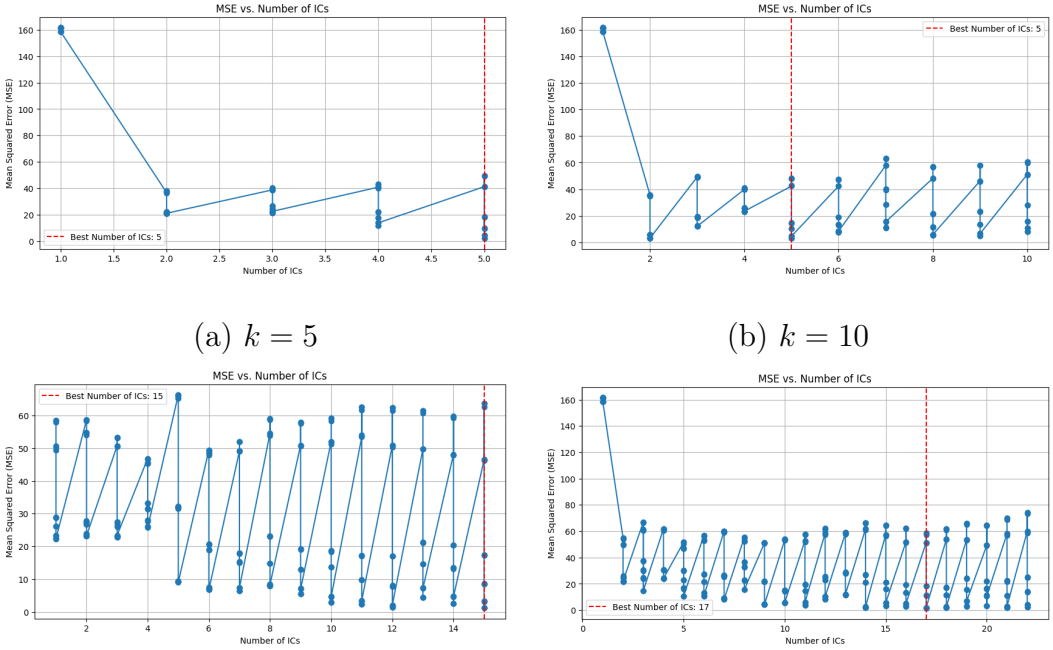
C	Epsilon	Gamma	Mean MSE	n_components
0.1	0.01	scale	161.4290	1
0.1	0.01	auto	161.4290	1
0.1	0.1	scale	161.5702	1
...	...	...	...	...
1	0.5	auto	18.8704	3
10	0.01	scale	12.4951	3
10	0.01	auto	12.2844	3
10	0.1	scale	12.4630	3
10	0.1	auto	12.2455	3
10	0.5	scale	12.2357	3
10	0.5	auto	12.0899	3
...	...	...	...	...
1	0.1	auto	14.5549	5
1	0.5	scale	10.4264	5
1	0.5	auto	14.9878	5
10	0.01	scale	3.1799	5
10	0.01	auto	4.8450	5
10	0.1	scale	3.0164	5
10	0.1	auto	4.7941	5
10	0.5	scale	3.1524	5
10	0.5	auto	4.6341	5
0.1	0.01	scale	42.3818	6
0.1	0.01	auto	47.3569	6
0.1	0.1	scale	42.4347	6
...	...	...	...	...
1	0.5	auto	21.2008	8
10	0.01	scale	5.5359	8
10	0.01	auto	5.9124	8
10	0.1	scale	5.5395	8
10	0.1	auto	5.9156	8
10	0.5	scale	5.5253	8
10	0.5	auto	5.8980	8
0.1	0.01	scale	46.2943	9
0.1	0.01	auto	58.0938	9
...	...	...	...	...
10	0.1	scale	7.9693	10
10	0.1	auto	10.6590	10
10	0.5	scale	7.9505	10
10	0.5	auto	10.6048	10

Table 8.12 : NLICA_SVR model hyperparameter tuning for different ICs using the top K = 15 external factors.

C	Epsilon	Gamma	Mean MSE	n_components
0.1	0.01	scale	58.5668	1
0.1	0.01	auto	50.7210	1
0.1	0.1	scale	58.4872	1
0.1	0.1	auto	50.4362	1
0.1	0.5	scale	57.9138	1
0.1	0.5	auto	49.4815	1
1	0.01	scale	28.8334	1
1	0.01	auto	26.2336	1
...	...	...	...	...
10	0.1	auto	9.1842	5
10	0.5	scale	9.2967	5
10	0.5	auto	9.0677	5
0.1	0.01	scale	49.4835	6
...	...	...	...	...
10	0.1	scale	7.9897	8
10	0.1	auto	8.3859	8
10	0.5	scale	7.9770	8
10	0.5	auto	8.3755	8
0.1	0.01	scale	50.9790	9
0.1	0.01	auto	58.0661	9
0.1	0.1	scale	50.7644	9
...	...	...	...	...
10	0.1	auto	1.9390	12
10	0.5	scale	1.3737	12
10	0.5	auto	2.0445	12
0.1	0.01	scale	49.7667	13
0.1	0.01	auto	61.5250	13
...	...	...	...	...
1	0.01	auto	17.4214	15
1	0.1	scale	8.5107	15
1	0.1	auto	17.3832	15
1	0.5	scale	8.5763	15
1	0.5	auto	17.1384	15
10	0.01	scale	1.1924	15
10	0.01	auto	3.2521	15
10	0.1	scale	1.1849	15
10	0.1	auto	3.2361	15
10	0.5	scale	1.2060	15
10	0.5	auto	3.0352	15

Table 8.13 : NLICA_SVR model hyperparameter tuning for different ICs using all K = 22 external factors.

C	Epsilon	Gamma	Mean MSE	n_components
0.1	0.01	scale	161.4290	1
0.1	0.01	auto	161.4290	1
0.1	0.1	scale	161.5702	1
0.1	0.1	auto	161.5702	1
...	...	...	...	...
0.1	0.1	auto	52.4577	8
0.1	0.5	scale	54.6879	8
0.1	0.5	auto	52.1428	8
1	0.1	scale	21.6411	9
...	...	...	...	...
10	0.5	scale	3.1352	16
10	0.5	auto	4.8672	16
0.1	0.01	scale	51.4122	17
0.1	0.01	auto	58.5094	17
0.1	0.1	scale	51.4939	17
0.1	0.1	auto	58.4311	17
0.1	0.5	scale	50.8787	17
0.1	0.5	auto	57.4001	17
1	0.01	scale	11.0850	17
1	0.01	auto	18.5179	17
10	0.01	scale	1.3432	17
10	0.01	auto	1.9535	17
10	0.1	scale	1.3176	17
10	0.1	auto	1.9534	17
10	0.5	scale	1.3768	17
10	0.5	auto	1.9223	17
0.1	0.01	scale	53.9299	18
0.1	0.01	auto	61.6949	18
...	...	...	...	...
1	0.1	auto	22.3946	21
1	0.5	scale	11.2432	21
1	0.5	auto	21.7053	21
...	...	...	...	...
10	0.1	auto	4.3057	22
10	0.5	scale	2.5302	22
10	0.5	auto	4.1600	22



(a) $k = 5$ (b) $k = 10$

(c) $k = 15$ (d) $k = 22$ (All factors)

Figure 8.19 : NLICA_SVR model hyperparameter tuning for different components using a different number of external factor configurations.

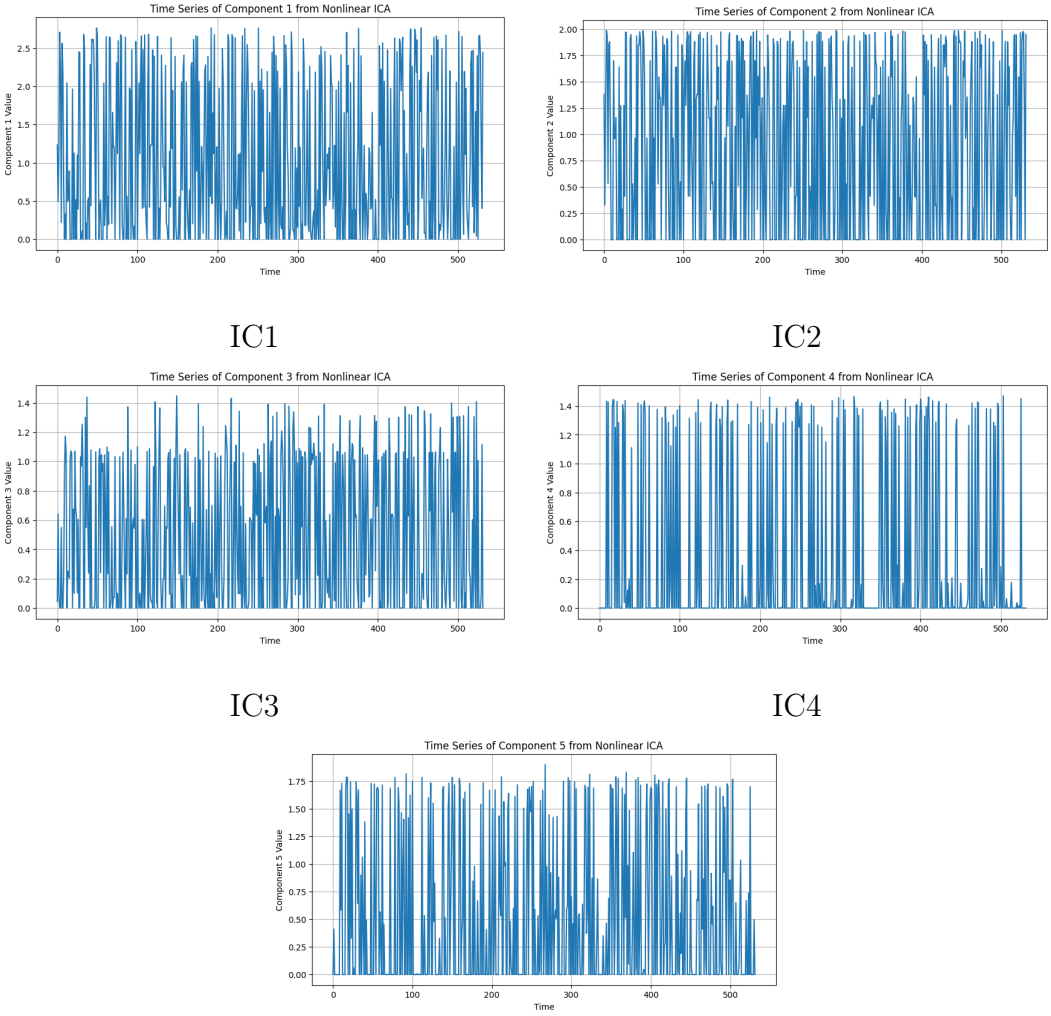
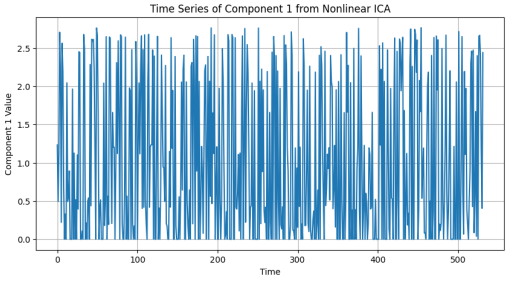
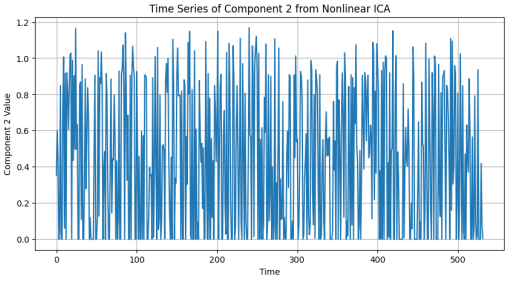


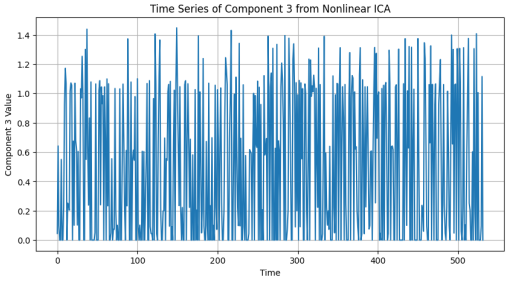
Figure 8.20 : Best extracted PC components by the proposed NLICA_SVR model using the top $K = 5$ factors.



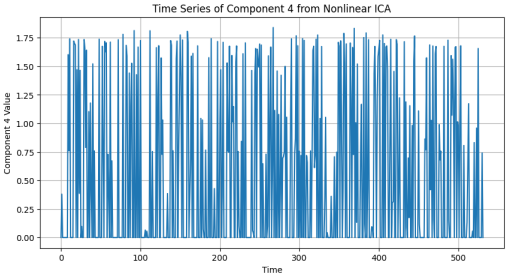
IC1



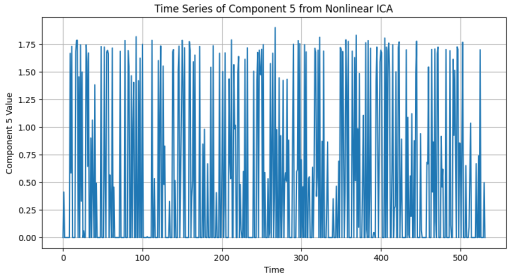
IC2



IC3



IC4



IC5

Figure 8.21 : Best extracted IC components by the proposed NLICA_SVR model using the top $K = 10$ factors.



Figure 8.22 : Best extracted IC components by the proposed NLICA_SVR model using the top $K = 15$ factors.

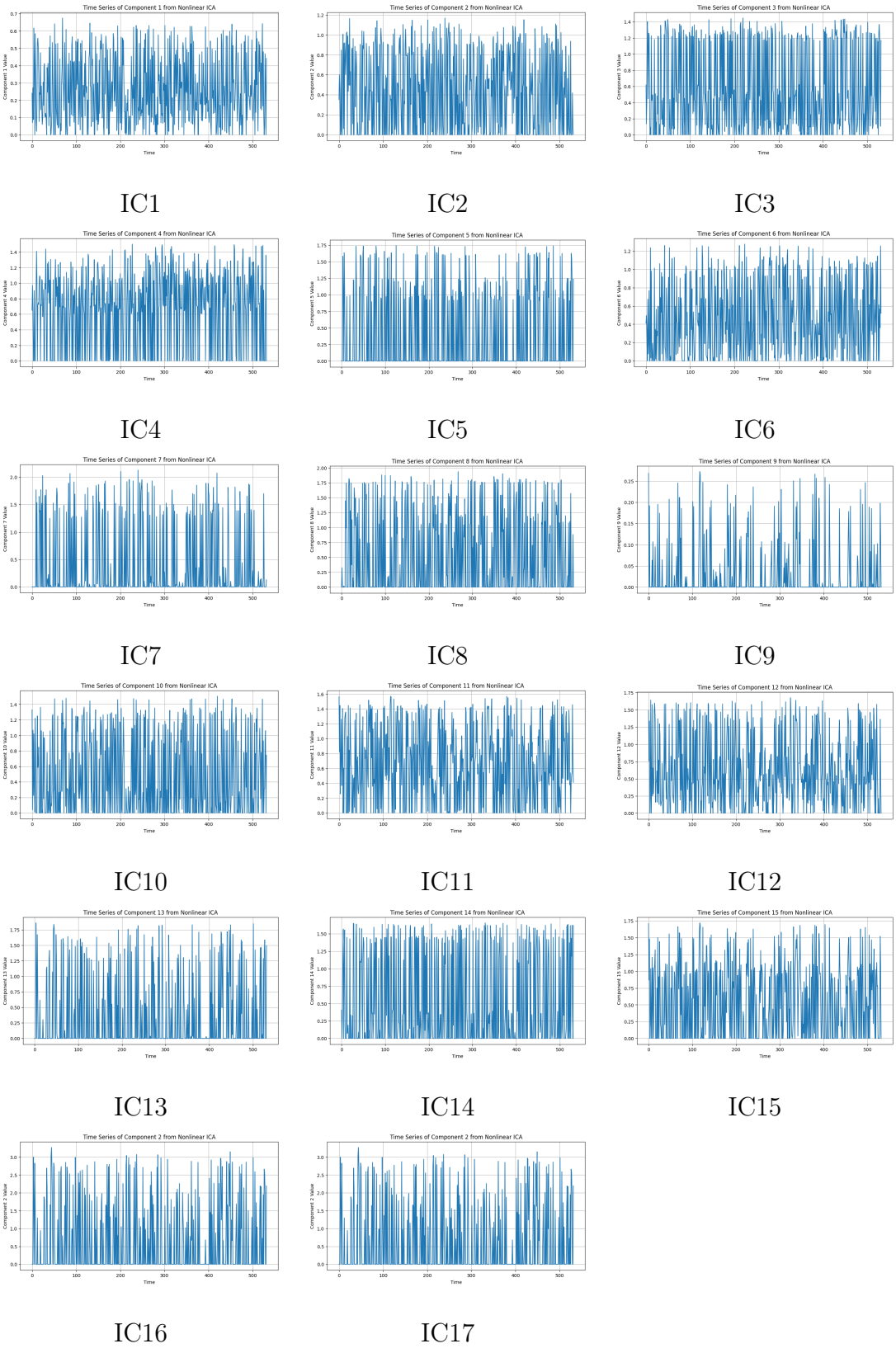
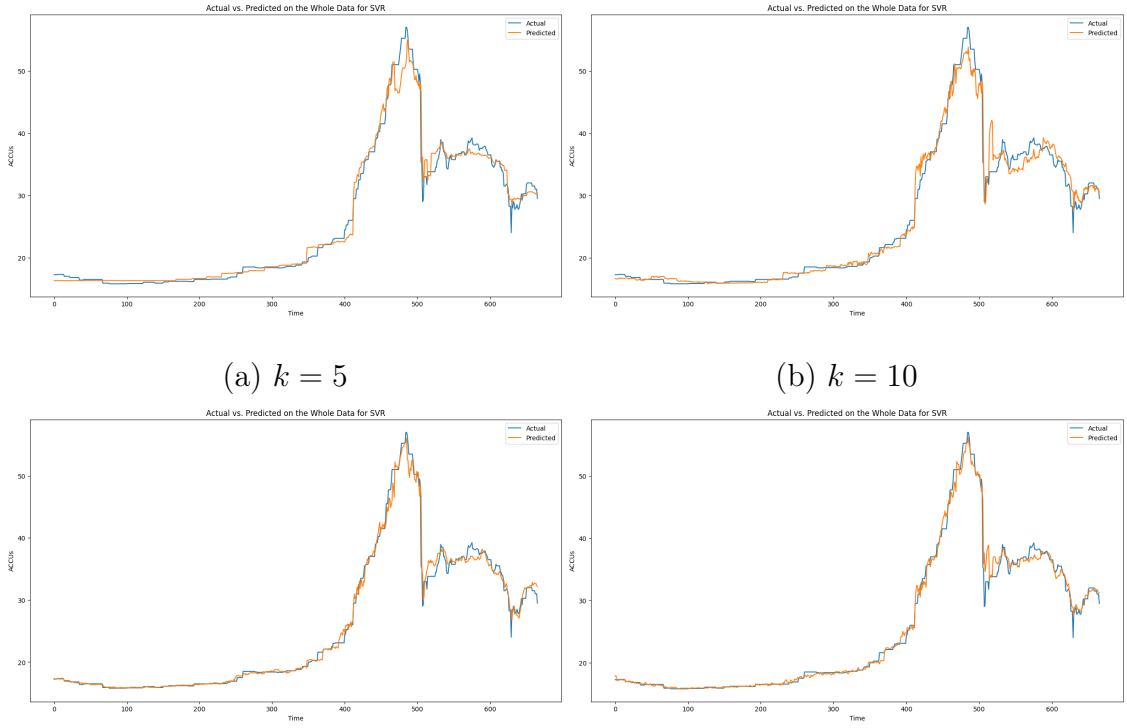


Figure 8.23 : Best extracted IC components by the proposed NLICA_SVR model using all $K = 22$ factors.



(a) $k = 5$ (b) $k = 10$

(c) $k = 15$ (d) $k = 22$ (All factors)

Figure 8.24 : Actual vs. predicted ACCUs price using the proposed NLICA_SVR model with different external factor configurations.

The optimal values of the hyperparameters obtained in our experiments for the SVR_based models using the feature extraction and reduction techniques were consistent. The details of the final optimal parameters for each model are shown in Table 8.14. The PCA_SVR, ICA_SVR and NLICA_SVR models all performed best with `C` set to 10, `epsilon` at 0.1, `gamma` set to `'scale'`, and using the `RBF` kernel. The ICA_SVR model differed slightly, achieving its best performance with `gamma` set to `'auto'`.

Table 8.14 : Hyperparameter tuning results for PCA_SVR, ICA_SVR, and NLICA_SVR models, showing mean MSE and optimal ICs based on a different number of external factors.

Model	**Kernel**	**C**	**Epsilon**	**Gamma**	**K=5**		**K=10**		**K=15**		**K=22**	
					MSE	**ICs**	**MSE**	**ICs**	**MSE**	**ICs**	**MSE**	**ICs**
PCA_SVR	RBF	10	0.1	'scale'	0.7561	3	0.8777	4	2.0387	6	2.6060	7
ICA_SVR	RBF	10	0.1	'auto'	0.7395	3	1.0393	9	1.0439	13	1.2375	9
NLICA_SVR	RBF	10	0.1	'scale'	2.2097	5	3.0164	5	1.1849	15	1.3176	17

8.6.2 Evaluation of the proposed CCMFP using the ICA_MLP, NLICA_MLP and PCA_MLP models

We conducted a series of experiments to evaluate the ICA_MLP, NLICA_MLP and PCA_MLP models, varying both the number of selected external factors K and the hyperparameters of the MLP_based models through a cross-validation grid search process.

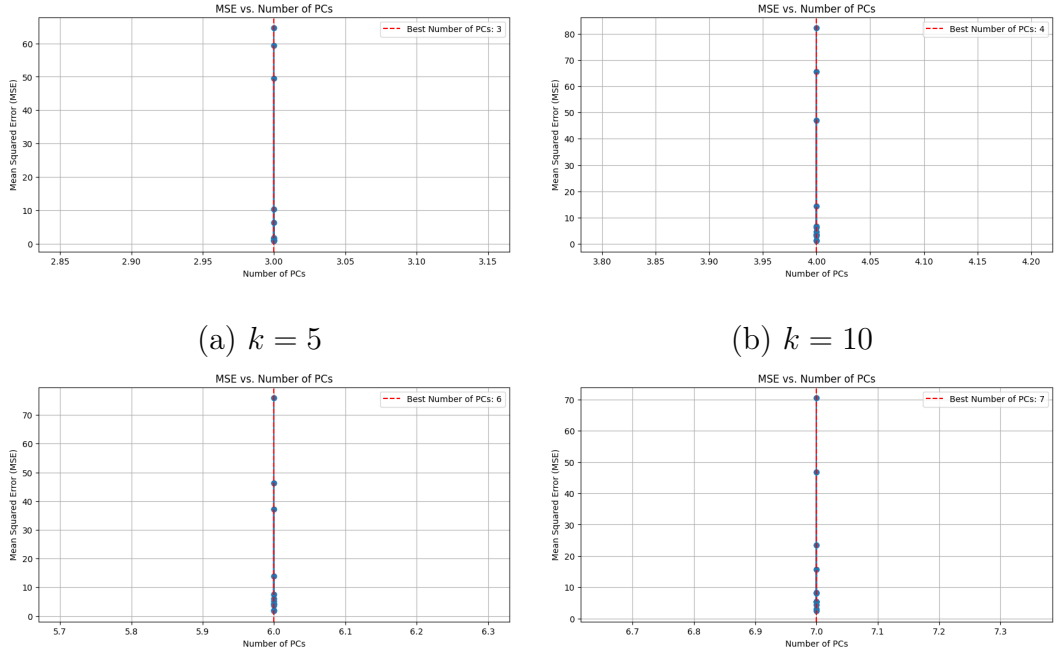
In the PCA_MLP model, the PCA technique and external factors are used to generate PCs, which represent the features of the input data and are used as new input variables to the MLP model. The results of the PCA_MLP model, including the optimal hyperparameters and the corresponding optimal number of PCs, are summarised in Table 8.15. The optimal hyperparameters for the PCA_MLP model were found to be `activation function` = `tanh`, `alpha` = 0.0001, `hidden_layer_sizes` = (50, 50), and `optimizer` = `sgd`. The minimum mean MSE values were 0.8017, 1.1357, 1.7564, and 2.4308 when the number of selected external factors K was 5, 10, 15, and 22, respectively.

Figure 8.25 provides additional insight into the tuning process conducted during the cross-validation grid search by illustrating the relationship between the MSE and the number of PCs for each configuration of K. The red dashed line in each

figure indicates the optimal number of PCs that minimises the MSE. The optimal number of PCs for each configuration of K was 3, 4, 6, and 7.

Table 8.15 : Parameters grid search result for PCA_MLP model using different external factor configurations.

Activation function	Alpha	Hidden layer size	Optimizer	Mean MSE				Optimal number of PCs			
				K=5	K=10	K=15	K=22	K=5	K=10	K=15	K=22
tanh	0.0001	(50,)	adam	59.3885	82.1603	75.9233	70.5078	3	4	6	7
tanh	0.0001	(50,)	sgd	1.4935	4.6444	5.2011	5.3873	3	4	6	7
tanh	0.0001	(100,)	adam	10.2413	14.3433	13.9787	15.7077	3	4	6	7
tanh	0.0001	(100,)	sgd	1.8314	6.2506	7.5579	8.2821	3	4	6	7
tanh	0.0001	(50, 50)	adam	49.5353	47.0294	46.4070	46.8083	3	4	6	7
tanh	0.0001	(50, 50)	sgd	0.8017	1.1357	1.7564	2.4308	3	4	6	7
tanh	0.001	(50,)	adam	59.3887	82.1604	75.9234	70.5079	3	4	6	7
tanh	0.001	(50,)	sgd	1.4936	4.6447	5.2015	5.3877	3	4	6	7
tanh	0.001	(100,)	adam	10.2415	14.3435	13.9788	15.7078	3	4	6	7
tanh	0.001	(100,)	sgd	1.8314	6.2508	7.5583	8.2827	3	4	6	7
tanh	0.001	(50, 50)	adam	49.5354	47.0295	46.4070	46.8084	3	4	6	7
tanh	0.001	(50, 50)	sgd	0.8189	1.1358	1.7565	2.4309	3	4	6	7
relu	0.0001	(50,)	adam	64.6320	65.4347	37.1879	23.4774	3	4	6	7
relu	0.0001	(50,)	sgd	1.0093	3.6390	4.2076	5.4737	3	4	6	7
relu	0.0001	(100,)	adam	6.3318	6.8550	6.0835	8.1662	3	4	6	7
relu	0.0001	(100,)	sgd	1.0024	3.5572	4.2834	5.4850	3	4	6	7
relu	0.0001	(50, 50)	adam	1.0327	3.0842	3.6154	4.4394	3	4	6	7
relu	0.0001	(50, 50)	sgd	0.8189	1.4613	1.9403	3.0554	3	4	6	7
relu	0.001	(50,)	adam	64.6317	65.4347	37.1907	23.4748	3	4	6	7
relu	0.001	(50,)	sgd	1.0094	3.6391	4.2077	5.4737	3	4	6	7
relu	0.001	(100,)	adam	6.3322	6.8562	6.0997	8.1662	3	4	6	7
relu	0.001	(100,)	sgd	1.0024	3.5573	4.2836	5.4852	3	4	6	7
relu	0.001	(50, 50)	adam	1.0473	3.0812	3.6087	4.4147	3	4	6	7
relu	0.001	(50, 50)	sgd	0.8018	1.4613	1.9398	3.0568	3	4	6	7

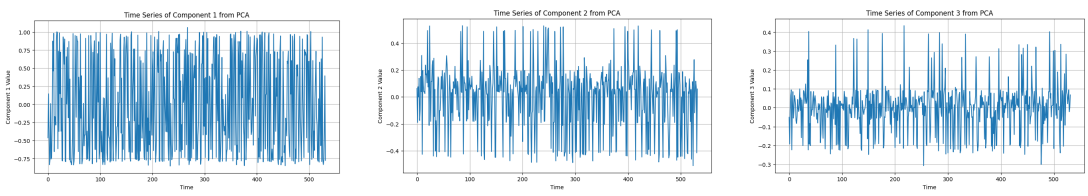


(a) $k = 5$ (b) $k = 10$

(c) $k = 15$ (d) $k = 22$ (All factors)

Figure 8.25 : PCA_MLP model hyperparameter tuning for different components using a different number of external factor configurations.

Figures 8.26 to 8.29 show the best-extracted PCs generated by the proposed PCA_MLP model from a different number of external factors ($K = 5$, $K = 10$, $K = 15$, and $K = 22$). Each PC represents a transformed version of the original input data, capturing the most significant variance within the dataset. These PCs are then used as new input variables for the MLP model.



PC1 PC2 PC3

Figure 8.26 : Best extracted PC components by the proposed PCA_MLP model using the top $K = 5$ factors.

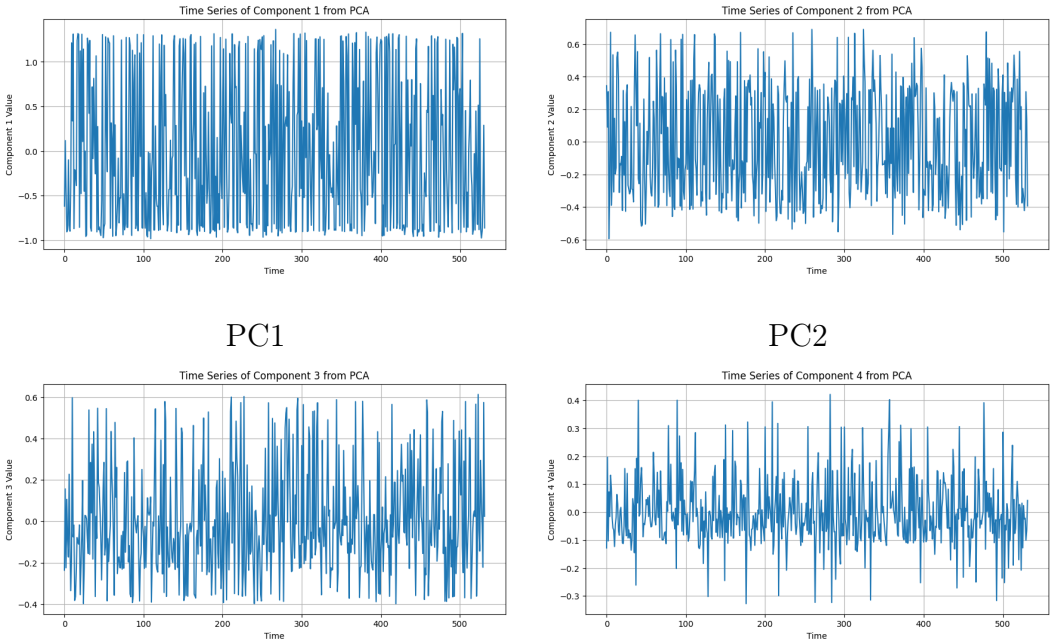


Figure 8.27 : Best extracted PC components by the proposed PCA_MLP model using the top $K = 10$ factors.

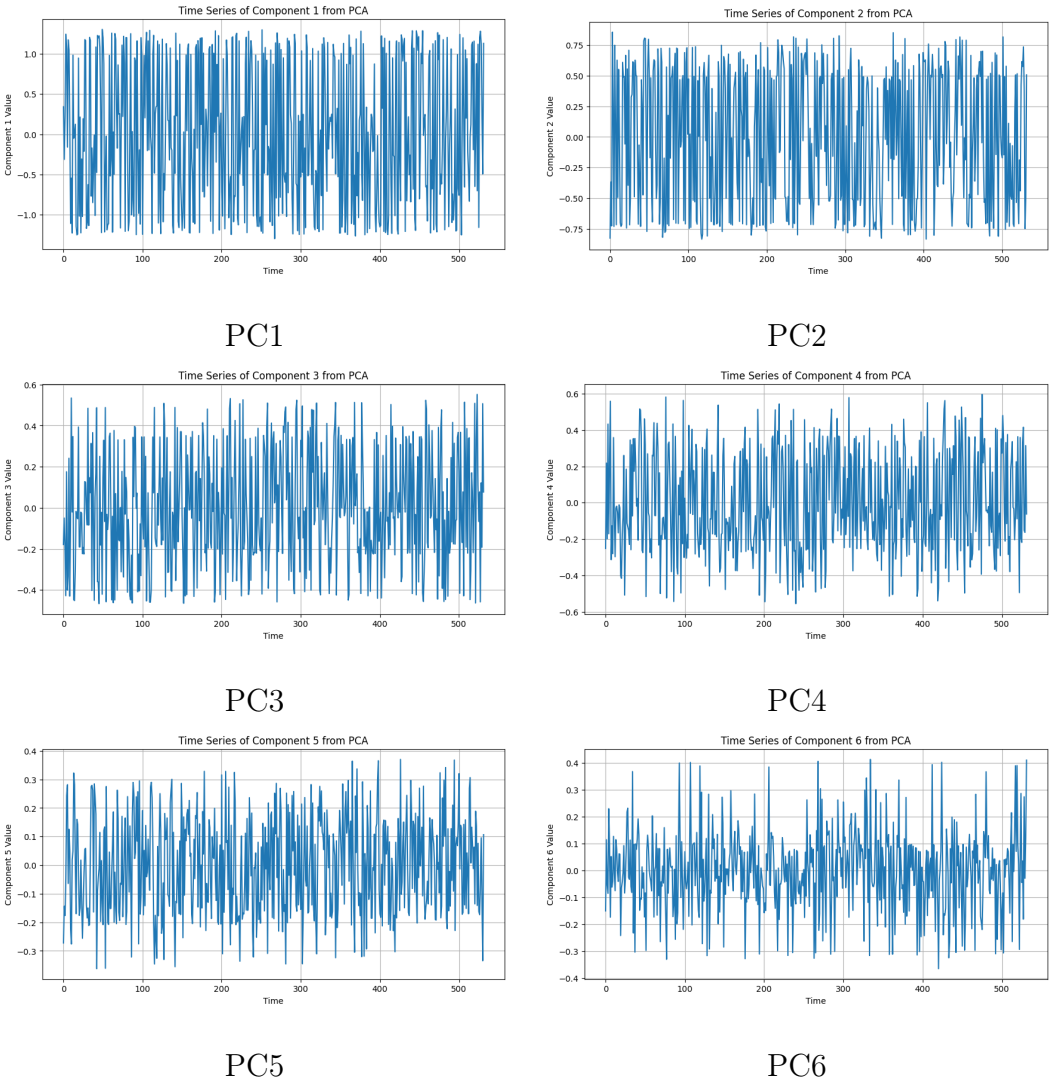
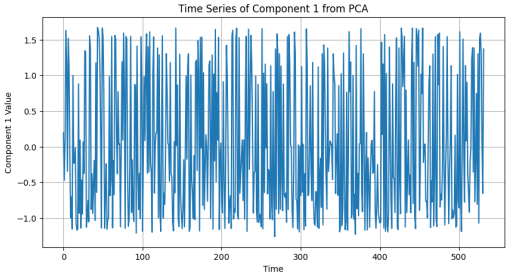
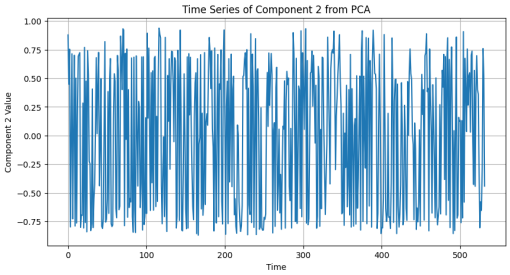


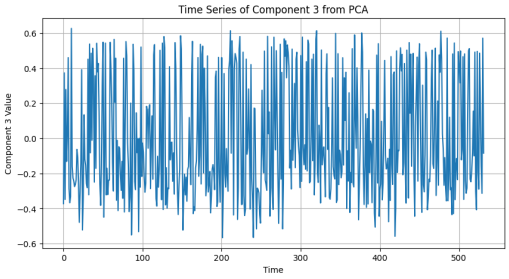
Figure 8.28 : Best extracted PC components by the proposed PCA_MLP model using the top $K = 15$ factors.



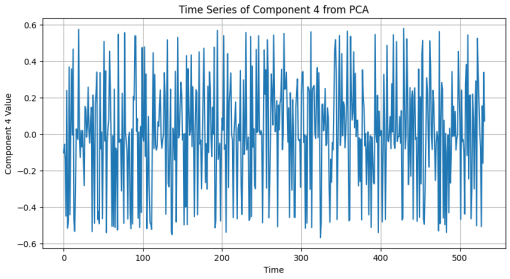
PC1



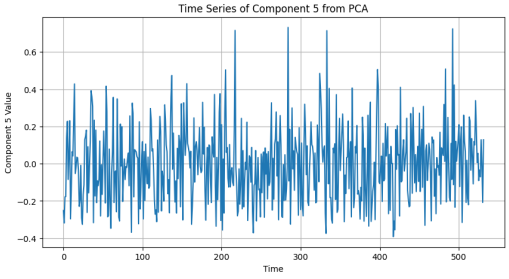
PC2



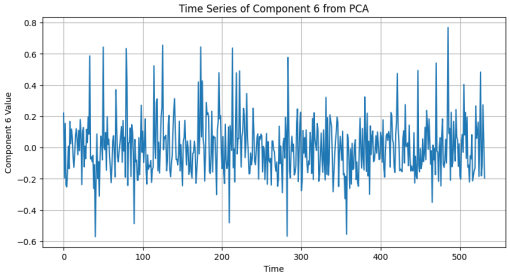
PC3



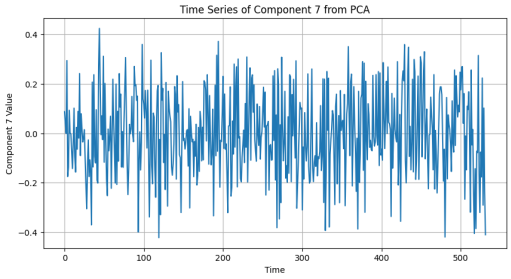
PC4



PC5



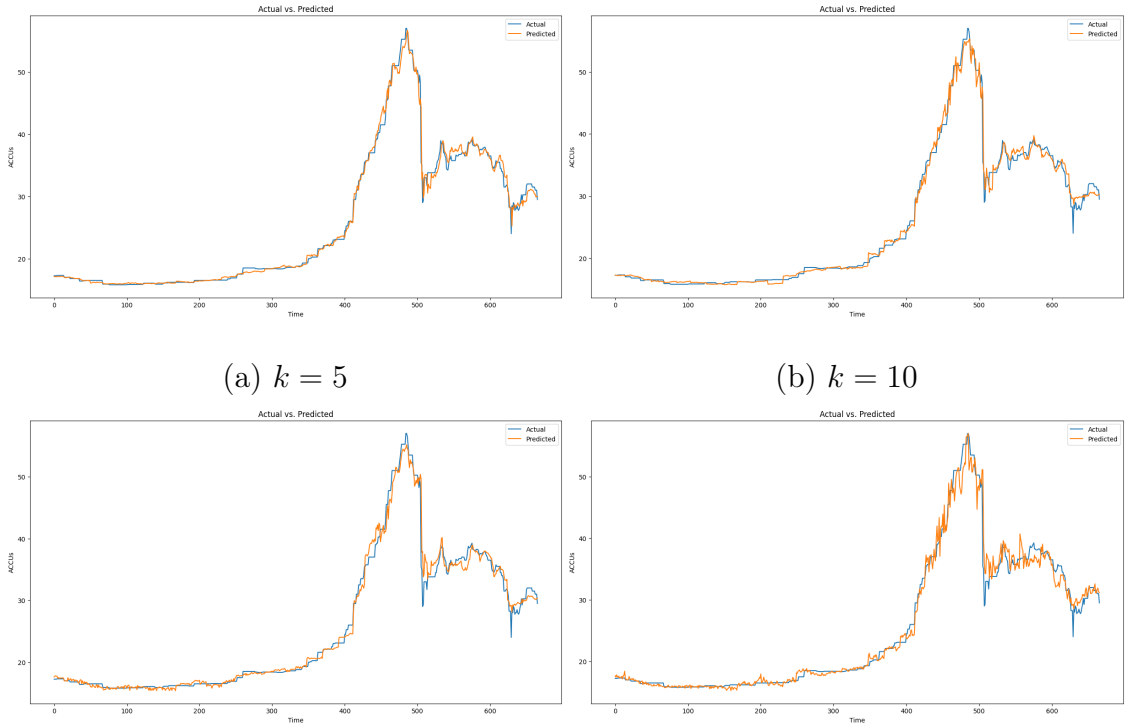
PC6



PC7

Figure 8.29 : Best extracted PC components by the proposed PCA_MLP model using all $K = 22$ factors.

Figure 8.30 compares the actual and predicted results for ACCUs prices using the proposed PCA_MLP model using a different number of external factors ($K = 5$, $K = 10$, $K = 15$, and $K = 22$).



(a) $k = 5$ (b) $k = 10$

(c) $k = 15$ (d) $k = 22$ (All factors)

Figure 8.30 : Actual vs. predicted ACCUs price using the proposed PCA_MLP model with different external factor configurations.

In the case of the ICA_MLP model, ICA was applied using the FastICA algorithm to extract ICs from the dataset. Unlike PCA, which selects components based on explained variance, ICA focusses on maximising statistical independence between the components. The number of ICs was not predefined; instead, a range was defined, from 1 to the total number of features in the dataset (22 external factors). For each possible number of components, the grid search was performed using cross-validation to optimise the MLP hyperparameters (`activation function`, `alpha`,

`hidden_layer_size` and `optimizer`) and the model was evaluated based on its MSE performance.

Tables 8.16 to 8.19 present the results of the hyperparameter tuning for the ICA_MLP model, using a different number of external factors K. For the sake of space, the tables have been summarised to show only the key configurations and the best-performing models with the optimal number of ICs. By systematically looping through different values for `n_components`, the lowest MSE is achieved with 5, 10, 14, 21 ICs when the number of external factors is ($K = 5$, $K = 10$, $K = 15$, and $K = 22$) respectively, as shown in Figure 8.31.

Figures 8.32 to 8.35 show the best IC components extracted by the proposed ICA_MLP model for a different number of external factors K. In addition, Figure 8.36 compares the actual and predicted results for ACCU prices using the proposed ICA_MLP model using a different number of external factors ($K = 5$, $K = 10$, $K = 15$, and $K = 22$).

Table 8.16 : ICA_MLP model hyperparameter tuning for different ICs using the top K = 5 external factors.

Activation function	Alpha	Hidden_layer_sizes	Optimizer	Mean MSE	n_components
tanh	0.0001	(50,)	adam	86.9545	1
tanh	0.0001	(50,)	sgd	22.7523	1
tanh	0.0001	(100,)	adam	33.0339	1
tanh	0.0001	(100,)	sgd	25.9032	1
...	...	...	...	...	
relu	0.0001	(50, 50)	adam	7.3859	2
relu	0.0001	(50, 50)	sgd	2.8754	2
relu	0.001	(50,)	adam	36.5403	2
relu	0.001	(50,)	sgd	4.5515	2
relu	0.001	(100,)	adam	10.1727	2
...	...	...	...	...	
relu	0.001	(100,)	adam	5.0660	3
relu	0.001	(100,)	sgd	0.7160	3
relu	0.001	(50, 50)	adam	1.1599	3
relu	0.001	(50, 50)	sgd	0.7551	3
tanh	0.0001	(50,)	adam	76.3427	4
tanh	0.0001	(50,)	sgd	1.1797	4
tanh	0.0001	(100,)	adam	5.2211	4
tanh	0.0001	(100,)	sgd	1.2982	4
...	...	...	...	...	
tanh	0.0001	(100,)	adam	3.1052	5
tanh	0.0001	(100,)	sgd	0.7505	5
tanh	0.0001	(50, 50)	adam	41.6864	5
tanh	0.0001	(50, 50)	sgd	0.6770	5
tanh	0.001	(50,)	adam	54.4130	5
tanh	0.001	(50,)	sgd	0.7820	5
tanh	0.001	(100,)	adam	3.1053	5
...	...	...	...	...	
relu	0.001	(100,)	adam	7.2184	5
relu	0.001	(100,)	sgd	0.6970	5
relu	0.001	(50, 50)	adam	3.0327	5
relu	0.001	(50, 50)	sgd	0.7255	5

Table 8.17 : ICA_MLP model hyperparameter tuning for different ICs using the top K = 10 external factors.

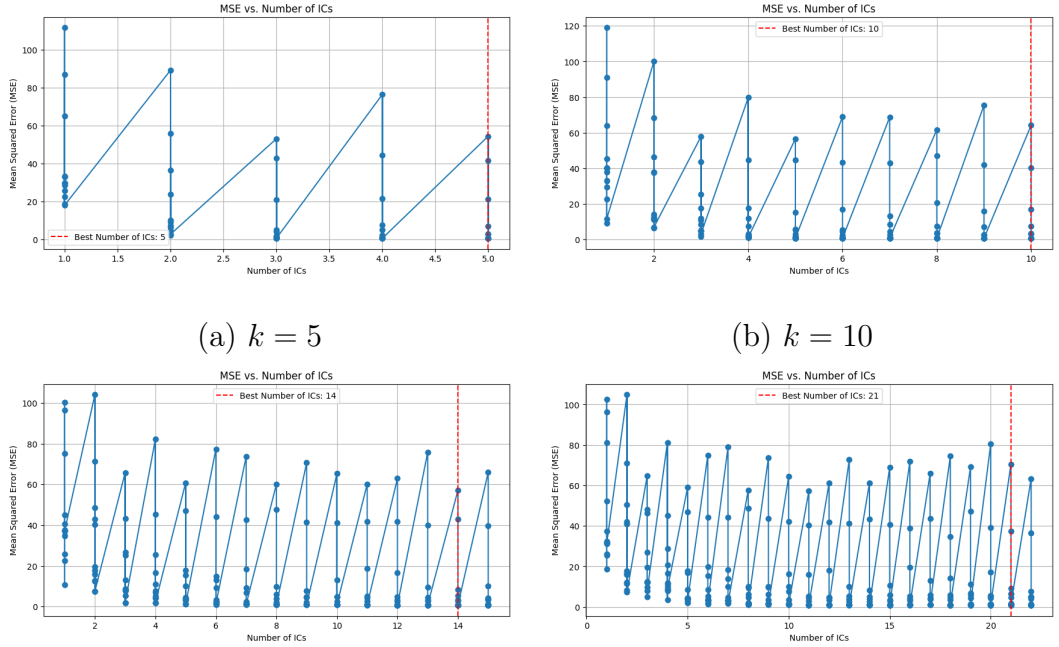
Activation function	Alpha	Hidden_layer_sizes	Optimizer	Mean MSE	n_components
tanh	0.0001	(50,)	adam	90.9647	1
tanh	0.0001	(50,)	sgd	22.7929	1
tanh	0.0001	(100,)	adam	37.9598	1
tanh	0.0001	(100,)	sgd	29.4569	1
...	...	...	...	...	
tanh	0.001	(50, 50)	adam	44.8219	4
tanh	0.001	(50, 50)	sgd	1.0734	4
relu	0.0001	(50,)	adam	17.5982	4
relu	0.0001	(50,)	sgd	1.4428	4
relu	0.0001	(100,)	adam	7.5256	4
relu	0.0001	(100,)	sgd	1.3802	4
...	...	...	...	...	
relu	0.001	(100,)	adam	7.5882	8
relu	0.001	(100,)	sgd	0.8968	8
relu	0.001	(50, 50)	adam	3.5902	8
relu	0.001	(50, 50)	sgd	0.8302	8
tanh	0.0001	(50,)	adam	75.4440	9
tanh	0.0001	(50,)	sgd	0.9547	9
tanh	0.0001	(100,)	adam	2.9632	9
tanh	0.0001	(100,)	sgd	0.9604	9
...	...	...	...	...	
tanh	0.0001	(50,)	sgd	0.8293	10
tanh	0.0001	(100,)	adam	3.4727	10
tanh	0.0001	(100,)	sgd	0.7241	10
tanh	0.0001	(50, 50)	adam	40.1967	10
tanh	0.0001	(50, 50)	sgd	0.7241	10
tanh	0.001	(50,)	adam	64.2525	10
tanh	0.001	(50,)	sgd	0.8293	10
tanh	0.001	(100,)	adam	3.4727	10
...	...	...	...	...	
relu	0.001	(100,)	sgd	0.7810	10
relu	0.001	(50, 50)	adam	3.0570	10
relu	0.001	(50, 50)	sgd	0.8436	10

Table 8.18 : ICA_MLP model hyperparameter tuning for different ICs using the top K = 15 external factors.

Activation function	Alpha	Hidden_layer_sizes	Optimizer	Mean MSE	n_components
tanh	0.0001	(50,)	adam	96.4619	1
tanh	0.0001	(50,)	sgd	22.5920	1
tanh	0.0001	(100,)	adam	45.0781	1
tanh	0.0001	(100,)	sgd	25.7811	1
...	...	...	...	...	
relu	0.0001	(100,)	adam	9.1630	6
relu	0.0001	(100,)	sgd	2.1185	6
relu	0.0001	(50, 50)	adam	3.3521	6
relu	0.0001	(50, 50)	sgd	1.4151	6
relu	0.001	(50,)	adam	14.9970	6
...	...	...	...	...	
tanh	0.001	(100,)	adam	4.7254	10
tanh	0.001	(100,)	sgd	1.0085	10
tanh	0.001	(50, 50)	adam	41.2403	10
tanh	0.001	(50, 50)	sgd	1.0904	10
relu	0.0001	(50,)	adam	13.1756	10
relu	0.0001	(50,)	sgd	1.0223	10
relu	0.0001	(100,)	adam	4.8805	10
...	...	...	...	...	
relu	0.001	(100,)	sgd	0.8722	13
relu	0.001	(50, 50)	adam	3.2425	13
relu	0.001	(50, 50)	sgd	0.8779	13
tanh	0.0001	(50,)	adam	57.1085	14
tanh	0.0001	(50,)	sgd	0.7649	14
tanh	0.0001	(100,)	adam	2.8784	14
tanh	0.0001	(100,)	sgd	0.6974	14
tanh	0.0001	(50, 50)	adam	42.7985	14
tanh	0.0001	(50, 50)	sgd	0.6313	14
tanh	0.001	(50,)	adam	57.1083	14
tanh	0.001	(50,)	sgd	0.7648	14
...	...	...	...	...	
relu	0.001	(100,)	adam	4.1983	15
relu	0.001	(100,)	sgd	1.0961	15
relu	0.001	(50, 50)	adam	3.5746	15
relu	0.001	(50, 50)	sgd	0.8658	15

Table 8.19 : ICA_MLP model hyperparameter tuning for different ICs using all K = 22 external factors.

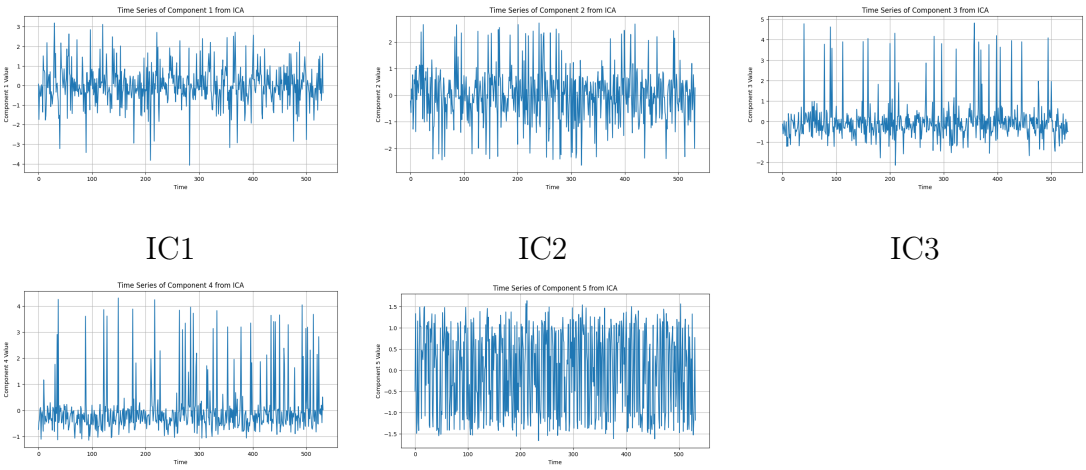
Activation function	Alpha	Hidden_layer_sizes	Optimizer	Mean MSE	n_components
tanh	0.0001	(50,)	adam	102.6125	1
tanh	0.0001	(50,)	sgd	25.5240	1
tanh	0.0001	(100,)	adam	52.3309	1
tanh	0.0001	(100,)	sgd	25.9488	1
...	...	...	...	...	
tanh	0.001	(100,)	adam	13.9915	7
tanh	0.001	(100,)	sgd	3.6177	7
tanh	0.001	(50, 50)	adam	44.2842	7
tanh	0.001	(50, 50)	sgd	1.8154	7
relu	0.0001	(50,)	adam	18.3370	7
...	...	...	...	...	
tanh	0.0001	(50,)	sgd	1.0307	14
tanh	0.0001	(100,)	adam	3.8895	14
tanh	0.0001	(100,)	sgd	0.9717	14
...	...	...	...	...	
tanh	0.001	(50, 50)	adam	34.5723	18
tanh	0.001	(50, 50)	sgd	0.8034	18
relu	0.0001	(50,)	adam	14.1198	18
relu	0.0001	(50,)	sgd	1.1210	18
relu	0.0001	(100,)	adam	5.8892	18
relu	0.0001	(100,)	sgd	1.1883	18
relu	0.0001	(50, 50)	adam	3.6256	18
...	...	...	...	...	
tanh	0.0001	(50,)	sgd	0.9046	21
tanh	0.0001	(100,)	adam	6.6053	21
tanh	0.0001	(100,)	sgd	0.7507	21
tanh	0.0001	(50, 50)	adam	37.4012	21
tanh	0.0001	(50, 50)	sgd	0.7507	21
tanh	0.001	(50,)	adam	70.5236	21
tanh	0.001	(50,)	sgd	0.9045	21
tanh	0.001	(100,)	adam	6.6052	21
...	...	...	...	...	
relu	0.001	(100,)	adam	4.8701	22
relu	0.001	(100,)	sgd	1.4665	22
relu	0.001	(50, 50)	adam	3.9918	22
relu	0.001	(50, 50)	sgd	1.2446	22



(a) $k = 5$ (b) $k = 10$

(c) $k = 15$ (d) $k = 22$ (All factors)

Figure 8.31 : ICA_MLP model hyperparameter tuning for different components using a different number of external factor configurations.



IC1 IC2 IC3

IC4 IC5

Figure 8.32 : Best extracted IC components by the proposed ICA_MLP model using the top $K = 5$ factors.

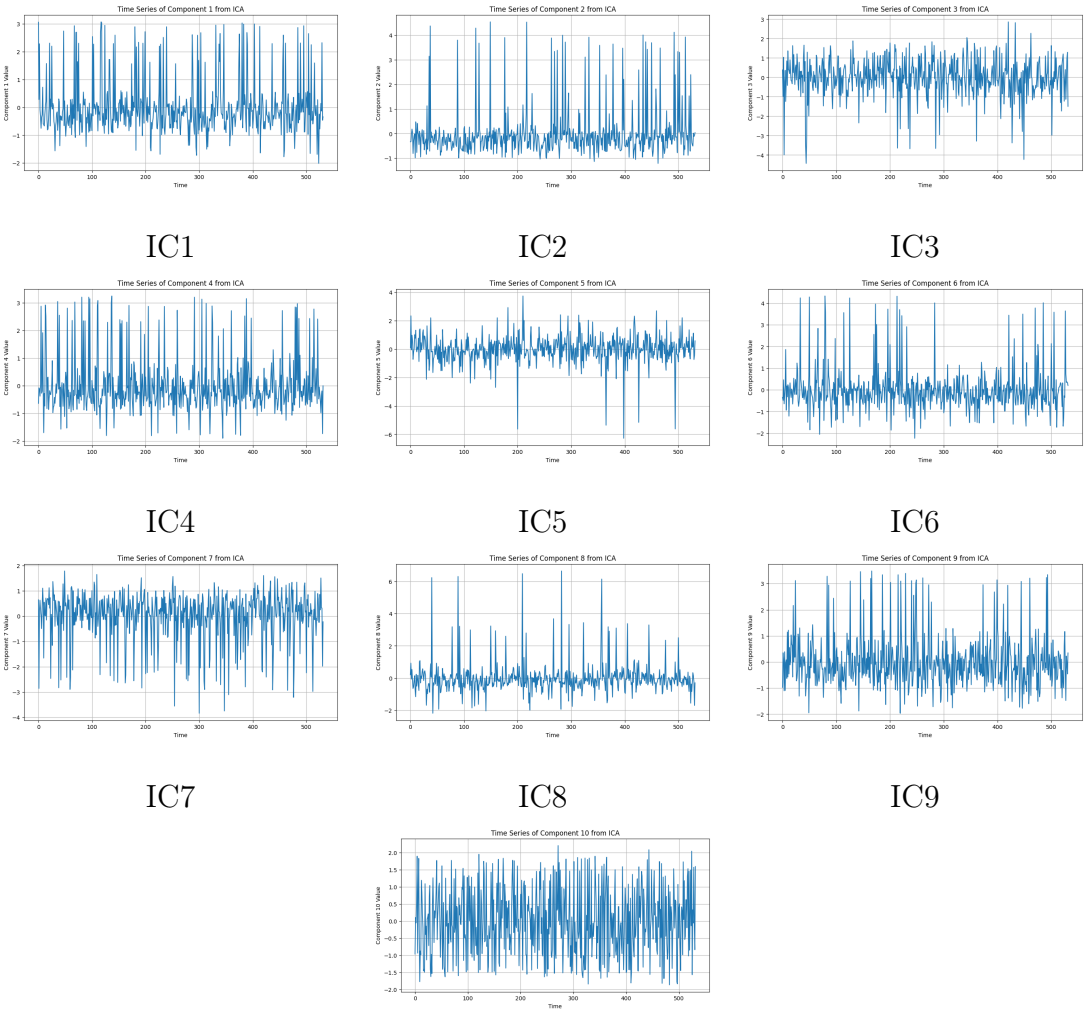


Figure 8.33 : Best extracted IC components by the proposed ICA_MLP model using the top $K = 10$ factors.

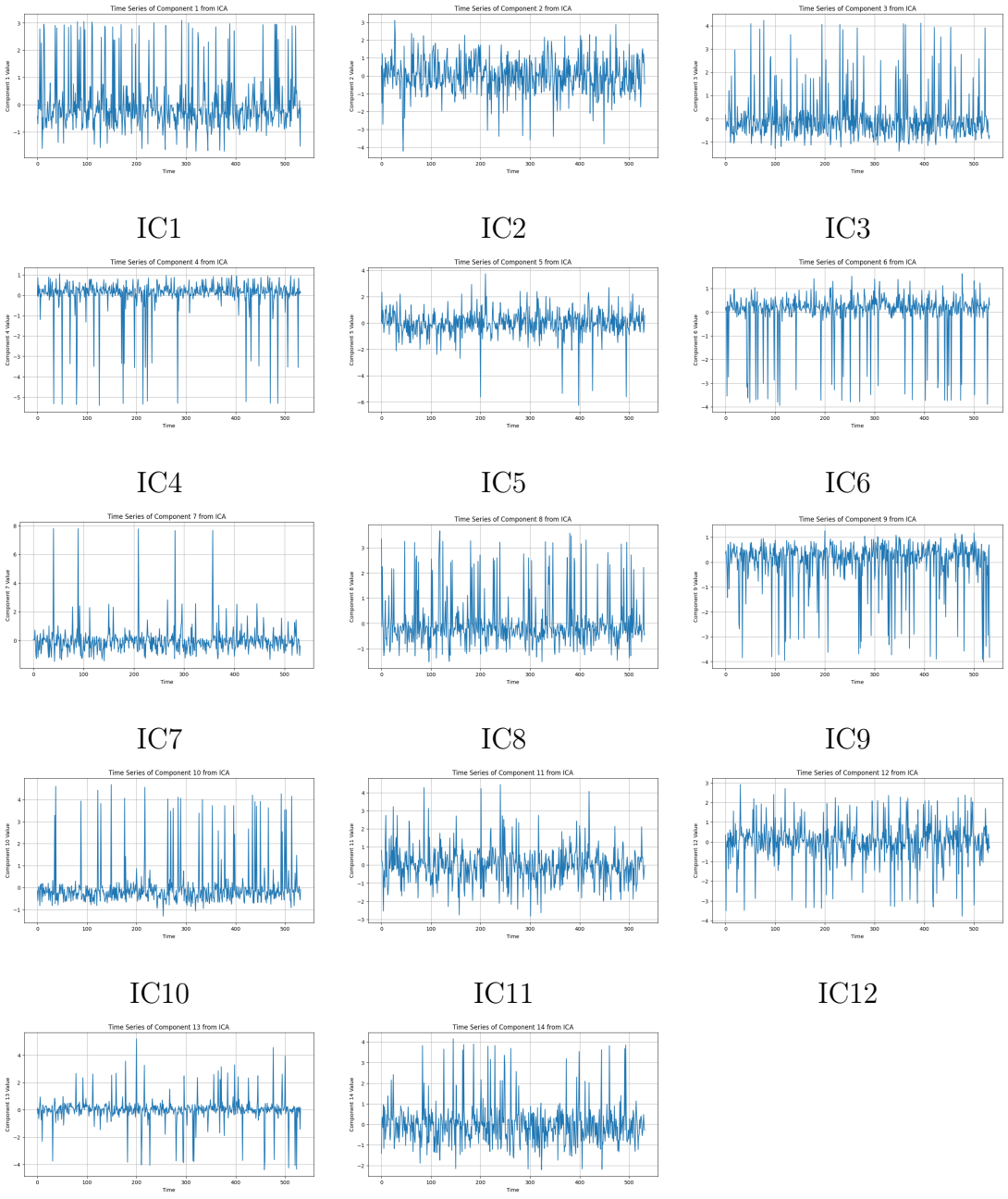
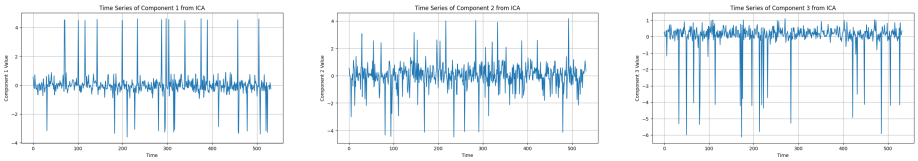
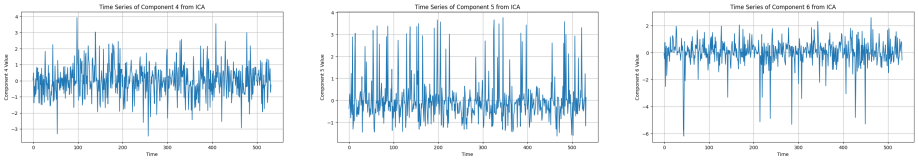


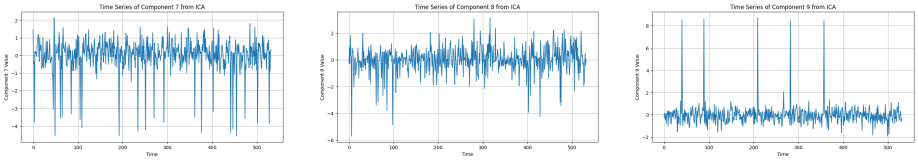
Figure 8.34 : Best extracted IC components by the proposed ICA_MLP model using the top $K = 15$ factors.



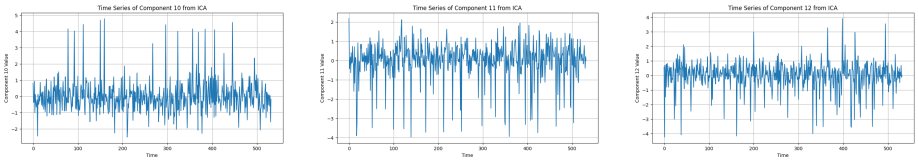
IC1 IC2 IC3



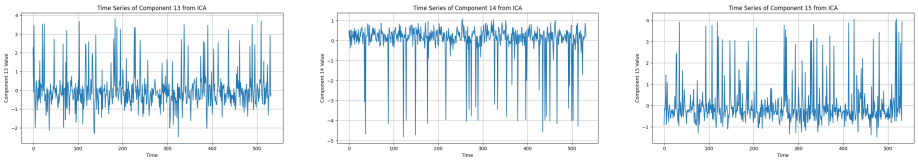
IC4 IC5 IC6



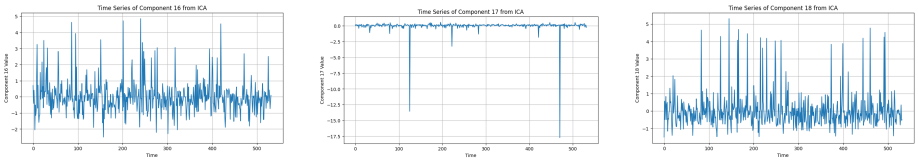
IC7 IC8 IC9



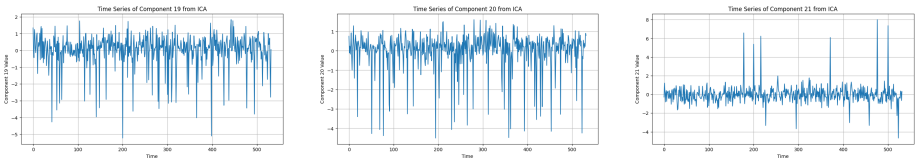
IC10 IC11 IC12



IC13 IC14 IC15

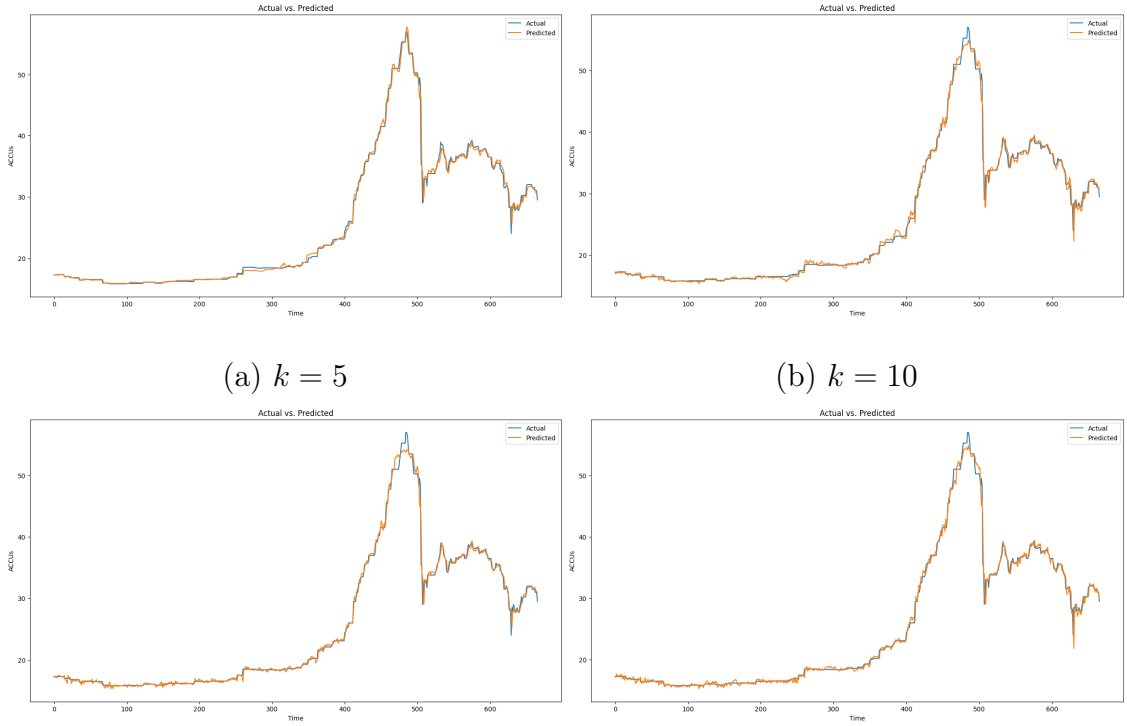


IC16 IC17 IC18



IC19 IC20 IC21

Figure 8.35 : Best extracted IC components by the proposed ICA_MLP model using all $K = 22$ factors.



(a) $k = 5$ (b) $k = 10$

(c) $k = 15$ (d) $k = 22$ (All factors)

Figure 8.36 : Actual vs. predicted ACCU price using the proposed ICA_MLP model with different external factor configurations.

In the case of the NLICA_MLP model, an autoencoder is used to compress the data into a lower-dimensional representation, capturing non-linear structures that linear methods PCA or ICA might miss. Similar to ICA, the number of independent components is not predefined. Instead, we define a range from 1 to the total number of features (22 factors) in the dataset and loop through this range. For each value, the autoencoder is configured to reduce the data to that number of components. The transformed data is then used to evaluate the model's performance based on MSE, and the number of components that result in the lowest MSE is selected as the optimal configuration.

For each possible number of components, the grid search was performed using cross-validation to optimise the MLP hyperparameters, and the model was evalu-

ated based on its MSE performance. Tables 8.20 to 8.23 present the results of the hyperparameter tuning for the NLICA_MLP model, using a different number of external factors K. For the sake of space, the tables have been summarised to show only the key configurations and the best-performing models with the optimal number of ICs. By systematically looping through different values for `n_components`, the lowest MSE was achieved with 5, 10, 9, 21 ICs when the number of external factors is ($K = 5$, $K = 10$, $K = 15$, and $K = 22$) respectively, as shown in Figure 8.37.

Figures 8.38 to 8.41 show the best IC components extracted by the proposed NLICA_MLP model for a different number of external factors K. In addition, Figure 8.42 compares the actual and predicted results for ACCU prices using the proposed NLICA_MLP model using a different numbers of external factors ($K = 5$, $K = 10$, $K = 15$, and $K = 22$).

Table 8.20 : NLICA_MLP model hyperparameter tuning for different ICs using the top K = 5 external factors.

Activation function	Alpha	Hidden_layer_sizes	Optimizer	Mean MSE	n_components
tanh	0.0001	(50,)	adam	53.2476	1
tanh	0.0001	(50,)	sgd	25.2054	1
tanh	0.0001	(100,)	adam	25.9481	1
tanh	0.0001	(100,)	sgd	25.0010	1
tanh	0.0001	(50, 50)	adam	54.7563	1
tanh	0.0001	(50, 50)	sgd	21.5151	1
...	...	...	...	...	
relu	0.001	(50,)	sgd	14.6644	2
relu	0.001	(100,)	adam	35.5527	2
relu	0.001	(100,)	sgd	15.1473	2
relu	0.001	(50, 50)	adam	22.7896	2
relu	0.001	(50, 50)	sgd	8.7731	2
...	...	...	...	...	
tanh	0.001	(50, 50)	sgd	9.0324	3
relu	0.0001	(50,)	adam	32.1109	3
relu	0.0001	(50,)	sgd	9.6890	3
relu	0.0001	(100,)	adam	12.4012	3
...	...	...	...	...	
relu	0.001	(100,)	sgd	27.5094	4
relu	0.001	(50, 50)	adam	28.8704	4
relu	0.001	(50, 50)	sgd	8.1043	4
tanh	0.0001	(50,)	adam	73.8986	5
tanh	0.0001	(50,)	sgd	12.6576	5
tanh	0.0001	(100,)	adam	24.9925	5
tanh	0.0001	(100,)	sgd	10.7728	5
tanh	0.0001	(50, 50)	adam	55.2361	5
tanh	0.0001	(50, 50)	sgd	18.1448	5
tanh	0.001	(50,)	adam	73.8987	5
tanh	0.001	(50,)	sgd	12.6583	5
...	...	...	...	...	
relu	0.0001	(50, 50)	adam	16.9239	5
relu	0.0001	(50, 50)	sgd	3.6594	5
relu	0.001	(50,)	adam	44.3277	5
relu	0.001	(50,)	sgd	7.6958	5
relu	0.001	(100,)	adam	28.2611	5
relu	0.001	(100,)	sgd	7.1746	5
relu	0.001	(50, 50)	adam	17.6072	5
relu	0.001	(50, 50)	sgd	3.6767	5

Table 8.21 : NLICA_MLP model hyperparameter tuning for different ICs using the top K = 10 external factors.

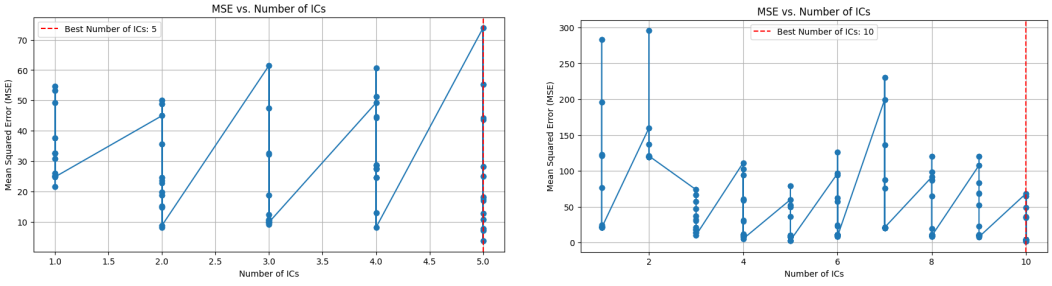
Activation function	Alpha	Hidden_layer_sizes	Optimizer	Mean MSE	n_components
tanh	0.0001	(50,)	adam	195.7280	1
tanh	0.0001	(50,)	sgd	21.6569	1
tanh	0.0001	(100,)	adam	76.8337	1
tanh	0.0001	(100,)	sgd	21.7155	1
...	...	...	...	...	
relu	0.001	(50,)	sgd	8.3366	4
relu	0.001	(100,)	adam	94.0933	4
relu	0.001	(100,)	sgd	7.9574	4
relu	0.001	(50, 50)	adam	29.8393	4
relu	0.001	(50, 50)	sgd	5.3727	4
tanh	0.0001	(50,)	adam	59.9072	5
tanh	0.0001	(50,)	sgd	10.4978	5
tanh	0.0001	(100,)	adam	36.5206	5
tanh	0.0001	(100,)	sgd	10.0165	5
tanh	0.0001	(50, 50)	adam	49.7406	5
tanh	0.0001	(50, 50)	sgd	3.0633	5
...	...	...	...	...	
tanh	0.001	(100,)	sgd	20.9316	7
tanh	0.001	(50, 50)	adam	87.3682	7
tanh	0.001	(50, 50)	sgd	20.6185	7
relu	0.0001	(50,)	adam	230.3729	7
relu	0.0001	(50,)	sgd	20.7279	7
...	...	...	...	...	
relu	0.0001	(100,)	sgd	4.1515	10
relu	0.0001	(50, 50)	adam	4.6288	10
relu	0.0001	(50, 50)	sgd	1.7691	10
relu	0.001	(50,)	adam	64.7199	10
relu	0.001	(50,)	sgd	4.0390	10
relu	0.001	(100,)	adam	35.0703	10
relu	0.001	(100,)	sgd	4.1520	10
relu	0.001	(50, 50)	adam	4.6814	10
relu	0.001	(50, 50)	sgd	1.7692	10

Table 8.22 : NLICA_MLP model hyperparameter tuning for different ICs using the top K = 15 external factors.

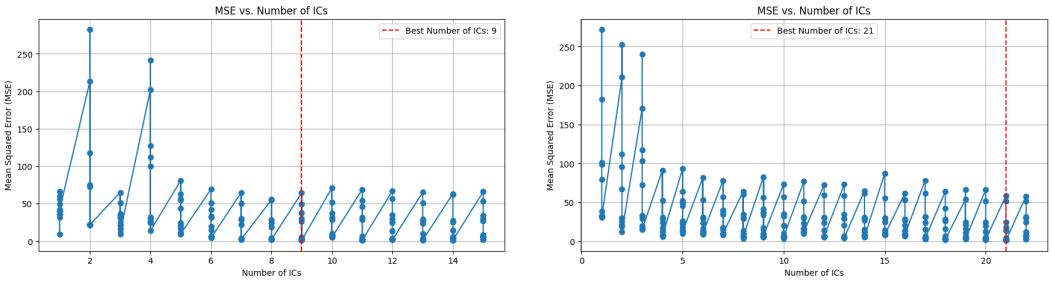
Activation function	Alpha	Hidden_layer_sizes	Optimizer	Mean MSE	n_components
tanh	0.0001	(50,)	adam	59.6202	1
tanh	0.0001	(50,)	sgd	31.9306	1
tanh	0.0001	(100,)	adam	35.4800	1
tanh	0.0001	(100,)	sgd	35.5034	1
...	...	...	...	...	
relu	0.0001	(50, 50)	sgd	4.3019	6
relu	0.001	(50,)	adam	42.1203	6
relu	0.001	(50,)	sgd	5.8851	6
relu	0.001	(100,)	adam	33.7870	6
relu	0.001	(100,)	sgd	5.9604	6
relu	0.001	(50, 50)	adam	14.5551	6
relu	0.001	(50, 50)	sgd	4.2533	6
tanh	0.0001	(50,)	adam	65.0202	7
tanh	0.0001	(50,)	sgd	2.9403	7
...	...	...	...	...	
relu	0.0001	(50,)	sgd	2.4842	9
relu	0.0001	(100,)	adam	27.7093	9
relu	0.0001	(100,)	sgd	2.5020	9
relu	0.0001	(50, 50)	adam	5.2407	9
relu	0.0001	(50, 50)	sgd	1.0862	9
relu	0.001	(50,)	adam	37.8710	9
relu	0.001	(50,)	sgd	2.4840	9
...	...	...	...	...	
tanh	0.0001	(50, 50)	sgd	1.7464	13
tanh	0.001	(50,)	adam	65.2967	13
tanh	0.001	(50,)	sgd	2.7671	13
tanh	0.001	(100,)	adam	27.6097	13
...	...	...	...	...	
relu	0.0001	(50, 50)	sgd	2.6541	15
relu	0.001	(50,)	adam	34.4033	15
relu	0.001	(50,)	sgd	6.1941	15
relu	0.001	(100,)	adam	27.3439	15
relu	0.001	(100,)	sgd	5.8118	15
relu	0.001	(50, 50)	adam	8.7144	15
relu	0.001	(50, 50)	sgd	2.6425	15

Table 8.23 : NLICA_MLP model hyperparameter tuning for different ICs using all K = 22 external factors.

Activation function	Alpha	Hidden_layer_sizes	Optimizer	Mean MSE	n_components
tanh	0.0001	(50,)	adam	182.0624	1
tanh	0.0001	(50,)	sgd	32.6723	1
tanh	0.0001	(100,)	adam	79.2115	1
tanh	0.0001	(100,)	sgd	33.3642	1
...	...	...	...	...	
relu	0.001	(50, 50)	adam	13.4733	8
relu	0.001	(50, 50)	sgd	6.1329	8
tanh	0.0001	(50,)	adam	82.7240	9
tanh	0.0001	(50,)	sgd	6.1307	9
tanh	0.0001	(100,)	adam	36.8157	9
...	...	...	...	...	
relu	0.001	(100,)	sgd	12.3240	14
relu	0.001	(50, 50)	adam	15.2802	14
relu	0.001	(50, 50)	sgd	5.5778	14
tanh	0.0001	(50,)	adam	87.2859	15
tanh	0.0001	(50,)	sgd	8.1009	15
...	...	...	...	...	
relu	0.001	(50, 50)	adam	6.5716	18
relu	0.001	(50, 50)	sgd	2.2328	18
tanh	0.0001	(50,)	adam	54.0271	19
tanh	0.0001	(50,)	sgd	5.6323	19
...	...	...	...	...	
relu	0.0001	(50,)	sgd	2.3462	21
relu	0.0001	(100,)	adam	13.3667	21
relu	0.0001	(100,)	sgd	2.2725	21
relu	0.0001	(50, 50)	adam	4.7690	21
relu	0.0001	(50, 50)	sgd	1.1922	21
relu	0.001	(50,)	adam	16.9129	21
relu	0.001	(50,)	sgd	2.3460	21
relu	0.001	(100,)	adam	13.9360	21
relu	0.001	(100,)	sgd	2.2726	21
relu	0.001	(50, 50)	adam	4.8251	21
...	...	...	...	...	
relu	0.001	(50,)	sgd	7.1853	22
relu	0.001	(100,)	adam	24.3833	22
relu	0.001	(100,)	sgd	7.3201	22
relu	0.001	(50, 50)	adam	11.8172	22
relu	0.001	(50, 50)	sgd	2.6371	22

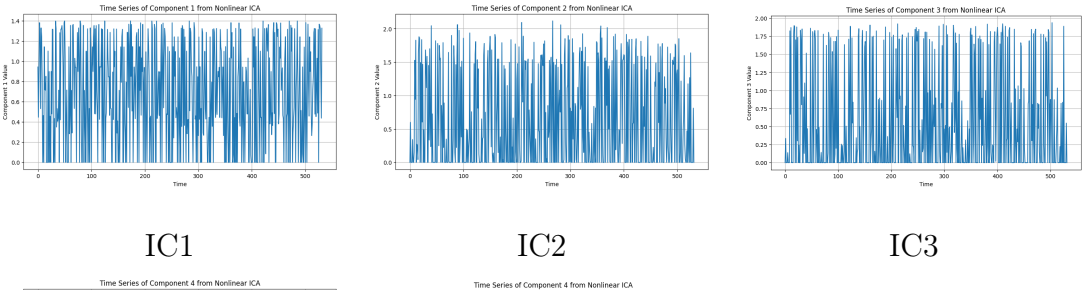


(a) $k = 5$ (b) $k = 10$



(c) $k = 15$ (d) $k = 22$ (All factors)

Figure 8.37 : NLICA_MLP model hyperparameter tuning for different components using a different number of external factor configurations.



IC1 IC2 IC3

IC4 IC5

Figure 8.38 : Best extracted IC components by the proposed NLICA_MLP model using the top $K = 5$ factors.

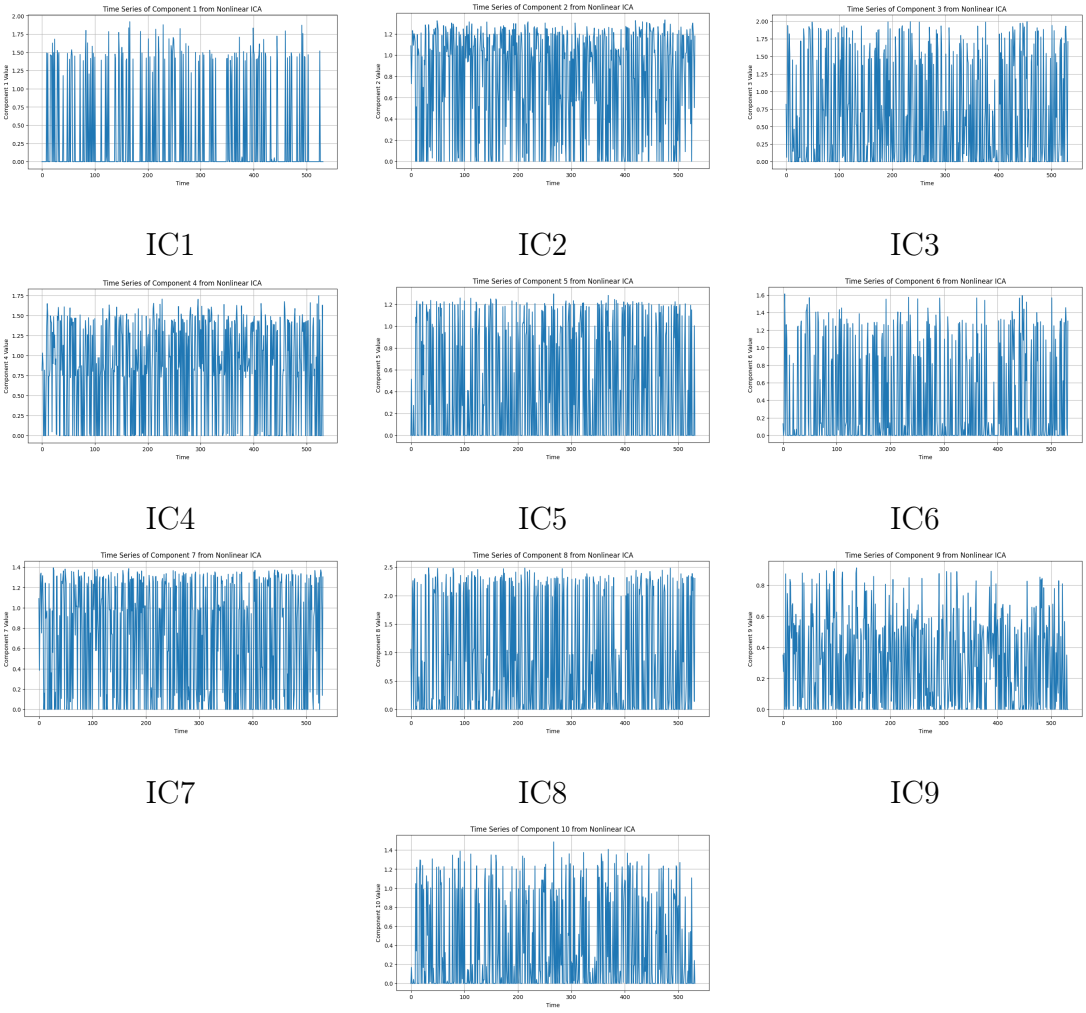


Figure 8.39 : Best extracted IC components by the proposed NLICA_MLP model using the top $K = 10$ factors.

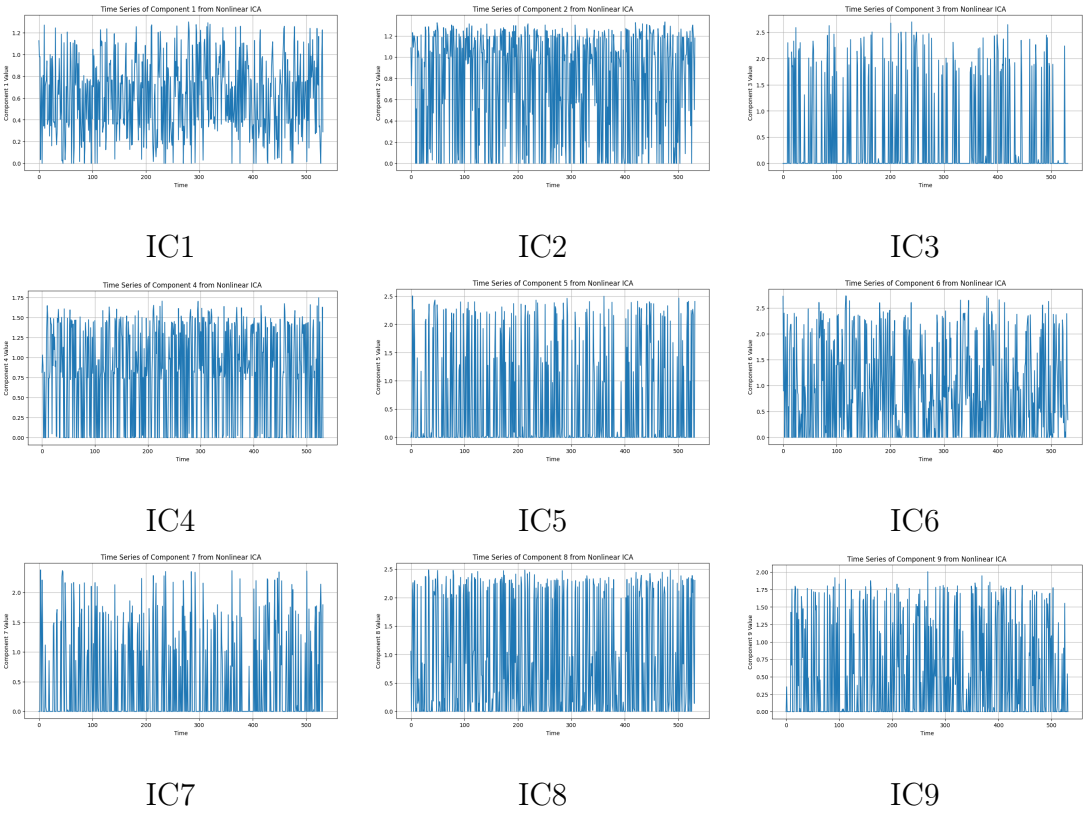
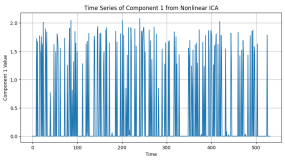
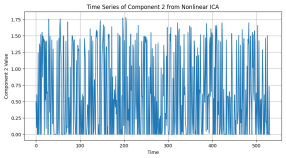


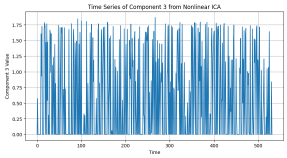
Figure 8.40 : Best extracted IC components by the proposed NLICA_MLP model using the top $K = 15$ factors.



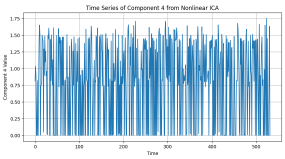
IC1



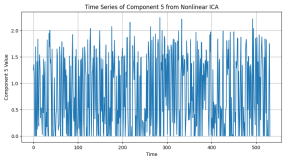
IC2



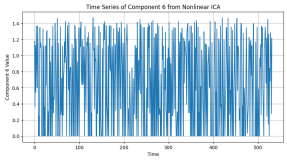
IC3



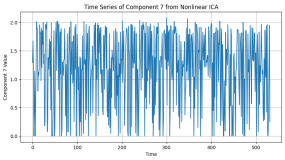
IC4



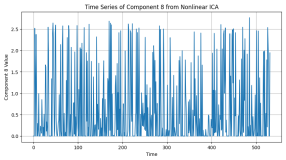
IC5



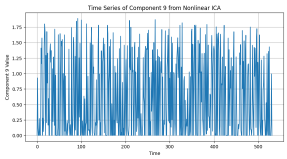
IC6



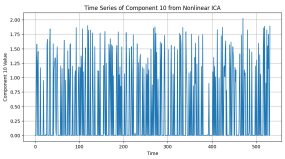
IC7



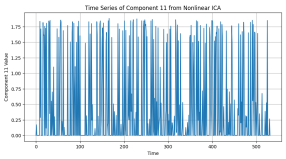
IC8



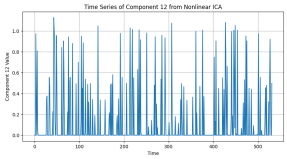
IC9



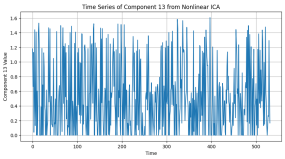
IC10



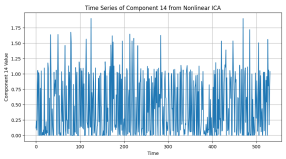
IC11



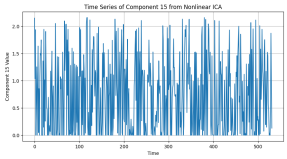
IC12



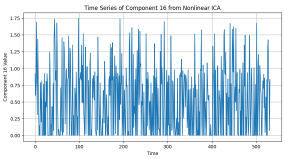
IC13



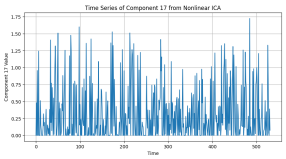
IC14



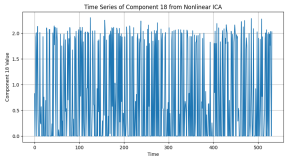
IC15



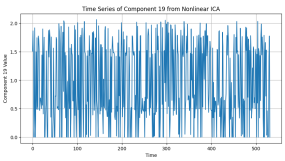
IC16



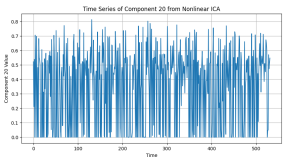
IC17



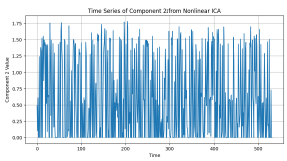
IC18



IC19

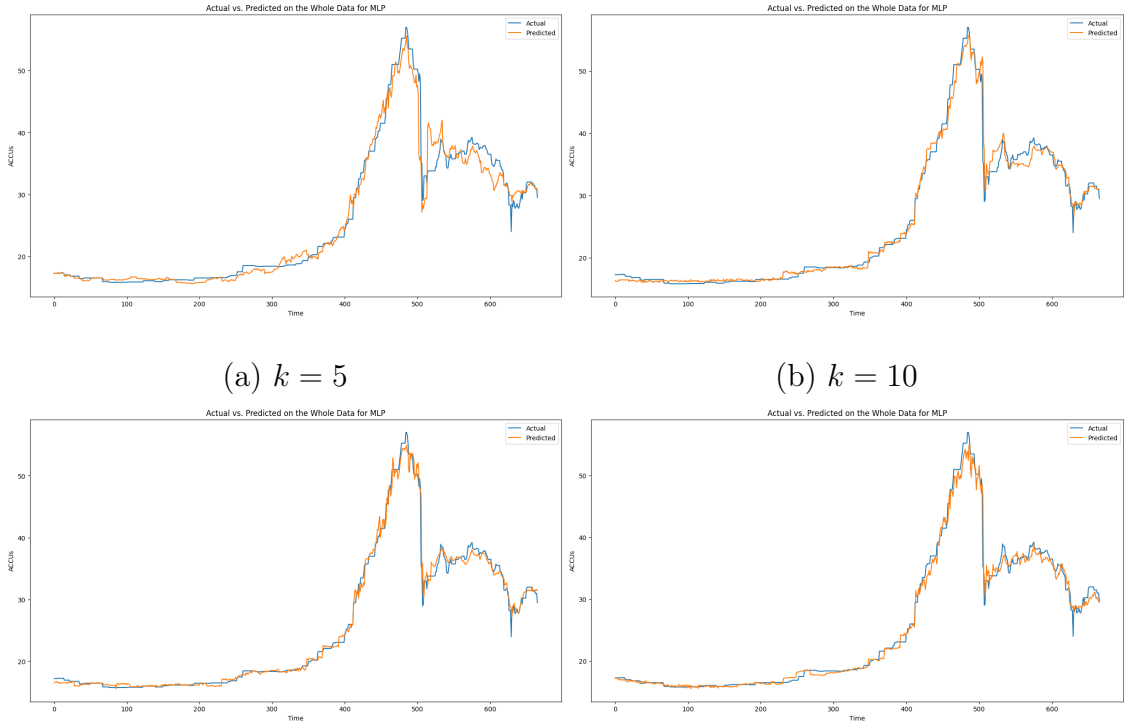


IC20



IC21

Figure 8.41 : Best extracted IC components by the proposed NLICA_MLP model using all $K = 22$ factors.



(a) $k = 5$ (b) $k = 10$

(c) $k = 15$ (d) $k = 22$ (All factors)

Figure 8.42 : Actual vs. predicted ACCU price using the proposed NLICA_MLP model with different external factor configurations.

The parameter tuning of the MLP models based on different feature extraction and reduction techniques highlighted varying optimal configurations. Both the PCA_MLP and ICA_MLP models achieved their best performance using the `tanh` activation function, while the NLICA_MLP model performed best with the `relu` activation function. The details of the final optimiser's parameters for each model are shown in Table 8.24.

Table 8.24 : Hyperparameter tuning results for PCA_MLP, ICA_MLP, and NLICA_MLP models, showing mean MSE and optimal ICs based on a different number of external factors.

Model	**Activation function**	**Alpha**	**Layers**	**Optimizer**	**K=5**		**K=10**		**K=15**		**K=22**	
					MSE	**ICs**	**MSE**	**ICs**	**MSE**	**ICs**	**MSE**	**ICs**
PCA_MLP	tanh	0.0001	(50, 50)	sgd	0.8017	3	1.1357	4	1.7564	6	2.4308	7
ICA_MLP	tanh	0.0001	(50, 50)	sgd	0.6770	5	0.7241	10	0.6313	14	0.7507	21
NLICA_MLP	relu	0.0001	(50, 50)	sgd	3.6594	5	1.7691	10	1.0862	9	1.1922	21

8.6.3 Comparative performance analysis of prediction models

This section presents a comprehensive analysis of the results generated from the proposed CCMFP models using standard evaluation metrics: MSE, MAE, RMSE, R^2, and execution time, highlighting the depth and rigour of our experimental evaluation. Three advanced feature extraction and reduction techniques, PCA, ICA, and NLICA, were applied to the SVR and MLP models, resulting in the development of six prediction models: PCA_SVR, ICA_SVR, NLICA_SVR, PCA_MLP, ICA_MLP, and NLICA_MLP. To further assess the performance of these models, we conducted extensive experiments by varying the number of top-selected factors ($K = 5, 10, 15, 22$) as input to the prediction models. This variation enabled us to systematically analyse how the number of input features impacts prediction accuracy and to identify the optimal set of features to achieve the most accurate predictions.

Based on the analysis of Table 8.25 and Figure 8.43, it was found that the ICA_SVR model with $K = 10$ factors and nine IC components is identified as the best model for accurate and efficient short-term prediction of ACCU prices in this study, followed by the ICA_MLP model with $K = 5$ factors and five IC components. The ICA_SVR model achieves a competitive MSE of 1.0393 and a high R^2 value of 0.9960 while maintaining a computational efficiency of 1.33 seconds. The

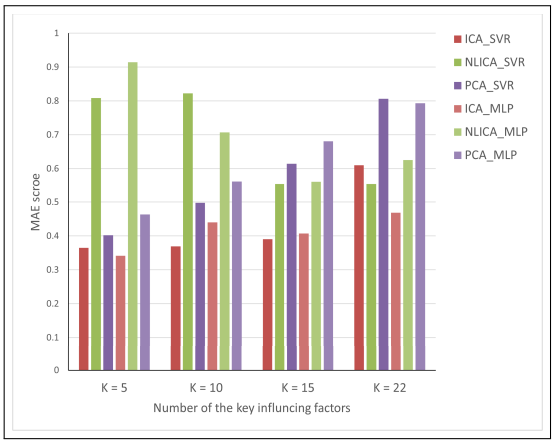
ICA_MLP model achieved an MSE of 0.6770 and an R^2 value of 0.9959; however, its computational time of 56.34 seconds is significantly higher, making it less efficient compared to the ICA_SVR model. It should be noted that all the proposed models achieved high prediction accuracy above 97.7%.

Figure 8.44 presents box-and-whisker plots aggregating the results of all K configurations to compare model performance. This visualisation provides a comprehensive assessment of consistency and robustness across evaluation metrics. In particular, ICA_SVR achieved the lowest average MAE and RMSE and the highest R^2, confirming the superior performance of the ICA-enhanced models.

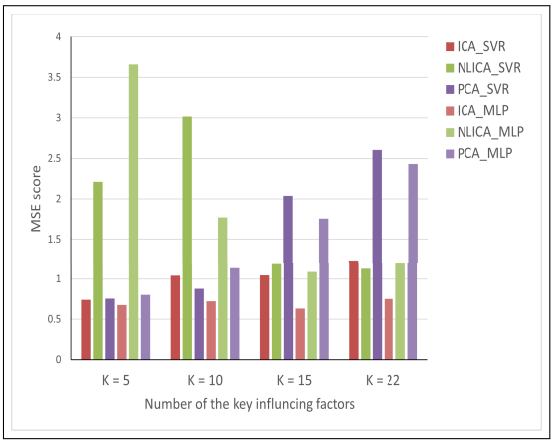
In terms of execution time, the SVR-based model consistently maintains a balance between accuracy and computational efficiency, making it the preferred model for predicting ACCU prices. However, the MLP-based model generally requires more computational resources, as reflected in the significantly longer time required for all K factor configurations. Moreover, ICA_SVR demonstrated nearly the fastest training time among all models, offering an optimal balance between accuracy and computational efficiency.

Table 8.25 : Comparison of the prediction models for different configurations of external factors (K).

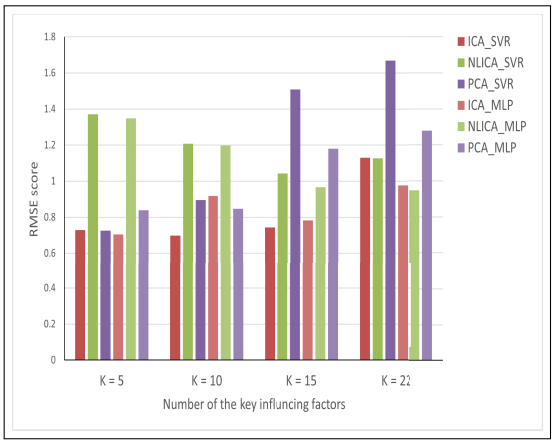
Factors	**Model**	**IC**	**MSE**	**MAE**	**RMSE**	R^2	**Time**
K = 5	PCA_SVR	3	0.7561	0.4021	0.7256	0.9956	10.46
	ICA_SVR	3	0.7395	0.3651	0.7296	0.9956	1.12
	NLICA_SVR	5	2.2097	0.8086	1.3705	0.9845	2.01
	PCA_MLP	3	0.8017	0.4634	0.8381	0.9942	72.81
	ICA_MLP	5	0.6770	0.3416	0.7045	0.9959	56.34
	NLICA_MLP	5	3.6594	0.9148	1.3478	0.9850	39.25
K = 10	PCA_SVR	4	0.8777	0.4974	0.8948	0.9934	2.94
	ICA_SVR	9	1.0393	0.3691	0.6975	0.9960	1.33
	NLICA_SVR	5	3.0164	0.8223	1.2080	0.9879	1.12
	PCA_MLP	4	1.1357	0.5605	0.8473	0.9941	57.83
	ICA_MLP	10	0.7241	0.4404	0.9186	0.9930	59.94
	NLICA_MLP	10	1.7691	0.7073	1.1978	0.9881	37.75
K = 15	PCA_SVR	6	2.0387	0.6148	1.5078	0.9812	1.14
	ICA_SVR	13	1.0439	0.3906	0.7438	0.9954	1.43
	NLICA_SVR	15	1.1849	0.5535	1.0426	0.9910	2.12
	PCA_MLP	6	1.7564	0.6811	1.1807	0.9885	63.57
	ICA_MLP	14	0.6313	0.4075	0.7819	0.9949	60.52
	NLICA_MLP	9	1.0862	0.5600	0.9671	0.9928	41.33
K = 22	PCA_SVR	7	2.6060	0.8064	1.6687	0.9770	1.83
	ICA_SVR	9	1.2375	0.6102	1.1303	0.9894	3.63
	NLICA_SVR	17	1.3176	0.5533	1.1262	0.9895	1.43
	PCA_MLP	7	2.4308	0.7932	1.2786	0.9865	63.09
	ICA_MLP	21	0.7507	0.4685	0.9761	0.9921	62.61
	NLICA_MLP	21	1.1922	0.6258	0.9464	0.9926	46.44



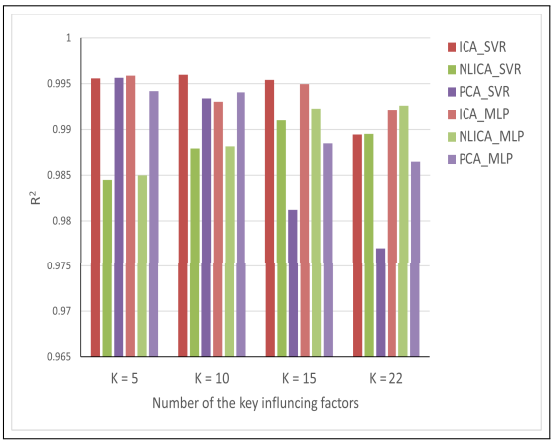
(a) MAE



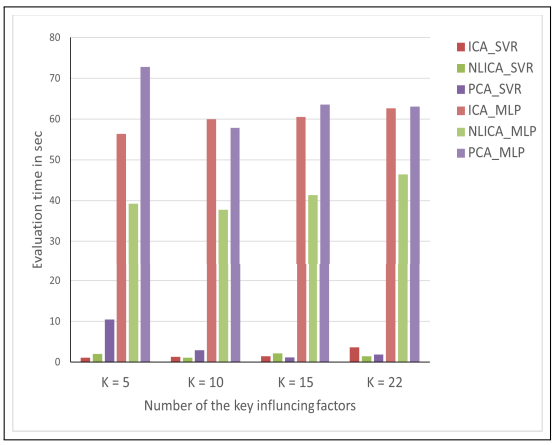
(b) MSE



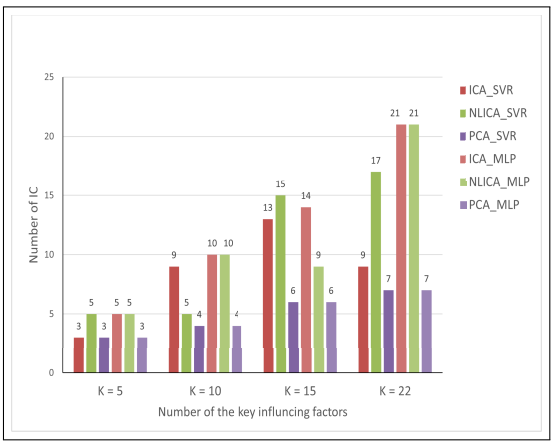
(c) RMSE



(d) R^2

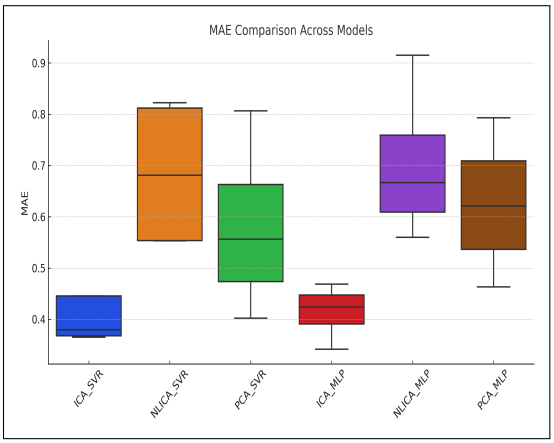


(e) Execution time

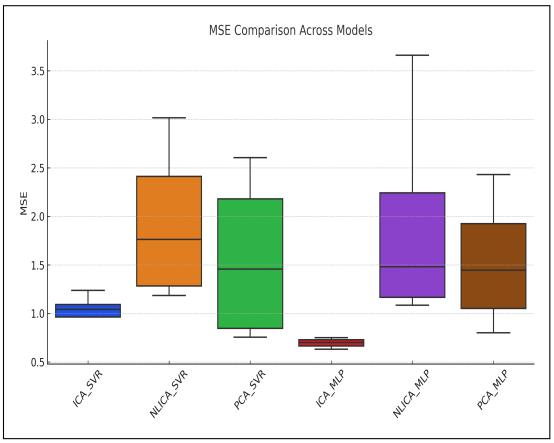


(f) The best-generated number of ICs

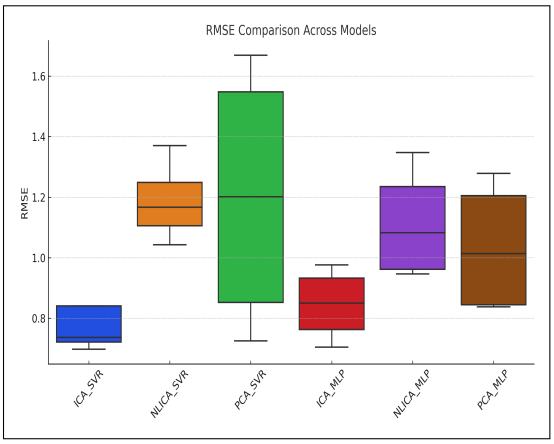
Figure 8.43 : Comparison of the proposed prediction models based on (a) MAE, (b) MSE, (c) RMSE, (d) R^2, (e) execution time and (f) the best-generated number of ICs, using different configurations of external factors.



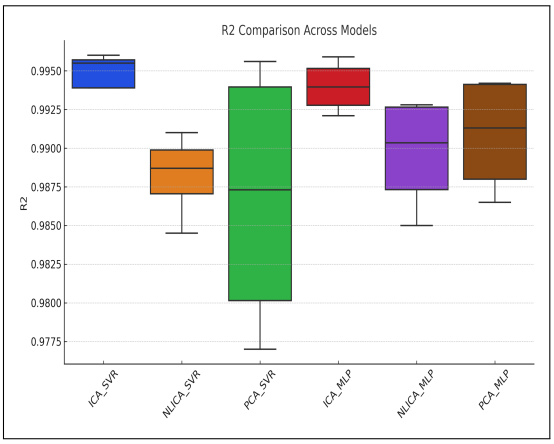
(a) MAE



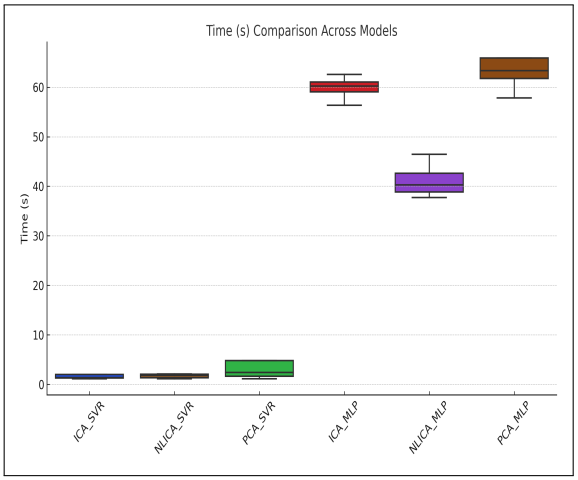
(b) MSE



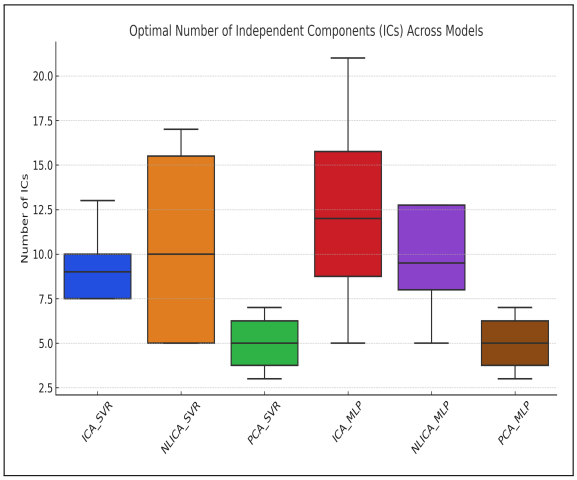
(c) RMSE



(d) R^2



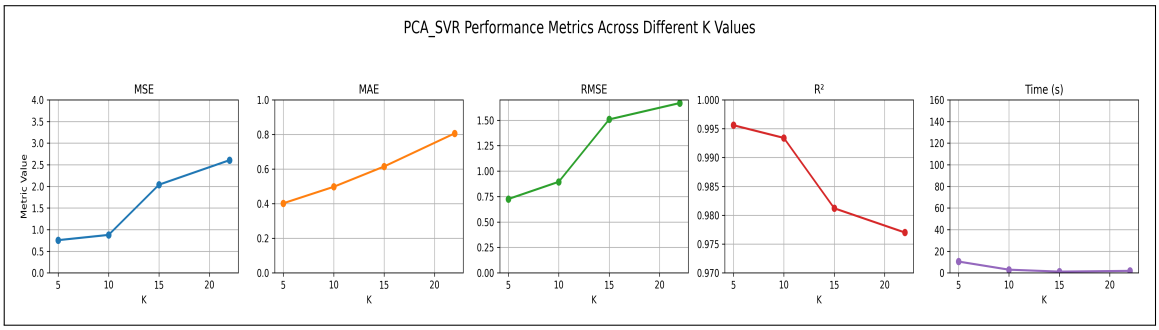
(e) Excution time



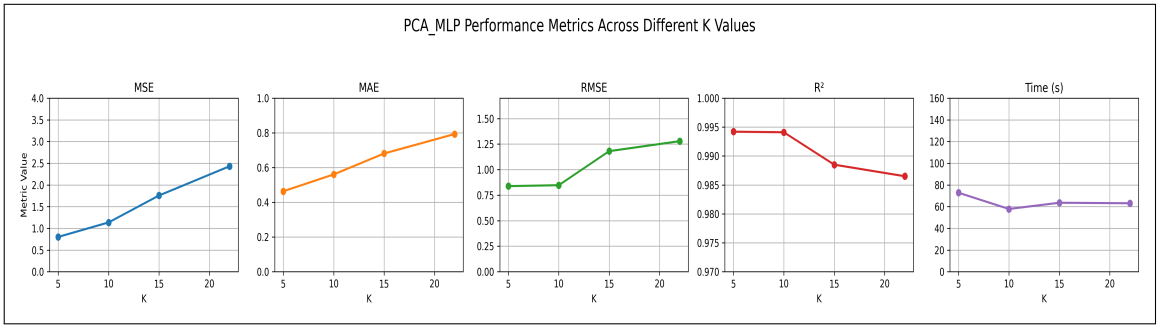
(f) The best-generated number of ICs

Figure 8.44 : Box-and-whisker plots comparing model performance based on (a) MAE, (b) MSE, (c) RMSE, (d) R^2, (e) execution time and (f) the best-generated number of ICs.

To further analyse the behaviour of each feature extraction and reduction technique, we examined performance trends across different K values, as shown in Figures 8.45 - 8.47. PCA-based models show a consistent decrease in accuracy as K increases, with increasing MSE, MAE, and RMSE, and decreasing R^2, suggesting that adding more features introduces irrelevant variance that reduces predictive performance. Although PCA_SVR is computationally efficient, its accuracy is limited, and PCA_MLP, though more accurate at lower K values, incurs significantly higher training costs. In contrast, ICA-based models perform robustly across all K values, with ICA_SVR demonstrating the best trade-off between accuracy and speed, and ICA_MLP maintaining stable error rates with slightly longer training times. ICA's ability to extract independent and informative components contributes to its consistent performance even with increased dimensionality. NLICA-based models show improvement with higher K, reaching peak performance at K = 15; however, they generally exhibit higher error rates and lower R^2 values compared to ICA-based models, and NLICA_MLP shows longer training times. Overall, ICA-based feature extraction significantly enhances model accuracy and robustness, with ICA_SVR emerging as the most efficient and reliable model for short-term carbon credit price prediction.

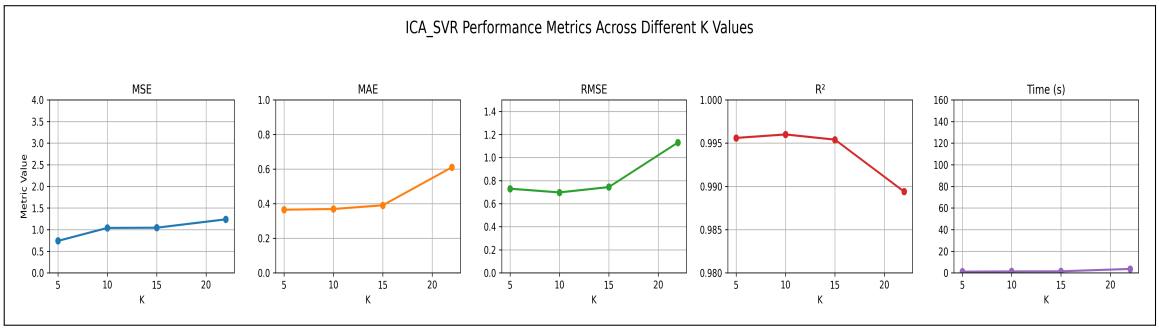


(a) PCA_SVR

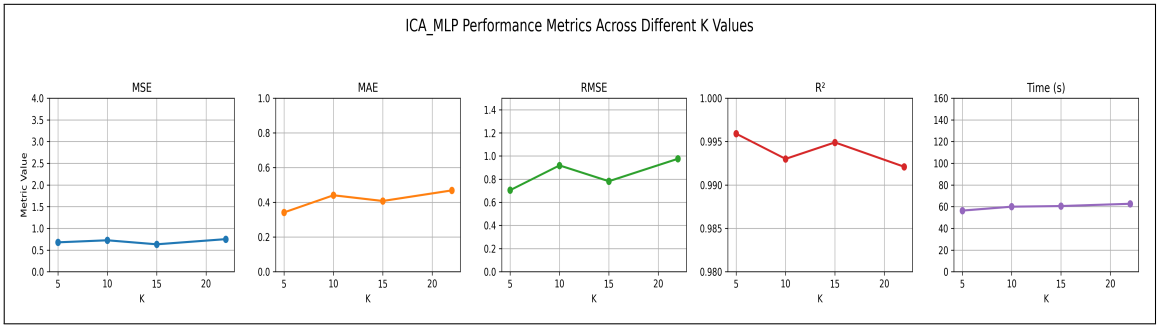


(b) PCA_MLP

Figure 8.45 : Performance evaluation of PCA-based models across different configurations of external factors.

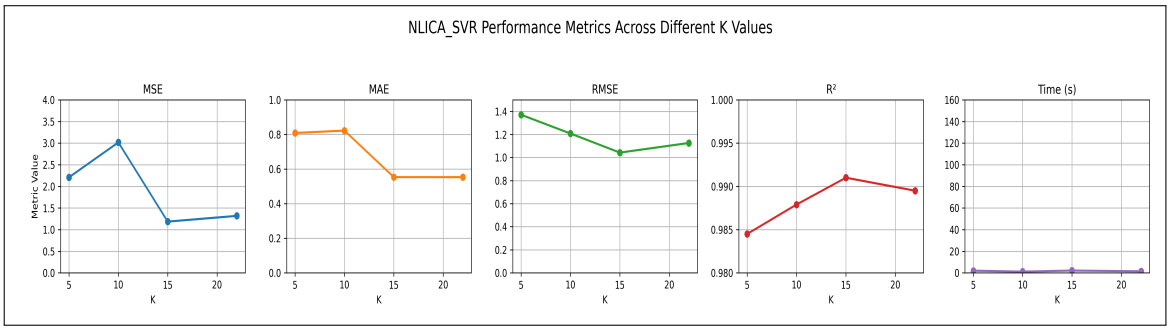


(c) ICA_SVR

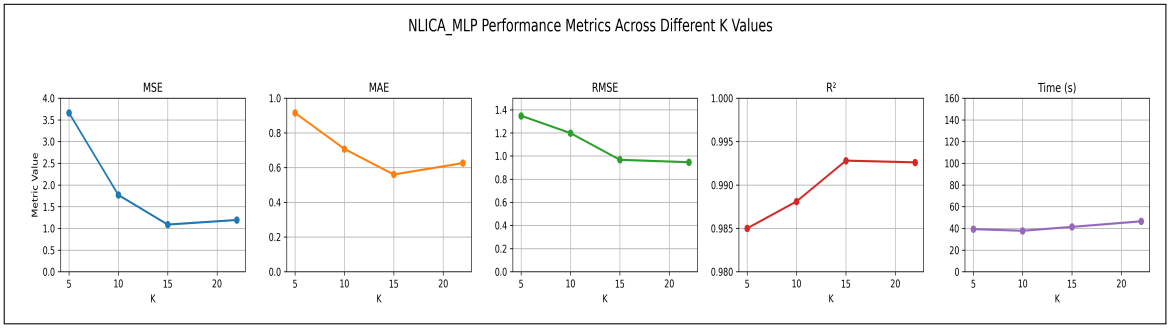


(d) ICA_MLP

Figure 8.46 : Performance evaluation of ICA-based models across different configurations of external factors.



(e) NLICA_SVR



(f) NLICA_MLP

Figure 8.47 : Performance evaluation of NLICA-based models across different configurations of external factors.

In summary, the results show that the ICA-based models consistently outperform their counterparts based on PCA and NLICA. This superior performance highlights the robustness of ICA in predictive modelling, especially under varying influencing factors. Although a slight decline in performance is observed as the number of factors increases, both ICA_SVR and ICA_MLP maintain high accuracy and stability, demonstrating strong generalisation capabilities. This continued effectiveness is attributed to the ability of ICA to extract independent source signals, especially when latent variables follow a non-Gaussian distribution. In addition, compared to PCA, ICA and NLICA generally require a larger number of independent components. This is because ICA seeks statistical independence beyond variance-based decorrelation, and NLICA models nonlinear dependencies, both of which often require a more

detailed latent representation [21].

Although the SVR and MLP models, combined with ICA, NLICA, and PCA, achieved very high prediction accuracy, with R^2 values exceeding 0.99 in some cases, these results should be interpreted with caution. The experiments were conducted as one-step-ahead forecasts, which are generally easier to predict than multi-step-ahead scenarios, where errors can accumulate over time. To minimise the risk of overfitting, redundant variables were removed and only the most relevant factors retained using three feature selection methods. Nevertheless, it is important to recognise that even carefully tuned complex models may face challenges when applied to real-world, multi-step prediction tasks.

The optimal set of features that yield the most accurate predictions using the ICA_SVR model is as follows:

- The PT factor indicates that advances in green technology in the category of sustainable industry have a strong influence on the price of carbon credits. The maturity of green technology serves as a metric to assess the level of progress in emission reduction technologies. As green technologies mature, their effectiveness in curbing emissions increases, leading to lower carbon trading prices [130].

- CO2 is one of the most significant GHG emissions and a critical factor representing the impact of the environmental factor, since higher emissions require more credits to meet regulatory requirements and sustainability goals [114].

- The coal (CP) and crude oil (CrP) markets are key influences that represent the impact of the energy market. These two factors directly affect the cost of carbon credits as companies seek to mitigate their carbon footprint in response to fluctuating energy costs [91], [148], [163].

- The RE factor reflects the influence of the adoption of renewable energy on the price of carbon credits. As more renewable energy sources (for example, wind and solar) are adopted, there is a significant reduction in carbon emissions, as renewables replace fossil fuels, which are the main sources of CO2 emissions [39], [61], [127]. This decrease in emissions can reduce the demand for carbon credits as fewer companies need to offset their emissions, potentially reducing carbon credit prices.

- The EUAs and CEAs were other critical factors highlighting the influence of international carbon markets. The EUAs represent the largest cap-and-trade system globally, and changes in their supply and demand directly affect the price of carbon credits worldwide [91], [163]. Similarly, CEA data provide insight into the Chinese carbon market, which is still developing and has significant implications for global carbon trading dynamics [160].

- The IO and NONF factors, in the economic category, have a strong impact on the price of carbon credits. The IO factor represents raw material prices that directly impact carbon credit prices because they affect production costs for industries that rely on carbon-intensive materials [1]. When raw material prices increase, the cost of production for these industries also increases, leading to higher carbon credit prices. Similarly, NONF represents product prices and also plays an important role in determining carbon credit prices. Industries that produce carbon-intensive products may pass the cost of carbon credits on to consumers, resulting in higher product prices [64]. This indicates that the price of the final product has a direct impact on the price of carbon credits.

8.7 Conclusion

In conclusion, this chapter has comprehensively solved the third research question by developing the CCMFP model for short-term carbon credit price forecasting. The CCMFP model used a two-stage hybrid approach: feature extraction and reduction using PCA, ICA, and NLICA in the first stage, followed by predictive modelling using SVR and MLP algorithms in the second stage. Through these steps, the issue of multicollinearity was effectively addressed, enhancing the quality of the input data and improving prediction accuracy. The chapter has detailed the methodology, experimental procedures, and evaluation of the results. Experiments were conducted and the results and performance were compared. This study identifies the ICA_SVR model with $K = 10$ factors and nine IC components as the best model for accurate and efficient short-term prediction of ACCU prices. In general, the results indicate that all the proposed models achieved a high prediction accuracy, exceeding 97.7%, as reflected in the presented comparative results. Thus, this chapter has validated the effectiveness of the CCMFP model and provided key insights into the prediction of carbon credit prices, setting the foundation for future research directions. The following chapter concludes the thesis, highlights the main contributions, and identifies specific areas for future research directions.

Chapter 9

Conclusion and Recommendations for Future Research

9.1 Overview

This chapter concludes the thesis by summarising the key findings and offering recommendations for future research directions. Predicting carbon credit prices is challenging due to the non-linear, non-stationary, and multicollinearity nature of the carbon market, influenced by multiple external factors from various domains, such as energy markets, economic indicators, and environmental factors. The influence of external factors can be considered separate channels of information, each contributing unique statistical insights that, when combined, provide a comprehensive view of the underlying carbon market behaviour. These characteristics make it difficult for traditional methods to capture the dynamic interactions and dependencies between factors. Accurate predictions of carbon credit prices are essential for guiding informed decision making in carbon markets, enabling better trading strategies and policy development. Therefore, a reliable prediction model is needed to efficiently and accurately predict short-term carbon credit prices.

To address these challenges, this thesis developed an intelligent, data-driven and accurate multi-factor prediction model to predict short-term carbon credit prices. The proposed model takes a multi-stage approach that combines key feature identification and optimised prediction algorithms to address the complex interplay of the 22 external factors identified and overcome the inherent complexities of carbon credit price data.

The next section of this chapter presents the thesis contributions aligned with the research objectives, followed by a discussion of the study limitations in Section 9.3. Section 9.4 then explores potential directions for future work.

9.2 Thesis contributions

Chapter 2 provides a comprehensive and in-depth review of the existing literature, focusing on the analysis and prediction of carbon credit prices. This review of the literature allowed the identification of research gaps, guiding the formulation of the research questions and objectives for this thesis. The objectives of this thesis are as follows:

1. Identify and analyse a comprehensive list of external factors that affect carbon credit prices.

2. Develop a model that leverages feature importance and feature engineering techniques to accurately and reliably analyse and select the most influential factors affecting carbon credit prices.

3. Develop a robust and intelligent multi-factor approach to predict short-term carbon credit prices, addressing non-linearity, non-stationary, and multicollinearity challenges.

4. Evaluate and test the methodologies developed in Objectives 2 and 3 to ensure their reliability and accuracy.

To achieve the research objectives, this thesis developed and evaluated various models for predicting short-term carbon credit prices. Of these, the most accurate and reliable model was selected as the optimal approach, leading to the development of the comprehensive and intelligent Carbon Credit Multi-Factor Prediction (CCMFP) model. The contributions of this thesis are summarised as follows:

Contribution 1: Comprehensive state-of-the-art review of the existing literature.

As detailed in this thesis, we conducted an extensive state-of-the-art systematic literature review (SLR) in the area of analysing and predicting the price of carbon credits. To the best of our knowledge, this SLR is the first attempt to systematically explore the methods and techniques in analysing and predicting carbon credits, focusing on models that consider external factors of influence. The primary focus of this review is to identify and evaluate relevant studies that propose prediction models, and address carbon credit price complexity, such as non-linearity, non-stationarity, and multicollinearity. This SLR is documented in Chapter 2 of this thesis.

The SLR was conducted through a series of structured steps to ensure a comprehensive and focused analysis. First, a well-defined search protocol was established, incorporating a predefined list of databases and search terms to systematically retrieve relevant studies. Next, inclusion and exclusion criteria were formulated to refine the scope of the review and prioritise relevant, high-quality research. The study selection process comprised a three-stage filtration approach involving the screening of titles, keywords, abstracts, and full texts to identify articles that met the criteria. Finally, a detailed data extraction process was performed to systematically capture and organise key information from the selected studies for further analysis. As it is difficult to ensure that all relevant studies are identified, backward citation was used in addition to the electronic data source search method. The same selection stages were applied to the articles found using the backward citation method. This process enabled us to collect and analyse state-of-the-art research articles published between 2005 and 2024, categorised into two areas: external factor analysis and price prediction models. Within price prediction models, further distinctions were made between statistical, regression-based, neural network-based, and hybrid methods,

creating a structured framework to understand current research approaches.

Furthermore, we critically analysed the 52 selected articles using 18 criteria to identify open issues and highlight areas where further research is necessary in the field of carbon credit price predictions. This approach addresses the complexities of carbon credit markets and the challenges associated with price prediction while providing insights into the strengths and weaknesses of various modelling techniques. The results of this SLR emphasise the need for more advanced models that address these challenges, providing a foundation for future research based on robust and comprehensive predictive models for carbon credit prices.

Contribution 2: Development of a comprehensive taxonomy of external influence factors.

This study makes a significant contribution by developing a comprehensive taxonomy of an extensive number of external factors that influence the price of carbon credits, representing the first systematic attempt to categorise these variables. To the best of our knowledge, no one has previously undertaken such an extensive effort to systematically classify and organise these factors. Through an in-depth analysis of the relevant literature, we identified and clarified the role of the key factors that affect carbon credit prices in various markets. This research integrates findings from multiple disciplines, revealing complex interconnections among factors and providing a holistic view of market dynamics. The resulting taxonomy organises the identified factors into nine categories: similar carbon markets, energy market prices, environmental, microeconomic, policy, social, economic, sustainable industry, and public awareness factors, providing a structured framework for future studies and enhancing the understanding of carbon credit price influences.

Contribution 3: Data collection of the external influence factors from reliable and authoritative online databases.

This thesis makes a significant contribution by collecting and compiling a dataset of 22 influencing factors. The collected data reflect seven key categories: similar factors, energy factors, environmental factors, macroeconomic factors, societal factors, economic factors, and sustainable industry factors. The dataset reflects Australia's unique market dynamics and conditions aligned with the case study ACCUS. The dataset represents the most up-to-date real-world records from January 2, 2020, to September 1, 2023, totalling 667 samples.

All data were obtained from reliable and authoritative databases. Data for international carbon markets were obtained from the World Bank. Energy market data, including natural gas, coal spot prices, crude oil, and electricity prices, were sourced from IndexMundi, the Australian Energy Regulator (AER), and the World Bank, ensuring coverage of energy-related factors critical to Australia. Environmental data, such as air temperature and the air quality index (AQI), were obtained from the Australian Bureau of Meteorology (BOM), while greenhouse gas emissions and renewable energy data were acquired from Macrotrends. Macroeconomic factors, including industrial development (business confidence indicator), stock market performance (S&P/ASX 200 Index), and the AUD/USD exchange rate, were sourced from OECD data, the Reserve Bank of Australia, and WSJ Markets, reflecting the economic conditions influencing ACCU prices. Societal factors, such as GDP and population growth, were collected from the World Bank. Economic variables, including iron ore prices and non-ferrous product prices, were retrieved from IndexMundi and the Australian Bureau of Statistics (ABS). Sustainable industry indicators, such as environmental technology patents, industrial emissions, and the use of renewable energy in industry, were sourced from the World Bank, Statista, and OECD data.

Contribution 4: Development of an intelligent and reliable factor selection model using machine learning and explainable AI algorithms to accurately identify and analyse the most influential factors affecting carbon credit prices.

Many researchers emphasise that selecting model input is an essential stage that significantly impacts model performance [46], [65], [66], [167]. Therefore, it is essential to understand how each external influencing factor affects carbon credit prices to determine the key factors. A valid feature selection method has been designed to identify the appropriate input variables for the carbon credit price prediction model. Importantly, this method considers both carbon price data and multiple influencing factors, which allows for the selection of more reasonable input variables and significantly enhances the ability to predict carbon credit prices accurately. This study identifies the key factors that influence carbon credit prices through a comprehensive multi-factor analysis based on three techniques.

In this thesis, we make a significant contribution by analysing the effect of 22 external factors on ACCU carbon credit prices using three methods: correlation analysis, MIC and the proposed CCMFI model. To the best of our knowledge, no one has previously used these methods to investigate the relationship between the 22 external factors identified and ACCUs in the Australian carbon market. We successfully developed the CCMFI model, which consists of three stages: analysis, ranking and selection of key factors. Using a comprehensive methodology that integrates RF regression and explainable AI SHAP values, the CCMFI model is shown to be a powerful tool for understanding and identifying the key drivers of carbon credit prices. The performance evaluation of the proposed CCMFI model resulted in an MSE of 1.44, which confirms its effectiveness. This is the first study to incorporate XAI using SHAP value techniques in carbon credit pricing research to select and reveal the impact of influencing factors, ensuring both accuracy and

transparency. The development process of the proposed models is given in detail in Chapter 7 of this thesis.

Contribution 5: Development of intelligent multi-factor prediction models to accurately predict the price of carbon credits.

This thesis makes a significant contribution by developing the CCMFP model to predict the short-term price of carbon credits. The proposed CCMFP model employs a hybrid approach by integrating feature extraction and reduction techniques, PCA, ICA and NLICA, with machine learning algorithms, SVR and MLP.

The proposed CCMFP model consists of two stages. In stage 1, we employed feature extraction and reduction techniques (PCA, ICA and NLICA) as a pre-processing step to improve the quality of input to prediction models by extracting essential features and mitigating the issue of multicollinearity within the set of independent variables influencing carbon credit prices [52]. We propose a systematic approach to determine the optimal number of ICs for ICA and NLICA, balancing predictive performance and computational efficiency. This optimisation ensures that the selected components provide the best possible input for the prediction models. The optimal number of IC components is systematically determined through a grid search with 10-fold cross-validation, ensuring the selection of the optimal number of components that minimises the MSE of prediction models.

Subsequently, in Stage 2, the extracted components are used as input for the SVR and LMP algorithms to build the prediction models. By identifying statistically independent components from the original variables, PCA, ICA and NLICA effectively reduce multicollinearity and generate components that are more suitable for machine learning applications [60]. This is one of the first studies in this domain to compare and integrate multiple feature extraction and reduction techniques, including PCA, ICA, and NLICA, to address the challenges of non-linearity, non-

stationarity, and multicollinearity in carbon credit price data. These techniques enhance feature representation, ensuring robust data analysis and forecasting.

The evaluation results show that all proposed models achieved high prediction accuracy, exceeding 97.7%, with the models using the ICA method performing better than the PCA and NLICA models in terms of both accuracy and time consumption. Specifically, the ICA_SVR model with $K = 10$ factors and nine IC components is identified as the best model for accurate and efficient short-term prediction of ACCUs prices in this study, followed by the ICA_MLP model with $K = 5$ factors and five IC components. The ICA_SVR model achieves a competitive MSE of 1.0393 and a high R^2 value of 0.9960 while maintaining a computational efficiency of 1.33 seconds. The ICA_MLP model achieved an MSE of 0.6770 and an R^2 value of 0.9959. Thus, this thesis has validated the effectiveness of the CCMFP model and provided key insights into the prediction of carbon credit prices. Chapter 8 provides a detailed explanation of the CCMFP framework and its implementation algorithms.

9.3 Limitations of the study

While this research provides significant insights into carbon credit price prediction, several limitations should be acknowledged. First, the study focuses on the Australian carbon market, and therefore, the findings may not fully capture the dynamics of other international markets. Second, the model uses mainly quantifiable external factors; qualitative influences such as sudden policy changes or public sentiment were outside the scope of this study.

9.4 Future research directions

- Develop a multi-horizon prediction model of the price of carbon credit to address both short- and long-term forecasting needs. By incorporating multiple time horizons, the models can better capture the varying influences of external factors over different periods, leading to more comprehensive and reliable predictions. This approach would support a variety of strategic goals, from immediate trading actions to long-term investment planning, making models more adaptable and practical for stakeholders in the dynamic carbon market.

- To evaluate the adaptability of the model, future research could test its performance on datasets from different carbon markets, such as the EU ETS and California Cap-and-Trade, as well as in diverse geographic regions. This would help establish the generalisability of the model and strengthen its application in a global context.

- Non-transactional factors, such as environmental policy changes and shifts in public awareness, significantly influence the price of carbon credit but are difficult to quantify and incorporate into predictive models. Future research could investigate approaches to capture and integrate these qualitative factors, such as using sentiment analysis, policy monitoring, or public opinion data, to improve the accuracy of the predictive model.

- Extend carbon credit price prediction models to include trading strategies. This involves developing simple rules for buying and selling carbon assets based on predicted price changes, considering real market conditions and transaction costs. Integrating these strategies would provide practical value, showing how investors and businesses can use predictions to manage risks and optimise returns.

These recommendations aim to enrich our understanding of carbon credit markets and contribute to the development of effective strategies for both policy design and investment decision-making. In addition, they offer opportunities for interdisciplinary research and collaboration as the field of carbon trading continues to evolve and adapt to new challenges and opportunities.

Appendix A

Appendix A: Tuning of Prediction Models

Table A.1 : ICA_SVR model hyperparameters tuning for different ICs using the top K = 5 external factors.

C	Epsilon	Gamma	Mean MSE	n_components
0.1	0.01	scale	35.27288666	1
0.1	0.01	auto	35.26116071	1
0.1	0.1	scale	35.30734972	1
0.1	0.1	auto	35.31176085	1
0.1	0.5	scale	35.26790954	1
0.1	0.5	auto	35.27071468	1
1	0.01	scale	22.85883997	1
1	0.01	auto	22.85014862	1
1	0.1	scale	22.87785202	1
1	0.1	auto	22.87382557	1
1	0.5	scale	22.85262952	1
1	0.5	auto	22.8425665	1
10	0.01	scale	20.49929465	1
10	0.01	auto	20.51877646	1
10	0.1	scale	20.49254449	1
10	0.1	auto	20.51238679	1
10	0.5	scale	20.26611597	1

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
10	0.5	auto	20.26578712	1
0.1	0.01	scale	38.49592096	2
0.1	0.01	auto	38.44768074	2
0.1	0.1	scale	38.45127197	2
0.1	0.1	auto	38.47682105	2
0.1	0.5	scale	37.6833192	2
0.1	0.5	auto	37.73921035	2
1	0.01	scale	8.154675067	2
1	0.01	auto	8.155139646	2
1	0.1	scale	8.135932163	2
1	0.1	auto	8.152255734	2
1	0.5	scale	8.191009151	2
1	0.5	auto	8.198202052	2
10	0.01	scale	2.726883186	2
10	0.01	auto	2.741554198	2
10	0.1	scale	2.72099442	2
10	0.1	auto	2.728619585	2
10	0.5	scale	2.834571055	2
10	0.5	auto	2.832360492	2
0.1	0.01	scale	53.64628196	3
0.1	0.01	auto	53.62233849	3
0.1	0.1	scale	53.59067399	3
0.1	0.1	auto	53.55815459	3
0.1	0.5	scale	52.34908891	3

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
0.1	0.5	auto	52.32998215	3
1	0.01	scale	1.821622115	3
1	0.01	auto	1.800564519	3
1	0.1	scale	1.853467963	3
1	0.1	auto	1.840196332	3
1	0.5	scale	2.002250733	3
1	0.5	auto	1.985569786	3
10	0.01	scale	0.746110547	3
10	0.01	auto	0.744114756	3
10	0.1	scale	0.741938871	3
10	0.1	auto	0.739545783	3
10	0.5	scale	0.77357409	3
10	0.5	auto	0.771959949	3
0.1	0.01	scale	57.38388768	4
0.1	0.01	auto	57.35959823	4
0.1	0.1	scale	57.59046399	4
0.1	0.1	auto	57.55884993	4
0.1	0.5	scale	55.80682697	4
0.1	0.5	auto	55.77903334	4
1	0.01	scale	2.521556141	4
1	0.01	auto	2.517645732	4
1	0.1	scale	2.511174568	4
1	0.1	auto	2.508207686	4
1	0.5	scale	2.596564297	4

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
1	0.5	auto	2.592843503	4
10	0.01	scale	1.2482814	4
10	0.01	auto	1.248420641	4
10	0.1	scale	1.234797946	4
10	0.1	auto	1.23437162	4
10	0.5	scale	1.238690621	4
10	0.5	auto	1.23938971	4
0.1	0.01	scale	64.78938959	5
0.1	0.01	auto	64.78528891	5
0.1	0.1	scale	64.63015513	5
0.1	0.1	auto	64.62386983	5
0.1	0.5	scale	64.185715	5
0.1	0.5	auto	64.15554956	5
1	0.01	scale	3.005150191	5
1	0.01	auto	3.002403937	5
1	0.1	scale	2.997624396	5
1	0.1	auto	2.995353385	5
1	0.5	scale	3.222479629	5
1	0.5	auto	3.223431586	5
10	0.01	scale	1.139940091	5
10	0.01	auto	1.139728705	5
10	0.1	scale	1.129250457	5
10	0.1	auto	1.128936683	5
10	0.5	scale	1.129075454	5

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
10	0.5	auto	1.128510199	5

Table A.2 : ICA_SVR model hyperparameters tuning for different ICs using the top K = 10 external factors.

C	Epsilon	Gamma	Mean MSE	n_components
0.1	0.01	scale	42.24079429	1
0.1	0.01	auto	42.25974318	1
0.1	0.1	scale	42.12821119	1
0.1	0.1	auto	42.1413488	1
0.1	0.5	scale	41.08066795	1
0.1	0.5	auto	41.08145518	1
1	0.01	scale	21.03680253	1
1	0.01	auto	20.96819867	1
1	0.1	scale	20.9509788	1
1	0.1	auto	20.8774293	1
1	0.5	scale	20.66418825	1
1	0.5	auto	20.6344078	1
10	0.01	scale	12.20113961	1
10	0.01	auto	12.20100363	1
10	0.1	scale	12.21015141	1
10	0.1	auto	12.2161889	1
10	0.5	scale	12.05653196	1
10	0.5	auto	12.08848712	1
0.1	0.01	scale	53.37494789	2

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
0.1	0.01	auto	53.35705382	2
0.1	0.1	scale	53.40200758	2
0.1	0.1	auto	53.40272338	2
0.1	0.5	scale	53.32386193	2
0.1	0.5	auto	53.32462387	2
1	0.01	scale	16.1748939	2
1	0.01	auto	16.11427733	2
1	0.1	scale	16.18734855	2
1	0.1	auto	16.12555742	2
1	0.5	scale	15.88642112	2
1	0.5	auto	15.86469566	2
10	0.01	scale	7.588306672	2
10	0.01	auto	7.579039849	2
10	0.1	scale	7.57080858	2
10	0.1	auto	7.553634516	2
10	0.5	scale	7.440282609	2
10	0.5	auto	7.416565794	2
0.1	0.01	scale	47.38969115	3
0.1	0.01	auto	47.39629541	3
0.1	0.1	scale	47.21693383	3
0.1	0.1	auto	47.22278704	3
0.1	0.5	scale	46.92223496	3
0.1	0.5	auto	46.91846008	3
1	0.01	scale	11.16637447	3

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
1	0.01	auto	11.16602005	3
1	0.1	scale	11.12893884	3
1	0.1	auto	11.13865209	3
1	0.5	scale	10.83631682	3
1	0.5	auto	10.83926647	3
10	0.01	scale	2.223238422	3
10	0.01	auto	2.219654826	3
10	0.1	scale	2.224430709	3
10	0.1	auto	2.224032705	3
10	0.5	scale	2.282008178	3
10	0.5	auto	2.287779809	3
0.1	0.01	scale	62.73645613	4
0.1	0.01	auto	62.71810969	4
0.1	0.1	scale	62.6616806	4
0.1	0.1	auto	62.65714074	4
0.1	0.5	scale	61.26600949	4
0.1	0.5	auto	61.2539393	4
1	0.01	scale	4.800966143	4
1	0.01	auto	4.797501233	4
1	0.1	scale	4.777664506	4
1	0.1	auto	4.773010868	4
1	0.5	scale	4.518105458	4
1	0.5	auto	4.512118457	4
10	0.01	scale	1.290899183	4

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
10	0.01	auto	1.291093375	4
10	0.1	scale	1.280573329	4
10	0.1	auto	1.280960412	4
10	0.5	scale	1.299615996	4
10	0.5	auto	1.29898472	4
0.1	0.01	scale	76.82179162	5
0.1	0.01	auto	76.79614916	5
0.1	0.1	scale	76.74619094	5
0.1	0.1	auto	76.71977362	5
0.1	0.5	scale	75.12648994	5
0.1	0.5	auto	75.08398051	5
1	0.01	scale	4.426405551	5
1	0.01	auto	4.390161753	5
1	0.1	scale	4.41192801	5
1	0.1	auto	4.378355889	5
1	0.5	scale	4.437790077	5
1	0.5	auto	4.409804675	5
10	0.01	scale	1.112566788	5
10	0.01	auto	1.111809309	5
10	0.1	scale	1.106203639	5
10	0.1	auto	1.105795509	5
10	0.5	scale	1.190297454	5
10	0.5	auto	1.188089007	5
0.1	0.01	scale	84.67024058	6

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
0.1	0.01	auto	84.65061228	6
0.1	0.1	scale	84.6418613	6
0.1	0.1	auto	84.62182308	6
0.1	0.5	scale	82.74108046	6
0.1	0.5	auto	82.74092745	6
1	0.01	scale	5.456190566	6
1	0.01	auto	5.424453297	6
1	0.1	scale	5.472925148	6
1	0.1	auto	5.441125175	6
1	0.5	scale	5.584979959	6
1	0.5	auto	5.547276242	6
10	0.01	scale	1.349113822	6
10	0.01	auto	1.34829065	6
10	0.1	scale	1.335341273	6
10	0.1	auto	1.333804183	6
10	0.5	scale	1.334346861	6
10	0.5	auto	1.33416412	6
0.1	0.01	scale	92.43848259	7
0.1	0.01	auto	92.43459602	7
0.1	0.1	scale	92.54024018	7
0.1	0.1	auto	92.53330363	7
0.1	0.5	scale	91.25039235	7
0.1	0.5	auto	91.24472716	7
1	0.01	scale	5.953147409	7

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
1	0.01	auto	5.943814462	7
1	0.1	scale	5.946314771	7
1	0.1	auto	5.935368115	7
1	0.5	scale	6.083390059	7
1	0.5	auto	6.075640344	7
10	0.01	scale	1.333743941	7
10	0.01	auto	1.335328854	7
10	0.1	scale	1.324575859	7
10	0.1	auto	1.3246216	7
10	0.5	scale	1.331959104	7
10	0.5	auto	1.332534147	7
0.1	0.01	scale	98.92587047	8
0.1	0.01	auto	98.91435462	8
0.1	0.1	scale	98.77438463	8
0.1	0.1	auto	98.76954689	8
0.1	0.5	scale	97.84454377	8
0.1	0.5	auto	97.84025994	8
1	0.01	scale	6.275201339	8
1	0.01	auto	6.258258105	8
1	0.1	scale	6.268172547	8
1	0.1	auto	6.250751724	8
1	0.5	scale	6.251301162	8
1	0.5	auto	6.239850808	8
10	0.01	scale	1.234905608	8

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
10	0.01	auto	1.232737717	8
10	0.1	scale	1.233372447	8
10	0.1	auto	1.23118787	8
10	0.5	scale	1.270665483	8
10	0.5	auto	1.269092155	8
0.1	0.01	scale	105.6786699	9
0.1	0.01	auto	105.668249	9
0.1	0.1	scale	105.4562624	9
0.1	0.1	auto	105.4305709	9
0.1	0.5	scale	103.8126839	9
0.1	0.5	auto	103.8045972	9
1	0.01	scale	7.696836848	9
1	0.01	auto	7.672380683	9
1	0.1	scale	7.67743868	9
1	0.1	auto	7.653117993	9
1	0.5	scale	7.665868339	9
1	0.5	auto	7.642398913	9
10	0.01	scale	1.044945017	9
10	0.01	auto	1.043328319	9
10	0.1	scale	1.040477328	9
10	0.1	auto	1.039252849	9
10	0.5	scale	1.088100764	9
10	0.5	auto	1.086364733	9
0.1	0.01	scale	109.2438841	10

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
0.1	0.01	auto	109.2383605	10
0.1	0.1	scale	108.9983685	10
0.1	0.1	auto	108.9928833	10
0.1	0.5	scale	107.3888533	10
0.1	0.5	auto	107.3889194	10
1	0.01	scale	9.451422717	10
1	0.01	auto	9.437703973	10
1	0.1	scale	9.419885469	10
1	0.1	auto	9.405787699	10
1	0.5	scale	9.590064554	10
1	0.5	auto	9.577683399	10
10	0.01	scale	1.094291768	10
10	0.01	auto	1.09314427	10
10	0.1	scale	1.072985136	10
10	0.1	auto	1.071908213	10
10	0.5	scale	1.089153592	10
10	0.5	auto	1.088071589	10

Table A.3 : ICA_SVR model hyperparameters tuning for different ICs using the top K = 15 external factors.

C	Epsilon	Gamma	Mean MSE	n_components
0.1	0.01	scale	45.16893693	1
0.1	0.01	auto	45.17061075	1
0.1	0.1	scale	45.3220115	1

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
0.1	0.1	auto	45.32546955	1
0.1	0.5	scale	44.77340166	1
0.1	0.5	auto	44.76752432	1
1	0.01	scale	28.90048629	1
1	0.01	auto	28.90741096	1
1	0.1	scale	28.8214479	1
1	0.1	auto	28.8284988	1
1	0.5	scale	28.75240152	1
1	0.5	auto	28.76147889	1
10	0.01	scale	22.01212897	1
10	0.01	auto	22.02851066	1
10	0.1	scale	21.91619513	1
10	0.1	auto	21.93854389	1
10	0.5	scale	21.79055106	1
10	0.5	auto	21.78785985	1
0.1	0.01	scale	55.5785587	2
0.1	0.01	auto	55.57578233	2
0.1	0.1	scale	55.65943703	2
0.1	0.1	auto	55.64256388	2
0.1	0.5	scale	55.64899932	2
0.1	0.5	auto	55.63417778	2
1	0.01	scale	19.79150109	2
1	0.01	auto	19.77351938	2
1	0.1	scale	19.52713582	2

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
1	0.1	auto	19.50154979	2
1	0.5	scale	19.49789937	2
1	0.5	auto	19.5196032	2
10	0.01	scale	8.028778236	2
10	0.01	auto	8.025358258	2
10	0.1	scale	8.029841223	2
10	0.1	auto	8.025767581	2
10	0.5	scale	8.057479006	2
10	0.5	auto	8.063478073	2
0.1	0.01	scale	53.81688429	3
0.1	0.01	auto	53.82697023	3
0.1	0.1	scale	53.89955379	3
0.1	0.1	auto	53.90063669	3
0.1	0.5	scale	52.56533742	3
0.1	0.5	auto	52.56313961	3
1	0.01	scale	10.91783865	3
1	0.01	auto	10.92261143	3
1	0.1	scale	10.83772436	3
1	0.1	auto	10.84396182	3
1	0.5	scale	10.76465231	3
1	0.5	auto	10.76039073	3
10	0.01	scale	2.379841835	3
10	0.01	auto	2.378606537	3
10	0.1	scale	2.391995846	3

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
10	0.1	auto	2.390451892	3
10	0.5	scale	2.367659521	3
10	0.5	auto	2.356917583	3
0.1	0.01	scale	53.80820776	4
0.1	0.01	auto	53.79514339	4
0.1	0.1	scale	53.82233441	4
0.1	0.1	auto	53.80037265	4
0.1	0.5	scale	53.24413874	4
0.1	0.5	auto	53.24302171	4
1	0.01	scale	12.10552764	4
1	0.01	auto	12.09206752	4
1	0.1	scale	12.09221412	4
1	0.1	auto	12.0899327	4
1	0.5	scale	11.80025199	4
1	0.5	auto	11.78806908	4
10	0.01	scale	1.793601155	4
10	0.01	auto	1.792753174	4
10	0.1	scale	1.800992192	4
10	0.1	auto	1.80689127	4
10	0.5	scale	1.746253801	4
10	0.5	auto	1.743458405	4
0.1	0.01	scale	74.39570075	5
0.1	0.01	auto	74.38413889	5
0.1	0.1	scale	74.15621375	5

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
0.1	0.1	auto	74.14671722	5
0.1	0.5	scale	72.10445053	5
0.1	0.5	auto	72.10646627	5
1	0.01	scale	8.867341025	5
1	0.01	auto	8.875922342	5
1	0.1	scale	8.799205731	5
1	0.1	auto	8.799263395	5
1	0.5	scale	8.7325913	5
1	0.5	auto	8.746392484	5
10	0.01	scale	1.674403845	5
10	0.01	auto	1.669751793	5
10	0.1	scale	1.689110534	5
10	0.1	auto	1.684265984	5
10	0.5	scale	1.657189719	5
10	0.5	auto	1.656165006	5
0.1	0.01	scale	87.51297833	6
0.1	0.01	auto	87.50490201	6
0.1	0.1	scale	87.24837767	6
0.1	0.1	auto	87.24461726	6
0.1	0.5	scale	85.40670847	6
0.1	0.5	auto	85.39996754	6
1	0.01	scale	8.61629574	6
1	0.01	auto	8.612126903	6
1	0.1	scale	8.546902388	6

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
1	0.1	auto	8.53911441	6
1	0.5	scale	8.39631823	6
1	0.5	auto	8.394640346	6
10	0.01	scale	1.189363126	6
10	0.01	auto	1.188083698	6
10	0.1	scale	1.172139612	6
10	0.1	auto	1.171136567	6
10	0.5	scale	1.177534287	6
10	0.5	auto	1.177005234	6
0.1	0.01	scale	94.30375051	7
0.1	0.01	auto	94.30567928	7
0.1	0.1	scale	94.22350954	7
0.1	0.1	auto	94.22668267	7
0.1	0.5	scale	93.16733823	7
0.1	0.5	auto	93.1662408	7
1	0.01	scale	6.266173039	7
1	0.01	auto	6.266339709	7
1	0.1	scale	6.221905627	7
1	0.1	auto	6.221779854	7
1	0.5	scale	5.990719983	7
1	0.5	auto	5.987844347	7
10	0.01	scale	1.096331241	7
10	0.01	auto	1.09582532	7
10	0.1	scale	1.108473541	7

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
10	0.1	auto	1.108111681	7
10	0.5	scale	1.141573868	7
10	0.5	auto	1.140948028	7
0.1	0.01	scale	99.29460481	8
0.1	0.01	auto	99.28741117	8
0.1	0.1	scale	99.2869254	8
0.1	0.1	auto	99.28561951	8
0.1	0.5	scale	97.93536645	8
0.1	0.5	auto	97.93148086	8
1	0.01	scale	6.065544195	8
1	0.01	auto	6.053203738	8
1	0.1	scale	5.938825513	8
1	0.1	auto	5.926337234	8
1	0.5	scale	5.522244	8
1	0.5	auto	5.505675253	8
10	0.01	scale	1.365514806	8
10	0.01	auto	1.362864543	8
10	0.1	scale	1.366709434	8
10	0.1	auto	1.365321634	8
10	0.5	scale	1.367526576	8
10	0.5	auto	1.365176385	8
0.1	0.01	scale	105.0288802	9
0.1	0.01	auto	105.0264724	9
0.1	0.1	scale	104.9212073	9

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
0.1	0.1	auto	104.9311874	9
0.1	0.5	scale	103.934938	9
0.1	0.5	auto	103.9369807	9
1	0.01	scale	8.457807333	9
1	0.01	auto	8.421270599	9
1	0.1	scale	8.451058125	9
1	0.1	auto	8.410993096	9
1	0.5	scale	8.353551829	9
1	0.5	auto	8.319855638	9
10	0.01	scale	1.324393326	9
10	0.01	auto	1.319694037	9
10	0.1	scale	1.318955459	9
10	0.1	auto	1.314716693	9
10	0.5	scale	1.346027709	9
10	0.5	auto	1.342342914	9
0.1	0.01	scale	108.0767047	10
0.1	0.01	auto	108.0770618	10
0.1	0.1	scale	108.0171237	10
0.1	0.1	auto	108.0134451	10
0.1	0.5	scale	107.0701246	10
0.1	0.5	auto	107.0604152	10
1	0.01	scale	8.895709592	10
1	0.01	auto	8.88735598	10
1	0.1	scale	8.875653311	10

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
1	0.1	auto	8.867754151	10
1	0.5	scale	8.732937189	10
1	0.5	auto	8.725965967	10
10	0.01	scale	1.333705412	10
10	0.01	auto	1.33311003	10
10	0.1	scale	1.320033837	10
10	0.1	auto	1.319855614	10
10	0.5	scale	1.333653715	10
10	0.5	auto	1.333582294	10
0.1	0.01	scale	112.0237203	11
0.1	0.01	auto	112.0228615	11
0.1	0.1	scale	111.927005	11
0.1	0.1	auto	111.9243572	11
0.1	0.5	scale	110.4489248	11
0.1	0.5	auto	110.442222	11
1	0.01	scale	9.719448986	11
1	0.01	auto	9.713648693	11
1	0.1	scale	9.662049624	11
1	0.1	auto	9.656493996	11
1	0.5	scale	9.78420439	11
1	0.5	auto	9.780507136	11
10	0.01	scale	1.243125585	11
10	0.01	auto	1.242794285	11
10	0.1	scale	1.226912076	11

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
10	0.1	auto	1.2268964	11
10	0.5	scale	1.264491056	11
10	0.5	auto	1.264635288	11
0.1	0.01	scale	116.251927	12
0.1	0.01	auto	116.2497901	12
0.1	0.1	scale	116.1351961	12
0.1	0.1	auto	116.1349127	12
0.1	0.5	scale	113.9082798	12
0.1	0.5	auto	113.9080819	12
1	0.01	scale	10.28818087	12
1	0.01	auto	10.28880584	12
1	0.1	scale	10.35906794	12
1	0.1	auto	10.36029011	12
1	0.5	scale	10.6162964	12
1	0.5	auto	10.61781451	12
10	0.01	scale	1.266085146	12
10	0.01	auto	1.267139341	12
10	0.1	scale	1.250436496	12
10	0.1	auto	1.251217094	12
10	0.5	scale	1.292262907	12
10	0.5	auto	1.292594389	12
0.1	0.01	scale	120.3488594	13
0.1	0.01	auto	120.3485592	13
0.1	0.1	scale	120.0615817	13

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
0.1	0.1	auto	120.0607179	13
0.1	0.5	scale	117.2812752	13
0.1	0.5	auto	117.2790257	13
1	0.01	scale	12.79060191	13
1	0.01	auto	12.78604999	13
1	0.1	scale	12.78681107	13
1	0.1	auto	12.78147912	13
1	0.5	scale	12.99983126	13
1	0.5	auto	12.99488865	13
10	0.01	scale	1.066786745	13
10	0.01	auto	1.065510134	13
10	0.1	scale	1.045377504	13
10	0.1	auto	1.043859621	13
10	0.5	scale	1.096258113	13
10	0.5	auto	1.095002233	13
0.1	0.01	scale	122.2138274	14
0.1	0.01	auto	122.2138513	14
0.1	0.1	scale	121.9858234	14
0.1	0.1	auto	121.9832925	14
0.1	0.5	scale	119.0066554	14
0.1	0.5	auto	119.006036	14
1	0.01	scale	15.60167924	14
1	0.01	auto	15.58270626	14
1	0.1	scale	15.54253298	14

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
1	0.1	auto	15.52498415	14
1	0.5	scale	15.70397968	14
1	0.5	auto	15.68874993	14
10	0.01	scale	1.218680299	14
10	0.01	auto	1.216164898	14
10	0.1	scale	1.185231353	14
10	0.1	auto	1.182599215	14
10	0.5	scale	1.234482063	14
10	0.5	auto	1.231955136	14
0.1	0.01	scale	124.331521	15
0.1	0.01	auto	124.3274641	15
0.1	0.1	scale	124.1359378	15
0.1	0.1	auto	124.1320338	15
0.1	0.5	scale	120.6802765	15
0.1	0.5	auto	120.6764962	15
1	0.01	scale	21.3173401	15
1	0.01	auto	21.29819252	15
1	0.1	scale	21.2734442	15
1	0.1	auto	21.25477861	15
1	0.5	scale	21.34119162	15
1	0.5	auto	21.32349666	15
10	0.01	scale	1.333597644	15
10	0.01	auto	1.329433892	15
10	0.1	scale	1.316026315	15

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
10	0.1	auto	1.311229231	15
10	0.5	scale	1.33898282	15
10	0.5	auto	1.334323586	15

Table A.4 : ICA_SVR model hyperparameters tuning for different ICs using all K = 22 external factors.

C	Epsilon	Gamma	Mean MSE	n_components
0.1	0.01	scale	45.84052588	1
0.1	0.01	auto	45.84130497	1
0.1	0.1	scale	45.87407295	1
0.1	0.1	auto	45.83398213	1
0.1	0.5	scale	45.86620013	1
0.1	0.5	auto	45.81132628	1
1	0.01	scale	31.46781525	1
1	0.01	auto	31.49233498	1
1	0.1	scale	31.46392078	1
1	0.1	auto	31.42655868	1
1	0.5	scale	31.22706237	1
1	0.5	auto	31.27254732	1
10	0.01	scale	28.13273415	1
10	0.01	auto	28.18036176	1
10	0.1	scale	28.22961709	1
10	0.1	auto	28.23204516	1
10	0.5	scale	27.9378299	1

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
10	0.5	auto	27.93366231	1
0.1	0.01	scale	57.25133693	2
0.1	0.01	auto	57.26342706	2
0.1	0.1	scale	57.13386124	2
0.1	0.1	auto	57.1170674	2
0.1	0.5	scale	57.08330677	2
0.1	0.5	auto	57.04228264	2
1	0.01	scale	20.82874215	2
1	0.01	auto	20.81239548	2
1	0.1	scale	20.75773296	2
1	0.1	auto	20.75440496	2
1	0.5	scale	20.4732681	2
1	0.5	auto	20.48734094	2
10	0.01	scale	9.320309807	2
10	0.01	auto	9.307966057	2
10	0.1	scale	9.254404031	2
10	0.1	auto	9.266763976	2
10	0.5	scale	9.205360078	2
10	0.5	auto	9.216688644	2
0.1	0.01	scale	54.71858749	3
0.1	0.01	auto	54.72509343	3
0.1	0.1	scale	54.72785837	3
0.1	0.1	auto	54.7465999	3
0.1	0.5	scale	54.45524827	3

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
0.1	0.5	auto	54.47277003	3
1	0.01	scale	17.99744361	3
1	0.01	auto	18.03813967	3
1	0.1	scale	17.92597433	3
1	0.1	auto	17.9733106	3
1	0.5	scale	17.74334694	3
1	0.5	auto	17.74519788	3
10	0.01	scale	6.272981831	3
10	0.01	auto	6.316235433	3
10	0.1	scale	6.248684847	3
10	0.1	auto	6.270112953	3
10	0.5	scale	6.114733263	3
10	0.5	auto	6.11266099	3
0.1	0.01	scale	55.6911023	4
0.1	0.01	auto	55.67657894	4
0.1	0.1	scale	55.75452688	4
0.1	0.1	auto	55.74477617	4
0.1	0.5	scale	55.44598749	4
0.1	0.5	auto	55.44311171	4
1	0.01	scale	16.60173663	4
1	0.01	auto	16.60206093	4
1	0.1	scale	16.68723998	4
1	0.1	auto	16.68180317	4
1	0.5	scale	16.55336694	4

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
1	0.5	auto	16.57649425	4
10	0.01	scale	4.854068157	4
10	0.01	auto	4.856661547	4
10	0.1	scale	4.879480979	4
10	0.1	auto	4.88024661	4
10	0.5	scale	4.982910065	4
10	0.5	auto	5.005129219	4
0.1	0.01	scale	74.47407577	5
0.1	0.01	auto	74.47761974	5
0.1	0.1	scale	74.43769305	5
0.1	0.1	auto	74.44001953	5
0.1	0.5	scale	72.2402676	5
0.1	0.5	auto	72.24289014	5
1	0.01	scale	9.913943869	5
1	0.01	auto	9.914760883	5
1	0.1	scale	9.878610953	5
1	0.1	auto	9.878962798	5
1	0.5	scale	9.642834683	5
1	0.5	auto	9.635147064	5
10	0.01	scale	2.615348923	5
10	0.01	auto	2.612408464	5
10	0.1	scale	2.619805229	5
10	0.1	auto	2.6123093	5
10	0.5	scale	2.714443613	5

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
10	0.5	auto	2.711754341	5
0.1	0.01	scale	84.18815386	6
0.1	0.01	auto	84.19873056	6
0.1	0.1	scale	84.06187365	6
0.1	0.1	auto	84.0702068	6
0.1	0.5	scale	82.27492487	6
0.1	0.5	auto	82.27064378	6
1	0.01	scale	10.65886566	6
1	0.01	auto	10.66777679	6
1	0.1	scale	10.60407689	6
1	0.1	auto	10.61086631	6
1	0.5	scale	10.33042188	6
1	0.5	auto	10.33580642	6
10	0.01	scale	2.064780663	6
10	0.01	auto	2.0637163	6
10	0.1	scale	2.055798224	6
10	0.1	auto	2.055403509	6
10	0.5	scale	2.114746883	6
10	0.5	auto	2.114374605	6
0.1	0.01	scale	93.13871702	7
0.1	0.01	auto	93.13603496	7
0.1	0.1	scale	93.03148935	7
0.1	0.1	auto	93.04005745	7
0.1	0.5	scale	91.82930947	7

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
0.1	0.5	auto	91.82973	7
1	0.01	scale	13.31247504	7
1	0.01	auto	13.3189099	7
1	0.1	scale	13.32059541	7
1	0.1	auto	13.32710923	7
1	0.5	scale	12.70066727	7
1	0.5	auto	12.7052706	7
10	0.01	scale	2.24510026	7
10	0.01	auto	2.245015478	7
10	0.1	scale	2.221811719	7
10	0.1	auto	2.221974188	7
10	0.5	scale	2.168356693	7
10	0.5	auto	2.16755465	7
0.1	0.01	scale	99.81244713	8
0.1	0.01	auto	99.80836357	8
0.1	0.1	scale	99.7979538	8
0.1	0.1	auto	99.81100155	8
0.1	0.5	scale	98.69279686	8
0.1	0.5	auto	98.69107004	8
1	0.01	scale	10.17937869	8
1	0.01	auto	10.18233385	8
1	0.1	scale	10.17401464	8
1	0.1	auto	10.1763042	8
1	0.5	scale	9.645221791	8

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
1	0.5	auto	9.644884918	8
10	0.01	scale	1.478062508	8
10	0.01	auto	1.476084068	8
10	0.1	scale	1.460325128	8
10	0.1	auto	1.459353481	8
10	0.5	scale	1.506214339	8
10	0.5	auto	1.505021915	8
0.1	0.01	scale	108.2627371	9
0.1	0.01	auto	108.2564243	9
0.1	0.1	scale	108.2779104	9
0.1	0.1	auto	108.2784475	9
0.1	0.5	scale	107.2580843	9
0.1	0.5	auto	107.2570779	9
1	0.01	scale	11.64123879	9
1	0.01	auto	11.64414692	9
1	0.1	scale	11.64871088	9
1	0.1	auto	11.6516129	9
1	0.5	scale	11.40709125	9
1	0.5	auto	11.41051541	9
10	0.01	scale	1.263180997	9
10	0.01	auto	1.237707097	9
10	0.1	scale	1.262729975	9
10	0.1	auto	1.23752345	9
10	0.5	scale	1.313118742	9

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
10	0.5	auto	1.313017696	9
0.1	0.01	scale	113.4911394	10
0.1	0.01	auto	113.4887716	10
0.1	0.1	scale	113.2917271	10
0.1	0.1	auto	113.2891026	10
0.1	0.5	scale	112.08999	10
0.1	0.5	auto	112.086577	10
1	0.01	scale	14.72655983	10
1	0.01	auto	14.72461835	10
1	0.1	scale	14.71788555	10
1	0.1	auto	14.71598679	10
1	0.5	scale	14.36463956	10
1	0.5	auto	14.36388237	10
10	0.01	scale	1.478588026	10
10	0.01	auto	1.477445531	10
10	0.1	scale	1.48942411	10
10	0.1	auto	1.487592087	10
10	0.5	scale	1.575430687	10
10	0.5	auto	1.573100934	10
0.1	0.01	scale	117.2938565	11
0.1	0.01	auto	117.2920031	11
0.1	0.1	scale	117.1957609	11
0.1	0.1	auto	117.1912597	11
0.1	0.5	scale	116.0100855	11

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
0.1	0.5	auto	116.004803	11
1	0.01	scale	14.48027372	11
1	0.01	auto	14.47818061	11
1	0.1	scale	14.43009654	11
1	0.1	auto	14.42267088	11
1	0.5	scale	14.10211981	11
1	0.5	auto	14.09250325	11
10	0.01	scale	1.55769328	11
10	0.01	auto	1.553728396	11
10	0.1	scale	1.536231313	11
10	0.1	auto	1.531556158	11
10	0.5	scale	1.565145034	11
10	0.5	auto	1.562767485	11
0.1	0.01	scale	120.4075943	12
0.1	0.01	auto	120.4064517	12
0.1	0.1	scale	120.5213241	12
0.1	0.1	auto	120.5213616	12
0.1	0.5	scale	118.9209087	12
0.1	0.5	auto	118.9180534	12
1	0.01	scale	17.4630386	12
1	0.01	auto	17.45768111	12
1	0.1	scale	17.51868952	12
1	0.1	auto	17.51428706	12
1	0.5	scale	17.40698017	12

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
1	0.5	auto	17.40052718	12
10	0.01	scale	1.941224562	12
10	0.01	auto	1.935755366	12
10	0.1	scale	1.89588719	12
10	0.1	auto	1.889660855	12
10	0.5	scale	1.879347153	12
10	0.5	auto	1.876660627	12
0.1	0.01	scale	122.4099537	13
0.1	0.01	auto	122.4084866	13
0.1	0.1	scale	122.4515096	13
0.1	0.1	auto	122.4520117	13
0.1	0.5	scale	120.779753	13
0.1	0.5	auto	120.7740207	13
1	0.01	scale	20.29369766	13
1	0.01	auto	20.27211995	13
1	0.1	scale	20.24989892	13
1	0.1	auto	20.23416785	13
1	0.5	scale	20.02961836	13
1	0.5	auto	20.01073196	13
10	0.01	scale	2.172224707	13
10	0.01	auto	2.169274522	13
10	0.1	scale	2.142855374	13
10	0.1	auto	2.14039824	13
10	0.5	scale	2.111336523	13

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
10	0.5	auto	2.110550809	13
0.1	0.01	scale	126.0566433	14
0.1	0.01	auto	126.0567102	14
0.1	0.1	scale	125.8189933	14
0.1	0.1	auto	125.8171374	14
0.1	0.5	scale	123.7707914	14
0.1	0.5	auto	123.7674957	14
1	0.01	scale	21.65815561	14
1	0.01	auto	21.64414156	14
1	0.1	scale	21.64931721	14
1	0.1	auto	21.63781236	14
1	0.5	scale	21.45320622	14
1	0.5	auto	21.4390671	14
10	0.01	scale	2.130319826	14
10	0.01	auto	2.129565427	14
10	0.1	scale	2.123746317	14
10	0.1	auto	2.123493103	14
10	0.5	scale	2.158959446	14
10	0.5	auto	2.158195942	14
0.1	0.01	scale	128.7210792	15
0.1	0.01	auto	128.7196243	15
0.1	0.1	scale	128.3363949	15
0.1	0.1	auto	128.3326561	15
0.1	0.5	scale	125.6930919	15

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
0.1	0.5	auto	125.6888097	15
1	0.01	scale	22.65673492	15
1	0.01	auto	22.64186097	15
1	0.1	scale	22.70079824	15
1	0.1	auto	22.68252262	15
1	0.5	scale	22.53407955	15
1	0.5	auto	22.51798525	15
10	0.01	scale	2.207802596	15
10	0.01	auto	2.199487751	15
10	0.1	scale	2.175837241	15
10	0.1	auto	2.167173426	15
10	0.5	scale	2.199281635	15
10	0.5	auto	2.192392988	15
0.1	0.01	scale	129.6085294	16
0.1	0.01	auto	129.6035948	16
0.1	0.1	scale	129.2135937	16
0.1	0.1	auto	129.2061993	16
0.1	0.5	scale	126.5006645	16
0.1	0.5	auto	126.4873314	16
1	0.01	scale	22.92179046	16
1	0.01	auto	22.8652032	16
1	0.1	scale	22.93110764	16
1	0.1	auto	22.87745795	16
1	0.5	scale	22.77353105	16

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
1	0.5	auto	22.72981235	16
10	0.01	scale	2.234598389	16
10	0.01	auto	2.22344222	16
10	0.1	scale	2.205780916	16
10	0.1	auto	2.194511163	16
10	0.5	scale	2.214486477	16
10	0.5	auto	2.208813161	16
0.1	0.01	scale	131.1897392	17
0.1	0.01	auto	131.1887779	17
0.1	0.1	scale	130.9407895	17
0.1	0.1	auto	130.9291166	17
0.1	0.5	scale	127.9915251	17
0.1	0.5	auto	127.9822285	17
1	0.01	scale	24.52368788	17
1	0.01	auto	24.44748868	17
1	0.1	scale	24.54956377	17
1	0.1	auto	24.47415159	17
1	0.5	scale	24.30210938	17
1	0.5	auto	24.2443095	17
10	0.01	scale	2.222725045	17
10	0.01	auto	2.213947311	17
10	0.1	scale	2.196895798	17
10	0.1	auto	2.188322953	17
10	0.5	scale	2.21800811	17

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
10	0.5	auto	2.20463527	17
0.1	0.01	scale	133.3838495	18
0.1	0.01	auto	133.3724424	18
0.1	0.1	scale	132.9088724	18
0.1	0.1	auto	132.9073545	18
0.1	0.5	scale	129.8096368	18
0.1	0.5	auto	129.7906441	18
1	0.01	scale	26.83799914	18
1	0.01	auto	26.74855041	18
1	0.1	scale	26.88276328	18
1	0.1	auto	26.79871704	18
1	0.5	scale	26.71033929	18
1	0.5	auto	26.63814621	18
10	0.01	scale	2.051481206	18
10	0.01	auto	2.034699497	18
10	0.1	scale	2.043908625	18
10	0.1	auto	2.027581648	18
10	0.5	scale	2.091017459	18
10	0.5	auto	2.070088852	18
0.1	0.01	scale	134.6072925	19
0.1	0.01	auto	134.598756	19
0.1	0.1	scale	134.1919764	19
0.1	0.1	auto	134.1801432	19
0.1	0.5	scale	130.8300419	19

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
0.1	0.5	auto	130.8190216	19
1	0.01	scale	29.0971849	19
1	0.01	auto	28.99859576	19
1	0.1	scale	29.04346224	19
1	0.1	auto	28.95372573	19
1	0.5	scale	28.83641912	19
1	0.5	auto	28.75847215	19
10	0.01	scale	2.110823978	19
10	0.01	auto	2.090401317	19
10	0.1	scale	2.099033889	19
10	0.1	auto	2.078135355	19
10	0.5	scale	2.185353151	19
10	0.5	auto	2.161255849	19
0.1	0.01	scale	135.6671868	20
0.1	0.01	auto	135.6601347	20
0.1	0.1	scale	135.1961399	20
0.1	0.1	auto	135.1821249	20
0.1	0.5	scale	131.88902	20
0.1	0.5	auto	131.8761983	20
1	0.01	scale	33.20947605	20
1	0.01	auto	33.1275298	20
1	0.1	scale	33.12751115	20
1	0.1	auto	33.04961593	20
1	0.5	scale	32.87285843	20

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
1	0.5	auto	32.8046008	20
10	0.01	scale	2.259810221	20
10	0.01	auto	2.237109679	20
10	0.1	scale	2.241043225	20
10	0.1	auto	2.216929932	20
10	0.5	scale	2.435430912	20
10	0.5	auto	2.409595748	20
0.1	0.01	scale	137.9355328	21
0.1	0.01	auto	137.9258401	21
0.1	0.1	scale	137.6138404	21
0.1	0.1	auto	137.6001169	21
0.1	0.5	scale	134.0852711	21
0.1	0.5	auto	134.0753492	21
1	0.01	scale	34.28311277	21
1	0.01	auto	34.21883384	21
1	0.1	scale	34.21023759	21
1	0.1	auto	34.14595458	21
1	0.5	scale	33.85076435	21
1	0.5	auto	33.77934463	21
10	0.01	scale	2.198632734	21
10	0.01	auto	2.177740461	21
10	0.1	scale	2.183546262	21
10	0.1	auto	2.161054296	21
10	0.5	scale	2.367377156	21

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
10	0.5	auto	2.341347957	21
0.1	0.01	scale	138.8954525	22
0.1	0.01	auto	138.8842214	22
0.1	0.1	scale	138.6826321	22
0.1	0.1	auto	138.6683617	22
0.1	0.5	scale	135.3523198	22
0.1	0.5	auto	135.3387449	22
1	0.01	scale	35.32996895	22
1	0.01	auto	35.26665305	22
1	0.1	scale	35.23895811	22
1	0.1	auto	35.17859236	22
1	0.5	scale	34.75079041	22
1	0.5	auto	34.69085476	22
10	0.01	scale	2.16525411	22
10	0.01	auto	2.147245527	22
10	0.1	scale	2.151281596	22
10	0.1	auto	2.13153856	22
10	0.5	scale	2.33107027	22
10	0.5	auto	2.306498567	22

Table A.5 : NLICA_SVR model hyperparameters tuning for different ICs using the top K = 5 external factors.

C	Epsilon	Gamma	Mean MSE	n_components
0.1	0.01	scale	161.4289577	1

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
0.1	0.01	auto	161.4289577	1
0.1	0.1	scale	161.5702249	1
0.1	0.1	auto	161.5702249	1
0.1	0.5	scale	158.3788166	1
0.1	0.5	auto	158.3788166	1
1	0.01	scale	161.4289577	1
1	0.01	auto	161.4289577	1
1	0.1	scale	161.5702249	1
1	0.1	auto	161.5702249	1
1	0.5	scale	158.3788166	1
1	0.5	auto	158.3788166	1
10	0.01	scale	161.4289577	1
10	0.01	auto	161.4289577	1
10	0.1	scale	161.5702249	1
10	0.1	auto	161.5702249	1
10	0.5	scale	158.3788166	1
10	0.5	auto	158.3788166	1
0.1	0.01	scale	36.81098194	2
0.1	0.01	auto	37.92010956	2
0.1	0.1	scale	36.63094882	2
0.1	0.1	auto	38.13225281	2
0.1	0.5	scale	36.71815733	2
0.1	0.5	auto	38.46001395	2
1	0.01	scale	21.84449261	2

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
1	0.01	auto	22.25357241	2
1	0.1	scale	21.84067998	2
1	0.1	auto	22.3064157	2
1	0.5	scale	22.0156895	2
1	0.5	auto	22.31989783	2
10	0.01	scale	20.94887862	2
10	0.01	auto	21.03895837	2
10	0.1	scale	21.00920415	2
10	0.1	auto	21.11336286	2
10	0.5	scale	21.09378071	2
10	0.5	auto	21.21130869	2
0.1	0.01	scale	38.89993776	3
0.1	0.01	auto	40.26922495	3
0.1	0.1	scale	38.69728188	3
0.1	0.1	auto	40.20622795	3
0.1	0.5	scale	38.64585635	3
0.1	0.5	auto	40.76986933	3
1	0.01	scale	24.38596934	3
1	0.01	auto	26.59708392	3
1	0.1	scale	24.46210926	3
1	0.1	auto	26.58088278	3
1	0.5	scale	24.59165441	3
1	0.5	auto	26.86284935	3
10	0.01	scale	21.37632737	3

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
10	0.01	auto	22.37158407	3
10	0.1	scale	21.33397351	3
10	0.1	auto	22.35602358	3
10	0.5	scale	21.56279109	3
10	0.5	auto	22.66517178	3
0.1	0.01	scale	40.93851265	4
0.1	0.01	auto	43.26681906	4
0.1	0.1	scale	40.97897121	4
0.1	0.1	auto	43.06916342	4
0.1	0.5	scale	40.06863512	4
0.1	0.5	auto	41.82153752	4
1	0.01	scale	17.85434674	4
1	0.01	auto	22.24496089	4
1	0.1	scale	17.85211763	4
1	0.1	auto	22.30698934	4
1	0.5	scale	17.86182039	4
1	0.5	auto	21.73410965	4
10	0.01	scale	11.93080837	4
10	0.01	auto	14.16139487	4
10	0.1	scale	11.90933421	4
10	0.1	auto	14.10035245	4
10	0.5	scale	11.81079387	4
10	0.5	auto	13.98657198	4
0.1	0.01	scale	41.31157977	5

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
0.1	0.01	auto	49.37729802	5
0.1	0.1	scale	41.29971345	5
0.1	0.1	auto	49.55714599	5
0.1	0.5	scale	40.85248468	5
0.1	0.5	auto	49.65969681	5
1	0.01	scale	9.706129752	5
1	0.01	auto	18.47694236	5
1	0.1	scale	9.67022682	5
1	0.1	auto	18.41335114	5
1	0.5	scale	9.973369927	5
1	0.5	auto	18.34060293	5
10	0.01	scale	2.325738406	5
10	0.01	auto	4.63602555	5
10	0.1	scale	2.209674528	5
10	0.1	auto	4.617314988	5
10	0.5	scale	2.311818045	5
10	0.5	auto	4.523024095	5

Table A.6 : NLICA_SVR model hyperparameters tuning for different ICs using the top K = 10 external factors.

C	Epsilon	Gamma	Mean MSE	n_components
0.1	0.01	scale	161.4289577	1
0.1	0.01	auto	161.4289577	1
0.1	0.1	scale	161.5702249	1

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
0.1	0.1	auto	161.5702249	1
0.1	0.5	scale	158.3788166	1
0.1	0.5	auto	158.3788166	1
1	0.01	scale	161.4289577	1
1	0.01	auto	161.4289577	1
1	0.1	scale	161.5702249	1
1	0.1	auto	161.5702249	1
1	0.5	scale	158.3788166	1
1	0.5	auto	158.3788166	1
10	0.01	scale	161.4289577	1
10	0.01	auto	161.4289577	1
10	0.1	scale	161.5702249	1
10	0.1	auto	161.5702249	1
10	0.5	scale	158.3788166	1
10	0.5	auto	158.3788166	1
0.1	0.01	scale	35.65144671	2
0.1	0.01	auto	34.83800146	2
0.1	0.1	scale	35.67654906	2
0.1	0.1	auto	34.87548812	2
0.1	0.5	scale	35.75044879	2
0.1	0.5	auto	35.0750286	2
1	0.01	scale	5.952181455	2
1	0.01	auto	5.854891262	2
1	0.1	scale	5.960598772	2

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
1	0.1	auto	5.836875665	2
1	0.5	scale	5.902448113	2
1	0.5	auto	5.792084431	2
10	0.01	scale	3.34415706	2
10	0.01	auto	3.265096146	2
10	0.1	scale	3.326745185	2
10	0.1	auto	3.252951181	2
10	0.5	scale	3.296812135	2
10	0.5	auto	3.241280023	2
0.1	0.01	scale	49.16374642	3
0.1	0.01	auto	48.94617389	3
0.1	0.1	scale	49.30834165	3
0.1	0.1	auto	49.06846976	3
0.1	0.5	scale	49.57613793	3
0.1	0.5	auto	49.35264826	3
1	0.01	scale	19.38702836	3
1	0.01	auto	18.66406631	3
1	0.1	scale	19.39651709	3
1	0.1	auto	18.6910439	3
1	0.5	scale	19.47975478	3
1	0.5	auto	18.87037715	3
10	0.01	scale	12.49506136	3
10	0.01	auto	12.28435077	3
10	0.1	scale	12.46302283	3

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
10	0.1	auto	12.24547863	3
10	0.5	scale	12.23572233	3
10	0.5	auto	12.08988084	3
0.1	0.01	scale	39.81611105	4
0.1	0.01	auto	40.42538285	4
0.1	0.1	scale	39.95879109	4
0.1	0.1	auto	40.50459937	4
0.1	0.5	scale	40.38771201	4
0.1	0.5	auto	40.97045544	4
1	0.01	scale	25.75294093	4
1	0.01	auto	26.05476386	4
1	0.1	scale	25.74841939	4
1	0.1	auto	25.96705574	4
1	0.5	scale	25.56952153	4
1	0.5	auto	25.82339301	4
10	0.01	scale	22.94281615	4
10	0.01	auto	23.50510084	4
10	0.1	scale	23.03356817	4
10	0.1	auto	23.61717849	4
10	0.5	scale	23.05126389	4
10	0.5	auto	23.43225854	4
0.1	0.01	scale	42.33984156	5
0.1	0.01	auto	47.9519223	5
0.1	0.1	scale	42.47359433	5

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
0.1	0.1	auto	47.94684851	5
0.1	0.5	scale	42.75945031	5
0.1	0.5	auto	48.54971349	5
1	0.01	scale	10.21135226	5
1	0.01	auto	14.50545309	5
1	0.1	scale	10.27460529	5
1	0.1	auto	14.55492221	5
1	0.5	scale	10.42642846	5
1	0.5	auto	14.98778905	5
10	0.01	scale	3.179937596	5
10	0.01	auto	4.845020372	5
10	0.1	scale	3.0163884	5
10	0.1	auto	4.794134993	5
10	0.5	scale	3.152356047	5
10	0.5	auto	4.634142385	5
0.1	0.01	scale	42.38175926	6
0.1	0.01	auto	47.35694165	6
0.1	0.1	scale	42.43473046	6
0.1	0.1	auto	47.32129089	6
0.1	0.5	scale	42.45088467	6
0.1	0.5	auto	47.03290708	6
1	0.01	scale	13.28081606	6
1	0.01	auto	18.93353373	6
1	0.1	scale	13.34864241	6

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
1	0.1	auto	18.91844432	6
1	0.5	scale	13.64713503	6
1	0.5	auto	18.89585437	6
10	0.01	scale	7.60636493	6
10	0.01	auto	8.388034559	6
10	0.1	scale	7.575653714	6
10	0.1	auto	8.391194027	6
10	0.5	scale	7.509623708	6
10	0.5	auto	8.491561509	6
0.1	0.01	scale	57.94399863	7
0.1	0.01	auto	63.01505549	7
0.1	0.1	scale	57.8530187	7
0.1	0.1	auto	63.15776842	7
0.1	0.5	scale	57.72786284	7
0.1	0.5	auto	63.08712152	7
1	0.01	scale	28.29507857	7
1	0.01	auto	40.07320379	7
1	0.1	scale	28.37341101	7
1	0.1	auto	40.01438567	7
1	0.5	scale	28.27138997	7
1	0.5	auto	39.37936469	7
10	0.01	scale	11.37174225	7
10	0.01	auto	15.70294909	7
10	0.1	scale	11.34465331	7

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
10	0.1	auto	15.70540336	7
10	0.5	scale	10.84942419	7
10	0.5	auto	15.55611466	7
0.1	0.01	scale	48.26977029	8
0.1	0.01	auto	57.18854135	8
0.1	0.1	scale	47.98996478	8
0.1	0.1	auto	57.11139035	8
0.1	0.5	scale	47.95792638	8
0.1	0.5	auto	56.73787979	8
1	0.01	scale	11.76286498	8
1	0.01	auto	21.50379904	8
1	0.1	scale	11.73619778	8
1	0.1	auto	21.53140704	8
1	0.5	scale	11.46182014	8
1	0.5	auto	21.20077538	8
10	0.01	scale	5.535878605	8
10	0.01	auto	5.912449	8
10	0.1	scale	5.539494479	8
10	0.1	auto	5.915591974	8
10	0.5	scale	5.52533426	8
10	0.5	auto	5.898014283	8
0.1	0.01	scale	46.29432285	9
0.1	0.01	auto	58.09381596	9
0.1	0.1	scale	46.33367826	9

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
0.1	0.1	auto	57.98092603	9
0.1	0.5	scale	45.8904031	9
0.1	0.5	auto	57.8885165	9
1	0.01	scale	13.46307923	9
1	0.01	auto	23.22484705	9
1	0.1	scale	13.56179562	9
1	0.1	auto	23.20316616	9
1	0.5	scale	13.50706842	9
1	0.5	auto	23.55356109	9
10	0.01	scale	4.848866989	9
10	0.01	auto	6.78748142	9
10	0.1	scale	4.807287732	9
10	0.1	auto	6.767036754	9
10	0.5	scale	4.774765269	9
10	0.5	auto	6.545090907	9
0.1	0.01	scale	51.1185535	10
0.1	0.01	auto	60.75167173	10
0.1	0.1	scale	51.13676355	10
0.1	0.1	auto	60.65145266	10
0.1	0.5	scale	50.86743673	10
0.1	0.5	auto	59.88223456	10
1	0.01	scale	15.76085209	10
1	0.01	auto	27.82952274	10
1	0.1	scale	15.77387855	10

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
1	0.1	auto	27.95667062	10
1	0.5	scale	15.83169725	10
1	0.5	auto	27.91806856	10
10	0.01	scale	7.945823189	10
10	0.01	auto	10.69975413	10
10	0.1	scale	7.969288745	10
10	0.1	auto	10.65898977	10
10	0.5	scale	7.950521322	10
10	0.5	auto	10.60476126	10

Table A.7 : NLICA_SVR model hyperparameters tuning for different ICs using the top K = 15 external factors.

C	Epsilon	Gamma	Mean MSE	n_components
0.1	0.01	scale	58.56676544	1
0.1	0.01	auto	50.72103543	1
0.1	0.1	scale	58.48722489	1
0.1	0.1	auto	50.43620062	1
0.1	0.5	scale	57.91378879	1
0.1	0.5	auto	49.48150607	1
1	0.01	scale	28.8333873	1
1	0.01	auto	26.23356566	1
1	0.1	scale	29.00187538	1
1	0.1	auto	26.18812935	1
1	0.5	scale	28.70640124	1

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
1	0.5	auto	26.09148369	1
10	0.01	scale	23.31650077	1
10	0.01	auto	22.36466729	1
10	0.1	scale	23.29573349	1
10	0.1	auto	22.33398258	1
10	0.5	scale	23.25655929	1
10	0.5	auto	22.38155633	1
0.1	0.01	scale	58.73425131	2
0.1	0.01	auto	54.7702358	2
0.1	0.1	scale	58.73322332	2
0.1	0.1	auto	54.79284014	2
0.1	0.5	scale	58.36845215	2
0.1	0.5	auto	54.03838269	2
1	0.01	scale	27.71839654	2
1	0.01	auto	27.11247137	2
1	0.1	scale	27.75991986	2
1	0.1	auto	27.02312017	2
1	0.5	scale	27.64905326	2
1	0.5	auto	26.67207749	2
10	0.01	scale	23.99317381	2
10	0.01	auto	23.20120936	2
10	0.1	scale	23.84941878	2
10	0.1	auto	23.166195	2
10	0.5	scale	23.58363425	2

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
10	0.5	auto	23.26179751	2
0.1	0.01	scale	50.8083656	3
0.1	0.01	auto	53.40117276	3
0.1	0.1	scale	50.79007797	3
0.1	0.1	auto	53.27482694	3
0.1	0.5	scale	50.45159329	3
0.1	0.5	auto	53.15782335	3
1	0.01	scale	26.0613944	3
1	0.01	auto	27.43905576	3
1	0.1	scale	26.18358126	3
1	0.1	auto	27.39248588	3
1	0.5	scale	26.62506144	3
1	0.5	auto	27.39431293	3
10	0.01	scale	22.86166832	3
10	0.01	auto	23.10201547	3
10	0.1	scale	22.86101044	3
10	0.1	auto	23.12162391	3
10	0.5	scale	23.04769489	3
10	0.5	auto	23.38837871	3
0.1	0.01	scale	46.90015121	4
0.1	0.01	auto	45.29672089	4
0.1	0.1	scale	46.82011815	4
0.1	0.1	auto	45.29406263	4
0.1	0.5	scale	46.65547	4

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
0.1	0.5	auto	45.6346927	4
1	0.01	scale	33.30036982	4
1	0.01	auto	31.64060101	4
1	0.1	scale	33.29265766	4
1	0.1	auto	31.62340453	4
1	0.5	scale	33.25448004	4
1	0.5	auto	31.36422595	4
10	0.01	scale	28.10193651	4
10	0.01	auto	25.99514533	4
10	0.1	scale	28.00926521	4
10	0.1	auto	26.11629612	4
10	0.5	scale	27.64331869	4
10	0.5	auto	25.8137164	4
0.1	0.01	scale	66.30429041	5
0.1	0.01	auto	66.06394269	5
0.1	0.1	scale	66.26853048	5
0.1	0.1	auto	66.16440748	5
0.1	0.5	scale	65.45017602	5
0.1	0.5	auto	65.29795961	5
1	0.01	scale	32.34676778	5
1	0.01	auto	31.84870464	5
1	0.1	scale	32.14408744	5
1	0.1	auto	31.78450073	5
1	0.5	scale	32.19851605	5

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
1	0.5	auto	31.62998903	5
10	0.01	scale	9.503803105	5
10	0.01	auto	9.179037284	5
10	0.1	scale	9.462975345	5
10	0.1	auto	9.184168617	5
10	0.5	scale	9.296722774	5
10	0.5	auto	9.067679285	5
0.1	0.01	scale	49.48348568	6
0.1	0.01	auto	48.48779844	6
0.1	0.1	scale	49.46723282	6
0.1	0.1	auto	48.57445784	6
0.1	0.5	scale	48.9839531	6
0.1	0.5	auto	47.92380104	6
1	0.01	scale	20.61171343	6
1	0.01	auto	18.99083348	6
1	0.1	scale	20.55684618	6
1	0.1	auto	18.99675147	6
1	0.5	scale	20.77252152	6
1	0.5	auto	18.94148704	6
10	0.01	scale	7.501944728	6
10	0.01	auto	6.727491783	6
10	0.1	scale	7.449506292	6
10	0.1	auto	6.717444229	6
10	0.5	scale	7.423984617	6

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
10	0.5	auto	6.773958095	6
0.1	0.01	scale	49.14657678	7
0.1	0.01	auto	52.0184246	7
0.1	0.1	scale	49.18943206	7
0.1	0.1	auto	52.04964559	7
0.1	0.5	scale	48.96272725	7
0.1	0.5	auto	51.99670802	7
1	0.01	scale	15.27371806	7
1	0.01	auto	17.94678099	7
1	0.1	scale	15.22354318	7
1	0.1	auto	17.86425862	7
1	0.5	scale	14.9681088	7
1	0.5	auto	18.04494835	7
10	0.01	scale	6.415695922	7
10	0.01	auto	7.365658148	7
10	0.1	scale	6.421756833	7
10	0.1	auto	7.403734861	7
10	0.5	scale	6.377678847	7
10	0.5	auto	7.329806836	7
0.1	0.01	scale	54.67123504	8
0.1	0.01	auto	58.84424	8
0.1	0.1	scale	54.41682688	8
0.1	0.1	auto	58.78122416	8
0.1	0.5	scale	53.80179638	8

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
0.1	0.5	auto	59.0982617	8
1	0.01	scale	14.80226521	8
1	0.01	auto	23.2185278	8
1	0.1	scale	14.82974732	8
1	0.1	auto	23.18571077	8
1	0.5	scale	14.89485991	8
1	0.5	auto	23.02191044	8
10	0.01	scale	7.988802613	8
10	0.01	auto	8.402946269	8
10	0.1	scale	7.989718809	8
10	0.1	auto	8.385948241	8
10	0.5	scale	7.976988508	8
10	0.5	auto	8.375479335	8
0.1	0.01	scale	50.97902374	9
0.1	0.01	auto	58.06605018	9
0.1	0.1	scale	50.764428	9
0.1	0.1	auto	57.83112919	9
0.1	0.5	scale	50.73370229	9
0.1	0.5	auto	57.54466483	9
1	0.01	scale	12.96819524	9
1	0.01	auto	19.13885121	9
1	0.1	scale	12.93868022	9
1	0.1	auto	19.2562863	9
1	0.5	scale	13.02004234	9

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
1	0.5	auto	19.17809161	9
10	0.01	scale	5.449429486	9
10	0.01	auto	7.106591559	9
10	0.1	scale	5.446046075	9
10	0.1	auto	7.123971235	9
10	0.5	scale	5.451940106	9
10	0.5	auto	6.908071979	9
0.1	0.01	scale	52.0778225	10
0.1	0.01	auto	59.15222439	10
0.1	0.1	scale	52.07483765	10
0.1	0.1	auto	59.24190897	10
0.1	0.5	scale	51.29942367	10
0.1	0.5	auto	58.34397318	10
1	0.01	scale	13.70684424	10
1	0.01	auto	18.75282482	10
1	0.1	scale	13.72523006	10
1	0.1	auto	18.62455742	10
1	0.5	scale	13.76030862	10
1	0.5	auto	18.40205838	10
10	0.01	scale	2.874533647	10
10	0.01	auto	4.721828046	10
10	0.1	scale	2.893979939	10
10	0.1	auto	4.65960664	10
10	0.5	scale	2.856563865	10

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
10	0.5	auto	4.564387212	10
0.1	0.01	scale	53.46484751	11
0.1	0.01	auto	62.5802125	11
0.1	0.1	scale	53.68754789	11
0.1	0.1	auto	62.53533693	11
0.1	0.5	scale	54.06856877	11
0.1	0.5	auto	61.63762644	11
1	0.01	scale	9.7917952	11
1	0.01	auto	17.22931979	11
1	0.1	scale	9.707879941	11
1	0.1	auto	17.23315681	11
1	0.5	scale	9.732230325	11
1	0.5	auto	17.11875653	11
10	0.01	scale	2.334265993	11
10	0.01	auto	3.299365094	11
10	0.1	scale	2.325447104	11
10	0.1	auto	3.281812615	11
10	0.5	scale	2.378108242	11
10	0.5	auto	3.477709383	11
0.1	0.01	scale	50.88025463	12
0.1	0.01	auto	62.3993279	12
0.1	0.1	scale	50.85478923	12
0.1	0.1	auto	62.53060597	12
0.1	0.5	scale	50.33335858	12

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
0.1	0.5	auto	61.55972823	12
1	0.01	scale	7.662923187	12
1	0.01	auto	17.11671159	12
1	0.1	scale	7.686871441	12
1	0.1	auto	17.09862601	12
1	0.5	scale	8.154241295	12
1	0.5	auto	17.00411693	12
10	0.01	scale	1.338915466	12
10	0.01	auto	1.949848965	12
10	0.1	scale	1.332784865	12
10	0.1	auto	1.939019265	12
10	0.5	scale	1.373720787	12
10	0.5	auto	2.044537807	12
0.1	0.01	scale	49.76669219	13
0.1	0.01	auto	61.52502368	13
0.1	0.1	scale	49.87788941	13
0.1	0.1	auto	61.56420763	13
0.1	0.5	scale	49.7981181	13
0.1	0.5	auto	60.73700507	13
1	0.01	scale	14.64409647	13
1	0.01	auto	21.04705965	13
1	0.1	scale	14.68333178	13
1	0.1	auto	21.074066	13
1	0.5	scale	14.53308725	13

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
1	0.5	auto	21.2534407	13
10	0.01	scale	4.302017057	13
10	0.01	auto	7.333112107	13
10	0.1	scale	4.308412311	13
10	0.1	auto	7.304604986	13
10	0.5	scale	4.33641403	13
10	0.5	auto	7.139439004	13
0.1	0.01	scale	48.16070386	14
0.1	0.01	auto	59.77345973	14
0.1	0.1	scale	47.98343794	14
0.1	0.1	auto	59.7042127	14
0.1	0.5	scale	47.84961192	14
0.1	0.5	auto	59.23996697	14
1	0.01	scale	13.50779894	14
1	0.01	auto	20.39687884	14
1	0.1	scale	13.46239973	14
1	0.1	auto	20.30011331	14
1	0.5	scale	13.0787278	14
1	0.5	auto	20.45135351	14
10	0.01	scale	2.498141492	14
10	0.01	auto	4.678945215	14
10	0.1	scale	2.476470189	14
10	0.1	auto	4.683189188	14
10	0.5	scale	2.521148374	14

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
10	0.5	auto	4.616567348	14
0.1	0.01	scale	46.45468263	15
0.1	0.01	auto	63.77331289	15
0.1	0.1	scale	46.53729825	15
0.1	0.1	auto	63.70932403	15
0.1	0.5	scale	46.17457659	15
0.1	0.5	auto	62.55144204	15
1	0.01	scale	8.57354556	15
1	0.01	auto	17.42135646	15
1	0.1	scale	8.510701559	15
1	0.1	auto	17.38321937	15
1	0.5	scale	8.57627459	15
1	0.5	auto	17.13842362	15
10	0.01	scale	1.192378154	15
10	0.01	auto	3.252087582	15
10	0.1	scale	1.184911498	15
10	0.1	auto	3.236126066	15
10	0.5	scale	1.205987672	15
10	0.5	auto	3.035191776	15

Table A.8 : NLICA_SVR model hyperparameters tuning for different ICs using all K = 22 external factors.

C	Epsilon	Gamma	Mean MSE	n_components
0.1	0.01	scale	161.4289577	1

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
0.1	0.01	auto	161.4289577	1
0.1	0.1	scale	161.5702249	1
0.1	0.1	auto	161.5702249	1
0.1	0.5	scale	158.3788166	1
0.1	0.5	auto	158.3788166	1
1	0.01	scale	161.4289577	1
1	0.01	auto	161.4289577	1
1	0.1	scale	161.5702249	1
1	0.1	auto	161.5702249	1
1	0.5	scale	158.3788166	1
1	0.5	auto	158.3788166	1
10	0.01	scale	161.4289577	1
10	0.01	auto	161.4289577	1
10	0.1	scale	161.5702249	1
10	0.1	auto	161.5702249	1
10	0.5	scale	158.3788166	1
10	0.5	auto	158.3788166	1
0.1	0.01	scale	54.80598156	2
0.1	0.01	auto	49.75845794	2
0.1	0.1	scale	54.73586302	2
0.1	0.1	auto	49.70607546	2
0.1	0.5	scale	53.99413101	2
0.1	0.5	auto	49.89405831	2
1	0.01	scale	26.0272451	2

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
1	0.01	auto	24.36359212	2
1	0.1	scale	25.85719472	2
1	0.1	auto	24.38625114	2
1	0.5	scale	25.71603874	2
1	0.5	auto	24.59525166	2
10	0.01	scale	21.92465965	2
10	0.01	auto	21.7009682	2
10	0.1	scale	21.99323698	2
10	0.1	auto	21.69617752	2
10	0.5	scale	22.19371392	2
10	0.5	auto	21.92422416	2
0.1	0.01	scale	66.8957795	3
0.1	0.01	auto	61.16451619	3
0.1	0.1	scale	66.90605256	3
0.1	0.1	auto	61.19719204	3
0.1	0.5	scale	66.53292273	3
0.1	0.5	auto	60.28367613	3
1	0.01	scale	37.60133668	3
1	0.01	auto	30.5353588	3
1	0.1	scale	37.43635668	3
1	0.1	auto	30.48857279	3
1	0.5	scale	37.16770577	3
1	0.5	auto	30.00112066	3
10	0.01	scale	24.52515414	3

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
10	0.01	auto	14.5711678	3
10	0.1	scale	24.5600488	3
10	0.1	auto	14.53958183	3
10	0.5	scale	24.05624763	3
10	0.5	auto	14.64333993	3
0.1	0.01	scale	61.82184727	4
0.1	0.01	auto	61.65746579	4
0.1	0.1	scale	61.58090197	4
0.1	0.1	auto	61.4212334	4
0.1	0.5	scale	60.88006182	4
0.1	0.5	auto	60.70877571	4
1	0.01	scale	30.51793224	4
1	0.01	auto	30.44107188	4
1	0.1	scale	30.50167989	4
1	0.1	auto	30.42346943	4
1	0.5	scale	30.56535924	4
1	0.5	auto	30.51453756	4
10	0.01	scale	24.25139775	4
10	0.01	auto	24.21394091	4
10	0.1	scale	24.26362041	4
10	0.1	auto	24.13191247	4
10	0.5	scale	24.07727805	4
10	0.5	auto	24.01307378	4
0.1	0.01	scale	51.77377498	5

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
0.1	0.01	auto	47.09813023	5
0.1	0.1	scale	51.59712604	5
0.1	0.1	auto	46.88665566	5
0.1	0.5	scale	50.54633292	5
0.1	0.5	auto	46.64892779	5
1	0.01	scale	30.37075387	5
1	0.01	auto	22.84878019	5
1	0.1	scale	30.21045363	5
1	0.1	auto	22.86255511	5
1	0.5	scale	29.94612144	5
1	0.5	auto	22.84016115	5
10	0.01	scale	16.9522511	5
10	0.01	auto	10.71212188	5
10	0.1	scale	16.78812042	5
10	0.1	auto	10.80235365	5
10	0.5	scale	16.29902485	5
10	0.5	auto	10.79049978	5
0.1	0.01	scale	56.83014792	6
0.1	0.01	auto	53.29169112	6
0.1	0.1	scale	56.74628409	6
0.1	0.1	auto	53.29497113	6
0.1	0.5	scale	55.90216452	6
0.1	0.5	auto	52.52138071	6
1	0.01	scale	27.29696523	6

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
1	0.01	auto	21.56965075	6
1	0.1	scale	27.27947876	6
1	0.1	auto	21.6426785	6
1	0.5	scale	26.85092188	6
1	0.5	auto	21.45546286	6
10	0.01	scale	12.90914078	6
10	0.01	auto	10.71886669	6
10	0.1	scale	12.92399107	6
10	0.1	auto	10.69870113	6
10	0.5	scale	13.10025441	6
10	0.5	auto	10.81285717	6
0.1	0.01	scale	59.97140174	7
0.1	0.01	auto	59.843717	7
0.1	0.1	scale	59.85244412	7
0.1	0.1	auto	59.74507953	7
0.1	0.5	scale	59.27015053	7
0.1	0.5	auto	59.09698146	7
1	0.01	scale	26.68175074	7
1	0.01	auto	26.06675138	7
1	0.1	scale	26.51321333	7
1	0.1	auto	25.92502266	7
1	0.5	scale	25.83491143	7
1	0.5	auto	25.38673256	7
10	0.01	scale	9.306553154	7

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
10	0.01	auto	8.303054151	7
10	0.1	scale	9.322806039	7
10	0.1	auto	8.259472432	7
10	0.5	scale	9.036601407	7
10	0.5	auto	8.060963288	7
0.1	0.01	scale	55.23585798	8
0.1	0.01	auto	52.7202778	8
0.1	0.1	scale	55.33806448	8
0.1	0.1	auto	52.4576874	8
0.1	0.5	scale	54.68790096	8
0.1	0.5	auto	52.14280451	8
1	0.01	scale	36.50607249	8
1	0.01	auto	32.8563665	8
1	0.1	scale	36.56516294	8
1	0.1	auto	32.79227325	8
1	0.5	scale	36.19592304	8
1	0.5	auto	32.74224731	8
10	0.01	scale	22.88096321	8
10	0.01	auto	15.69020306	8
10	0.1	scale	22.81777074	8
10	0.1	auto	15.5785345	8
10	0.5	scale	22.25687233	8
10	0.5	auto	15.54088625	8
0.1	0.01	scale	51.03622702	9

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
0.1	0.01	auto	51.14721288	9
0.1	0.1	scale	51.14740524	9
0.1	0.1	auto	51.33914592	9
0.1	0.5	scale	50.79265553	9
0.1	0.5	auto	50.97959225	9
1	0.01	scale	21.57022344	9
1	0.01	auto	21.76309546	9
1	0.1	scale	21.64110255	9
1	0.1	auto	21.86128531	9
1	0.5	scale	21.82895059	9
1	0.5	auto	22.09078396	9
10	0.01	scale	4.366317094	9
10	0.01	auto	4.422426132	9
10	0.1	scale	4.379362606	9
10	0.1	auto	4.438576674	9
10	0.5	scale	4.431728262	9
10	0.5	auto	4.496771511	9
0.1	0.01	scale	53.43439352	10
0.1	0.01	auto	53.98529446	10
0.1	0.1	scale	53.41264175	10
0.1	0.1	auto	53.88781642	10
0.1	0.5	scale	52.98936861	10
0.1	0.5	auto	53.37258683	10
1	0.01	scale	14.43803373	10

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
1	0.01	auto	14.96102084	10
1	0.1	scale	14.50410702	10
1	0.1	auto	14.96143967	10
1	0.5	scale	14.60395822	10
1	0.5	auto	14.93210298	10
10	0.01	scale	5.567290011	10
10	0.01	auto	5.627673514	10
10	0.1	scale	5.581763254	10
10	0.1	auto	5.671846429	10
10	0.5	scale	5.738655038	10
10	0.5	auto	5.790707915	10
0.1	0.01	scale	52.83498442	11
0.1	0.01	auto	57.59617129	11
0.1	0.1	scale	52.70789284	11
0.1	0.1	auto	57.52139534	11
0.1	0.5	scale	51.98303389	11
0.1	0.5	auto	57.13403955	11
1	0.01	scale	14.58680875	11
1	0.01	auto	19.27709718	11
1	0.1	scale	14.56438336	11
1	0.1	auto	19.4256397	11
1	0.5	scale	14.57622322	11
1	0.5	auto	19.74176495	11
10	0.01	scale	3.85677026	11

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
10	0.01	auto	5.820457143	11
10	0.1	scale	3.84946157	11
10	0.1	auto	5.859013647	11
10	0.5	scale	3.870999328	11
10	0.5	auto	5.716140718	11
0.1	0.01	scale	58.49432727	12
0.1	0.01	auto	62.16546959	12
0.1	0.1	scale	58.5120978	12
0.1	0.1	auto	62.02579234	12
0.1	0.5	scale	57.47805883	12
0.1	0.5	auto	61.5588147	12
1	0.01	scale	22.77262262	12
1	0.01	auto	24.90035928	12
1	0.1	scale	22.72033169	12
1	0.1	auto	25.03416009	12
1	0.5	scale	22.76467371	12
1	0.5	auto	25.59317128	12
10	0.01	scale	8.510329332	12
10	0.01	auto	10.37096734	12
10	0.1	scale	8.448436544	12
10	0.1	auto	10.31612072	12
10	0.5	scale	8.491217555	12
10	0.5	auto	10.1235583	12
0.1	0.01	scale	58.35710937	13

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
0.1	0.01	auto	58.95144738	13
0.1	0.1	scale	58.43615545	13
0.1	0.1	auto	59.1150307	13
0.1	0.5	scale	57.62568753	13
0.1	0.5	auto	58.20114789	13
1	0.01	scale	28.01692218	13
1	0.01	auto	28.75656742	13
1	0.1	scale	28.00818399	13
1	0.1	auto	28.75609668	13
1	0.5	scale	27.87803584	13
1	0.5	auto	28.50393328	13
10	0.01	scale	11.65702304	13
10	0.01	auto	11.95601458	13
10	0.1	scale	11.67301974	13
10	0.1	auto	12.01393767	13
10	0.5	scale	11.54564757	13
10	0.5	auto	11.95036205	13
0.1	0.01	scale	61.6629273	14
0.1	0.01	auto	66.34304394	14
0.1	0.1	scale	61.67591484	14
0.1	0.1	auto	66.25411966	14
0.1	0.5	scale	61.02751473	14
0.1	0.5	auto	65.89811159	14
1	0.01	scale	21.03049525	14

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
1	0.01	auto	27.15103665	14
1	0.1	scale	21.02895331	14
1	0.1	auto	27.10989629	14
1	0.5	scale	21.0721837	14
1	0.5	auto	26.72963323	14
10	0.01	scale	1.836899971	14
10	0.01	auto	2.583351946	14
10	0.1	scale	1.813290853	14
10	0.1	auto	2.550800859	14
10	0.5	scale	1.8216291	14
10	0.5	auto	2.5590838	14
0.1	0.01	scale	57.38582495	15
0.1	0.01	auto	64.70082273	15
0.1	0.1	scale	57.44638843	15
0.1	0.1	auto	64.6065187	15
0.1	0.5	scale	56.32562998	15
0.1	0.5	auto	63.95575457	15
1	0.01	scale	16.02734917	15
1	0.01	auto	21.0020666	15
1	0.1	scale	15.99528893	15
1	0.1	auto	20.9648411	15
1	0.5	scale	16.29719132	15
1	0.5	auto	21.0222635	15
10	0.01	scale	3.335451968	15

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
10	0.01	auto	5.459206168	15
10	0.1	scale	3.32027329	15
10	0.1	auto	5.504076755	15
10	0.5	scale	3.344399944	15
10	0.5	auto	5.572559192	15
0.1	0.01	scale	51.92078314	16
0.1	0.01	auto	62.47514279	16
0.1	0.1	scale	51.83261347	16
0.1	0.1	auto	62.37793773	16
0.1	0.5	scale	51.34161535	16
0.1	0.5	auto	62.01226466	16
1	0.01	scale	13.42986774	16
1	0.01	auto	19.31290637	16
1	0.1	scale	13.50681406	16
1	0.1	auto	19.25161329	16
1	0.5	scale	13.31381792	16
1	0.5	auto	19.27128397	16
10	0.01	scale	2.984131902	16
10	0.01	auto	4.879922912	16
10	0.1	scale	3.005722114	16
10	0.1	auto	4.830925468	16
10	0.5	scale	3.135219454	16
10	0.5	auto	4.867198002	16
0.1	0.01	scale	51.41215893	17

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
0.1	0.01	auto	58.50935291	17
0.1	0.1	scale	51.49391458	17
0.1	0.1	auto	58.43108725	17
0.1	0.5	scale	50.87866361	17
0.1	0.5	auto	57.40013826	17
1	0.01	scale	11.08495646	17
1	0.01	auto	18.51788093	17
1	0.1	scale	11.09876857	17
1	0.1	auto	18.45256867	17
1	0.5	scale	11.28399845	17
1	0.5	auto	18.28415418	17
10	0.01	scale	1.343153302	17
10	0.01	auto	1.953543845	17
10	0.1	scale	1.317589598	17
10	0.1	auto	1.953383868	17
10	0.5	scale	1.37677275	17
10	0.5	auto	1.922310741	17
0.1	0.01	scale	53.92993919	18
0.1	0.01	auto	61.69494655	18
0.1	0.1	scale	53.87278293	18
0.1	0.1	auto	61.86072425	18
0.1	0.5	scale	53.55877926	18
0.1	0.5	auto	61.13104412	18
1	0.01	scale	11.48520973	18

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
1	0.01	auto	16.93788651	18
1	0.1	scale	11.42654636	18
1	0.1	auto	16.97465759	18
1	0.5	scale	11.63302091	18
1	0.5	auto	17.49518901	18
10	0.01	scale	1.983345512	18
10	0.01	auto	2.504088416	18
10	0.1	scale	1.985480751	18
10	0.1	auto	2.505842219	18
10	0.5	scale	2.055730242	18
10	0.5	auto	2.581422615	18
0.1	0.01	scale	53.63067144	19
0.1	0.01	auto	65.95216782	19
0.1	0.1	scale	53.62778207	19
0.1	0.1	auto	65.80591558	19
0.1	0.5	scale	53.12867558	19
0.1	0.5	auto	65.00215641	19
1	0.01	scale	15.63721625	19
1	0.01	auto	24.44434435	19
1	0.1	scale	15.72167867	19
1	0.1	auto	24.38276215	19
1	0.5	scale	15.74971059	19
1	0.5	auto	23.92718605	19
10	0.01	scale	2.99679475	19

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
10	0.01	auto	6.967911905	19
10	0.1	scale	3.008739908	19
10	0.1	auto	6.885867944	19
10	0.5	scale	3.173572809	19
10	0.5	auto	6.768199681	19
0.1	0.01	scale	49.37341408	20
0.1	0.01	auto	64.64683843	20
0.1	0.1	scale	49.29477	20
0.1	0.1	auto	64.62029705	20
0.1	0.5	scale	48.92623996	20
0.1	0.5	auto	64.17388587	20
1	0.01	scale	16.56574265	20
1	0.01	auto	22.03159441	20
1	0.1	scale	16.49645403	20
1	0.1	auto	22.03323659	20
1	0.5	scale	16.11422543	20
1	0.5	auto	22.05007315	20
10	0.01	scale	3.06265094	20
10	0.01	auto	11.26403298	20
10	0.1	scale	3.073393063	20
10	0.1	auto	11.22087999	20
10	0.5	scale	3.203964871	20
10	0.5	auto	10.64365002	20
0.1	0.01	scale	58.202272	21

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
0.1	0.01	auto	70.09101858	21
0.1	0.1	scale	58.01552383	21
0.1	0.1	auto	69.84250548	21
0.1	0.5	scale	56.6843782	21
0.1	0.5	auto	68.74140002	21
1	0.01	scale	11.20659519	21
1	0.01	auto	22.54308531	21
1	0.1	scale	11.1907347	21
1	0.1	auto	22.39464923	21
1	0.5	scale	11.24323667	21
1	0.5	auto	21.70531266	21
10	0.01	scale	1.706822336	21
10	0.01	auto	2.787464001	21
10	0.1	scale	1.695871532	21
10	0.1	auto	2.804053148	21
10	0.5	scale	1.76979828	21
10	0.5	auto	2.862183958	21
0.1	0.01	scale	59.86551736	22
0.1	0.01	auto	74.40615082	22
0.1	0.1	scale	59.63982652	22
0.1	0.1	auto	74.17627712	22
0.1	0.5	scale	58.55829874	22
0.1	0.5	auto	73.42897295	22
1	0.01	scale	13.88275318	22

Table continues on the next page

C	epsilon	gamma	Mean MSE	n_components
1	0.01	auto	25.0191239	22
1	0.1	scale	13.84685827	22
1	0.1	auto	24.83044779	22
1	0.5	scale	14.18971022	22
1	0.5	auto	24.87763142	22
10	0.01	scale	2.391931868	22
10	0.01	auto	4.283488873	22
10	0.1	scale	2.401364551	22
10	0.1	auto	4.305740706	22
10	0.5	scale	2.530233642	22
10	0.5	auto	4.159976173	22

Table A.9 : ICA_MLP model hyperparameters tuning for different ICs using the top K = 5 external factors.

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.0001	(50,)	adam	86.95450344	1
tanh	0.0001	(50,)	sgd	22.75230909	1
tanh	0.0001	(100,)	adam	33.03386472	1
tanh	0.0001	(100,)	sgd	25.9031724	1
tanh	0.0001	(50, 50)	adam	65.07279019	1
tanh	0.0001	(50, 50)	sgd	18.73111778	1
tanh	0.001	(50,)	adam	86.9545938	1
tanh	0.001	(50,)	sgd	22.75271982	1
tanh	0.001	(100,)	adam	33.03387939	1

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.001	(100,)	sgd	25.90356644	1
tanh	0.001	(50, 50)	adam	65.07269284	1
tanh	0.001	(50, 50)	sgd	18.73125866	1
relu	0.0001	(50,)	adam	111.5286903	1
relu	0.0001	(50,)	sgd	29.42261812	1
relu	0.0001	(100,)	adam	33.36934791	1
relu	0.0001	(100,)	sgd	29.87949293	1
relu	0.0001	(50, 50)	adam	28.72470628	1
relu	0.0001	(50, 50)	sgd	18.15771476	1
relu	0.001	(50,)	adam	111.5286567	1
relu	0.001	(50,)	sgd	29.42262208	1
relu	0.001	(100,)	adam	33.36929286	1
relu	0.001	(100,)	sgd	29.87948915	1
relu	0.001	(50, 50)	adam	28.67907811	1
relu	0.001	(50, 50)	sgd	18.14341954	1
tanh	0.0001	(50,)	adam	89.1671814	2
tanh	0.0001	(50,)	sgd	6.351687397	2
tanh	0.0001	(100,)	adam	23.99993934	2
tanh	0.0001	(100,)	sgd	9.308789392	2
tanh	0.0001	(50, 50)	adam	55.97072016	2
tanh	0.0001	(50, 50)	sgd	2.444156953	2
tanh	0.001	(50,)	adam	89.16721611	2
tanh	0.001	(50,)	sgd	6.351978614	2
tanh	0.001	(100,)	adam	23.99995753	2

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.001	(100,)	sgd	9.30934177	2
tanh	0.001	(50, 50)	adam	55.97074271	2
tanh	0.001	(50, 50)	sgd	2.444220266	2
relu	0.0001	(50,)	adam	36.54017426	2
relu	0.0001	(50,)	sgd	4.551399665	2
relu	0.0001	(100,)	adam	10.17274983	2
relu	0.0001	(100,)	sgd	4.54349396	2
relu	0.0001	(50, 50)	adam	7.385877029	2
relu	0.0001	(50, 50)	sgd	2.875421538	2
relu	0.001	(50,)	adam	36.54025697	2
relu	0.001	(50,)	sgd	4.551495247	2
relu	0.001	(100,)	adam	10.17267862	2
relu	0.001	(100,)	sgd	4.543696971	2
relu	0.001	(50, 50)	adam	7.404082466	2
relu	0.001	(50, 50)	sgd	2.873865417	2
tanh	0.0001	(50,)	adam	53.08223209	3
tanh	0.0001	(50,)	sgd	1.654178751	3
tanh	0.0001	(100,)	adam	4.136375801	3
tanh	0.0001	(100,)	sgd	1.872068985	3
tanh	0.0001	(50, 50)	adam	43.00980643	3
tanh	0.0001	(50, 50)	sgd	0.776122853	3
tanh	0.001	(50,)	adam	53.08228963	3
tanh	0.001	(50,)	sgd	1.654315584	3
tanh	0.001	(100,)	adam	4.136400146	3

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
tanh	0.001	(100,)	sgd	1.872122503	3
tanh	0.001	(50, 50)	adam	43.00967779	3
tanh	0.001	(50, 50)	sgd	0.776154044	3
relu	0.0001	(50,)	adam	20.86940234	3
relu	0.0001	(50,)	sgd	0.71624002	3
relu	0.0001	(100,)	adam	5.066081538	3
relu	0.0001	(100,)	sgd	0.716003035	3
relu	0.0001	(50, 50)	adam	1.152315032	3
relu	0.0001	(50, 50)	sgd	0.75516695	3
relu	0.001	(50,)	adam	20.86947509	3
relu	0.001	(50,)	sgd	0.716214819	3
relu	0.001	(100,)	adam	5.065964511	3
relu	0.001	(100,)	sgd	0.716028894	3
relu	0.001	(50, 50)	adam	1.159944642	3
relu	0.001	(50, 50)	sgd	0.755096828	3
tanh	0.0001	(50,)	adam	76.34267402	4
tanh	0.0001	(50,)	sgd	1.179716896	4
tanh	0.0001	(100,)	adam	5.221124536	4
tanh	0.0001	(100,)	sgd	1.298151218	4
tanh	0.0001	(50, 50)	adam	44.43254228	4
tanh	0.0001	(50, 50)	sgd	0.852731241	4
tanh	0.001	(50,)	adam	76.34268757	4
tanh	0.001	(50,)	sgd	1.179755994	4
tanh	0.001	(100,)	adam	5.221163007	4

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.001	(100,)	sgd	1.298158396	4
tanh	0.001	(50, 50)	adam	44.43245085	4
tanh	0.001	(50, 50)	sgd	0.852768842	4
relu	0.0001	(50,)	adam	21.62981356	4
relu	0.0001	(50,)	sgd	0.814038802	4
relu	0.0001	(100,)	adam	7.74891819	4
relu	0.0001	(100,)	sgd	0.768225544	4
relu	0.0001	(50, 50)	adam	2.413925056	4
relu	0.0001	(50, 50)	sgd	0.710744375	4
relu	0.001	(50,)	adam	21.62857433	4
relu	0.001	(50,)	sgd	0.814091224	4
relu	0.001	(100,)	adam	7.748846094	4
relu	0.001	(100,)	sgd	0.768468261	4
relu	0.001	(50, 50)	adam	2.422532104	4
relu	0.001	(50, 50)	sgd	0.710839666	4
tanh	0.0001	(50,)	adam	54.4130063	5
tanh	0.0001	(50,)	sgd	0.781989676	5
tanh	0.0001	(100,)	adam	3.105247697	5
tanh	0.0001	(100,)	sgd	0.750503178	5
tanh	0.0001	(50, 50)	adam	41.68642074	5
tanh	0.0001	(50, 50)	sgd	0.677030133	5
tanh	0.001	(50,)	adam	54.41303126	5
tanh	0.001	(50,)	sgd	0.781992853	5
tanh	0.001	(100,)	adam	3.105280209	5

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.001	(100,)	sgd	0.750509685	5
tanh	0.001	(50, 50)	adam	41.68640857	5
tanh	0.001	(50, 50)	sgd	0.752923723	5
relu	0.0001	(50,)	adam	21.44070954	5
relu	0.0001	(50,)	sgd	0.752892917	5
relu	0.0001	(100,)	adam	7.218440262	5
relu	0.0001	(100,)	sgd	0.696892557	5
relu	0.0001	(50, 50)	adam	3.043478751	5
relu	0.0001	(50, 50)	sgd	0.725274241	5
relu	0.001	(50,)	adam	21.44080913	5
relu	0.001	(50,)	sgd	0.677637031	5
relu	0.001	(100,)	adam	7.218377429	5
relu	0.001	(100,)	sgd	0.696958541	5
relu	0.001	(50, 50)	adam	3.032745689	5
relu	0.001	(50, 50)	sgd	0.725476547	5

Table A.10 : ICA_MLP model hyperparameters tuning for different ICs using the top K = 10 external factors.

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.0001	(50,)	adam	90.96466306	1
tanh	0.0001	(50,)	sgd	22.79290927	1
tanh	0.0001	(100,)	adam	37.95981306	1
tanh	0.0001	(100,)	sgd	29.45689945	1
tanh	0.0001	(50, 50)	adam	63.86585622	1

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.0001	(50, 50)	sgd	9.318743272	1
tanh	0.001	(50,)	adam	90.96475746	1
tanh	0.001	(50,)	sgd	22.79376841	1
tanh	0.001	(100,)	adam	37.95983154	1
tanh	0.001	(100,)	sgd	29.4578545	1
tanh	0.001	(50, 50)	adam	63.86567182	1
tanh	0.001	(50, 50)	sgd	9.318843047	1
relu	0.0001	(50,)	adam	118.9986127	1
relu	0.0001	(50,)	sgd	40.039809	1
relu	0.0001	(100,)	adam	45.38720608	1
relu	0.0001	(100,)	sgd	40.22641991	1
relu	0.0001	(50, 50)	adam	32.91323871	1
relu	0.0001	(50, 50)	sgd	11.71640247	1
relu	0.001	(50,)	adam	118.9984985	1
relu	0.001	(50,)	sgd	40.0398035	1
relu	0.001	(100,)	adam	45.38718501	1
relu	0.001	(100,)	sgd	40.22641857	1
relu	0.001	(50, 50)	adam	33.20900152	1
relu	0.001	(50, 50)	sgd	11.7170773	1
tanh	0.0001	(50,)	adam	100.047067	2
tanh	0.0001	(50,)	sgd	12.29566316	2
tanh	0.0001	(100,)	adam	37.95527602	2
tanh	0.0001	(100,)	sgd	14.14708954	2
tanh	0.0001	(50, 50)	adam	46.30151742	2

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.0001	(50, 50)	sgd	6.549265277	2
tanh	0.001	(50,)	adam	100.0471032	2
tanh	0.001	(50,)	sgd	12.29612529	2
tanh	0.001	(100,)	adam	37.95529776	2
tanh	0.001	(100,)	sgd	14.14737034	2
tanh	0.001	(50, 50)	adam	46.30170152	2
tanh	0.001	(50, 50)	sgd	6.549383221	2
relu	0.0001	(50,)	adam	68.26576497	2
relu	0.0001	(50,)	sgd	11.24036743	2
relu	0.0001	(100,)	adam	37.54584211	2
relu	0.0001	(100,)	sgd	11.71451153	2
relu	0.0001	(50, 50)	adam	12.6026952	2
relu	0.0001	(50, 50)	sgd	6.982802978	2
relu	0.001	(50,)	adam	68.26579511	2
relu	0.001	(50,)	sgd	11.2408011	2
relu	0.001	(100,)	adam	37.54603111	2
relu	0.001	(100,)	sgd	11.71552903	2
relu	0.001	(50, 50)	adam	12.68618986	2
relu	0.001	(50, 50)	sgd	6.982466877	2
tanh	0.0001	(50,)	adam	57.95326333	3
tanh	0.0001	(50,)	sgd	8.726446544	3
tanh	0.0001	(100,)	adam	17.74338802	3
tanh	0.0001	(100,)	sgd	12.03707447	3
tanh	0.0001	(50, 50)	adam	43.61105194	3

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
tanh	0.0001	(50, 50)	sgd	1.82147768	3
tanh	0.001	(50,)	adam	57.95333517	3
tanh	0.001	(50,)	sgd	8.726866967	3
tanh	0.001	(100,)	adam	17.7434084	3
tanh	0.001	(100,)	sgd	12.03733038	3
tanh	0.001	(50, 50)	adam	43.61112561	3
tanh	0.001	(50, 50)	sgd	1.821556938	3
relu	0.0001	(50,)	adam	25.37213521	3
relu	0.0001	(50,)	sgd	4.968810684	3
relu	0.0001	(100,)	adam	11.04825492	3
relu	0.0001	(100,)	sgd	5.050512355	3
relu	0.0001	(50, 50)	adam	4.716603858	3
relu	0.0001	(50, 50)	sgd	3.298373267	3
relu	0.001	(50,)	adam	25.37217407	3
relu	0.001	(50,)	sgd	4.968998056	3
relu	0.001	(100,)	adam	11.04842619	3
relu	0.001	(100,)	sgd	5.050578766	3
relu	0.001	(50, 50)	adam	4.727787729	3
relu	0.001	(50, 50)	sgd	3.296781075	3
tanh	0.0001	(50,)	adam	79.73241239	4
tanh	0.0001	(50,)	sgd	2.099111761	4
tanh	0.0001	(100,)	adam	12.05864501	4
tanh	0.0001	(100,)	sgd	3.290944987	4
tanh	0.0001	(50, 50)	adam	44.8218625	4

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
tanh	0.0001	(50, 50)	sgd	1.073400758	4
tanh	0.001	(50,)	adam	79.73242753	4
tanh	0.001	(50,)	sgd	2.099310236	4
tanh	0.001	(100,)	adam	12.05869107	4
tanh	0.001	(100,)	sgd	3.291177581	4
tanh	0.001	(50, 50)	adam	44.82189068	4
tanh	0.001	(50, 50)	sgd	1.073426005	4
relu	0.0001	(50,)	adam	17.59818985	4
relu	0.0001	(50,)	sgd	1.44282529	4
relu	0.0001	(100,)	adam	7.525623903	4
relu	0.0001	(100,)	sgd	1.380176016	4
relu	0.0001	(50, 50)	adam	2.413440518	4
relu	0.0001	(50, 50)	sgd	1.234984419	4
relu	0.001	(50,)	adam	17.59791202	4
relu	0.001	(50,)	sgd	1.442836379	4
relu	0.001	(100,)	adam	7.525579021	4
relu	0.001	(100,)	sgd	1.3801892	4
relu	0.001	(50, 50)	adam	2.376041578	4
relu	0.001	(50, 50)	sgd	1.236008712	4
tanh	0.0001	(50,)	adam	56.57406103	5
tanh	0.0001	(50,)	sgd	1.453209952	5
tanh	0.0001	(100,)	adam	5.401321562	5
tanh	0.0001	(100,)	sgd	1.668341918	5
tanh	0.0001	(50, 50)	adam	44.72222204	5

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.0001	(50, 50)	sgd	0.846495845	5
tanh	0.001	(50,)	adam	56.57407928	5
tanh	0.001	(50,)	sgd	1.453270403	5
tanh	0.001	(100,)	adam	5.401355429	5
tanh	0.001	(100,)	sgd	1.668378694	5
tanh	0.001	(50, 50)	adam	44.72216086	5
tanh	0.001	(50, 50)	sgd	0.846525478	5
relu	0.0001	(50,)	adam	15.46081337	5
relu	0.0001	(50,)	sgd	0.796935192	5
relu	0.0001	(100,)	adam	5.80612643	5
relu	0.0001	(100,)	sgd	0.847439839	5
relu	0.0001	(50, 50)	adam	2.978322784	5
relu	0.0001	(50, 50)	sgd	0.79244823	5
relu	0.001	(50,)	adam	15.460774	5
relu	0.001	(50,)	sgd	0.796957665	5
relu	0.001	(100,)	adam	5.806095457	5
relu	0.001	(100,)	sgd	0.847503394	5
relu	0.001	(50, 50)	adam	3.041598492	5
relu	0.001	(50, 50)	sgd	0.779073076	5
tanh	0.0001	(50,)	adam	68.95388083	6
tanh	0.0001	(50,)	sgd	1.268275788	6
tanh	0.0001	(100,)	adam	4.416589113	6
tanh	0.0001	(100,)	sgd	1.510866603	6
tanh	0.0001	(50, 50)	adam	43.52696913	6

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.0001	(50, 50)	sgd	0.927492566	6
tanh	0.001	(50,)	adam	68.95388073	6
tanh	0.001	(50,)	sgd	1.268323695	6
tanh	0.001	(100,)	adam	4.416617311	6
tanh	0.001	(100,)	sgd	1.510889349	6
tanh	0.001	(50, 50)	adam	43.52689011	6
tanh	0.001	(50, 50)	sgd	0.927532115	6
relu	0.0001	(50,)	adam	17.18118942	6
relu	0.0001	(50,)	sgd	0.885974891	6
relu	0.0001	(100,)	adam	5.617251708	6
relu	0.0001	(100,)	sgd	0.865209418	6
relu	0.0001	(50, 50)	adam	2.548195366	6
relu	0.0001	(50, 50)	sgd	0.876008675	6
relu	0.001	(50,)	adam	17.18122448	6
relu	0.001	(50,)	sgd	0.886021605	6
relu	0.001	(100,)	adam	5.617188052	6
relu	0.001	(100,)	sgd	0.856150097	6
relu	0.001	(50, 50)	adam	2.570699115	6
relu	0.001	(50, 50)	sgd	0.876067008	6
tanh	0.0001	(50,)	adam	68.74732999	7
tanh	0.0001	(50,)	sgd	1.183655778	7
tanh	0.0001	(100,)	adam	4.570046102	7
tanh	0.0001	(100,)	sgd	1.350213926	7
tanh	0.0001	(50, 50)	adam	42.99932665	7

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.0001	(50, 50)	sgd	0.842915711	7
tanh	0.001	(50,)	adam	68.74718341	7
tanh	0.001	(50,)	sgd	1.183686938	7
tanh	0.001	(100,)	adam	4.570074361	7
tanh	0.001	(100,)	sgd	1.350253677	7
tanh	0.001	(50, 50)	adam	42.99926753	7
tanh	0.001	(50, 50)	sgd	0.842944606	7
relu	0.0001	(50,)	adam	13.25821171	7
relu	0.0001	(50,)	sgd	0.875896735	7
relu	0.0001	(100,)	adam	8.500478091	7
relu	0.0001	(100,)	sgd	0.870800542	7
relu	0.0001	(50, 50)	adam	2.667467881	7
relu	0.0001	(50, 50)	sgd	0.824708063	7
relu	0.001	(50,)	adam	13.25832565	7
relu	0.001	(50,)	sgd	0.875814502	7
relu	0.001	(100,)	adam	8.500161852	7
relu	0.001	(100,)	sgd	0.870566585	7
relu	0.001	(50, 50)	adam	2.668142394	7
relu	0.001	(50, 50)	sgd	0.825380133	7
tanh	0.0001	(50,)	adam	61.75635091	8
tanh	0.0001	(50,)	sgd	1.191714247	8
tanh	0.0001	(100,)	adam	3.728997016	8
tanh	0.0001	(100,)	sgd	1.242146008	8
tanh	0.0001	(50, 50)	adam	47.04118984	8

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
tanh	0.0001	(50, 50)	sgd	0.843549313	8
tanh	0.001	(50,)	adam	61.75635083	8
tanh	0.001	(50,)	sgd	1.191745854	8
tanh	0.001	(100,)	adam	3.729031306	8
tanh	0.001	(100,)	sgd	1.242168159	8
tanh	0.001	(50, 50)	adam	47.04113904	8
tanh	0.001	(50, 50)	sgd	0.843604819	8
relu	0.0001	(50,)	adam	20.6399632	8
relu	0.0001	(50,)	sgd	0.870640558	8
relu	0.0001	(100,)	adam	7.588282634	8
relu	0.0001	(100,)	sgd	0.89682232	8
relu	0.0001	(50, 50)	adam	3.536308791	8
relu	0.0001	(50, 50)	sgd	0.830161265	8
relu	0.001	(50,)	adam	20.63994713	8
relu	0.001	(50,)	sgd	0.870522561	8
relu	0.001	(100,)	adam	7.588198487	8
relu	0.001	(100,)	sgd	0.896790184	8
relu	0.001	(50, 50)	adam	3.590163711	8
relu	0.001	(50, 50)	sgd	0.830195561	8
tanh	0.0001	(50,)	adam	75.44402438	9
tanh	0.0001	(50,)	sgd	0.954736335	9
tanh	0.0001	(100,)	adam	2.963224915	9
tanh	0.0001	(100,)	sgd	0.960356258	9
tanh	0.0001	(50, 50)	adam	42.02344098	9

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.0001	(50, 50)	sgd	0.843094762	9
tanh	0.001	(50,)	adam	75.44365558	9
tanh	0.001	(50,)	sgd	0.954751419	9
tanh	0.001	(100,)	adam	2.963250743	9
tanh	0.001	(100,)	sgd	0.960365337	9
tanh	0.001	(50, 50)	adam	42.02342988	9
tanh	0.001	(50, 50)	sgd	0.843121533	9
relu	0.0001	(50,)	adam	15.99642695	9
relu	0.0001	(50,)	sgd	0.809906007	9
relu	0.0001	(100,)	adam	7.297819425	9
relu	0.0001	(100,)	sgd	0.84789724	9
relu	0.0001	(50, 50)	adam	2.587273551	9
relu	0.0001	(50, 50)	sgd	0.771511587	9
relu	0.001	(50,)	adam	15.99631142	9
relu	0.001	(50,)	sgd	0.80979401	9
relu	0.001	(100,)	adam	7.297721087	9
relu	0.001	(100,)	sgd	0.848336087	9
relu	0.001	(50, 50)	adam	2.574543605	9
relu	0.001	(50, 50)	sgd	0.77051071	9
tanh	0.0001	(50,)	adam	64.25256023	10
tanh	0.0001	(50,)	sgd	0.829285599	10
tanh	0.0001	(100,)	adam	3.472657367	10
tanh	0.0001	(100,)	sgd	0.724095578	10
tanh	0.0001	(50, 50)	adam	40.19669689	10

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.0001	(50, 50)	sgd	0.724090491	10
tanh	0.001	(50,)	adam	64.25250946	10
tanh	0.001	(50,)	sgd	0.829287991	10
tanh	0.001	(100,)	adam	3.472675333	10
tanh	0.001	(100,)	sgd	0.824726321	10
tanh	0.001	(50, 50)	adam	40.19672473	10
tanh	0.001	(50, 50)	sgd	0.824745669	10
relu	0.0001	(50,)	adam	16.90567251	10
relu	0.0001	(50,)	sgd	0.891833825	10
relu	0.0001	(100,)	adam	7.518162856	10
relu	0.0001	(100,)	sgd	0.781334995	10
relu	0.0001	(50, 50)	adam	3.056732499	10
relu	0.0001	(50, 50)	sgd	0.843706479	10
relu	0.001	(50,)	adam	16.90545985	10
relu	0.001	(50,)	sgd	0.863850359	10
relu	0.001	(100,)	adam	7.518150542	10
relu	0.001	(100,)	sgd	0.781016124	10
relu	0.001	(50, 50)	adam	3.056992174	10
relu	0.001	(50, 50)	sgd	0.843588435	10

Table A.11 : ICA_MLP model hyperparameters tuning for different ICs using the top K = 15 external factors.

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.0001	(50,)	adam	96.46194021	1

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
tanh	0.0001	(50,)	sgd	22.59199189	1
tanh	0.0001	(100,)	adam	45.07814036	1
tanh	0.0001	(100,)	sgd	25.7811302	1
tanh	0.0001	(50, 50)	adam	75.10988655	1
tanh	0.0001	(50, 50)	sgd	10.61368861	1
tanh	0.001	(50,)	adam	96.46202282	1
tanh	0.001	(50,)	sgd	22.59277803	1
tanh	0.001	(100,)	adam	45.07815436	1
tanh	0.001	(100,)	sgd	25.78181304	1
tanh	0.001	(50, 50)	adam	75.10969524	1
tanh	0.001	(50, 50)	sgd	10.61395201	1
relu	0.0001	(50,)	adam	100.2033614	1
relu	0.0001	(50,)	sgd	37.68887075	1
relu	0.0001	(100,)	adam	40.43320293	1
relu	0.0001	(100,)	sgd	36.95404069	1
relu	0.0001	(50, 50)	adam	34.88261494	1
relu	0.0001	(50, 50)	sgd	37.27625771	1
relu	0.001	(50,)	adam	100.2033814	1
relu	0.001	(50,)	sgd	37.68887738	1
relu	0.001	(100,)	adam	40.43324104	1
relu	0.001	(100,)	sgd	36.95404908	1
relu	0.001	(50, 50)	adam	34.65577024	1
relu	0.001	(50, 50)	sgd	37.27625861	1
tanh	0.0001	(50,)	adam	104.0065895	2

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.0001	(50,)	sgd	12.50427333	2
tanh	0.0001	(100,)	adam	40.39840417	2
tanh	0.0001	(100,)	sgd	12.85417417	2
tanh	0.0001	(50, 50)	adam	48.50755668	2
tanh	0.0001	(50, 50)	sgd	7.401410719	2
tanh	0.001	(50,)	adam	104.0066133	2
tanh	0.001	(50,)	sgd	12.50444942	2
tanh	0.001	(100,)	adam	40.39842909	2
tanh	0.001	(100,)	sgd	12.85433643	2
tanh	0.001	(50, 50)	adam	48.50784584	2
tanh	0.001	(50, 50)	sgd	7.401453231	2
relu	0.0001	(50,)	adam	71.22025597	2
relu	0.0001	(50,)	sgd	17.87357873	2
relu	0.0001	(100,)	adam	42.94199158	2
relu	0.0001	(100,)	sgd	19.73796009	2
relu	0.0001	(50, 50)	adam	16.11894343	2
relu	0.0001	(50, 50)	sgd	7.570011333	2
relu	0.001	(50,)	adam	71.22024137	2
relu	0.001	(50,)	sgd	17.8737872	2
relu	0.001	(100,)	adam	42.94169384	2
relu	0.001	(100,)	sgd	19.73845877	2
relu	0.001	(50, 50)	adam	15.85703384	2
relu	0.001	(50, 50)	sgd	7.57095175	2
tanh	0.0001	(50,)	adam	65.72327648	3

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.0001	(50,)	sgd	5.375846272	3
tanh	0.0001	(100,)	adam	26.57272135	3
tanh	0.0001	(100,)	sgd	7.725204284	3
tanh	0.0001	(50, 50)	adam	43.33323657	3
tanh	0.0001	(50, 50)	sgd	1.735482608	3
tanh	0.001	(50,)	adam	65.72333668	3
tanh	0.001	(50,)	sgd	5.37625166	3
tanh	0.001	(100,)	adam	26.57275276	3
tanh	0.001	(100,)	sgd	7.725567963	3
tanh	0.001	(50, 50)	adam	43.33329584	3
tanh	0.001	(50, 50)	sgd	1.735621894	3
relu	0.0001	(50,)	adam	25.31554863	3
relu	0.0001	(50,)	sgd	8.768288172	3
relu	0.0001	(100,)	adam	13.14693796	3
relu	0.0001	(100,)	sgd	8.791640191	3
relu	0.0001	(50, 50)	adam	7.788545614	3
relu	0.0001	(50, 50)	sgd	1.915986441	3
relu	0.001	(50,)	adam	25.31564828	3
relu	0.001	(50,)	sgd	8.768237919	3
relu	0.001	(100,)	adam	13.14694668	3
relu	0.001	(100,)	sgd	8.791720586	3
relu	0.001	(50, 50)	adam	8.103193527	3
relu	0.001	(50, 50)	sgd	1.907599176	3
tanh	0.0001	(50,)	adam	82.27945738	4

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.0001	(50,)	sgd	5.044143717	4
tanh	0.0001	(100,)	adam	16.73896886	4
tanh	0.0001	(100,)	sgd	6.864311179	4
tanh	0.0001	(50, 50)	adam	45.17618763	4
tanh	0.0001	(50, 50)	sgd	1.745418512	4
tanh	0.001	(50,)	adam	82.27947421	4
tanh	0.001	(50,)	sgd	5.044599323	4
tanh	0.001	(100,)	adam	16.73902695	4
tanh	0.001	(100,)	sgd	6.864847707	4
tanh	0.001	(50, 50)	adam	45.17628379	4
tanh	0.001	(50, 50)	sgd	1.745538634	4
relu	0.0001	(50,)	adam	25.60865469	4
relu	0.0001	(50,)	sgd	4.285192038	4
relu	0.0001	(100,)	adam	10.93706372	4
relu	0.0001	(100,)	sgd	4.286470129	4
relu	0.0001	(50, 50)	adam	7.598159299	4
relu	0.0001	(50, 50)	sgd	2.00110558	4
relu	0.001	(50,)	adam	25.60862359	4
relu	0.001	(50,)	sgd	4.285333012	4
relu	0.001	(100,)	adam	10.94404596	4
relu	0.001	(100,)	sgd	4.286477176	4
relu	0.001	(50, 50)	adam	7.662549074	4
relu	0.001	(50, 50)	sgd	2.000237504	4
tanh	0.0001	(50,)	adam	60.78168083	5

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
tanh	0.0001	(50,)	sgd	4.073660694	5
tanh	0.0001	(100,)	adam	15.41987459	5
tanh	0.0001	(100,)	sgd	4.503339521	5
tanh	0.0001	(50, 50)	adam	47.05348087	5
tanh	0.0001	(50, 50)	sgd	1.226522783	5
tanh	0.001	(50,)	adam	60.78172089	5
tanh	0.001	(50,)	sgd	4.073943489	5
tanh	0.001	(100,)	adam	15.41991321	5
tanh	0.001	(100,)	sgd	4.503646392	5
tanh	0.001	(50, 50)	adam	47.05355941	5
tanh	0.001	(50, 50)	sgd	1.226605692	5
relu	0.0001	(50,)	adam	17.94231906	5
relu	0.0001	(50,)	sgd	3.308402231	5
relu	0.0001	(100,)	adam	10.04413027	5
relu	0.0001	(100,)	sgd	3.082981001	5
relu	0.0001	(50, 50)	adam	4.007351638	5
relu	0.0001	(50, 50)	sgd	1.218605345	5
relu	0.001	(50,)	adam	17.94248779	5
relu	0.001	(50,)	sgd	3.308676441	5
relu	0.001	(100,)	adam	10.04410354	5
relu	0.001	(100,)	sgd	3.083060014	5
relu	0.001	(50, 50)	adam	4.018984935	5
relu	0.001	(50, 50)	sgd	1.218680325	5
tanh	0.0001	(50,)	adam	77.34078164	6

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
tanh	0.0001	(50,)	sgd	2.139051021	6
tanh	0.0001	(100,)	adam	13.22793998	6
tanh	0.0001	(100,)	sgd	2.490513918	6
tanh	0.0001	(50, 50)	adam	44.05530783	6
tanh	0.0001	(50, 50)	sgd	1.088352341	6
tanh	0.001	(50,)	adam	77.34066204	6
tanh	0.001	(50,)	sgd	2.139232743	6
tanh	0.001	(100,)	adam	13.22797682	6
tanh	0.001	(100,)	sgd	2.490688593	6
tanh	0.001	(50, 50)	adam	44.0553321	6
tanh	0.001	(50, 50)	sgd	1.088379351	6
relu	0.0001	(50,)	adam	14.99691032	6
relu	0.0001	(50,)	sgd	2.167077545	6
relu	0.0001	(100,)	adam	9.163032988	6
relu	0.0001	(100,)	sgd	2.118515366	6
relu	0.0001	(50, 50)	adam	3.352126096	6
relu	0.0001	(50, 50)	sgd	1.415093293	6
relu	0.001	(50,)	adam	14.99703279	6
relu	0.001	(50,)	sgd	2.16706212	6
relu	0.001	(100,)	adam	9.16300434	6
relu	0.001	(100,)	sgd	2.118463494	6
relu	0.001	(50, 50)	adam	3.352112566	6
relu	0.001	(50, 50)	sgd	1.415443092	6
tanh	0.0001	(50,)	adam	73.65148217	7

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.0001	(50,)	sgd	1.39267671	7
tanh	0.0001	(100,)	adam	9.133248053	7
tanh	0.0001	(100,)	sgd	1.728613449	7
tanh	0.0001	(50, 50)	adam	42.76460815	7
tanh	0.0001	(50, 50)	sgd	1.023687262	7
tanh	0.001	(50,)	adam	73.65136178	7
tanh	0.001	(50,)	sgd	1.39282104	7
tanh	0.001	(100,)	adam	9.133312727	7
tanh	0.001	(100,)	sgd	1.728782248	7
tanh	0.001	(50, 50)	adam	42.76469675	7
tanh	0.001	(50, 50)	sgd	1.023723473	7
relu	0.0001	(50,)	adam	18.36007336	7
relu	0.0001	(50,)	sgd	1.353140211	7
relu	0.0001	(100,)	adam	6.76807071	7
relu	0.0001	(100,)	sgd	1.411945216	7
relu	0.0001	(50, 50)	adam	2.358112596	7
relu	0.0001	(50, 50)	sgd	1.151533261	7
relu	0.001	(50,)	adam	18.36011731	7
relu	0.001	(50,)	sgd	1.353159752	7
relu	0.001	(100,)	adam	6.767498564	7
relu	0.001	(100,)	sgd	1.411889268	7
relu	0.001	(50, 50)	adam	2.366026852	7
relu	0.001	(50, 50)	sgd	1.151887507	7
tanh	0.0001	(50,)	adam	60.07652565	8

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
tanh	0.0001	(50,)	sgd	1.166009604	8
tanh	0.0001	(100,)	adam	5.987843114	8
tanh	0.0001	(100,)	sgd	1.243887516	8
tanh	0.0001	(50, 50)	adam	47.77776957	8
tanh	0.0001	(50, 50)	sgd	0.997475399	8
tanh	0.001	(50,)	adam	60.0764992	8
tanh	0.001	(50,)	sgd	1.166062545	8
tanh	0.001	(100,)	adam	5.987874889	8
tanh	0.001	(100,)	sgd	1.243970587	8
tanh	0.001	(50, 50)	adam	47.7777895	8
tanh	0.001	(50, 50)	sgd	0.99748001	8
relu	0.0001	(50,)	adam	9.815689678	8
relu	0.0001	(50,)	sgd	1.050397716	8
relu	0.0001	(100,)	adam	4.071476382	8
relu	0.0001	(100,)	sgd	1.101435291	8
relu	0.0001	(50, 50)	adam	2.12105984	8
relu	0.0001	(50, 50)	sgd	1.00087253	8
relu	0.001	(50,)	adam	9.815656152	8
relu	0.001	(50,)	sgd	1.050402742	8
relu	0.001	(100,)	adam	4.071382036	8
relu	0.001	(100,)	sgd	1.101019404	8
relu	0.001	(50, 50)	adam	2.116745649	8
relu	0.001	(50, 50)	sgd	1.000865739	8
tanh	0.0001	(50,)	adam	70.64395922	9

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
tanh	0.0001	(50,)	sgd	1.074152059	9
tanh	0.0001	(100,)	adam	4.571698772	9
tanh	0.0001	(100,)	sgd	1.101309059	9
tanh	0.0001	(50, 50)	adam	41.42566756	9
tanh	0.0001	(50, 50)	sgd	1.029667027	9
tanh	0.001	(50,)	adam	70.64386703	9
tanh	0.001	(50,)	sgd	1.074161604	9
tanh	0.001	(100,)	adam	4.571713981	9
tanh	0.001	(100,)	sgd	1.101336471	9
tanh	0.001	(50, 50)	adam	41.4256705	9
tanh	0.001	(50, 50)	sgd	1.029673992	9
relu	0.0001	(50,)	adam	7.859373822	9
relu	0.0001	(50,)	sgd	0.966801673	9
relu	0.0001	(100,)	adam	4.809509721	9
relu	0.0001	(100,)	sgd	1.057169058	9
relu	0.0001	(50, 50)	adam	2.241065808	9
relu	0.0001	(50, 50)	sgd	0.997128953	9
relu	0.001	(50,)	adam	7.859377292	9
relu	0.001	(50,)	sgd	0.966729434	9
relu	0.001	(100,)	adam	4.8094205	9
relu	0.001	(100,)	sgd	1.057207407	9
relu	0.001	(50, 50)	adam	2.240292112	9
relu	0.001	(50, 50)	sgd	0.997107293	9
tanh	0.0001	(50,)	adam	65.24631209	10

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.0001	(50,)	sgd	0.998638764	10
tanh	0.0001	(100,)	adam	4.725382907	10
tanh	0.0001	(100,)	sgd	1.008460248	10
tanh	0.0001	(50, 50)	adam	41.24021371	10
tanh	0.0001	(50, 50)	sgd	1.090400314	10
tanh	0.001	(50,)	adam	65.24638697	10
tanh	0.001	(50,)	sgd	0.998659824	10
tanh	0.001	(100,)	adam	4.725407704	10
tanh	0.001	(100,)	sgd	1.008486061	10
tanh	0.001	(50, 50)	adam	41.2402742	10
tanh	0.001	(50, 50)	sgd	1.09038286	10
relu	0.0001	(50,)	adam	13.17561501	10
relu	0.0001	(50,)	sgd	1.022337041	10
relu	0.0001	(100,)	adam	4.880451067	10
relu	0.0001	(100,)	sgd	0.953270149	10
relu	0.0001	(50, 50)	adam	2.531022269	10
relu	0.0001	(50, 50)	sgd	1.006260302	10
relu	0.001	(50,)	adam	13.17564277	10
relu	0.001	(50,)	sgd	1.02226148	10
relu	0.001	(100,)	adam	4.88041972	10
relu	0.001	(100,)	sgd	0.95535964	10
relu	0.001	(50, 50)	adam	2.530654039	10
relu	0.001	(50, 50)	sgd	1.006079194	10
tanh	0.0001	(50,)	adam	60.12461029	11

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
tanh	0.0001	(50,)	sgd	0.962402412	11
tanh	0.0001	(100,)	adam	4.198070373	11
tanh	0.0001	(100,)	sgd	0.986863174	11
tanh	0.0001	(50, 50)	adam	41.69146209	11
tanh	0.0001	(50, 50)	sgd	0.886664614	11
tanh	0.001	(50,)	adam	60.12436196	11
tanh	0.001	(50,)	sgd	0.962426304	11
tanh	0.001	(100,)	adam	4.19809423	11
tanh	0.001	(100,)	sgd	0.98689069	11
tanh	0.001	(50, 50)	adam	41.69146607	11
tanh	0.001	(50, 50)	sgd	0.886687688	11
relu	0.0001	(50,)	adam	18.73878961	11
relu	0.0001	(50,)	sgd	0.841597029	11
relu	0.0001	(100,)	adam	5.198159715	11
relu	0.0001	(100,)	sgd	0.902951845	11
relu	0.0001	(50, 50)	adam	2.616819863	11
relu	0.0001	(50, 50)	sgd	0.918191691	11
relu	0.001	(50,)	adam	18.73545596	11
relu	0.001	(50,)	sgd	0.841555512	11
relu	0.001	(100,)	adam	5.198095928	11
relu	0.001	(100,)	sgd	0.902911004	11
relu	0.001	(50, 50)	adam	2.61637388	11
relu	0.001	(50, 50)	sgd	0.91815051	11
tanh	0.0001	(50,)	adam	62.88470878	12

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.0001	(50,)	sgd	0.815547166	12
tanh	0.0001	(100,)	adam	3.386045102	12
tanh	0.0001	(100,)	sgd	0.739510124	12
tanh	0.0001	(50, 50)	adam	41.63872059	12
tanh	0.0001	(50, 50)	sgd	0.70769267	12
tanh	0.001	(50,)	adam	62.88453314	12
tanh	0.001	(50,)	sgd	0.815544125	12
tanh	0.001	(100,)	adam	3.386066821	12
tanh	0.001	(100,)	sgd	0.739510252	12
tanh	0.001	(50, 50)	adam	41.6387746	12
tanh	0.001	(50, 50)	sgd	0.707714659	12
relu	0.0001	(50,)	adam	16.56917481	12
relu	0.0001	(50,)	sgd	0.843867269	12
relu	0.0001	(100,)	adam	4.833030326	12
relu	0.0001	(100,)	sgd	0.853220443	12
relu	0.0001	(50, 50)	adam	2.602500173	12
relu	0.0001	(50, 50)	sgd	0.762672162	12
relu	0.001	(50,)	adam	16.53939747	12
relu	0.001	(50,)	sgd	0.843549865	12
relu	0.001	(100,)	adam	4.832982447	12
relu	0.001	(100,)	sgd	0.853304817	12
relu	0.001	(50, 50)	adam	2.602916905	12
relu	0.001	(50, 50)	sgd	0.762768712	12
tanh	0.0001	(50,)	adam	75.76902299	13

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
tanh	0.0001	(50,)	sgd	0.722063209	13
tanh	0.0001	(100,)	adam	2.667301942	13
tanh	0.0001	(100,)	sgd	0.743593731	13
tanh	0.0001	(50, 50)	adam	40.05104878	13
tanh	0.0001	(50, 50)	sgd	0.668557209	13
tanh	0.001	(50,)	adam	75.76884681	13
tanh	0.001	(50,)	sgd	0.722071633	13
tanh	0.001	(100,)	adam	2.66731547	13
tanh	0.001	(100,)	sgd	0.743595387	13
tanh	0.001	(50, 50)	adam	40.05110732	13
tanh	0.001	(50, 50)	sgd	0.668583329	13
relu	0.0001	(50,)	adam	9.65363528	13
relu	0.0001	(50,)	sgd	0.84620994	13
relu	0.0001	(100,)	adam	4.492845829	13
relu	0.0001	(100,)	sgd	0.872224157	13
relu	0.0001	(50, 50)	adam	3.264721937	13
relu	0.0001	(50, 50)	sgd	0.873599553	13
relu	0.001	(50,)	adam	9.653899671	13
relu	0.001	(50,)	sgd	0.846241474	13
relu	0.001	(100,)	adam	4.492835526	13
relu	0.001	(100,)	sgd	0.872160468	13
relu	0.001	(50, 50)	adam	3.242513318	13
relu	0.001	(50, 50)	sgd	0.877868652	13
tanh	0.0001	(50,)	adam	57.10848355	14

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.0001	(50,)	sgd	0.764851132	14
tanh	0.0001	(100,)	adam	2.878373812	14
tanh	0.0001	(100,)	sgd	0.697374258	14
tanh	0.0001	(50, 50)	adam	42.79848578	14
tanh	0.0001	(50, 50)	sgd	0.631338381	14
tanh	0.001	(50,)	adam	57.10825049	14
tanh	0.001	(50,)	sgd	0.764839307	14
tanh	0.001	(100,)	adam	2.878395845	14
tanh	0.001	(100,)	sgd	0.69737553	14
tanh	0.001	(50, 50)	adam	42.79852795	14
tanh	0.001	(50, 50)	sgd	0.631359632	14
relu	0.0001	(50,)	adam	8.346585437	14
relu	0.0001	(50,)	sgd	1.047973848	14
relu	0.0001	(100,)	adam	5.343744001	14
relu	0.0001	(100,)	sgd	0.909904061	14
relu	0.0001	(50, 50)	adam	3.217706529	14
relu	0.0001	(50, 50)	sgd	0.833890494	14
relu	0.001	(50,)	adam	8.346683374	14
relu	0.001	(50,)	sgd	1.047773595	14
relu	0.001	(100,)	adam	5.343731965	14
relu	0.001	(100,)	sgd	0.909889365	14
relu	0.001	(50, 50)	adam	3.218897153	14
relu	0.001	(50, 50)	sgd	0.833815821	14
tanh	0.0001	(50,)	adam	65.91481518	15

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
tanh	0.0001	(50,)	sgd	0.831994663	15
tanh	0.0001	(100,)	adam	3.22101054	15
tanh	0.0001	(100,)	sgd	0.649293436	15
tanh	0.0001	(50, 50)	adam	39.64232592	15
tanh	0.0001	(50, 50)	sgd	0.708013502	15
tanh	0.001	(50,)	adam	65.91448907	15
tanh	0.001	(50,)	sgd	0.831981043	15
tanh	0.001	(100,)	adam	3.221020911	15
tanh	0.001	(100,)	sgd	0.649295222	15
tanh	0.001	(50, 50)	adam	39.64235349	15
tanh	0.001	(50, 50)	sgd	0.708043575	15
relu	0.0001	(50,)	adam	10.05280506	15
relu	0.0001	(50,)	sgd	1.189077067	15
relu	0.0001	(100,)	adam	4.198404029	15
relu	0.0001	(100,)	sgd	1.096081122	15
relu	0.0001	(50, 50)	adam	3.574927604	15
relu	0.0001	(50, 50)	sgd	0.866017341	15
relu	0.001	(50,)	adam	10.05284325	15
relu	0.001	(50,)	sgd	1.189017114	15
relu	0.001	(100,)	adam	4.198261088	15
relu	0.001	(100,)	sgd	1.096121078	15
relu	0.001	(50, 50)	adam	3.574637176	15
relu	0.001	(50, 50)	sgd	0.865752847	15

Table A.12 : ICA_MLP model hyperparameters tuning for different ICs using all K = 22 external factors.

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
tanh	0.0001	(50,)	adam	102.6124902	1
tanh	0.0001	(50,)	sgd	25.52399399	1
tanh	0.0001	(100,)	adam	52.33085943	1
tanh	0.0001	(100,)	sgd	25.94875085	1
tanh	0.0001	(50, 50)	adam	81.07806356	1
tanh	0.0001	(50, 50)	sgd	18.61135378	1
tanh	0.001	(50,)	adam	102.6125657	1
tanh	0.001	(50,)	sgd	25.52435122	1
tanh	0.001	(100,)	adam	52.3308725	1
tanh	0.001	(100,)	sgd	25.94901705	1
tanh	0.001	(50, 50)	adam	81.07805292	1
tanh	0.001	(50, 50)	sgd	18.6116496	1
relu	0.0001	(50,)	adam	96.15378283	1
relu	0.0001	(50,)	sgd	32.44420928	1
relu	0.0001	(100,)	adam	37.44621814	1
relu	0.0001	(100,)	sgd	31.67906396	1
relu	0.0001	(50, 50)	adam	31.30674074	1
relu	0.0001	(50, 50)	sgd	31.36545137	1
relu	0.001	(50,)	adam	96.15378733	1
relu	0.001	(50,)	sgd	32.44420926	1
relu	0.001	(100,)	adam	37.44625223	1
relu	0.001	(100,)	sgd	31.6790655	1

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
relu	0.001	(50, 50)	adam	31.14037123	1
relu	0.001	(50, 50)	sgd	31.36555638	1
tanh	0.0001	(50,)	adam	104.8336453	2
tanh	0.0001	(50,)	sgd	11.43791095	2
tanh	0.0001	(100,)	adam	41.08681996	2
tanh	0.0001	(100,)	sgd	12.03473389	2
tanh	0.0001	(50, 50)	adam	50.4601361	2
tanh	0.0001	(50, 50)	sgd	8.119107726	2
tanh	0.001	(50,)	adam	104.8336663	2
tanh	0.001	(50,)	sgd	11.43811336	2
tanh	0.001	(100,)	adam	41.08684395	2
tanh	0.001	(100,)	sgd	12.03492257	2
tanh	0.001	(50, 50)	adam	50.460409	2
tanh	0.001	(50, 50)	sgd	8.119153468	2
relu	0.0001	(50,)	adam	71.02580475	2
relu	0.0001	(50,)	sgd	16.65580711	2
relu	0.0001	(100,)	adam	42.28975935	2
relu	0.0001	(100,)	sgd	17.88115091	2
relu	0.0001	(50, 50)	adam	15.96906265	2
relu	0.0001	(50, 50)	sgd	7.24032454	2
relu	0.001	(50,)	adam	71.0258418	2
relu	0.001	(50,)	sgd	16.65627054	2
relu	0.001	(100,)	adam	41.65675784	2
relu	0.001	(100,)	sgd	17.88166004	2

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
relu	0.001	(50, 50)	adam	15.90623198	2
relu	0.001	(50, 50)	sgd	7.241141332	2
tanh	0.0001	(50,)	adam	64.77861024	3
tanh	0.0001	(50,)	sgd	9.654615261	3
tanh	0.0001	(100,)	adam	26.84127803	3
tanh	0.0001	(100,)	sgd	12.16272734	3
tanh	0.0001	(50, 50)	adam	46.19896732	3
tanh	0.0001	(50, 50)	sgd	4.958709574	3
tanh	0.001	(50,)	adam	64.77866746	3
tanh	0.001	(50,)	sgd	9.655058323	3
tanh	0.001	(100,)	adam	26.8413145	3
tanh	0.001	(100,)	sgd	12.16331174	3
tanh	0.001	(50, 50)	adam	46.19911217	3
tanh	0.001	(50, 50)	sgd	4.959012072	3
relu	0.0001	(50,)	adam	48.19474186	3
relu	0.0001	(50,)	sgd	12.46097861	3
relu	0.0001	(100,)	adam	19.55182443	3
relu	0.0001	(100,)	sgd	11.9903297	3
relu	0.0001	(50, 50)	adam	12.54052888	3
relu	0.0001	(50, 50)	sgd	7.932647052	3
relu	0.001	(50,)	adam	48.19463756	3
relu	0.001	(50,)	sgd	12.46143087	3
relu	0.001	(100,)	adam	19.5520213	3
relu	0.001	(100,)	sgd	11.99054416	3

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
relu	0.001	(50, 50)	adam	12.47291201	3
relu	0.001	(50, 50)	sgd	7.935142404	3
tanh	0.0001	(50,)	adam	81.15277011	4
tanh	0.0001	(50,)	sgd	7.942890803	4
tanh	0.0001	(100,)	adam	20.78203644	4
tanh	0.0001	(100,)	sgd	9.119489501	4
tanh	0.0001	(50, 50)	adam	45.10579991	4
tanh	0.0001	(50, 50)	sgd	3.380331634	4
tanh	0.001	(50,)	adam	81.15282489	4
tanh	0.001	(50,)	sgd	7.943363228	4
tanh	0.001	(100,)	adam	20.78207788	4
tanh	0.001	(100,)	sgd	9.119889185	4
tanh	0.001	(50, 50)	adam	45.10593115	4
tanh	0.001	(50, 50)	sgd	3.380636632	4
relu	0.0001	(50,)	adam	28.68090592	4
relu	0.0001	(50,)	sgd	11.20488214	4
relu	0.0001	(100,)	adam	16.64733231	4
relu	0.0001	(100,)	sgd	11.77404007	4
relu	0.0001	(50, 50)	adam	10.17631259	4
relu	0.0001	(50, 50)	sgd	8.635790602	4
relu	0.001	(50,)	adam	28.6812981	4
relu	0.001	(50,)	sgd	11.20480432	4
relu	0.001	(100,)	adam	16.64735538	4
relu	0.001	(100,)	sgd	11.77410474	4

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
relu	0.001	(50, 50)	adam	10.26778118	4
relu	0.001	(50, 50)	sgd	8.6300652	4
tanh	0.0001	(50,)	adam	58.99344174	5
tanh	0.0001	(50,)	sgd	4.201864945	5
tanh	0.0001	(100,)	adam	16.74150491	5
tanh	0.0001	(100,)	sgd	4.387182579	5
tanh	0.0001	(50, 50)	adam	46.86622123	5
tanh	0.0001	(50, 50)	sgd	2.044829495	5
tanh	0.001	(50,)	adam	58.99351777	5
tanh	0.001	(50,)	sgd	4.202068534	5
tanh	0.001	(100,)	adam	16.74154855	5
tanh	0.001	(100,)	sgd	4.387344005	5
tanh	0.001	(50, 50)	adam	46.86627916	5
tanh	0.001	(50, 50)	sgd	2.044998727	5
relu	0.0001	(50,)	adam	17.7731275	5
relu	0.0001	(50,)	sgd	4.160646816	5
relu	0.0001	(100,)	adam	8.421908952	5
relu	0.0001	(100,)	sgd	3.89726743	5
relu	0.0001	(50, 50)	adam	4.453731955	5
relu	0.0001	(50, 50)	sgd	2.933882506	5
relu	0.001	(50,)	adam	17.7733155	5
relu	0.001	(50,)	sgd	4.16074642	5
relu	0.001	(100,)	adam	8.421894114	5
relu	0.001	(100,)	sgd	3.897347208	5

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
relu	0.001	(50, 50)	adam	4.44704445	5
relu	0.001	(50, 50)	sgd	2.934016314	5
tanh	0.0001	(50,)	adam	74.89698806	6
tanh	0.0001	(50,)	sgd	3.230564331	6
tanh	0.0001	(100,)	adam	15.43749246	6
tanh	0.0001	(100,)	sgd	3.708362556	6
tanh	0.0001	(50, 50)	adam	44.12584241	6
tanh	0.0001	(50, 50)	sgd	1.485333766	6
tanh	0.001	(50,)	adam	74.8969727	6
tanh	0.001	(50,)	sgd	3.230781462	6
tanh	0.001	(100,)	adam	15.43756786	6
tanh	0.001	(100,)	sgd	3.708558983	6
tanh	0.001	(50, 50)	adam	44.1260786	6
tanh	0.001	(50, 50)	sgd	1.48546225	6
relu	0.0001	(50,)	adam	19.8071939	6
relu	0.0001	(50,)	sgd	3.198987011	6
relu	0.0001	(100,)	adam	8.41585575	6
relu	0.0001	(100,)	sgd	3.17542523	6
relu	0.0001	(50, 50)	adam	5.419796508	6
relu	0.0001	(50, 50)	sgd	1.895696566	6
relu	0.001	(50,)	adam	19.80736974	6
relu	0.001	(50,)	sgd	3.199369279	6
relu	0.001	(100,)	adam	8.415661097	6
relu	0.001	(100,)	sgd	3.175387707	6

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
relu	0.001	(50, 50)	adam	5.42283397	6
relu	0.001	(50, 50)	sgd	1.896980718	6
tanh	0.0001	(50,)	adam	78.95790741	7
tanh	0.0001	(50,)	sgd	3.343391494	7
tanh	0.0001	(100,)	adam	13.99144453	7
tanh	0.0001	(100,)	sgd	3.617452994	7
tanh	0.0001	(50, 50)	adam	44.28406868	7
tanh	0.0001	(50, 50)	sgd	1.815238432	7
tanh	0.001	(50,)	adam	78.95786318	7
tanh	0.001	(50,)	sgd	3.343596488	7
tanh	0.001	(100,)	adam	13.99150191	7
tanh	0.001	(100,)	sgd	3.617672342	7
tanh	0.001	(50, 50)	adam	44.28415366	7
tanh	0.001	(50, 50)	sgd	1.815365153	7
relu	0.0001	(50,)	adam	18.33703047	7
relu	0.0001	(50,)	sgd	3.130967007	7
relu	0.0001	(100,)	adam	10.09041544	7
relu	0.0001	(100,)	sgd	3.010271288	7
relu	0.0001	(50, 50)	adam	4.75744829	7
relu	0.0001	(50, 50)	sgd	1.962784107	7
relu	0.001	(50,)	adam	18.33782183	7
relu	0.001	(50,)	sgd	3.130880217	7
relu	0.001	(100,)	adam	10.09313902	7
relu	0.001	(100,)	sgd	3.010218947	7

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
relu	0.001	(50, 50)	adam	4.781663529	7
relu	0.001	(50, 50)	sgd	1.96269448	7
tanh	0.0001	(50,)	adam	57.70466057	8
tanh	0.0001	(50,)	sgd	1.723562429	8
tanh	0.0001	(100,)	adam	9.464561466	8
tanh	0.0001	(100,)	sgd	1.924270639	8
tanh	0.0001	(50, 50)	adam	48.58394892	8
tanh	0.0001	(50, 50)	sgd	1.230171507	8
tanh	0.001	(50,)	adam	57.70458902	8
tanh	0.001	(50,)	sgd	1.72368614	8
tanh	0.001	(100,)	adam	9.464594694	8
tanh	0.001	(100,)	sgd	1.924420885	8
tanh	0.001	(50, 50)	adam	48.58397175	8
tanh	0.001	(50, 50)	sgd	1.230181983	8
relu	0.0001	(50,)	adam	10.1635331	8
relu	0.0001	(50,)	sgd	1.827649235	8
relu	0.0001	(100,)	adam	6.1847094	8
relu	0.0001	(100,)	sgd	1.75640138	8
relu	0.0001	(50, 50)	adam	4.544845501	8
relu	0.0001	(50, 50)	sgd	1.209767712	8
relu	0.001	(50,)	adam	10.16353359	8
relu	0.001	(50,)	sgd	1.827551191	8
relu	0.001	(100,)	adam	6.184639091	8
relu	0.001	(100,)	sgd	1.756138862	8

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
relu	0.001	(50, 50)	adam	4.542376417	8
relu	0.001	(50, 50)	sgd	1.209708656	8
tanh	0.0001	(50,)	adam	73.69499002	9
tanh	0.0001	(50,)	sgd	1.53659403	9
tanh	0.0001	(100,)	adam	9.970145785	9
tanh	0.0001	(100,)	sgd	1.742161621	9
tanh	0.0001	(50, 50)	adam	43.61744268	9
tanh	0.0001	(50, 50)	sgd	1.069476648	9
tanh	0.001	(50,)	adam	73.69490919	9
tanh	0.001	(50,)	sgd	1.536698994	9
tanh	0.001	(100,)	adam	9.970157087	9
tanh	0.001	(100,)	sgd	1.742309047	9
tanh	0.001	(50, 50)	adam	43.61743779	9
tanh	0.001	(50, 50)	sgd	1.069500869	9
relu	0.0001	(50,)	adam	9.616885969	9
relu	0.0001	(50,)	sgd	1.668566091	9
relu	0.0001	(100,)	adam	6.115117987	9
relu	0.0001	(100,)	sgd	1.647153453	9
relu	0.0001	(50, 50)	adam	3.576474807	9
relu	0.0001	(50, 50)	sgd	1.31263282	9
relu	0.001	(50,)	adam	9.616952385	9
relu	0.001	(50,)	sgd	1.668214299	9
relu	0.001	(100,)	adam	6.115046896	9
relu	0.001	(100,)	sgd	1.647292558	9

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
relu	0.001	(50, 50)	adam	3.575936734	9
relu	0.001	(50, 50)	sgd	1.311802509	9
tanh	0.0001	(50,)	adam	64.32448759	10
tanh	0.0001	(50,)	sgd	1.41966551	10
tanh	0.0001	(100,)	adam	9.973645869	10
tanh	0.0001	(100,)	sgd	1.608835918	10
tanh	0.0001	(50, 50)	adam	42.26569543	10
tanh	0.0001	(50, 50)	sgd	1.081927445	10
tanh	0.001	(50,)	adam	64.32442046	10
tanh	0.001	(50,)	sgd	1.419760984	10
tanh	0.001	(100,)	adam	9.97366022	10
tanh	0.001	(100,)	sgd	1.608950093	10
tanh	0.001	(50, 50)	adam	42.26576616	10
tanh	0.001	(50, 50)	sgd	1.081942336	10
relu	0.0001	(50,)	adam	16.34315908	10
relu	0.0001	(50,)	sgd	1.426315941	10
relu	0.0001	(100,)	adam	7.675963406	10
relu	0.0001	(100,)	sgd	1.408554999	10
relu	0.0001	(50, 50)	adam	3.554897273	10
relu	0.0001	(50, 50)	sgd	1.221068237	10
relu	0.001	(50,)	adam	16.34352818	10
relu	0.001	(50,)	sgd	1.426254249	10
relu	0.001	(100,)	adam	7.675932183	10
relu	0.001	(100,)	sgd	1.408493331	10

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
relu	0.001	(50, 50)	adam	3.557411638	10
relu	0.001	(50, 50)	sgd	1.221238504	10
tanh	0.0001	(50,)	adam	57.21495508	11
tanh	0.0001	(50,)	sgd	0.950900557	11
tanh	0.0001	(100,)	adam	5.352194403	11
tanh	0.0001	(100,)	sgd	1.067211411	11
tanh	0.0001	(50, 50)	adam	40.48465238	11
tanh	0.0001	(50, 50)	sgd	0.883754375	11
tanh	0.001	(50,)	adam	57.21458919	11
tanh	0.001	(50,)	sgd	0.950953246	11
tanh	0.001	(100,)	adam	5.352208932	11
tanh	0.001	(100,)	sgd	1.067277998	11
tanh	0.001	(50, 50)	adam	40.48469518	11
tanh	0.001	(50, 50)	sgd	0.883762473	11
relu	0.0001	(50,)	adam	16.10529848	11
relu	0.0001	(50,)	sgd	0.97322471	11
relu	0.0001	(100,)	adam	4.268572866	11
relu	0.0001	(100,)	sgd	1.045689458	11
relu	0.0001	(50, 50)	adam	2.985528086	11
relu	0.0001	(50, 50)	sgd	0.991417036	11
relu	0.001	(50,)	adam	16.10557925	11
relu	0.001	(50,)	sgd	0.973285594	11
relu	0.001	(100,)	adam	4.268550014	11
relu	0.001	(100,)	sgd	1.045516653	11

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
relu	0.001	(50, 50)	adam	2.984439124	11
relu	0.001	(50, 50)	sgd	0.991446724	11
tanh	0.0001	(50,)	adam	61.20551264	12
tanh	0.0001	(50,)	sgd	0.972362957	12
tanh	0.0001	(100,)	adam	4.622097096	12
tanh	0.0001	(100,)	sgd	1.073212838	12
tanh	0.0001	(50, 50)	adam	41.71337001	12
tanh	0.0001	(50, 50)	sgd	0.876584052	12
tanh	0.001	(50,)	adam	61.20522282	12
tanh	0.001	(50,)	sgd	0.972391297	12
tanh	0.001	(100,)	adam	4.622108522	12
tanh	0.001	(100,)	sgd	1.073266003	12
tanh	0.001	(50, 50)	adam	41.71336099	12
tanh	0.001	(50, 50)	sgd	0.87659252	12
relu	0.0001	(50,)	adam	18.09705986	12
relu	0.0001	(50,)	sgd	1.129147162	12
relu	0.0001	(100,)	adam	4.664331072	12
relu	0.0001	(100,)	sgd	1.078133557	12
relu	0.0001	(50, 50)	adam	3.495423344	12
relu	0.0001	(50, 50)	sgd	0.964397959	12
relu	0.001	(50,)	adam	18.09722768	12
relu	0.001	(50,)	sgd	1.129246724	12
relu	0.001	(100,)	adam	4.664300501	12
relu	0.001	(100,)	sgd	1.078024856	12

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
relu	0.001	(50, 50)	adam	3.499413395	12
relu	0.001	(50, 50)	sgd	1.006431644	12
tanh	0.0001	(50,)	adam	72.81191829	13
tanh	0.0001	(50,)	sgd	0.975285912	13
tanh	0.0001	(100,)	adam	3.87883376	13
tanh	0.0001	(100,)	sgd	1.047355759	13
tanh	0.0001	(50, 50)	adam	41.32365315	13
tanh	0.0001	(50, 50)	sgd	0.95629803	13
tanh	0.001	(50,)	adam	72.81169657	13
tanh	0.001	(50,)	sgd	0.975295756	13
tanh	0.001	(100,)	adam	3.878849333	13
tanh	0.001	(100,)	sgd	1.047385041	13
tanh	0.001	(50, 50)	adam	41.32363537	13
tanh	0.001	(50, 50)	sgd	0.956318045	13
relu	0.0001	(50,)	adam	9.973695244	13
relu	0.0001	(50,)	sgd	1.135125366	13
relu	0.0001	(100,)	adam	5.272379566	13
relu	0.0001	(100,)	sgd	1.131771208	13
relu	0.0001	(50, 50)	adam	2.725761413	13
relu	0.0001	(50, 50)	sgd	1.080043558	13
relu	0.001	(50,)	adam	9.973795067	13
relu	0.001	(50,)	sgd	1.134853941	13
relu	0.001	(100,)	adam	5.272377955	13
relu	0.001	(100,)	sgd	1.131783224	13

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
relu	0.001	(50, 50)	adam	2.717531273	13
relu	0.001	(50, 50)	sgd	1.080020756	13
tanh	0.0001	(50,)	adam	61.25289201	14
tanh	0.0001	(50,)	sgd	1.030729839	14
tanh	0.0001	(100,)	adam	3.889466184	14
tanh	0.0001	(100,)	sgd	0.971711094	14
tanh	0.0001	(50, 50)	adam	43.47265907	14
tanh	0.0001	(50, 50)	sgd	0.979972695	14
tanh	0.001	(50,)	adam	61.25245373	14
tanh	0.001	(50,)	sgd	1.030729286	14
tanh	0.001	(100,)	adam	3.889474095	14
tanh	0.001	(100,)	sgd	0.971717018	14
tanh	0.001	(50, 50)	adam	43.47259527	14
tanh	0.001	(50, 50)	sgd	0.979978839	14
relu	0.0001	(50,)	adam	8.156816988	14
relu	0.0001	(50,)	sgd	1.036291313	14
relu	0.0001	(100,)	adam	4.957330511	14
relu	0.0001	(100,)	sgd	0.982346166	14
relu	0.0001	(50, 50)	adam	2.707602106	14
relu	0.0001	(50, 50)	sgd	0.932973574	14
relu	0.001	(50,)	adam	8.157197563	14
relu	0.001	(50,)	sgd	1.03639362	14
relu	0.001	(100,)	adam	4.957317281	14
relu	0.001	(100,)	sgd	0.982251458	14

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
relu	0.001	(50, 50)	adam	2.709223476	14
relu	0.001	(50, 50)	sgd	0.938461779	14
tanh	0.0001	(50,)	adam	68.81254558	15
tanh	0.0001	(50,)	sgd	1.008910247	15
tanh	0.0001	(100,)	adam	3.494237099	15
tanh	0.0001	(100,)	sgd	0.963970458	15
tanh	0.0001	(50, 50)	adam	40.60309639	15
tanh	0.0001	(50, 50)	sgd	0.909164459	15
tanh	0.001	(50,)	adam	68.81228956	15
tanh	0.001	(50,)	sgd	1.008916701	15
tanh	0.001	(100,)	adam	3.494241269	15
tanh	0.001	(100,)	sgd	0.963974824	15
tanh	0.001	(50, 50)	adam	40.60312384	15
tanh	0.001	(50, 50)	sgd	0.909169953	15
relu	0.0001	(50,)	adam	10.54888967	15
relu	0.0001	(50,)	sgd	1.20032514	15
relu	0.0001	(100,)	adam	5.75437964	15
relu	0.0001	(100,)	sgd	1.132124167	15
relu	0.0001	(50, 50)	adam	2.937414921	15
relu	0.0001	(50, 50)	sgd	0.964469558	15
relu	0.001	(50,)	adam	10.54874234	15
relu	0.001	(50,)	sgd	1.200133432	15
relu	0.001	(100,)	adam	5.754289156	15
relu	0.001	(100,)	sgd	1.132055809	15

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
relu	0.001	(50, 50)	adam	2.937613565	15
relu	0.001	(50, 50)	sgd	0.964310808	15
tanh	0.0001	(50,)	adam	71.84260711	16
tanh	0.0001	(50,)	sgd	1.04200752	16
tanh	0.0001	(100,)	adam	4.175112267	16
tanh	0.0001	(100,)	sgd	0.995095418	16
tanh	0.0001	(50, 50)	adam	38.88206276	16
tanh	0.0001	(50, 50)	sgd	1.065325088	16
tanh	0.001	(50,)	adam	71.84234257	16
tanh	0.001	(50,)	sgd	1.04200645	16
tanh	0.001	(100,)	adam	4.175070018	16
tanh	0.001	(100,)	sgd	0.995097962	16
tanh	0.001	(50, 50)	adam	38.8821025	16
tanh	0.001	(50, 50)	sgd	1.065320344	16
relu	0.0001	(50,)	adam	19.54061618	16
relu	0.0001	(50,)	sgd	1.080897596	16
relu	0.0001	(100,)	adam	5.319738396	16
relu	0.0001	(100,)	sgd	1.141698114	16
relu	0.0001	(50, 50)	adam	3.089719393	16
relu	0.0001	(50, 50)	sgd	0.910171785	16
relu	0.001	(50,)	adam	19.53959567	16
relu	0.001	(50,)	sgd	1.081658406	16
relu	0.001	(100,)	adam	5.31961446	16
relu	0.001	(100,)	sgd	1.141462577	16

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
relu	0.001	(50, 50)	adam	3.089699819	16
relu	0.001	(50, 50)	sgd	0.910055396	16
tanh	0.0001	(50,)	adam	65.80257156	17
tanh	0.0001	(50,)	sgd	1.012680598	17
tanh	0.0001	(100,)	adam	5.724955389	17
tanh	0.0001	(100,)	sgd	0.972138165	17
tanh	0.0001	(50, 50)	adam	43.76568583	17
tanh	0.0001	(50, 50)	sgd	1.003851028	17
tanh	0.001	(50,)	adam	65.8023685	17
tanh	0.001	(50,)	sgd	1.012675807	17
tanh	0.001	(100,)	adam	5.724837978	17
tanh	0.001	(100,)	sgd	0.972140632	17
tanh	0.001	(50, 50)	adam	43.76569309	17
tanh	0.001	(50, 50)	sgd	1.003842311	17
relu	0.0001	(50,)	adam	12.90673176	17
relu	0.0001	(50,)	sgd	1.147623049	17
relu	0.0001	(100,)	adam	4.417874681	17
relu	0.0001	(100,)	sgd	1.100155231	17
relu	0.0001	(50, 50)	adam	3.75566064	17
relu	0.0001	(50, 50)	sgd	1.177835521	17
relu	0.001	(50,)	adam	12.90668197	17
relu	0.001	(50,)	sgd	1.147475246	17
relu	0.001	(100,)	adam	4.417823041	17
relu	0.001	(100,)	sgd	1.099847112	17

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
relu	0.001	(50, 50)	adam	3.756500633	17
relu	0.001	(50, 50)	sgd	1.177831868	17
tanh	0.0001	(50,)	adam	74.50945262	18
tanh	0.0001	(50,)	sgd	0.828956722	18
tanh	0.0001	(100,)	adam	4.724368651	18
tanh	0.0001	(100,)	sgd	0.765186082	18
tanh	0.0001	(50, 50)	adam	34.57229582	18
tanh	0.0001	(50, 50)	sgd	0.803391042	18
tanh	0.001	(50,)	adam	74.50877851	18
tanh	0.001	(50,)	sgd	0.828953025	18
tanh	0.001	(100,)	adam	4.724301142	18
tanh	0.001	(100,)	sgd	0.765178734	18
tanh	0.001	(50, 50)	adam	34.57230992	18
tanh	0.001	(50, 50)	sgd	0.803388995	18
relu	0.0001	(50,)	adam	14.11982357	18
relu	0.0001	(50,)	sgd	1.121000321	18
relu	0.0001	(100,)	adam	5.889179975	18
relu	0.0001	(100,)	sgd	1.18827624	18
relu	0.0001	(50, 50)	adam	3.62560789	18
relu	0.0001	(50, 50)	sgd	0.953881283	18
relu	0.001	(50,)	adam	14.1198742	18
relu	0.001	(50,)	sgd	1.121018554	18
relu	0.001	(100,)	adam	5.889078704	18
relu	0.001	(100,)	sgd	1.187994723	18

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
relu	0.001	(50, 50)	adam	3.606996052	18
relu	0.001	(50, 50)	sgd	0.953917114	18
tanh	0.0001	(50,)	adam	69.18112525	19
tanh	0.0001	(50,)	sgd	0.880062111	19
tanh	0.0001	(100,)	adam	6.811700107	19
tanh	0.0001	(100,)	sgd	0.822320456	19
tanh	0.0001	(50, 50)	adam	47.32858655	19
tanh	0.0001	(50, 50)	sgd	0.783843587	19
tanh	0.001	(50,)	adam	69.18080809	19
tanh	0.001	(50,)	sgd	0.880060841	19
tanh	0.001	(100,)	adam	6.81149106	19
tanh	0.001	(100,)	sgd	0.822310733	19
tanh	0.001	(50, 50)	adam	47.32858367	19
tanh	0.001	(50, 50)	sgd	0.783820515	19
relu	0.0001	(50,)	adam	11.29807176	19
relu	0.0001	(50,)	sgd	1.423411923	19
relu	0.0001	(100,)	adam	5.926219342	19
relu	0.0001	(100,)	sgd	1.234124119	19
relu	0.0001	(50, 50)	adam	4.418685687	19
relu	0.0001	(50, 50)	sgd	1.343714324	19
relu	0.001	(50,)	adam	11.29822407	19
relu	0.001	(50,)	sgd	1.421128919	19
relu	0.001	(100,)	adam	5.926195954	19
relu	0.001	(100,)	sgd	1.234105287	19

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
relu	0.001	(50, 50)	adam	4.419141618	19
relu	0.001	(50, 50)	sgd	1.343401735	19
tanh	0.0001	(50,)	adam	80.64040396	20
tanh	0.0001	(50,)	sgd	0.766760364	20
tanh	0.0001	(100,)	adam	5.256444423	20
tanh	0.0001	(100,)	sgd	0.771304342	20
tanh	0.0001	(50, 50)	adam	39.23613818	20
tanh	0.0001	(50, 50)	sgd	0.786195856	20
tanh	0.001	(50,)	adam	80.64000609	20
tanh	0.001	(50,)	sgd	0.766761435	20
tanh	0.001	(100,)	adam	5.256361215	20
tanh	0.001	(100,)	sgd	0.771298566	20
tanh	0.001	(50, 50)	adam	39.23614973	20
tanh	0.001	(50, 50)	sgd	0.786171283	20
relu	0.0001	(50,)	adam	17.16892652	20
relu	0.0001	(50,)	sgd	1.387546152	20
relu	0.0001	(100,)	adam	5.702822201	20
relu	0.0001	(100,)	sgd	1.555834515	20
relu	0.0001	(50, 50)	adam	4.263415986	20
relu	0.0001	(50, 50)	sgd	1.295525943	20
relu	0.001	(50,)	adam	17.16886803	20
relu	0.001	(50,)	sgd	1.387192672	20
relu	0.001	(100,)	adam	5.702786321	20
relu	0.001	(100,)	sgd	1.556109422	20

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
relu	0.001	(50, 50)	adam	4.262818034	20
relu	0.001	(50, 50)	sgd	1.296132119	20
tanh	0.0001	(50,)	adam	70.52407615	21
tanh	0.0001	(50,)	sgd	0.904561653	21
tanh	0.0001	(100,)	adam	6.605329581	21
tanh	0.0001	(100,)	sgd	0.750671474	21
tanh	0.0001	(50, 50)	adam	37.40124817	21
tanh	0.0001	(50, 50)	sgd	0.750657649	21
tanh	0.001	(50,)	adam	70.52360507	21
tanh	0.001	(50,)	sgd	0.904528351	21
tanh	0.001	(100,)	adam	6.605230651	21
tanh	0.001	(100,)	sgd	0.816886424	21
tanh	0.001	(50, 50)	adam	37.40125097	21
tanh	0.001	(50, 50)	sgd	0.816878684	21
relu	0.0001	(50,)	adam	9.029437392	21
relu	0.0001	(50,)	sgd	1.62961518	21
relu	0.0001	(100,)	adam	6.591357188	21
relu	0.0001	(100,)	sgd	1.351110435	21
relu	0.0001	(50, 50)	adam	4.800023026	21
relu	0.0001	(50, 50)	sgd	1.294871734	21
relu	0.001	(50,)	adam	9.028501302	21
relu	0.001	(50,)	sgd	1.628830873	21
relu	0.001	(100,)	adam	6.591316586	21
relu	0.001	(100,)	sgd	1.350543744	21

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
relu	0.001	(50, 50)	adam	4.799906419	21
relu	0.001	(50, 50)	sgd	1.293812551	21
tanh	0.0001	(50,)	adam	63.39468763	22
tanh	0.0001	(50,)	sgd	0.786416116	22
tanh	0.0001	(100,)	adam	5.065439666	22
tanh	0.0001	(100,)	sgd	0.760273595	22
tanh	0.0001	(50, 50)	adam	36.63298705	22
tanh	0.0001	(50, 50)	sgd	0.911697403	22
tanh	0.001	(50,)	adam	63.39439006	22
tanh	0.001	(50,)	sgd	0.786417586	22
tanh	0.001	(100,)	adam	5.065326691	22
tanh	0.001	(100,)	sgd	0.760255771	22
tanh	0.001	(50, 50)	adam	36.63299066	22
tanh	0.001	(50, 50)	sgd	0.911670566	22
relu	0.0001	(50,)	adam	7.694735071	22
relu	0.0001	(50,)	sgd	1.248911406	22
relu	0.0001	(100,)	adam	4.870179921	22
relu	0.0001	(100,)	sgd	1.4664348	22
relu	0.0001	(50, 50)	adam	3.991424859	22
relu	0.0001	(50, 50)	sgd	1.244671466	22
relu	0.001	(50,)	adam	7.694692981	22
relu	0.001	(50,)	sgd	1.24890554	22
relu	0.001	(100,)	adam	4.870085587	22
relu	0.001	(100,)	sgd	1.466518077	22

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
relu	0.001	(50, 50)	adam	3.991805992	22
relu	0.001	(50, 50)	sgd	1.244560621	22

Table A.13 : NLICA_MLP model hyperparameters tuning for different ICs using the top K = 5 external factors.

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.0001	(50,)	adam	53.247633	1
tanh	0.0001	(50,)	sgd	25.20544857	1
tanh	0.0001	(100,)	adam	25.94813195	1
tanh	0.0001	(100,)	sgd	25.00101552	1
tanh	0.0001	(50, 50)	adam	54.75630313	1
tanh	0.0001	(50, 50)	sgd	21.51507093	1
tanh	0.001	(50,)	adam	53.2477161	1
tanh	0.001	(50,)	sgd	25.20545113	1
tanh	0.001	(100,)	adam	25.94814378	1
tanh	0.001	(100,)	sgd	25.0010335	1
tanh	0.001	(50, 50)	adam	54.75623368	1
tanh	0.001	(50, 50)	sgd	21.51513621	1
relu	0.0001	(50,)	adam	49.22934655	1
relu	0.0001	(50,)	sgd	32.5661459	1
relu	0.0001	(100,)	adam	37.57611365	1
relu	0.0001	(100,)	sgd	32.58226203	1
relu	0.0001	(50, 50)	adam	30.88126388	1
relu	0.0001	(50, 50)	sgd	24.76147558	1

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
relu	0.001	(50,)	adam	49.22939839	1
relu	0.001	(50,)	sgd	32.56613749	1
relu	0.001	(100,)	adam	37.57655128	1
relu	0.001	(100,)	sgd	32.58225714	1
relu	0.001	(50, 50)	adam	30.84306457	1
relu	0.001	(50, 50)	sgd	24.76205399	1
tanh	0.0001	(50,)	adam	44.96524808	2
tanh	0.0001	(50,)	sgd	19.67536378	2
tanh	0.0001	(100,)	adam	24.65401222	2
tanh	0.0001	(100,)	sgd	18.71689713	2
tanh	0.0001	(50, 50)	adam	48.93981963	2
tanh	0.0001	(50, 50)	sgd	8.218084884	2
tanh	0.001	(50,)	adam	44.96532445	2
tanh	0.001	(50,)	sgd	19.67572977	2
tanh	0.001	(100,)	adam	24.65401534	2
tanh	0.001	(100,)	sgd	18.71716363	2
tanh	0.001	(50, 50)	adam	48.93971993	2
tanh	0.001	(50, 50)	sgd	8.218652702	2
relu	0.0001	(50,)	adam	49.96632908	2
relu	0.0001	(50,)	sgd	14.66432371	2
relu	0.0001	(100,)	adam	35.54476855	2
relu	0.0001	(100,)	sgd	15.14721834	2
relu	0.0001	(50, 50)	adam	23.55002627	2
relu	0.0001	(50, 50)	sgd	8.772316498	2

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
relu	0.001	(50,)	adam	49.96639833	2
relu	0.001	(50,)	sgd	14.66437818	2
relu	0.001	(100,)	adam	35.55266005	2
relu	0.001	(100,)	sgd	15.14733367	2
relu	0.001	(50, 50)	adam	22.78964201	2
relu	0.001	(50, 50)	sgd	8.773076024	2
tanh	0.0001	(50,)	adam	61.53307609	3
tanh	0.0001	(50,)	sgd	10.56794898	3
tanh	0.0001	(100,)	adam	18.6854117	3
tanh	0.0001	(100,)	sgd	10.74625949	3
tanh	0.0001	(50, 50)	adam	47.50540448	3
tanh	0.0001	(50, 50)	sgd	9.032176514	3
tanh	0.001	(50,)	adam	61.53314544	3
tanh	0.001	(50,)	sgd	10.5680568	3
tanh	0.001	(100,)	adam	18.68543383	3
tanh	0.001	(100,)	sgd	10.74637251	3
tanh	0.001	(50, 50)	adam	47.50541003	3
tanh	0.001	(50, 50)	sgd	9.032400904	3
relu	0.0001	(50,)	adam	32.11094487	3
relu	0.0001	(50,)	sgd	9.688955216	3
relu	0.0001	(100,)	adam	12.40120673	3
relu	0.0001	(100,)	sgd	9.711394821	3
relu	0.0001	(50, 50)	adam	10.24032107	3
relu	0.0001	(50, 50)	sgd	9.828106508	3

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
relu	0.001	(50,)	adam	32.50787582	3
relu	0.001	(50,)	sgd	9.688997287	3
relu	0.001	(100,)	adam	12.35164678	3
relu	0.001	(100,)	sgd	9.711398953	3
relu	0.001	(50, 50)	adam	10.23430073	3
relu	0.001	(50, 50)	sgd	9.828183129	3
tanh	0.0001	(50,)	adam	49.21000752	4
tanh	0.0001	(50,)	sgd	24.62249876	4
tanh	0.0001	(100,)	adam	28.57592097	4
tanh	0.0001	(100,)	sgd	24.53918231	4
tanh	0.0001	(50, 50)	adam	51.22268502	4
tanh	0.0001	(50, 50)	sgd	13.03168151	4
tanh	0.001	(50,)	adam	49.21009627	4
tanh	0.001	(50,)	sgd	24.62311514	4
tanh	0.001	(100,)	adam	28.57592455	4
tanh	0.001	(100,)	sgd	24.53970851	4
tanh	0.001	(50, 50)	adam	51.22275856	4
tanh	0.001	(50, 50)	sgd	13.0325526	4
relu	0.0001	(50,)	adam	60.68222759	4
relu	0.0001	(50,)	sgd	27.40477473	4
relu	0.0001	(100,)	adam	44.70761873	4
relu	0.0001	(100,)	sgd	27.50931755	4
relu	0.0001	(50, 50)	adam	28.57996339	4
relu	0.0001	(50, 50)	sgd	8.111065426	4

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
relu	0.001	(50,)	adam	60.69069907	4
relu	0.001	(50,)	sgd	27.40483969	4
relu	0.001	(100,)	adam	44.30533541	4
relu	0.001	(100,)	sgd	27.50938025	4
relu	0.001	(50, 50)	adam	28.87036534	4
relu	0.001	(50, 50)	sgd	8.104304461	4
tanh	0.0001	(50,)	adam	73.89864741	5
tanh	0.0001	(50,)	sgd	12.65755161	5
tanh	0.0001	(100,)	adam	24.99253132	5
tanh	0.0001	(100,)	sgd	10.7728475	5
tanh	0.0001	(50, 50)	adam	55.23612609	5
tanh	0.0001	(50, 50)	sgd	18.14478459	5
tanh	0.001	(50,)	adam	73.89869539	5
tanh	0.001	(50,)	sgd	12.65834991	5
tanh	0.001	(100,)	adam	24.9925751	5
tanh	0.001	(100,)	sgd	10.77349968	5
tanh	0.001	(50, 50)	adam	55.23605772	5
tanh	0.001	(50, 50)	sgd	18.1446595	5
relu	0.0001	(50,)	adam	43.71779267	5
relu	0.0001	(50,)	sgd	7.69574558	5
relu	0.0001	(100,)	adam	28.19378341	5
relu	0.0001	(100,)	sgd	7.174410302	5
relu	0.0001	(50, 50)	adam	16.92388149	5
relu	0.0001	(50, 50)	sgd	3.65937294	5

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
relu	0.001	(50,)	adam	44.32773366	5
relu	0.001	(50,)	sgd	7.695849628	5
relu	0.001	(100,)	adam	28.26105022	5
relu	0.001	(100,)	sgd	7.174561608	5
relu	0.001	(50, 50)	adam	17.60716522	5
relu	0.001	(50, 50)	sgd	3.676660712	5

Table A.14 : NLICA_MLP model hyperparameters tuning for different ICs using the top K = 10 external factors.

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.0001	(50,)	adam	195.7279955	1
tanh	0.0001	(50,)	sgd	21.65691539	1
tanh	0.0001	(100,)	adam	76.83374912	1
tanh	0.0001	(100,)	sgd	21.71548016	1
tanh	0.0001	(50, 50)	adam	120.8160437	1
tanh	0.0001	(50, 50)	sgd	21.28437689	1
tanh	0.001	(50,)	adam	195.7279863	1
tanh	0.001	(50,)	sgd	21.65701081	1
tanh	0.001	(100,)	adam	76.83377752	1
tanh	0.001	(100,)	sgd	21.71557346	1
tanh	0.001	(50, 50)	adam	120.8161142	1
tanh	0.001	(50, 50)	sgd	21.28437613	1
relu	0.0001	(50,)	adam	283.3962765	1
relu	0.0001	(50,)	sgd	22.06444889	1

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
relu	0.0001	(100,)	adam	123.0627021	1
relu	0.0001	(100,)	sgd	21.81406337	1
relu	0.0001	(50, 50)	adam	25.00344943	1
relu	0.0001	(50, 50)	sgd	20.97201532	1
relu	0.001	(50,)	adam	283.3963453	1
relu	0.001	(50,)	sgd	22.06437117	1
relu	0.001	(100,)	adam	123.0631399	1
relu	0.001	(100,)	sgd	21.81420723	1
relu	0.001	(50, 50)	adam	24.68456762	1
relu	0.001	(50, 50)	sgd	20.97245163	1
tanh	0.0001	(50,)	adam	159.6927966	2
tanh	0.0001	(50,)	sgd	119.2894451	2
tanh	0.0001	(100,)	adam	119.9976485	2
tanh	0.0001	(100,)	sgd	120.5422958	2
tanh	0.0001	(50, 50)	adam	119.8191761	2
tanh	0.0001	(50, 50)	sgd	119.4822139	2
tanh	0.001	(50,)	adam	159.6928681	2
tanh	0.001	(50,)	sgd	119.2894369	2
tanh	0.001	(100,)	adam	119.9976602	2
tanh	0.001	(100,)	sgd	120.5422942	2
tanh	0.001	(50, 50)	adam	119.8191827	2
tanh	0.001	(50, 50)	sgd	119.4822094	2
relu	0.0001	(50,)	adam	295.5095774	2
relu	0.0001	(50,)	sgd	119.9925294	2

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
relu	0.0001	(100,)	adam	137.3314974	2
relu	0.0001	(100,)	sgd	120.5797728	2
relu	0.0001	(50, 50)	adam	119.8761528	2
relu	0.0001	(50, 50)	sgd	120.8466834	2
relu	0.001	(50,)	adam	295.5096225	2
relu	0.001	(50,)	sgd	119.992527	2
relu	0.001	(100,)	adam	137.3315318	2
relu	0.001	(100,)	sgd	120.5797637	2
relu	0.001	(50, 50)	adam	119.8830476	2
relu	0.001	(50, 50)	sgd	120.8466877	2
tanh	0.0001	(50,)	adam	73.79877842	3
tanh	0.0001	(50,)	sgd	17.84570067	3
tanh	0.0001	(100,)	adam	37.03506739	3
tanh	0.0001	(100,)	sgd	18.64792948	3
tanh	0.0001	(50, 50)	adam	46.943435	3
tanh	0.0001	(50, 50)	sgd	20.98933927	3
tanh	0.001	(50,)	adam	73.79885466	3
tanh	0.001	(50,)	sgd	17.8462849	3
tanh	0.001	(100,)	adam	37.0350871	3
tanh	0.001	(100,)	sgd	18.64873721	3
tanh	0.001	(50, 50)	adam	46.9434658	3
tanh	0.001	(50, 50)	sgd	20.9894619	3
relu	0.0001	(50,)	adam	66.38545604	3
relu	0.0001	(50,)	sgd	14.76064784	3

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
relu	0.0001	(100,)	adam	57.13981418	3
relu	0.0001	(100,)	sgd	13.92436558	3
relu	0.0001	(50, 50)	adam	31.83492658	3
relu	0.0001	(50, 50)	sgd	10.55442503	3
relu	0.001	(50,)	adam	66.43760176	3
relu	0.001	(50,)	sgd	14.76064372	3
relu	0.001	(100,)	adam	57.19224114	3
relu	0.001	(100,)	sgd	13.92450892	3
relu	0.001	(50, 50)	adam	30.8435569	3
relu	0.001	(50, 50)	sgd	10.54463454	3
tanh	0.0001	(50,)	adam	111.4247639	4
tanh	0.0001	(50,)	sgd	11.41302052	4
tanh	0.0001	(100,)	adam	60.62450975	4
tanh	0.0001	(100,)	sgd	11.99117778	4
tanh	0.0001	(50, 50)	adam	59.38016826	4
tanh	0.0001	(50, 50)	sgd	10.62697538	4
tanh	0.001	(50,)	adam	111.4247825	4
tanh	0.001	(50,)	sgd	11.41359718	4
tanh	0.001	(100,)	adam	60.62452838	4
tanh	0.001	(100,)	sgd	11.9918634	4
tanh	0.001	(50, 50)	adam	59.37977734	4
tanh	0.001	(50, 50)	sgd	10.62734643	4
relu	0.0001	(50,)	adam	102.6367761	4
relu	0.0001	(50,)	sgd	8.336223527	4

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
relu	0.0001	(100,)	adam	94.09616731	4
relu	0.0001	(100,)	sgd	7.95688596	4
relu	0.0001	(50, 50)	adam	31.02988307	4
relu	0.0001	(50, 50)	sgd	5.360241815	4
relu	0.001	(50,)	adam	102.79839	4
relu	0.001	(50,)	sgd	8.336573568	4
relu	0.001	(100,)	adam	94.09329462	4
relu	0.001	(100,)	sgd	7.957366362	4
relu	0.001	(50, 50)	adam	29.83925225	4
relu	0.001	(50, 50)	sgd	5.372748704	4
tanh	0.0001	(50,)	adam	59.90716385	5
tanh	0.0001	(50,)	sgd	10.49775832	5
tanh	0.0001	(100,)	adam	36.52063581	5
tanh	0.0001	(100,)	sgd	10.01653831	5
tanh	0.0001	(50, 50)	adam	49.74058992	5
tanh	0.0001	(50, 50)	sgd	3.063295225	5
tanh	0.001	(50,)	adam	59.90723	5
tanh	0.001	(50,)	sgd	10.49791299	5
tanh	0.001	(100,)	adam	36.52065111	5
tanh	0.001	(100,)	sgd	10.01657916	5
tanh	0.001	(50, 50)	adam	49.74050854	5
tanh	0.001	(50, 50)	sgd	3.063371023	5
relu	0.0001	(50,)	adam	79.19750895	5
relu	0.0001	(50,)	sgd	7.751008823	5

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
relu	0.0001	(100,)	adam	52.29238239	5
relu	0.0001	(100,)	sgd	7.69904155	5
relu	0.0001	(50, 50)	adam	8.629812031	5
relu	0.0001	(50, 50)	sgd	3.105525576	5
relu	0.001	(50,)	adam	79.33024957	5
relu	0.001	(50,)	sgd	7.750915364	5
relu	0.001	(100,)	adam	52.41064879	5
relu	0.001	(100,)	sgd	7.698993727	5
relu	0.001	(50, 50)	adam	8.595370333	5
relu	0.001	(50, 50)	sgd	3.105226411	5
tanh	0.0001	(50,)	adam	96.41514601	6
tanh	0.0001	(50,)	sgd	10.1448056	6
tanh	0.0001	(100,)	adam	57.49959691	6
tanh	0.0001	(100,)	sgd	10.95561067	6
tanh	0.0001	(50, 50)	adam	61.99182134	6
tanh	0.0001	(50, 50)	sgd	8.575917284	6
tanh	0.001	(50,)	adam	96.41513275	6
tanh	0.001	(50,)	sgd	10.14520871	6
tanh	0.001	(100,)	adam	57.49959975	6
tanh	0.001	(100,)	sgd	10.9560097	6
tanh	0.001	(50, 50)	adam	61.99072346	6
tanh	0.001	(50, 50)	sgd	8.57620474	6
relu	0.0001	(50,)	adam	126.0759529	6
relu	0.0001	(50,)	sgd	11.16002308	6

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
relu	0.0001	(100,)	adam	93.9125309	6
relu	0.0001	(100,)	sgd	10.93278364	6
relu	0.0001	(50, 50)	adam	22.83593585	6
relu	0.0001	(50, 50)	sgd	8.551849676	6
relu	0.001	(50,)	adam	126.1589392	6
relu	0.001	(50,)	sgd	11.15732204	6
relu	0.001	(100,)	adam	93.91078404	6
relu	0.001	(100,)	sgd	10.93736835	6
relu	0.001	(50, 50)	adam	24.3325743	6
relu	0.001	(50, 50)	sgd	8.950016783	6
tanh	0.0001	(50,)	adam	199.5414445	7
tanh	0.0001	(50,)	sgd	20.78694861	7
tanh	0.0001	(100,)	adam	75.53952495	7
tanh	0.0001	(100,)	sgd	20.93159649	7
tanh	0.0001	(50, 50)	adam	87.38127006	7
tanh	0.0001	(50, 50)	sgd	20.61855512	7
tanh	0.001	(50,)	adam	199.5414712	7
tanh	0.001	(50,)	sgd	20.78694486	7
tanh	0.001	(100,)	adam	75.53954245	7
tanh	0.001	(100,)	sgd	20.93158317	7
tanh	0.001	(50, 50)	adam	87.36820177	7
tanh	0.001	(50, 50)	sgd	20.61854692	7
relu	0.0001	(50,)	adam	230.3728884	7
relu	0.0001	(50,)	sgd	20.72791246	7

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
relu	0.0001	(100,)	adam	135.9129894	7
relu	0.0001	(100,)	sgd	21.19293661	7
relu	0.0001	(50, 50)	adam	20.90593333	7
relu	0.0001	(50, 50)	sgd	21.17076701	7
relu	0.001	(50,)	adam	230.3731465	7
relu	0.001	(50,)	sgd	20.72792628	7
relu	0.001	(100,)	adam	135.9433278	7
relu	0.001	(100,)	sgd	21.19294643	7
relu	0.001	(50, 50)	adam	20.90343469	7
relu	0.001	(50, 50)	sgd	21.17073615	7
tanh	0.0001	(50,)	adam	91.61866689	8
tanh	0.0001	(50,)	sgd	8.694150834	8
tanh	0.0001	(100,)	adam	64.77489876	8
tanh	0.0001	(100,)	sgd	9.449894239	8
tanh	0.0001	(50, 50)	adam	120.600635	8
tanh	0.0001	(50, 50)	sgd	8.309691296	8
tanh	0.001	(50,)	adam	91.61870487	8
tanh	0.001	(50,)	sgd	8.694363521	8
tanh	0.001	(100,)	adam	64.77490628	8
tanh	0.001	(100,)	sgd	9.450176181	8
tanh	0.001	(50, 50)	adam	120.6006937	8
tanh	0.001	(50, 50)	sgd	8.309763254	8
relu	0.0001	(50,)	adam	98.20880847	8
relu	0.0001	(50,)	sgd	11.39636555	8

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
relu	0.0001	(100,)	adam	87.09217943	8
relu	0.0001	(100,)	sgd	11.40268462	8
relu	0.0001	(50, 50)	adam	19.2952931	8
relu	0.0001	(50, 50)	sgd	9.650949232	8
relu	0.001	(50,)	adam	98.24967519	8
relu	0.001	(50,)	sgd	11.39974808	8
relu	0.001	(100,)	adam	87.1570114	8
relu	0.001	(100,)	sgd	11.39734767	8
relu	0.001	(50, 50)	adam	19.29983822	8
relu	0.001	(50, 50)	sgd	9.599518556	8
tanh	0.0001	(50,)	adam	107.768056	9
tanh	0.0001	(50,)	sgd	10.5826673	9
tanh	0.0001	(100,)	adam	52.29913191	9
tanh	0.0001	(100,)	sgd	11.28564581	9
tanh	0.0001	(50, 50)	adam	119.9584129	9
tanh	0.0001	(50, 50)	sgd	10.53929407	9
tanh	0.001	(50,)	adam	107.768091	9
tanh	0.001	(50,)	sgd	10.58310844	9
tanh	0.001	(100,)	adam	52.29917707	9
tanh	0.001	(100,)	sgd	11.28606608	9
tanh	0.001	(50, 50)	adam	119.9582828	9
tanh	0.001	(50, 50)	sgd	10.53953218	9
relu	0.0001	(50,)	adam	83.62958295	9
relu	0.0001	(50,)	sgd	9.98008918	9

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
relu	0.0001	(100,)	adam	69.09144917	9
relu	0.0001	(100,)	sgd	9.406333176	9
relu	0.0001	(50, 50)	adam	23.10277449	9
relu	0.0001	(50, 50)	sgd	8.090533895	9
relu	0.001	(50,)	adam	83.66763312	9
relu	0.001	(50,)	sgd	9.980323331	9
relu	0.001	(100,)	adam	67.84548984	9
relu	0.001	(100,)	sgd	9.406614107	9
relu	0.001	(50, 50)	adam	23.1667718	9
relu	0.001	(50, 50)	sgd	8.029377295	9
tanh	0.0001	(50,)	adam	68.65908849	10
tanh	0.0001	(50,)	sgd	3.559125262	10
tanh	0.0001	(100,)	adam	35.32726392	10
tanh	0.0001	(100,)	sgd	4.310086291	10
tanh	0.0001	(50, 50)	adam	48.91462137	10
tanh	0.0001	(50, 50)	sgd	2.132620298	10
tanh	0.001	(50,)	adam	68.65912265	10
tanh	0.001	(50,)	sgd	3.55931846	10
tanh	0.001	(100,)	adam	35.32725199	10
tanh	0.001	(100,)	sgd	4.310327552	10
tanh	0.001	(50, 50)	adam	48.91466075	10
tanh	0.001	(50, 50)	sgd	2.132712528	10
relu	0.0001	(50,)	adam	64.5084099	10
relu	0.0001	(50,)	sgd	4.038883701	10

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
relu	0.0001	(100,)	adam	36.35551799	10
relu	0.0001	(100,)	sgd	4.151488075	10
relu	0.0001	(50, 50)	adam	4.628822434	10
relu	0.0001	(50, 50)	sgd	1.769130188	10
relu	0.001	(50,)	adam	64.71991273	10
relu	0.001	(50,)	sgd	4.038955369	10
relu	0.001	(100,)	adam	35.07032979	10
relu	0.001	(100,)	sgd	4.151975594	10
relu	0.001	(50, 50)	adam	4.681430207	10
relu	0.001	(50, 50)	sgd	1.769174435	10

Table A.15 : NLICA_MLP model hyperparameters tuning for different ICs using the top K = 15 external factors.

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.0001	(50,)	adam	59.62015344	1
tanh	0.0001	(50,)	sgd	31.93061156	1
tanh	0.0001	(100,)	adam	35.47999307	1
tanh	0.0001	(100,)	sgd	35.50343876	1
tanh	0.0001	(50, 50)	adam	65.98228573	1
tanh	0.0001	(50, 50)	sgd	9.698301364	1
tanh	0.001	(50,)	adam	59.62020801	1
tanh	0.001	(50,)	sgd	31.93083365	1
tanh	0.001	(100,)	adam	35.47998651	1
tanh	0.001	(100,)	sgd	35.50343443	1

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
tanh	0.001	(50, 50)	adam	65.98214544	1
tanh	0.001	(50, 50)	sgd	9.698485057	1
relu	0.0001	(50,)	adam	56.03640196	1
relu	0.0001	(50,)	sgd	41.12067132	1
relu	0.0001	(100,)	adam	49.35685559	1
relu	0.0001	(100,)	sgd	40.06529514	1
relu	0.0001	(50, 50)	adam	40.1810824	1
relu	0.0001	(50, 50)	sgd	42.07799298	1
relu	0.001	(50,)	adam	56.08042835	1
relu	0.001	(50,)	sgd	41.12066363	1
relu	0.001	(100,)	adam	49.38387276	1
relu	0.001	(100,)	sgd	40.06530555	1
relu	0.001	(50, 50)	adam	40.0483127	1
relu	0.001	(50, 50)	sgd	42.07854621	1
tanh	0.0001	(50,)	adam	213.2190638	2
tanh	0.0001	(50,)	sgd	21.85722489	2
tanh	0.0001	(100,)	adam	72.97807754	2
tanh	0.0001	(100,)	sgd	22.13782769	2
tanh	0.0001	(50, 50)	adam	74.88748347	2
tanh	0.0001	(50, 50)	sgd	21.29110961	2
tanh	0.001	(50,)	adam	213.2191103	2
tanh	0.001	(50,)	sgd	21.85724684	2
tanh	0.001	(100,)	adam	72.97815313	2
tanh	0.001	(100,)	sgd	22.13783821	2

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
tanh	0.001	(50, 50)	adam	74.88480226	2
tanh	0.001	(50, 50)	sgd	21.29110727	2
relu	0.0001	(50,)	adam	282.3441191	2
relu	0.0001	(50,)	sgd	22.24280827	2
relu	0.0001	(100,)	adam	117.7983543	2
relu	0.0001	(100,)	sgd	22.150991	2
relu	0.0001	(50, 50)	adam	22.48240228	2
relu	0.0001	(50, 50)	sgd	22.14174198	2
relu	0.001	(50,)	adam	282.3441315	2
relu	0.001	(50,)	sgd	22.24281096	2
relu	0.001	(100,)	adam	117.8018326	2
relu	0.001	(100,)	sgd	22.15097875	2
relu	0.001	(50, 50)	adam	22.55177051	2
relu	0.001	(50, 50)	sgd	22.14173964	2
tanh	0.0001	(50,)	adam	64.8949535	3
tanh	0.0001	(50,)	sgd	21.69136573	3
tanh	0.0001	(100,)	adam	30.89011415	3
tanh	0.0001	(100,)	sgd	25.10319115	3
tanh	0.0001	(50, 50)	adam	51.23106921	3
tanh	0.0001	(50, 50)	sgd	16.20520805	3
tanh	0.001	(50,)	adam	64.89497374	3
tanh	0.001	(50,)	sgd	21.69207698	3
tanh	0.001	(100,)	adam	30.89012593	3
tanh	0.001	(100,)	sgd	25.10370598	3

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
tanh	0.001	(50, 50)	adam	51.23114067	3
tanh	0.001	(50, 50)	sgd	16.20526002	3
relu	0.0001	(50,)	adam	36.91871818	3
relu	0.0001	(50,)	sgd	33.7405091	3
relu	0.0001	(100,)	adam	35.06240382	3
relu	0.0001	(100,)	sgd	32.0844067	3
relu	0.0001	(50, 50)	adam	32.1806286	3
relu	0.0001	(50, 50)	sgd	9.063126156	3
relu	0.001	(50,)	adam	36.79343526	3
relu	0.001	(50,)	sgd	33.74068906	3
relu	0.001	(100,)	adam	35.02707451	3
relu	0.001	(100,)	sgd	32.08462826	3
relu	0.001	(50, 50)	adam	31.78591918	3
relu	0.001	(50, 50)	sgd	11.31144935	3
tanh	0.0001	(50,)	adam	202.1955801	4
tanh	0.0001	(50,)	sgd	25.59715868	4
tanh	0.0001	(100,)	adam	100.4432089	4
tanh	0.0001	(100,)	sgd	26.85043168	4
tanh	0.0001	(50, 50)	adam	111.8790542	4
tanh	0.0001	(50, 50)	sgd	24.7072505	4
tanh	0.001	(50,)	adam	202.1956207	4
tanh	0.001	(50,)	sgd	25.59759693	4
tanh	0.001	(100,)	adam	100.4432339	4
tanh	0.001	(100,)	sgd	26.85087569	4

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.001	(50, 50)	adam	111.8764469	4
tanh	0.001	(50, 50)	sgd	24.70726551	4
relu	0.0001	(50,)	adam	241.4147947	4
relu	0.0001	(50,)	sgd	32.09726501	4
relu	0.0001	(100,)	adam	127.651917	4
relu	0.0001	(100,)	sgd	31.83732786	4
relu	0.0001	(50, 50)	adam	29.49723271	4
relu	0.0001	(50, 50)	sgd	13.88021076	4
relu	0.001	(50,)	adam	241.3875558	4
relu	0.001	(50,)	sgd	32.09731279	4
relu	0.001	(100,)	adam	127.6603463	4
relu	0.001	(100,)	sgd	31.83733063	4
relu	0.001	(50, 50)	adam	29.45878512	4
relu	0.001	(50, 50)	sgd	13.88078391	4
tanh	0.0001	(50,)	adam	80.8442621	5
tanh	0.0001	(50,)	sgd	11.44693153	5
tanh	0.0001	(100,)	adam	43.52698616	5
tanh	0.0001	(100,)	sgd	11.47045445	5
tanh	0.0001	(50, 50)	adam	54.78781825	5
tanh	0.0001	(50, 50)	sgd	9.810714268	5
tanh	0.001	(50,)	adam	80.84433512	5
tanh	0.001	(50,)	sgd	11.44711248	5
tanh	0.001	(100,)	adam	43.52705814	5
tanh	0.001	(100,)	sgd	11.47059252	5

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.001	(50, 50)	adam	54.78794149	5
tanh	0.001	(50, 50)	sgd	9.810837412	5
relu	0.0001	(50,)	adam	62.86561882	5
relu	0.0001	(50,)	sgd	17.23947842	5
relu	0.0001	(100,)	adam	56.26230383	5
relu	0.0001	(100,)	sgd	21.39436287	5
relu	0.0001	(50, 50)	adam	24.14953783	5
relu	0.0001	(50, 50)	sgd	9.638447764	5
relu	0.001	(50,)	adam	63.17458639	5
relu	0.001	(50,)	sgd	17.23291002	5
relu	0.001	(100,)	adam	56.35705424	5
relu	0.001	(100,)	sgd	21.30285158	5
relu	0.001	(50, 50)	adam	24.1865069	5
relu	0.001	(50, 50)	sgd	9.64212135	5
tanh	0.0001	(50,)	adam	69.19758467	6
tanh	0.0001	(50,)	sgd	7.213631464	6
tanh	0.0001	(100,)	adam	31.89676182	6
tanh	0.0001	(100,)	sgd	6.504243009	6
tanh	0.0001	(50, 50)	adam	50.74598425	6
tanh	0.0001	(50, 50)	sgd	19.75891856	6
tanh	0.001	(50,)	adam	69.19761169	6
tanh	0.001	(50,)	sgd	7.213991448	6
tanh	0.001	(100,)	adam	31.89679378	6
tanh	0.001	(100,)	sgd	6.504517725	6

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.001	(50, 50)	adam	50.7461173	6
tanh	0.001	(50, 50)	sgd	19.75972323	6
relu	0.0001	(50,)	adam	42.00986461	6
relu	0.0001	(50,)	sgd	5.884988829	6
relu	0.0001	(100,)	adam	33.69013226	6
relu	0.0001	(100,)	sgd	5.960379236	6
relu	0.0001	(50, 50)	adam	14.68992729	6
relu	0.0001	(50, 50)	sgd	4.301864693	6
relu	0.001	(50,)	adam	42.12030187	6
relu	0.001	(50,)	sgd	5.885117286	6
relu	0.001	(100,)	adam	33.78696645	6
relu	0.001	(100,)	sgd	5.960430909	6
relu	0.001	(50, 50)	adam	14.55514117	6
relu	0.001	(50, 50)	sgd	4.253290003	6
tanh	0.0001	(50,)	adam	65.02018506	7
tanh	0.0001	(50,)	sgd	2.940317184	7
tanh	0.0001	(100,)	adam	28.65228525	7
tanh	0.0001	(100,)	sgd	3.159743197	7
tanh	0.0001	(50, 50)	adam	50.47175798	7
tanh	0.0001	(50, 50)	sgd	2.382263096	7
tanh	0.001	(50,)	adam	65.02018877	7
tanh	0.001	(50,)	sgd	2.940420868	7
tanh	0.001	(100,)	adam	28.65230969	7
tanh	0.001	(100,)	sgd	3.1598159	7

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.001	(50, 50)	adam	50.4718508	7
tanh	0.001	(50, 50)	sgd	2.382313981	7
relu	0.0001	(50,)	adam	30.22699887	7
relu	0.0001	(50,)	sgd	3.342389144	7
relu	0.0001	(100,)	adam	22.10443018	7
relu	0.0001	(100,)	sgd	3.285647873	7
relu	0.0001	(50, 50)	adam	3.880706944	7
relu	0.0001	(50, 50)	sgd	2.661808341	7
relu	0.001	(50,)	adam	29.90667485	7
relu	0.001	(50,)	sgd	3.342371118	7
relu	0.001	(100,)	adam	21.9582217	7
relu	0.001	(100,)	sgd	3.286753338	7
relu	0.001	(50, 50)	adam	3.791425477	7
relu	0.001	(50, 50)	sgd	2.621569216	7
tanh	0.0001	(50,)	adam	54.96198891	8
tanh	0.0001	(50,)	sgd	3.144165497	8
tanh	0.0001	(100,)	adam	24.20418642	8
tanh	0.0001	(100,)	sgd	3.416234577	8
tanh	0.0001	(50, 50)	adam	55.9336407	8
tanh	0.0001	(50, 50)	sgd	1.920430187	8
tanh	0.001	(50,)	adam	54.96204651	8
tanh	0.001	(50,)	sgd	3.144315544	8
tanh	0.001	(100,)	adam	24.20421493	8
tanh	0.001	(100,)	sgd	3.416398963	8

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.001	(50, 50)	adam	55.93364763	8
tanh	0.001	(50, 50)	sgd	1.920514863	8
relu	0.0001	(50,)	adam	28.68394297	8
relu	0.0001	(50,)	sgd	2.890579596	8
relu	0.0001	(100,)	adam	19.1133383	8
relu	0.0001	(100,)	sgd	2.962574742	8
relu	0.0001	(50, 50)	adam	4.363040089	8
relu	0.0001	(50, 50)	sgd	1.997768344	8
relu	0.001	(50,)	adam	28.74712097	8
relu	0.001	(50,)	sgd	2.890610467	8
relu	0.001	(100,)	adam	18.76251098	8
relu	0.001	(100,)	sgd	2.962709526	8
relu	0.001	(50, 50)	adam	4.044537273	8
relu	0.001	(50, 50)	sgd	1.997794687	8
tanh	0.0001	(50,)	adam	64.51131182	9
tanh	0.0001	(50,)	sgd	2.755494271	9
tanh	0.0001	(100,)	adam	29.95891728	9
tanh	0.0001	(100,)	sgd	3.485344725	9
tanh	0.0001	(50, 50)	adam	49.41521089	9
tanh	0.0001	(50, 50)	sgd	1.328219524	9
tanh	0.001	(50,)	adam	64.51136871	9
tanh	0.001	(50,)	sgd	2.75566109	9
tanh	0.001	(100,)	adam	29.9589482	9
tanh	0.001	(100,)	sgd	3.485503459	9

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.001	(50, 50)	adam	49.41519884	9
tanh	0.001	(50, 50)	sgd	1.328310483	9
relu	0.0001	(50,)	adam	37.73227718	9
relu	0.0001	(50,)	sgd	2.484180931	9
relu	0.0001	(100,)	adam	27.70929114	9
relu	0.0001	(100,)	sgd	2.501985502	9
relu	0.0001	(50, 50)	adam	5.24072952	9
relu	0.0001	(50, 50)	sgd	1.086228962	9
relu	0.001	(50,)	adam	37.87098606	9
relu	0.001	(50,)	sgd	2.484008127	9
relu	0.001	(100,)	adam	26.21512243	9
relu	0.001	(100,)	sgd	2.502129192	9
relu	0.001	(50, 50)	adam	5.638157595	9
relu	0.001	(50, 50)	sgd	1.086247297	9
tanh	0.0001	(50,)	adam	71.27484986	10
tanh	0.0001	(50,)	sgd	6.7063792	10
tanh	0.0001	(100,)	adam	31.66896989	10
tanh	0.0001	(100,)	sgd	7.353298954	10
tanh	0.0001	(50, 50)	adam	51.97268801	10
tanh	0.0001	(50, 50)	sgd	4.930993403	10
tanh	0.001	(50,)	adam	71.27486486	10
tanh	0.001	(50,)	sgd	6.706603529	10
tanh	0.001	(100,)	adam	31.66901351	10
tanh	0.001	(100,)	sgd	7.353509371	10

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
tanh	0.001	(50, 50)	adam	51.9728361	10
tanh	0.001	(50, 50)	sgd	4.931160968	10
relu	0.0001	(50,)	adam	37.58446387	10
relu	0.0001	(50,)	sgd	7.444639627	10
relu	0.0001	(100,)	adam	29.27862984	10
relu	0.0001	(100,)	sgd	7.208030882	10
relu	0.0001	(50, 50)	adam	9.771854321	10
relu	0.0001	(50, 50)	sgd	5.008789972	10
relu	0.001	(50,)	adam	37.60511724	10
relu	0.001	(50,)	sgd	7.444978017	10
relu	0.001	(100,)	adam	28.9223296	10
relu	0.001	(100,)	sgd	7.208928318	10
relu	0.001	(50, 50)	adam	9.805221781	10
relu	0.001	(50, 50)	sgd	5.02053666	10
tanh	0.0001	(50,)	adam	69.08501027	11
tanh	0.0001	(50,)	sgd	2.339064284	11
tanh	0.0001	(100,)	adam	32.10875446	11
tanh	0.0001	(100,)	sgd	3.026777276	11
tanh	0.0001	(50, 50)	adam	54.55482405	11
tanh	0.0001	(50, 50)	sgd	1.165586472	11
tanh	0.001	(50,)	adam	69.08502413	11
tanh	0.001	(50,)	sgd	2.339329134	11
tanh	0.001	(100,)	adam	32.10879918	11
tanh	0.001	(100,)	sgd	3.027052566	11

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.001	(50, 50)	adam	54.55463046	11
tanh	0.001	(50, 50)	sgd	1.165615154	11
relu	0.0001	(50,)	adam	46.47093279	11
relu	0.0001	(50,)	sgd	2.724828379	11
relu	0.0001	(100,)	adam	28.8039163	11
relu	0.0001	(100,)	sgd	2.750898551	11
relu	0.0001	(50, 50)	adam	6.365295589	11
relu	0.0001	(50, 50)	sgd	1.140470753	11
relu	0.001	(50,)	adam	46.43762501	11
relu	0.001	(50,)	sgd	2.724741645	11
relu	0.001	(100,)	adam	29.04992063	11
relu	0.001	(100,)	sgd	2.751160545	11
relu	0.001	(50, 50)	adam	6.513625678	11
relu	0.001	(50, 50)	sgd	1.140488859	11
tanh	0.0001	(50,)	adam	67.29710824	12
tanh	0.0001	(50,)	sgd	3.55649116	12
tanh	0.0001	(100,)	adam	29.72285185	12
tanh	0.0001	(100,)	sgd	3.427551507	12
tanh	0.0001	(50, 50)	adam	56.38945469	12
tanh	0.0001	(50, 50)	sgd	1.736245688	12
tanh	0.001	(50,)	adam	67.29701305	12
tanh	0.001	(50,)	sgd	3.556722543	12
tanh	0.001	(100,)	adam	29.72288986	12
tanh	0.001	(100,)	sgd	3.427831088	12

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
tanh	0.001	(50, 50)	adam	56.38934398	12
tanh	0.001	(50, 50)	sgd	1.736311153	12
relu	0.0001	(50,)	adam	35.16723987	12
relu	0.0001	(50,)	sgd	3.034739327	12
relu	0.0001	(100,)	adam	25.05733609	12
relu	0.0001	(100,)	sgd	3.083752965	12
relu	0.0001	(50, 50)	adam	14.36128311	12
relu	0.0001	(50, 50)	sgd	1.755408508	12
relu	0.001	(50,)	adam	35.17207331	12
relu	0.001	(50,)	sgd	3.034484762	12
relu	0.001	(100,)	adam	24.56385353	12
relu	0.001	(100,)	sgd	3.084064648	12
relu	0.001	(50, 50)	adam	13.12791655	12
relu	0.001	(50, 50)	sgd	1.756524065	12
tanh	0.0001	(50,)	adam	65.29670543	13
tanh	0.0001	(50,)	sgd	2.766822803	13
tanh	0.0001	(100,)	adam	27.60964838	13
tanh	0.0001	(100,)	sgd	2.871064639	13
tanh	0.0001	(50, 50)	adam	51.45797386	13
tanh	0.0001	(50, 50)	sgd	1.746359667	13
tanh	0.001	(50,)	adam	65.29672624	13
tanh	0.001	(50,)	sgd	2.767077068	13
tanh	0.001	(100,)	adam	27.60971316	13
tanh	0.001	(100,)	sgd	2.871260005	13

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
tanh	0.001	(50, 50)	adam	51.45803414	13
tanh	0.001	(50, 50)	sgd	1.746358034	13
relu	0.0001	(50,)	adam	29.43104267	13
relu	0.0001	(50,)	sgd	3.793710496	13
relu	0.0001	(100,)	adam	23.79636588	13
relu	0.0001	(100,)	sgd	3.677992147	13
relu	0.0001	(50, 50)	adam	10.4499187	13
relu	0.0001	(50, 50)	sgd	1.37518077	13
relu	0.001	(50,)	adam	29.04367678	13
relu	0.001	(50,)	sgd	3.793886885	13
relu	0.001	(100,)	adam	23.51974684	13
relu	0.001	(100,)	sgd	3.678339828	13
relu	0.001	(50, 50)	adam	10.50408235	13
relu	0.001	(50, 50)	sgd	1.374079294	13
tanh	0.0001	(50,)	adam	63.51197592	14
tanh	0.0001	(50,)	sgd	2.045077745	14
tanh	0.0001	(100,)	adam	25.59904981	14
tanh	0.0001	(100,)	sgd	2.433101295	14
tanh	0.0001	(50, 50)	adam	62.33187705	14
tanh	0.0001	(50, 50)	sgd	1.35843824	14
tanh	0.001	(50,)	adam	63.51192477	14
tanh	0.001	(50,)	sgd	2.045262468	14
tanh	0.001	(100,)	adam	25.59908475	14
tanh	0.001	(100,)	sgd	2.433292037	14

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
tanh	0.001	(50, 50)	adam	62.33138924	14
tanh	0.001	(50, 50)	sgd	1.358489923	14
relu	0.0001	(50,)	adam	27.75056218	14
relu	0.0001	(50,)	sgd	4.189850961	14
relu	0.0001	(100,)	adam	15.01437958	14
relu	0.0001	(100,)	sgd	4.118306118	14
relu	0.0001	(50, 50)	adam	5.733275889	14
relu	0.0001	(50, 50)	sgd	1.506710841	14
relu	0.001	(50,)	adam	25.91378946	14
relu	0.001	(50,)	sgd	4.190258858	14
relu	0.001	(100,)	adam	14.36706842	14
relu	0.001	(100,)	sgd	4.11815866	14
relu	0.001	(50, 50)	adam	5.622983187	14
relu	0.001	(50, 50)	sgd	1.507677407	14
tanh	0.0001	(50,)	adam	66.01674384	15
tanh	0.0001	(50,)	sgd	5.908920865	15
tanh	0.0001	(100,)	adam	30.19356277	15
tanh	0.0001	(100,)	sgd	5.596103785	15
tanh	0.0001	(50, 50)	adam	53.8274348	15
tanh	0.0001	(50, 50)	sgd	2.127573819	15
tanh	0.001	(50,)	adam	66.01658757	15
tanh	0.001	(50,)	sgd	5.90917577	15
tanh	0.001	(100,)	adam	30.19358653	15
tanh	0.001	(100,)	sgd	5.596434886	15

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.001	(50, 50)	adam	53.82730368	15
tanh	0.001	(50, 50)	sgd	2.127701936	15
relu	0.0001	(50,)	adam	34.36092038	15
relu	0.0001	(50,)	sgd	6.194413848	15
relu	0.0001	(100,)	adam	27.81342697	15
relu	0.0001	(100,)	sgd	5.81127682	15
relu	0.0001	(50, 50)	adam	8.642475916	15
relu	0.0001	(50, 50)	sgd	2.654097208	15
relu	0.001	(50,)	adam	34.40325239	15
relu	0.001	(50,)	sgd	6.19411104	15
relu	0.001	(100,)	adam	27.34385146	15
relu	0.001	(100,)	sgd	5.811847452	15
relu	0.001	(50, 50)	adam	8.714388506	15
relu	0.001	(50, 50)	sgd	2.642453118	15

Table A.16 : NLICA_MLP model hyperparameters tuning for different ICs using all K = 22 external factors.

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.0001	(50,)	adam	182.062435	1
tanh	0.0001	(50,)	sgd	32.67228535	1
tanh	0.0001	(100,)	adam	79.21146056	1
tanh	0.0001	(100,)	sgd	33.36418149	1
tanh	0.0001	(50, 50)	adam	99.05733613	1
tanh	0.0001	(50, 50)	sgd	31.36407349	1

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.001	(50,)	adam	182.062464	1
tanh	0.001	(50,)	sgd	32.67256941	1
tanh	0.001	(100,)	adam	79.21146565	1
tanh	0.001	(100,)	sgd	33.36442683	1
tanh	0.001	(50, 50)	adam	99.05133814	1
tanh	0.001	(50, 50)	sgd	31.36410816	1
relu	0.0001	(50,)	adam	271.8637879	1
relu	0.0001	(50,)	sgd	31.94117816	1
relu	0.0001	(100,)	adam	101.0723578	1
relu	0.0001	(100,)	sgd	32.10880257	1
relu	0.0001	(50, 50)	adam	38.76032033	1
relu	0.0001	(50, 50)	sgd	30.93548401	1
relu	0.001	(50,)	adam	271.8639047	1
relu	0.001	(50,)	sgd	31.94154396	1
relu	0.001	(100,)	adam	101.0724333	1
relu	0.001	(100,)	sgd	32.10914399	1
relu	0.001	(50, 50)	adam	38.3420954	1
relu	0.001	(50, 50)	sgd	30.93544146	1
tanh	0.0001	(50,)	adam	210.5970094	2
tanh	0.0001	(50,)	sgd	20.20591457	2
tanh	0.0001	(100,)	adam	95.70980468	2
tanh	0.0001	(100,)	sgd	20.37174751	2
tanh	0.0001	(50, 50)	adam	67.11295688	2
tanh	0.0001	(50, 50)	sgd	19.25467476	2

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.001	(50,)	adam	210.5970336	2
tanh	0.001	(50,)	sgd	20.20660126	2
tanh	0.001	(100,)	adam	95.70990299	2
tanh	0.001	(100,)	sgd	20.37260447	2
tanh	0.001	(50, 50)	adam	67.11242876	2
tanh	0.001	(50, 50)	sgd	19.25469025	2
relu	0.0001	(50,)	adam	252.5488639	2
relu	0.0001	(50,)	sgd	29.78618571	2
relu	0.0001	(100,)	adam	111.9818379	2
relu	0.0001	(100,)	sgd	30.02926516	2
relu	0.0001	(50, 50)	adam	25.2253123	2
relu	0.0001	(50, 50)	sgd	12.47128233	2
relu	0.001	(50,)	adam	252.5416392	2
relu	0.001	(50,)	sgd	29.7861792	2
relu	0.001	(100,)	adam	111.9829729	2
relu	0.001	(100,)	sgd	30.02926918	2
relu	0.001	(50, 50)	adam	25.20827685	2
relu	0.001	(50, 50)	sgd	12.47112366	2
tanh	0.0001	(50,)	adam	170.9391976	3
tanh	0.0001	(50,)	sgd	17.23937158	3
tanh	0.0001	(100,)	adam	72.27522407	3
tanh	0.0001	(100,)	sgd	19.79871953	3
tanh	0.0001	(50, 50)	adam	117.7050644	3
tanh	0.0001	(50, 50)	sgd	15.63472187	3

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
tanh	0.001	(50,)	adam	170.9392621	3
tanh	0.001	(50,)	sgd	17.24050288	3
tanh	0.001	(100,)	adam	72.27533362	3
tanh	0.001	(100,)	sgd	19.79984388	3
tanh	0.001	(50, 50)	adam	117.7053802	3
tanh	0.001	(50, 50)	sgd	15.63427149	3
relu	0.0001	(50,)	adam	240.0639503	3
relu	0.0001	(50,)	sgd	33.37010968	3
relu	0.0001	(100,)	adam	103.3243609	3
relu	0.0001	(100,)	sgd	33.31320959	3
relu	0.0001	(50, 50)	adam	29.27521658	3
relu	0.0001	(50, 50)	sgd	15.38210174	3
relu	0.001	(50,)	adam	240.073241	3
relu	0.001	(50,)	sgd	33.37015945	3
relu	0.001	(100,)	adam	103.3062503	3
relu	0.001	(100,)	sgd	33.31320375	3
relu	0.001	(50, 50)	adam	29.45959143	3
relu	0.001	(50, 50)	sgd	15.37726876	3
tanh	0.0001	(50,)	adam	91.19276203	4
tanh	0.0001	(50,)	sgd	7.89358198	4
tanh	0.0001	(100,)	adam	30.62048718	4
tanh	0.0001	(100,)	sgd	7.773874225	4
tanh	0.0001	(50, 50)	adam	52.73214774	4
tanh	0.0001	(50, 50)	sgd	8.358371117	4

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
tanh	0.001	(50,)	adam	91.19280141	4
tanh	0.001	(50,)	sgd	7.893612695	4
tanh	0.001	(100,)	adam	30.62052751	4
tanh	0.001	(100,)	sgd	7.773895308	4
tanh	0.001	(50, 50)	adam	52.73204311	4
tanh	0.001	(50, 50)	sgd	8.358293601	4
relu	0.0001	(50,)	adam	28.71718222	4
relu	0.0001	(50,)	sgd	14.66577264	4
relu	0.0001	(100,)	adam	23.61102037	4
relu	0.0001	(100,)	sgd	12.3216539	4
relu	0.0001	(50, 50)	adam	17.5201045	4
relu	0.0001	(50, 50)	sgd	6.26556768	4
relu	0.001	(50,)	adam	28.22526277	4
relu	0.001	(50,)	sgd	14.66627768	4
relu	0.001	(100,)	adam	23.6656721	4
relu	0.001	(100,)	sgd	12.32189484	4
relu	0.001	(50, 50)	adam	16.70711974	4
relu	0.001	(50, 50)	sgd	6.269893827	4
tanh	0.0001	(50,)	adam	93.15405732	5
tanh	0.0001	(50,)	sgd	13.69922222	5
tanh	0.0001	(100,)	adam	50.02638248	5
tanh	0.0001	(100,)	sgd	12.85142903	5
tanh	0.0001	(50, 50)	adam	64.11119288	5
tanh	0.0001	(50, 50)	sgd	17.64734294	5

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.001	(50,)	adam	93.1541052	5
tanh	0.001	(50,)	sgd	13.69943202	5
tanh	0.001	(100,)	adam	50.02646424	5
tanh	0.001	(100,)	sgd	12.8517376	5
tanh	0.001	(50, 50)	adam	64.11110935	5
tanh	0.001	(50, 50)	sgd	17.64653454	5
relu	0.0001	(50,)	adam	52.41441187	5
relu	0.0001	(50,)	sgd	25.80466134	5
relu	0.0001	(100,)	adam	45.20063205	5
relu	0.0001	(100,)	sgd	24.15709861	5
relu	0.0001	(50, 50)	adam	22.25359466	5
relu	0.0001	(50, 50)	sgd	9.988499363	5
relu	0.001	(50,)	adam	51.6803096	5
relu	0.001	(50,)	sgd	25.80316493	5
relu	0.001	(100,)	adam	48.21806902	5
relu	0.001	(100,)	sgd	24.15710723	5
relu	0.001	(50, 50)	adam	22.18023616	5
relu	0.001	(50, 50)	sgd	12.09773061	5
tanh	0.0001	(50,)	adam	81.59358787	6
tanh	0.0001	(50,)	sgd	9.746432161	6
tanh	0.0001	(100,)	adam	30.82986903	6
tanh	0.0001	(100,)	sgd	9.457708106	6
tanh	0.0001	(50, 50)	adam	53.1644803	6
tanh	0.0001	(50, 50)	sgd	8.983571166	6

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.001	(50,)	adam	81.59361198	6
tanh	0.001	(50,)	sgd	9.746541596	6
tanh	0.001	(100,)	adam	30.82990777	6
tanh	0.001	(100,)	sgd	9.457865928	6
tanh	0.001	(50, 50)	adam	53.16421383	6
tanh	0.001	(50, 50)	sgd	8.983640128	6
relu	0.0001	(50,)	adam	30.85132873	6
relu	0.0001	(50,)	sgd	12.36929342	6
relu	0.0001	(100,)	adam	23.69039991	6
relu	0.0001	(100,)	sgd	12.85238017	6
relu	0.0001	(50, 50)	adam	14.83813447	6
relu	0.0001	(50, 50)	sgd	10.16191619	6
relu	0.001	(50,)	adam	29.46249006	6
relu	0.001	(50,)	sgd	12.3698301	6
relu	0.001	(100,)	adam	23.71261147	6
relu	0.001	(100,)	sgd	12.84636531	6
relu	0.001	(50, 50)	adam	16.79924443	6
relu	0.001	(50, 50)	sgd	10.08180134	6
tanh	0.0001	(50,)	adam	78.03583835	7
tanh	0.0001	(50,)	sgd	9.652285855	7
tanh	0.0001	(100,)	adam	39.85421619	7
tanh	0.0001	(100,)	sgd	9.827606566	7
tanh	0.0001	(50, 50)	adam	57.32201799	7
tanh	0.0001	(50, 50)	sgd	8.353310302	7

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.001	(50,)	adam	78.03585413	7
tanh	0.001	(50,)	sgd	9.652499926	7
tanh	0.001	(100,)	adam	39.85426938	7
tanh	0.001	(100,)	sgd	9.827821853	7
tanh	0.001	(50, 50)	adam	57.3219587	7
tanh	0.001	(50, 50)	sgd	8.353347922	7
relu	0.0001	(50,)	adam	37.38547125	7
relu	0.0001	(50,)	sgd	17.84291171	7
relu	0.0001	(100,)	adam	34.85231814	7
relu	0.0001	(100,)	sgd	16.84410243	7
relu	0.0001	(50, 50)	adam	15.29161862	7
relu	0.0001	(50, 50)	sgd	9.813705905	7
relu	0.001	(50,)	adam	37.05046215	7
relu	0.001	(50,)	sgd	17.8429168	7
relu	0.001	(100,)	adam	34.6269666	7
relu	0.001	(100,)	sgd	16.84649507	7
relu	0.001	(50, 50)	adam	14.89606281	7
relu	0.001	(50, 50)	sgd	9.708668624	7
tanh	0.0001	(50,)	adam	64.26288395	8
tanh	0.0001	(50,)	sgd	6.105715989	8
tanh	0.0001	(100,)	adam	30.43710974	8
tanh	0.0001	(100,)	sgd	6.375317164	8
tanh	0.0001	(50, 50)	adam	60.06596506	8
tanh	0.0001	(50, 50)	sgd	3.832481174	8

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.001	(50,)	adam	64.26289409	8
tanh	0.001	(50,)	sgd	6.106040072	8
tanh	0.001	(100,)	adam	30.43715551	8
tanh	0.001	(100,)	sgd	6.375623027	8
tanh	0.001	(50, 50)	adam	60.06605271	8
tanh	0.001	(50, 50)	sgd	3.832841044	8
relu	0.0001	(50,)	adam	34.62090179	8
relu	0.0001	(50,)	sgd	9.941543367	8
relu	0.0001	(100,)	adam	29.22693295	8
relu	0.0001	(100,)	sgd	9.868885337	8
relu	0.0001	(50, 50)	adam	13.510237	8
relu	0.0001	(50, 50)	sgd	6.111280905	8
relu	0.001	(50,)	adam	34.62530301	8
relu	0.001	(50,)	sgd	9.941889089	8
relu	0.001	(100,)	adam	28.93622251	8
relu	0.001	(100,)	sgd	9.867619057	8
relu	0.001	(50, 50)	adam	13.47333806	8
relu	0.001	(50, 50)	sgd	6.132882646	8
tanh	0.0001	(50,)	adam	82.72403455	9
tanh	0.0001	(50,)	sgd	6.130722844	9
tanh	0.0001	(100,)	adam	36.81571768	9
tanh	0.0001	(100,)	sgd	5.168059255	9
tanh	0.0001	(50, 50)	adam	56.3453389	9
tanh	0.0001	(50, 50)	sgd	6.677019294	9

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.001	(50,)	adam	82.72406801	9
tanh	0.001	(50,)	sgd	6.130834877	9
tanh	0.001	(100,)	adam	36.81578922	9
tanh	0.001	(100,)	sgd	5.168162727	9
tanh	0.001	(50, 50)	adam	56.34549983	9
tanh	0.001	(50, 50)	sgd	6.676918224	9
relu	0.0001	(50,)	adam	41.43123034	9
relu	0.0001	(50,)	sgd	8.905802858	9
relu	0.0001	(100,)	adam	36.02183574	9
relu	0.0001	(100,)	sgd	9.648046465	9
relu	0.0001	(50, 50)	adam	17.80253671	9
relu	0.0001	(50, 50)	sgd	6.898520602	9
relu	0.001	(50,)	adam	40.94071287	9
relu	0.001	(50,)	sgd	8.898934631	9
relu	0.001	(100,)	adam	34.05310935	9
relu	0.001	(100,)	sgd	9.646677911	9
relu	0.001	(50, 50)	adam	15.48108645	9
relu	0.001	(50, 50)	sgd	6.898056868	9
tanh	0.0001	(50,)	adam	73.45146819	10
tanh	0.0001	(50,)	sgd	6.186502504	10
tanh	0.0001	(100,)	adam	31.9348635	10
tanh	0.0001	(100,)	sgd	6.109298188	10
tanh	0.0001	(50, 50)	adam	57.30765557	10
tanh	0.0001	(50, 50)	sgd	3.99104324	10

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
tanh	0.001	(50,)	adam	73.45136358	10
tanh	0.001	(50,)	sgd	6.186680812	10
tanh	0.001	(100,)	adam	31.93488551	10
tanh	0.001	(100,)	sgd	6.109522255	10
tanh	0.001	(50, 50)	adam	57.30750798	10
tanh	0.001	(50, 50)	sgd	3.991210224	10
relu	0.0001	(50,)	adam	34.99904332	10
relu	0.0001	(50,)	sgd	6.773953405	10
relu	0.0001	(100,)	adam	20.35752787	10
relu	0.0001	(100,)	sgd	5.881951939	10
relu	0.0001	(50, 50)	adam	11.00655026	10
relu	0.0001	(50, 50)	sgd	5.64103248	10
relu	0.001	(50,)	adam	35.36125417	10
relu	0.001	(50,)	sgd	6.773738053	10
relu	0.001	(100,)	adam	19.95412755	10
relu	0.001	(100,)	sgd	5.882406691	10
relu	0.001	(50, 50)	adam	9.980746338	10
relu	0.001	(50, 50)	sgd	5.647849658	10
tanh	0.0001	(50,)	adam	76.87280893	11
tanh	0.0001	(50,)	sgd	11.58135335	11
tanh	0.0001	(100,)	adam	29.87937707	11
tanh	0.0001	(100,)	sgd	12.54391971	11
tanh	0.0001	(50, 50)	adam	51.62437596	11
tanh	0.0001	(50, 50)	sgd	9.931651722	11

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
tanh	0.001	(50,)	adam	76.87270205	11
tanh	0.001	(50,)	sgd	11.58156645	11
tanh	0.001	(100,)	adam	29.87943646	11
tanh	0.001	(100,)	sgd	12.54420314	11
tanh	0.001	(50, 50)	adam	51.62443066	11
tanh	0.001	(50, 50)	sgd	9.931768319	11
relu	0.0001	(50,)	adam	31.64292409	11
relu	0.0001	(50,)	sgd	15.92484361	11
relu	0.0001	(100,)	adam	23.34878019	11
relu	0.0001	(100,)	sgd	15.36445517	11
relu	0.0001	(50, 50)	adam	16.90723615	11
relu	0.0001	(50, 50)	sgd	12.18833877	11
relu	0.001	(50,)	adam	29.77310526	11
relu	0.001	(50,)	sgd	15.92468219	11
relu	0.001	(100,)	adam	23.10506435	11
relu	0.001	(100,)	sgd	15.36154967	11
relu	0.001	(50, 50)	adam	16.67075439	11
relu	0.001	(50, 50)	sgd	12.05781713	11
tanh	0.0001	(50,)	adam	72.3679999	12
tanh	0.0001	(50,)	sgd	6.120443138	12
tanh	0.0001	(100,)	adam	30.99617568	12
tanh	0.0001	(100,)	sgd	6.26228012	12
tanh	0.0001	(50, 50)	adam	59.31370819	12
tanh	0.0001	(50, 50)	sgd	5.524886298	12

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.001	(50,)	adam	72.3679585	12
tanh	0.001	(50,)	sgd	6.120564206	12
tanh	0.001	(100,)	adam	30.99620528	12
tanh	0.001	(100,)	sgd	6.262401197	12
tanh	0.001	(50, 50)	adam	59.31357672	12
tanh	0.001	(50, 50)	sgd	5.525027334	12
relu	0.0001	(50,)	adam	30.40386131	12
relu	0.0001	(50,)	sgd	6.767496967	12
relu	0.0001	(100,)	adam	23.69630177	12
relu	0.0001	(100,)	sgd	6.517121353	12
relu	0.0001	(50, 50)	adam	14.27950318	12
relu	0.0001	(50, 50)	sgd	6.292893785	12
relu	0.001	(50,)	adam	30.12787316	12
relu	0.001	(50,)	sgd	6.767757453	12
relu	0.001	(100,)	adam	23.46584372	12
relu	0.001	(100,)	sgd	6.517236485	12
relu	0.001	(50, 50)	adam	14.10106202	12
relu	0.001	(50, 50)	sgd	6.211917846	12
tanh	0.0001	(50,)	adam	73.08437338	13
tanh	0.0001	(50,)	sgd	5.112667926	13
tanh	0.0001	(100,)	adam	28.97875288	13
tanh	0.0001	(100,)	sgd	4.790382401	13
tanh	0.0001	(50, 50)	adam	58.34083348	13
tanh	0.0001	(50, 50)	sgd	4.299396383	13

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
tanh	0.001	(50,)	adam	73.08433502	13
tanh	0.001	(50,)	sgd	5.112753355	13
tanh	0.001	(100,)	adam	28.97878188	13
tanh	0.001	(100,)	sgd	4.790475607	13
tanh	0.001	(50, 50)	adam	58.3405136	13
tanh	0.001	(50, 50)	sgd	4.299533987	13
relu	0.0001	(50,)	adam	34.51534604	13
relu	0.0001	(50,)	sgd	4.677630701	13
relu	0.0001	(100,)	adam	20.93805087	13
relu	0.0001	(100,)	sgd	4.610727749	13
relu	0.0001	(50, 50)	adam	6.420919318	13
relu	0.0001	(50, 50)	sgd	4.417718613	13
relu	0.001	(50,)	adam	34.75299382	13
relu	0.001	(50,)	sgd	4.677768226	13
relu	0.001	(100,)	adam	20.40578767	13
relu	0.001	(100,)	sgd	4.610437929	13
relu	0.001	(50, 50)	adam	6.447628936	13
relu	0.001	(50, 50)	sgd	4.431750896	13
tanh	0.0001	(50,)	adam	64.67824882	14
tanh	0.0001	(50,)	sgd	7.79473486	14
tanh	0.0001	(100,)	adam	29.38615362	14
tanh	0.0001	(100,)	sgd	8.500936929	14
tanh	0.0001	(50, 50)	adam	61.46293195	14
tanh	0.0001	(50, 50)	sgd	6.586757172	14

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.001	(50,)	adam	64.67812514	14
tanh	0.001	(50,)	sgd	7.795085532	14
tanh	0.001	(100,)	adam	29.38619838	14
tanh	0.001	(100,)	sgd	8.501314914	14
tanh	0.001	(50, 50)	adam	61.46232943	14
tanh	0.001	(50, 50)	sgd	6.587126705	14
relu	0.0001	(50,)	adam	31.11876884	14
relu	0.0001	(50,)	sgd	12.14498065	14
relu	0.0001	(100,)	adam	27.30180889	14
relu	0.0001	(100,)	sgd	12.32394606	14
relu	0.0001	(50, 50)	adam	15.66345026	14
relu	0.0001	(50, 50)	sgd	7.037368988	14
relu	0.001	(50,)	adam	31.07867748	14
relu	0.001	(50,)	sgd	12.14500806	14
relu	0.001	(100,)	adam	26.98999639	14
relu	0.001	(100,)	sgd	12.32403273	14
relu	0.001	(50, 50)	adam	15.28017461	14
relu	0.001	(50, 50)	sgd	5.57777831	14
tanh	0.0001	(50,)	adam	87.28587216	15
tanh	0.0001	(50,)	sgd	8.100906593	15
tanh	0.0001	(100,)	adam	29.86384442	15
tanh	0.0001	(100,)	sgd	8.751678662	15
tanh	0.0001	(50, 50)	adam	55.48060285	15
tanh	0.0001	(50, 50)	sgd	9.163274084	15

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
tanh	0.001	(50,)	adam	87.28579413	15
tanh	0.001	(50,)	sgd	8.10124258	15
tanh	0.001	(100,)	adam	29.86389165	15
tanh	0.001	(100,)	sgd	8.751990017	15
tanh	0.001	(50, 50)	adam	55.48066626	15
tanh	0.001	(50, 50)	sgd	9.163142576	15
relu	0.0001	(50,)	adam	29.60387206	15
relu	0.0001	(50,)	sgd	10.53003782	15
relu	0.0001	(100,)	adam	24.48596473	15
relu	0.0001	(100,)	sgd	10.40556962	15
relu	0.0001	(50, 50)	adam	14.28414504	15
relu	0.0001	(50, 50)	sgd	8.211579607	15
relu	0.001	(50,)	adam	29.64945663	15
relu	0.001	(50,)	sgd	10.52940053	15
relu	0.001	(100,)	adam	25.20200095	15
relu	0.001	(100,)	sgd	10.40204994	15
relu	0.001	(50, 50)	adam	14.08709572	15
relu	0.001	(50, 50)	sgd	8.213108389	15
tanh	0.0001	(50,)	adam	61.66501507	16
tanh	0.0001	(50,)	sgd	7.885018306	16
tanh	0.0001	(100,)	adam	26.6503096	16
tanh	0.0001	(100,)	sgd	8.264051148	16
tanh	0.0001	(50, 50)	adam	52.86320928	16
tanh	0.0001	(50, 50)	sgd	7.120385441	16

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
tanh	0.001	(50,)	adam	61.66491937	16
tanh	0.001	(50,)	sgd	7.885187009	16
tanh	0.001	(100,)	adam	26.65031212	16
tanh	0.001	(100,)	sgd	8.264248284	16
tanh	0.001	(50, 50)	adam	52.86307049	16
tanh	0.001	(50, 50)	sgd	7.120558463	16
relu	0.0001	(50,)	adam	28.7843538	16
relu	0.0001	(50,)	sgd	9.209566724	16
relu	0.0001	(100,)	adam	20.88823804	16
relu	0.0001	(100,)	sgd	9.237337771	16
relu	0.0001	(50, 50)	adam	13.50440162	16
relu	0.0001	(50, 50)	sgd	8.432805428	16
relu	0.001	(50,)	adam	28.79389676	16
relu	0.001	(50,)	sgd	9.208886922	16
relu	0.001	(100,)	adam	20.89223368	16
relu	0.001	(100,)	sgd	9.225793791	16
relu	0.001	(50, 50)	adam	13.45377094	16
relu	0.001	(50, 50)	sgd	8.43306627	16
tanh	0.0001	(50,)	adam	78.25897822	17
tanh	0.0001	(50,)	sgd	4.775585137	17
tanh	0.0001	(100,)	adam	31.90764335	17
tanh	0.0001	(100,)	sgd	4.658684415	17
tanh	0.0001	(50, 50)	adam	61.39199776	17
tanh	0.0001	(50, 50)	sgd	2.730540148	17

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.001	(50,)	adam	78.25897254	17
tanh	0.001	(50,)	sgd	4.775763068	17
tanh	0.001	(100,)	adam	31.9076774	17
tanh	0.001	(100,)	sgd	4.658938745	17
tanh	0.001	(50, 50)	adam	61.39199079	17
tanh	0.001	(50, 50)	sgd	2.730664131	17
relu	0.0001	(50,)	adam	31.88418166	17
relu	0.0001	(50,)	sgd	5.386460374	17
relu	0.0001	(100,)	adam	24.06012099	17
relu	0.0001	(100,)	sgd	6.474194177	17
relu	0.0001	(50, 50)	adam	15.05731704	17
relu	0.0001	(50, 50)	sgd	3.527612366	17
relu	0.001	(50,)	adam	31.90097197	17
relu	0.001	(50,)	sgd	5.387809077	17
relu	0.001	(100,)	adam	24.25004394	17
relu	0.001	(100,)	sgd	6.474650298	17
relu	0.001	(50, 50)	adam	15.05145487	17
relu	0.001	(50, 50)	sgd	3.527866382	17
tanh	0.0001	(50,)	adam	64.10885141	18
tanh	0.0001	(50,)	sgd	3.967945413	18
tanh	0.0001	(100,)	adam	26.4395461	18
tanh	0.0001	(100,)	sgd	3.543046601	18
tanh	0.0001	(50, 50)	adam	41.99322779	18
tanh	0.0001	(50, 50)	sgd	1.963695552	18

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.001	(50,)	adam	64.1088216	18
tanh	0.001	(50,)	sgd	3.968190122	18
tanh	0.001	(100,)	adam	26.43962978	18
tanh	0.001	(100,)	sgd	3.543243229	18
tanh	0.001	(50, 50)	adam	41.99326078	18
tanh	0.001	(50, 50)	sgd	1.963799284	18
relu	0.0001	(50,)	adam	28.21894345	18
relu	0.0001	(50,)	sgd	2.927318293	18
relu	0.0001	(100,)	adam	16.29731071	18
relu	0.0001	(100,)	sgd	3.017106361	18
relu	0.0001	(50, 50)	adam	6.622191488	18
relu	0.0001	(50, 50)	sgd	2.230253528	18
relu	0.001	(50,)	adam	28.23247915	18
relu	0.001	(50,)	sgd	2.927961515	18
relu	0.001	(100,)	adam	16.04638846	18
relu	0.001	(100,)	sgd	3.017672037	18
relu	0.001	(50, 50)	adam	6.571636399	18
relu	0.001	(50, 50)	sgd	2.232796613	18
tanh	0.0001	(50,)	adam	54.02706401	19
tanh	0.0001	(50,)	sgd	5.632268289	19
tanh	0.0001	(100,)	adam	24.83501204	19
tanh	0.0001	(100,)	sgd	5.29240034	19
tanh	0.0001	(50, 50)	adam	66.31926667	19
tanh	0.0001	(50, 50)	sgd	3.340566794	19

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
tanh	0.001	(50,)	adam	54.02700212	19
tanh	0.001	(50,)	sgd	5.632509863	19
tanh	0.001	(100,)	adam	24.83507491	19
tanh	0.001	(100,)	sgd	5.292662605	19
tanh	0.001	(50, 50)	adam	66.31870223	19
tanh	0.001	(50, 50)	sgd	3.340817727	19
relu	0.0001	(50,)	adam	21.68667198	19
relu	0.0001	(50,)	sgd	5.885529977	19
relu	0.0001	(100,)	adam	16.3590227	19
relu	0.0001	(100,)	sgd	4.809759417	19
relu	0.0001	(50, 50)	adam	10.68292321	19
relu	0.0001	(50, 50)	sgd	2.604785839	19
relu	0.001	(50,)	adam	21.6927769	19
relu	0.001	(50,)	sgd	5.886771304	19
relu	0.001	(100,)	adam	16.06265533	19
relu	0.001	(100,)	sgd	4.809699067	19
relu	0.001	(50, 50)	adam	10.65783585	19
relu	0.001	(50, 50)	sgd	2.6049568	19
tanh	0.0001	(50,)	adam	66.29168398	20
tanh	0.0001	(50,)	sgd	4.280032774	20
tanh	0.0001	(100,)	adam	24.38218077	20
tanh	0.0001	(100,)	sgd	4.006305992	20
tanh	0.0001	(50, 50)	adam	51.61339693	20
tanh	0.0001	(50, 50)	sgd	1.907890137	20

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.001	(50,)	adam	66.29156446	20
tanh	0.001	(50,)	sgd	4.280163775	20
tanh	0.001	(100,)	adam	24.3822002	20
tanh	0.001	(100,)	sgd	4.006433571	20
tanh	0.001	(50, 50)	adam	51.61335453	20
tanh	0.001	(50, 50)	sgd	1.908067551	20
relu	0.0001	(50,)	adam	20.08881546	20
relu	0.0001	(50,)	sgd	3.793560307	20
relu	0.0001	(100,)	adam	12.71919778	20
relu	0.0001	(100,)	sgd	3.852329212	20
relu	0.0001	(50, 50)	adam	5.107387516	20
relu	0.0001	(50, 50)	sgd	2.045046625	20
relu	0.001	(50,)	adam	20.21379826	20
relu	0.001	(50,)	sgd	3.794706091	20
relu	0.001	(100,)	adam	12.50771823	20
relu	0.001	(100,)	sgd	3.852441873	20
relu	0.001	(50, 50)	adam	5.246381054	20
relu	0.001	(50, 50)	sgd	2.045416265	20
tanh	0.0001	(50,)	adam	58.66241652	21
tanh	0.0001	(50,)	sgd	2.779401932	21
tanh	0.0001	(100,)	adam	24.90966917	21
tanh	0.0001	(100,)	sgd	2.509158911	21
tanh	0.0001	(50, 50)	adam	51.43153835	21
tanh	0.0001	(50, 50)	sgd	2.635116615	21

Table continues on the next page

Activation	Alpha	Hidden_layer	Solver	Mean MSE	n_components
tanh	0.001	(50,)	adam	58.6622046	21
tanh	0.001	(50,)	sgd	2.77955965	21
tanh	0.001	(100,)	adam	24.90970919	21
tanh	0.001	(100,)	sgd	2.509294894	21
tanh	0.001	(50, 50)	adam	51.43152335	21
tanh	0.001	(50, 50)	sgd	1.272335944	21
relu	0.0001	(50,)	adam	18.06724034	21
relu	0.0001	(50,)	sgd	2.346150342	21
relu	0.0001	(100,)	adam	13.36670417	21
relu	0.0001	(100,)	sgd	2.272487476	21
relu	0.0001	(50, 50)	adam	4.769017214	21
relu	0.0001	(50, 50)	sgd	1.192216333	21
relu	0.001	(50,)	adam	16.91289043	21
relu	0.001	(50,)	sgd	2.346013716	21
relu	0.001	(100,)	adam	13.9360097	21
relu	0.001	(100,)	sgd	2.272649047	21
relu	0.001	(50, 50)	adam	4.82513124	21
relu	0.001	(50, 50)	sgd	1.274975385	21
tanh	0.0001	(50,)	adam	58.12964334	22
tanh	0.0001	(50,)	sgd	6.156120236	22
tanh	0.0001	(100,)	adam	29.96062121	22
tanh	0.0001	(100,)	sgd	6.770758486	22
tanh	0.0001	(50, 50)	adam	51.66346395	22
tanh	0.0001	(50, 50)	sgd	3.086067197	22

Table continues on the next page

Activation	**Alpha**	**Hidden_layer**	**Solver**	**Mean MSE**	**n_components**
tanh	0.001	(50,)	adam	58.12949182	22
tanh	0.001	(50,)	sgd	6.156347555	22
tanh	0.001	(100,)	adam	29.96068411	22
tanh	0.001	(100,)	sgd	6.771045473	22
tanh	0.001	(50, 50)	adam	51.66334673	22
tanh	0.001	(50, 50)	sgd	3.086308987	22
relu	0.0001	(50,)	adam	31.69380884	22
relu	0.0001	(50,)	sgd	7.18514807	22
relu	0.0001	(100,)	adam	24.37123338	22
relu	0.0001	(100,)	sgd	7.319747165	22
relu	0.0001	(50, 50)	adam	11.99694926	22
relu	0.0001	(50, 50)	sgd	1.19233211	22
relu	0.001	(50,)	adam	31.69864663	22
relu	0.001	(50,)	sgd	7.185325131	22
relu	0.001	(100,)	adam	24.38334631	22
relu	0.001	(100,)	sgd	7.320098052	22
relu	0.001	(50, 50)	adam	11.81720694	22
relu	0.001	(50, 50)	sgd	2.637120262	22

Appendix B

Appendix B: Code links

- The code used for data pre-processing and EDA can be found at:

  https://drive.google.com/file/d/1Dxuq_wgPdcIBb5JqElHWbA8YokLBwlBc/view?usp=sharing

- The code used for building the CCMFI Model can be found at:

  https://drive.google.com/file/d/1t12EGiRyWE_OeYLLPsZI1bTa-QSkWygQ/view?usp=sharing

- The code used for building the CCMFP prediction model based on the ICA_SVR model can be found at:

  https://colab.research.google.com/drive/1p2iOrfghsq97im1UBhm_KM7MG7TNmo5v?usp=drive_link

- The code used for building the CCMFP prediction model based on the NLICA_SVR model can be found at:

  https://colab.research.google.com/drive/1BB_aAGjZstE9nIeXrakcinrPBgRCU9pI?usp=drive_link

- The code used for building the CCMFP prediction model based on the PCA_SVR model can be found at:

  https://colab.research.google.com/drive/12Pczt61GPpFtGYC5sgVmbjT6SlOJfTyR?usp=drive_link

- The code used for building the CCMFP prediction model based on the ICA_MLP can be found at:

  `https://colab.research.google.com/drive/1lL-UXvr6meMiaRdchqNkWiBmJkocifs0?usp=drive_link`

- The code used for building the CCMFP prediction model based on the NLICA_MLP can be found at:

  `https://colab.research.google.com/drive/14SeO3LB3n0Q57GZ7y1Q7w50ugIvQrPHO?usp=drive_link`

- The code used for building the CCMFP prediction model based on the PCA_MLP models can be found at:

  `https://colab.research.google.com/drive/16rZwoFYp__CA5cfE5ptw08RZzE1JVW99?usp=drive_link`

Bibliography

[1] Aatola, P., Ollikainen, M., and Toppinen, A. (2013). Price determination in the EU ETS market: Theory and econometric analysis with market fundamentals. *Energy Economics*, 36:380–395.

[2] Abedin, M., Hasan, M., Hassan, M. K., and Hájek, P. (2021). Deep learning-based exchange rate prediction during the covid-19 pandemic. *Annals of Operations Research*.

[3] Adams, S., Adedoyin, F., Olaniran, E., and Bekun, F. V. (2020). Energy consumption, economic policy uncertainty and carbon emissions; causality evidence from resource rich economies. *Economic Analysis and Policy*, 68:179–190.

[4] Adekoya, O. B. (2021). Predicting carbon allowance prices with energy prices: A new approach. *Journal of Cleaner Production*, 282:124519.

[5] Ağbulut, Ü., Gürel, A. E., and Biçen, Y. (2021). Prediction of daily global solar radiation using different machine learning algorithms: Evaluation and comparison. *Renewable and Sustainable Energy Reviews*, 135:110114.

[6] Aït-Sahalia, Y. and Jacod, J. (2014). High-frequency financial econometrics. In *High-Frequency Financial Econometrics*. Princeton University Press.

[7] Alanazi, I., Al-Doghman, F., Alsubhi, A., and Hussain, F. (2024). Carbon credits price prediction model (ccppm). In *International Conference on Advanced Information Networking and Applications*, pages 143–150. Springer.

[8] Alberola, E., Chevallier, J., and Chèze, B. (2008). Price drivers and structural breaks in European carbon prices 2005–2007. *Energy Policy*, 36(2):787–797.

[9] Awad, M. and Khanna, R. (2015). *Support Vector Regression*, pages 67–80. Apress, Berkeley, CA.

[10] Bakay, M. S. and Ağbulut, Ü. (2021). Electricity production based forecasting of greenhouse gas emissions in turkey with deep learning, support vector machine and artificial neural network algorithms. *Journal of Cleaner Production*, 285:125324.

[11] Benz, E. and Trück, S. (2009). Modeling the price dynamics of co2 emission allowances. *Energy Economics*, 31(1):4–15.

[12] Berthold, M. R., Borgelt, C., Höppner, F., Klawonn, F., and Silipo, R. (2020). *Guide to intelligent data science*. Springer.

[13] Bisong, E. et al. (2019). *Building machine learning and deep learning models on Google cloud platform*. Springer.

[14] Breiman, L. (2001). Random forests. *Machine learning*, 45:5–32.

[15] Brigato, L. and Iocchi, L. (2021). A close look at deep learning with small data. In *2020 25th international conference on pattern recognition (ICPR)*, pages 2490–2497. IEEE.

[16] Brownlee, J. (2016). *Machine learning mastery with Python: understand your data, create accurate models, and work projects end-to-end*. Machine Learning Mastery.

[17] Bruninx, K., Ovaere, M., and Delarue, E. (2020). The long-term impact of the market stability reserve on the EU emission trading system. *Energy Economics*, 89:104746.

[18] Cadavid, J. P. U., Lamouri, S., Grabot, B., and Fortin, A. (2019). Machine learning in production planning and control: A review of empirical literature. *IFAC-PapersOnLine*, 52(13):385–390. 9th IFAC Conference on Manufacturing Modelling, Management and Control MIM 2019.

[19] Cao, J., Li, Z., and Li, J. (2019). Financial time series forecasting model based on CEEMDAN and LSTM. *Physica A: Statistical Mechanics and its Applications*, 519:127–139.

[20] Cao, Y., Zha, D., Wang, Q., and Wen, L. (2023). Probabilistic carbon price prediction with quantile temporal convolutional network considering uncertain factors. *Journal of Environmental Management*, 342:118137.

[21] Cardoso, J.-F. (2003). Dependence, correlation and gaussianity in independent component analysis. *Journal of Machine Learning Research*, 4(Dec):1177–1203.

[22] Che, J., Wang, J., and Tang, Y. (2012). Optimal training subset in a support vector regression electric load forecasting model. *Applied Soft Computing*, 12(5):1523–1531.

[23] Chevallier, J. (2009). Carbon futures and macroeconomic risk factors: A view from the EU ETS. *Energy Economics*, 31(4):614–625.

[24] Chevallier, J. (2011). A model of carbon price interactions with macroeconomic and energy dynamics. *Energy Economics*, 33(6):1295–1312.

[25] Clean Energy Regulator (2023). Australian carbon credit units. Website. Accessed: 2023-05-29.

[26] Commonwealth of Australia (n.d.). Australian carbon credit unit (accu) scheme. Website. Accessed: 2023-05-29.

[27] Comon, P. (1994). Independent component analysis, a new concept? *Signal processing*, 36(3):287–314.

[28] Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative finance*, 1(2):223.

[29] Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20:273–297.

[30] Craglia, M., Hradec, J., Nativi, S., and Santoro, M. (2017). Exploring the depths of the global earth observation system of systems. *Big Earth Data*, 1(1-2):21–46.

[31] Creti, A., Jouvet, P.-A., and Mignon, V. (2012). Carbon price drivers: Phase i versus phase ii equilibrium? *Energy Economics*, 34(1):327–334.

[32] Cristea, A. (2013). Trade and the greenhouse gas emissions from international freight transport. *Journal of Environmental Economics and Management*, 65(1):153–173.

[33] Cun, Y., Bottou, L., Orr, G., and Muller, K. (1998). Efficient backprop, neural networks: tricks of the trade. *Lecture notes in computer sciences*, 1524:5–50.

[34] Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a):427–431.

[35] Dietzenbacher, E., Cazcarro, I., and Arto, I. (2020). Towards a more effective climate policy on international trade. *Nature communications*, 11(1):1130.

[36] Dong, H., Hu, Y., Yang, Y., and Jiang, W. (2023). A multi-strategy integration prediction model for carbon price. *Energies*, 16(12):4613.

[37] Dutta, A. (2018). Modeling and forecasting the volatility of carbon emission market: The role of outliers, time-varying jumps and oil price risk. *Journal of Cleaner Production*, 172:2773–2781.

[38] El-Masry, A. A. (2016). Capital structure and performance of middle east and north africa (mena) banks: an assessment of credit rating. *Banks & bank systems*, 11(11, Iss. 1):77–91.

[39] Ellabban, O., Abu-Rub, H., and Blaabjerg, F. (2014). Renewable energy resources: Current status, future prospects and their enabling technology. *Renewable and sustainable energy reviews*, 39:748–764.

[40] Evans, J. D. (1996). *Straightforward statistics for the behavioral sciences.* Thomson Brooks/Cole Publishing Co.

[41] Fan, J., Wu, L., Ma, X., Zhou, H., and Zhang, F. (2020). Hybrid support vector machines with heuristic algorithms for prediction of daily diffuse solar radiation in air-polluted regions. *Renewable Energy*, 145:2034–2045.

[42] Faris, H., Aljarah, I., Al-Madi, N., and Mirjalili, S. (2016). Optimizing the learning process of feedforward neural networks using lightning search algorithm. *International Journal of Artificial Intelligence Tools*, 25.

[43] Feng, M., Duan, Y., Wang, X., Zhang, J., and Ma, L. (2023). Carbon price prediction based on decomposition technique and extreme gradient boosting optimized by the grey wolf optimizer algorithm. *Scientific Reports*, 13(1):18447.

[44] Feng, Z.-H., Liu, C.-F., and Wei, Y.-M. (2010). How does carbon price change? evidences from EU ETS. CEEP-BIT Working Papers 11, Center for Energy and Environmental Policy Research (CEEP), Beijing Institute of Technology.

[45] Gavard, C. and Kirat, D. (2018). Flexibility in the market for international

carbon credits and price dynamics difference with European allowances. *Energy Economics*, 76:504–518.

[46] Ghannam, S. and Hussain, F. (2023). Comparison of deep learning approaches for forecasting urban short-term water demand a greater Sydney region case study. *Knowledge-Based Systems*, 275:110660.

[47] Ghosh, P., Jha, A., and Sharma, R. (2020). Managing carbon footprint for a sustainable supply chain: a systematic literature review. *Modern Supply Chain Research and Applications*.

[48] Goh, W. W. B., Foo, R. J. K., and Wong, L. (2024). What can scatterplots teach us about doing data science better? *International Journal of Data Science and Analytics*, 17(1):111–125.

[49] Guan, H. (2021). The determinants of carbon emission price—empirical analysis based on the vec model. In *2021 2nd International Conference on Big Data Economy and Information Management (BDEIM)*, pages 68–71. IEEE.

[50] Guðbrandsdóttir, H. N. and Haraldsson, H. Ó. (2011). Predicting the price of EU ETS carbon credits. *Systems Engineering Procedia*, 1:481–489.

[51] Guyon, I. and Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar):1157–1182.

[52] Hair, J. F., Black, W. C., Babin, B. J., and Anderson, R. E. (2009). *Multivariate data analysis*. Prentice Hall, Upper Saddle River, NJ, 7th edition.

[53] Hammoudeh, S., Nguyen, D. K., and Sousa, R. M. (2014). Energy prices and co2 emission allowance prices: A quantile regression approach. *Energy policy*, 70:201–206.

[54] Hao, Y. and Tian, C. (2020). A hybrid framework for carbon trading price forecasting: The role of multiple influence factor. *Journal of Cleaner Production*, 262:120378.

[55] Heidari, A. A., Faris, H., Mirjalili, S., Aljarah, I., and Mafarja, M. (2020). Ant lion optimizer: Theory, literature review, and application in multi-layer perceptron neural networks. In Mirjalili, S., Dong, J., and Lewis, A., editors, *Nature-Inspired Optimizers Theories, Literature Reviews, and Applications*, pages 23–46. Springer.

[56] Hevner, A. R., March, S. T., Park, J., and Ram, S. (2004). Design science in information systems research. *MIS Q.*, 28(1):75–105.

[57] Hintermann, B. (2010). Allowance price drivers in the first phase of the EU ETS. *Journal of Environmental Economics and Management*, 59(1):43–56.

[58] Horizon Grand View Research (2025). Australia voluntary carbon credit market outlook. `https://www.grandviewresearch.com/horizon/outlook/voluntary-carbon-credit-market/australia`. Accessed: January 23, 2025.

[59] Hu, X., Rousseau, R., and Chen, J. (2011). On the definition of forward and backward citation generations. *Journal of Informetrics*, 5(1):27–36.

[60] Hyvärinen, A. and Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural Networks*, 13(4):411–430.

[61] Jacobson, M. Z. and Delucchi, M. A. (2011). Providing all global energy with wind, water, and solar power, part i: Technologies, energy resources, quantities and areas of infrastructure, and materials. *Energy policy*, 39(3):1154–1169.

[62] Jambu, M. (1991). *Exploratory and multivariate data analysis*. Elsevier.

[63] Jarden (n.d.). JARDEN website. Website. Accessed: 2023-03-31.

[64] Ji, C.-J., Hu, Y.-J., Tang, B.-J., and Qu, S. (2021). Price drivers in the carbon emissions trading scheme: Evidence from Chinese emissions trading scheme pilots. *Journal of Cleaner Production*, 278:123469.

[65] Jiang, H. (2018). Sparse estimation based on square root nonconvex optimization in high-dimensional data. *Neurocomputing*, 282:122–135.

[66] Jiang, H. and Dong, Y. (2019). Structural regularization in quadratic logistic regression model. *Knowledge-Based Systems*, 163:842–857.

[67] Jiang, Q., Khattak, S. I., and Rahman, Z. U. (2021). Measuring the simultaneous effects of electricity consumption and production on carbon dioxide emissions (co2e) in China: New evidence from an ekc-based assessment. *Energy*, 229:120616.

[68] Jiang, Y., Lei, Y.-L., Yang, Y.-Z., and Wang, F. (2018). Factors affecting the pilot trading market of carbon emissions in China. *Petroleum Science*, 15:412–420.

[69] Jolliffe, I. T. and Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374.

[70] Judi, D., Rakowski, C., Waichler, S., Feng, Y., and Wigmosta, M. (2018). Integrated modeling approach for the development of climate-informed, actionable information. *Water*, 10:775.

[71] Kazem, Hussein A., Quteishat, Anas, and Younis, Mahmoud A. (2021). Techno-economical study of solar water pumping system: optimum design, evaluation, and comparison. *Renew. Energy Environ. Sustain.*, 6:41.

[72] Keppler, J. H. and Mansanet-Bataller, M. (2010). Causalities between co2, electricity, and other energy variables during phase i and phase ii of the EU ETS. *Energy Policy*, 38(7):3329–3341. Large-scale wind power in electricity markets with Regular Papers.

[73] Khan, M. Y., Qayoom, A., Nizami, M. S., Siddiqui, M. S., Wasi, S., and Raazi, S. M. K.-u.-R. (2021). Automated prediction of good dictionary examples (gdex): a comprehensive experiment with distant supervision, machine learning, and word embedding-based deep learning techniques. *Complexity*, 2021:1–18.

[74] Kim, H., Kim, Y., Ko, Y., and Han, S. (2022). Performance comparison of predictive methodologies for carbon emission credit price in the Korea emission trading system. *Sustainability*, 14(13):8177.

[75] Kim, H. S. and Koo, W. W. (2010). Factors affecting the carbon allowance market in the us. *Energy Policy*, 38(4):1879–1884.

[76] Kitchenham, B. and Brereton, P. (2013). A systematic review of systematic review process research in software engineering. *Information and software technology*, 55(12):2049–2075.

[77] Kitchenham, B. A. and Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. Technical Report EBSE 2007-001, Keele University and Durham University Joint Report.

[78] Knapič, S., Malhi, A., Saluja, R., and Främling, K. (2021). Explainable artificial intelligence for human decision support system in the medical domain. *Machine Learning and Knowledge Extraction*, 3(3):740–770.

[79] Koch, N., Fuss, S., Grosjean, G., and Edenhofer, O. (2014). Causes of the EU ETS price drop: Recession, cdm, renewable policies or a bit of everything?—new evidence. *Energy Policy*, 73:676–685.

[80] Kothari, S. and Oh, H. (1993). Title of the chapter or section. In Yovits, M. C., editor, *Neural Networks for Pattern Recognition*, volume 37 of *Advances in Computers*, pages 119–166. Elsevier.

[81] Kotishwar, A. (2024). Impact of global recession on carbon credits with special reference to India. *International Journal of Business and Globalisation*.

[82] Kuhn, M. and Johnson, K. (2019). *Feature engineering and selection: A practical approach for predictive models*. Chapman and Hall/CRC.

[83] Kuizinienė, D., Savickas, P., Kunickaitė, R., Juozaitienė, R., Damaševičius, R., Maskeliūnas, R., and Krilavičius, T. (2024). A comparative study of feature selection and feature extraction methods for financial distress identification. *PeerJ Computer Science*, 10:e1956.

[84] Kumar, R., Kumar, P., and Kumar, Y. (2020). Time series data prediction using iot and machine learning technique. *Procedia computer science*, 167:373–381.

[85] Lazarou, S., Christodoulou, C., and Vita, V. (2019). Global change assessment model (GCAM) considerations of the primary sources energy mix for an energetic scenario that could meet paris agreement. In *2019 54th International Universities Power Engineering Conference (UPEC)*, pages 1–5.

[86] Lepot, M., Aubin, J.-B., and Clemens, F. H. (2017). Interpolation in time series: An introductive overview of existing methods, their performance criteria and uncertainty assessment. *Water*, 9(10).

[87] Li, B., Geng, Y., Xia, X., and Qiao, D. (2021). The impact of government subsidies on the low-carbon supply chain based on carbon emission reduction level. *International Journal of Environmental Research and Public Health*, 18(14):7603.

[88] Li, H. and Lei, M. (2018). The influencing factors of China carbon price: a study based on carbon trading market in hubei province. *IOP Conference Series: Earth and Environmental Science*, 121(5):052073.

[89] Li, H., Li, Q., Huang, X., and Guo, L. (2023a). Do green bonds and economic policy uncertainty matter for carbon price? new insights from a tvp-var framework. *International Review of Financial Analysis*, 86:102502.

[90] Li, J., Song, L., Wu, D., Shui, J., and Wang, T. (2023b). Lagging problem in financial time series forecasting. *Neural Comput. Appl.*, 35(28):20819–20839.

[91] Li, M.-F., Hu, H., and Zhao, L.-T. (2022a). Key factors affecting carbon prices from a time-varying perspective. *Environmental Science and Pollution Research*, 29(43):65144–65160.

[92] Li, Q., Wang, H., Li, Z., and Yuan, S. (2022b). A comparative study of the effect of different carbon-reduction policies on outsourcing remanufacturing. *International Journal of Environmental Research and Public Health*, 19(6):3590.

[93] Lin, B. and Jia, Z. (2019). What are the main factors affecting carbon price in emission trading scheme? a case study in China. *Science of The Total Environment*, 654:525–534.

[94] Lin, B. and Xu, B. (2021). A non-parametric analysis of the driving factors of China's carbon prices. *Energy Economics*, 104:105684.

[95] Lipton, Z. C. (2018). The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*, 16(3):31–57.

[96] Littell, J. H., Corcoran, J., and Pillai, V. (2008). *Systematic Reviews and Meta-Analysis*. Oxford University Press.

[97] Liu, B., Wu, Q., and Cao, Q. (2020). An improved elman network for stock price prediction service. *Security and Communication Networks*, 2020:1–9.

[98] Liu, N. and Song, F. (2022). Carbon price prediction under output uncertainty. *Environmental Science and Pollution Research*, 29(15):21577–21590.

[99] Lones, M. A. (2024). Avoiding common machine learning pitfalls. *Patterns*, 5(10).

[100] Lovcha, Y., Perez-Laborda, A., and Sikora, I. (2022). The determinants of co2 prices in the EU emission trading system. *Applied Energy*, 305:117903.

[101] Lu, C.-J., Chiu, C.-C., and Yang, J.-L. (2009). Integrating nonlinear independent component analysis and neural network in stock price prediction. In *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*.

[102] Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S.-I. (2020). From local explanations to global understanding with explainable ai for trees. *Nature Machine Intelligence*, 2(1):56–67.

[103] Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.

[104] Luo, R., Li, Y., Wang, Z., and Sun, M. (2022). Co-movement between carbon prices and energy prices in time and frequency domains: A wavelet-based analysis for beijing carbon emission trading system. *International Journal of Environmental Research and Public Health*, 19(9).

[105] Lutz, B. J., Pigorsch, U., and Rotfuß, W. (2013). Nonlinearity in cap-and-trade systems: The EUA price and its fundamentals. *Energy Economics*, 40:222–232.

[106] Montgomery, D. C., Jennings, C. L., and Kulahci, M. (2015). *Introduction to time series analysis and forecasting*. John Wiley & Sons.

[107] Mori, H. and Jiang, W. (2008). A risk analysis method for carbon price prediction with hybrid intelligent model in consideration of variable selection of

graphical modeling. In *2008 IEEE International Conference on Sustainable Energy Technologies*, pages 1019–1024. IEEE.

[108] Mu, G., Dai, L., Ju, X., Chen, Y., and Huang, X. (2024). MS-IHHO-LSTM: Carbon price prediction model of multi-source data based on improved swarm intelligence algorithm and deep learning method. *IEEE Access*, 12:80754–80769.

[109] Mukonza, C. and Nhamo, G. (2016). Trade implications of grid emission factors under climate change and the green economy: comparative study of African and Asian continents. *Journal of Economic and Financial Sciences*, 9(1):13–27.

[110] Müller, A. C. and Guido, S. (2016). *Introduction to machine learning with Python: a guide for data scientists*. " O'Reilly Media, Inc.".

[111] Nair, V. and Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814.

[112] Ng, A. (2016). What artificial intelligence can and can't do right now. *Harvard Business Review*, 9(11):1–4.

[113] Niu, L., Lu, J., and Zhang, G. (2009). Cognition-driven decision support for business intelligence. *Models, Techniques, Systems and Applications. Studies in Computational Intelligence, Springer, Berlin*, pages 4–5.

[114] Ozturk, F., Keles, M., and Evrendilek, F. (2016). Quantifying rates and drivers of change in long-term sector- and country-specific trends of carbon dioxide-equivalent greenhouse gas emissions. *Renewable and Sustainable Energy Reviews*, 65:823–831.

[115] O'brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & quantity*, 41:673–690.

[116] Pan, W. (2021). Feature selection algorithm based on maximum information coefficient. In *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, pages 2600–2603.

[117] Paris Agreement (2015). Paris Agreement. In *Report of the Conference of the Parties to the United Nations Framework Convention on Climate Change (21st Session, 2015: Paris). Retrieved December*, volume 4, page 2017. HeinOnline.

[118] Patel, D., Britto, B., Sharma, S., Gaikwad, K., Dusing, Y., and Gupta, M. (2020). Carbon Credits on Blockchain. In *2020 International Conference on Innovative Trends in Information Technology (ICITIIT)*, pages 1–5.

[119] Pu, H., Wu, Y., Hua, C., Song, Z., Zhang, W., Zhang, L., Chu, Q., and Qiu, Y. (2022). Carbon trading based on quadratic modal decomposition and recurrent neural network price prediction model. In *2022 5th International Conference on Pattern Recognition and Artificial Intelligence (PRAI)*, pages 26–35. IEEE.

[120] Raheem, I., Mubarak, N. M., Karri, R. R., Manoj, T., Ibrahim, S. M., Mazari, S. A., and Nizamuddin, S. (2022). Forecasting of energy consumption by g20 countries using an adjacent accumulation grey model. *Scientific Reports*, 12(1):13417.

[121] Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., Lander, E. S., Mitzenmacher, M., and Sabeti, P. C. (2011). Detecting novel associations in large data sets. *Science*, 334(6062):1518–1524.

[122] Sadawi, A. A., Madani, B., Saboor, S., Ndiaye, M., and Abu-Lebdeh, G. (2021). A comprehensive hierarchical blockchain system for carbon emission trading utilizing blockchain of things and smart contract. *Technological Forecasting and Social Change*, 173:121124.

[123] Salmerón-Gómez, R., García-García, C. B., and García-Pérez, J. (2024). A

redefined variance inflation factor: overcoming the limitations of the variance inflation factor. *Computational Economics*, pages 1–27.

[124] Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J., and Müller, K.-R. (2021). Explaining deep neural networks and beyond: A review of methods and applications. *Proceedings of the IEEE*, 109(3):247–278.

[125] Schell, A. and Oberhauser, H. (2023). Nonlinear independent component analysis for discrete-time and continuous-time signals. *The Annals of Statistics*, 51(2):487–518.

[126] Shapley, L. S. et al. (1953). A value for n-person games. *Contributions to the Theory of Games II*, pages 307–317.

[127] Sims, R. E., Rogner, H.-H., and Gregory, K. (2003). Carbon emission and mitigation cost comparisons between fossil fuel, nuclear and renewable energy resources for electricity generation. *Energy policy*, 31(13):1315–1326.

[128] Singh, G. (2009). *Understanding Carbon Credits*. Aditya Books Pvt. Ltd., New Delhi, India.

[129] Smola, A. J. and Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, 14:199–222.

[130] Song, X., Zhang, W., Ge, Z., Huang, S., Huang, Y., and Xiong, S. (2022). A study of the influencing factors on the carbon emission trading price in China based on the improved gray relational analysis model. *Sustainability*, 14(13).

[131] Song, Y., Liu, T., Liang, D., Li, Y., and Song, X. (2019). A fuzzy stochastic model for carbon price prediction under the effect of demand-related policy in China's carbon market. *Ecological Economics*, 157:253–265.

[132] Sun, G., Song, Z., Liu, J., Zhu, S., and He, Y. (2017). Feature selection method based on maximum information coefficient and approximate Markov blanket. *Zidonghua Xuebao/acta Automatica Sinica*, 43(5):795–805.

[133] Sun, W. and Duan, M. (2019). Analysis and forecasting of the carbon price in China's regional carbon markets based on fast ensemble empirical mode decomposition, phase space reconstruction, and an improved extreme learning machine. *Energies*, 12(2):277.

[134] Sun, W. and Wang, Y. (2020). Factor analysis and carbon price prediction based on empirical mode decomposition and least squares support vector machine optimized by improved particle swarm optimization. *Carbon Management*, 11(3):315–329.

[135] Ta, V.-D., Liu, C.-M., and Tadesse, D. A. (2020). Portfolio optimization-based stock prediction using long-short term memory network in quantitative trading. *Applied Sciences*, 10(2):437.

[136] Tan, X.-P. and Wang, X.-Y. (2017). Dependence changes between the carbon price and its fundamentals: A quantile regression approach. *Applied Energy*, 190:306–325.

[137] Tang, B.-j., Gong, P.-q., and Shen, C. (2017). Factors of carbon price volatility in a comparative analysis of the eua and scer. *Annals of Operations Research*, 255(1-2):157–168.

[138] Tercan, A. and Sohrabian, B. (2013). Multivariate geostatistical simulation of coal quality data by independent components. *International Journal of Coal Geology*, 112:53–66.

[139] Tsay, R. S. (2005). *Analysis of Financial Time Series*. John Wiley & Sons.

[140] Tuerk, A. and , Zelljadt, E. (2000). Climate policy info hub. `https://climatepolicyinfohub.eu/global-rise-emissions-trading`. Accessed: 2022-05-23.

[141] United Nations (1998). Kyoto protocol to the United Nations framework convention on climate change. Accessed: 2022-05-23.

[142] Vapnik, V. (2013). *The nature of statistical learning theory*. Springer science & business media.

[143] Wang, J., Sun, X., Cheng, Q., and Cui, Q. (2021). An innovative random forest-based nonlinear ensemble paradigm of improved feature extraction and deep learning for carbon price forecasting. *Science of The Total Environment*, 762:143099.

[144] Wang, P., Liu, J., Tao, Z., and Chen, H. (2022). A novel carbon price combination forecasting approach based on multi-source information fusion and hybrid multi-scale decomposition. *Engineering Applications of Artificial Intelligence*, 114:105172.

[145] Wang, P., Zhou, X., and Zeng, Z. (2024). LightGBM-BES-BiLSTM carbon price prediction based on environmental impact factors. *Computational Economics*, pages 1–27.

[146] Wang, R., Zhao, X., Wu, K., and et al. (2023a). Examination of the transmission mechanism of energy prices influencing carbon prices: An analysis of mediating effects based on demand heterogeneity. *Environmental Science and Pollution Research*, 30:59567–59578.

[147] Wang, Y., Qin, L., Wang, Q., Chen, Y., Yang, Q., Xing, L., and Ba, S. (2023b). A novel deep learning carbon price short-term prediction model with dual-stage attention mechanism. *Applied Energy*, 347:121380.

[148] Wang, Z.-J. and Zhao, L.-T. (2021). The impact of the global stock and energy market on EU ETS: A structural equation modelling approach. *Journal of Cleaner Production*, 289:125140.

[149] Wei, X. and Ouyang, H. (2024). Carbon price prediction based on a scaled pca approach. *Plos one*, 19(1):e0296105.

[150] Wei, Qi, Bian, Yuanyuan, and Yang, Xuejuan (2020). Influencing factors of price fluctuation in China´s carbon market. *E3S Web Conf.*, 218:01044.

[151] Williams, J. R. (2009). What is the carbon market: Is there a final answer? *Journal of Soil and Water Conservation*, 64(1).

[152] Wu, Z., Zhang, W., and Zeng, X. (2023). Exploring the short-term and long-term linkages between carbon price and influence factors considering COVID-19 impact. *Environmental Science and Pollution Research*, 30(22):61479–61495.

[153] Xiao, C. and Liu, Y. (2024). A multifrequency data fusion deep learning model for carbon price prediction. *Journal of Forecasting*.

[154] Xie, Q., Hao, J., Li, J., and Zheng, X. (2022). Carbon price prediction considering climate change: A text-based framework. *Economic Analysis and Policy*, 74:382–401.

[155] Xu, W., Wang, J., Zhang, Y., Li, J., and Wei, L. (2022). An optimized decomposition integration framework for carbon price prediction based on multi-factor two-stage feature dimension reduction. *Annals of Operations Research*, pages 1–38.

[156] Yang, L., Li, F., and Zhang, X. (2016). Chinese companies' awareness and perceptions of the emissions trading scheme (ets): Evidence from a national survey in China. *Energy Policy*, 98:254–265.

[157] Yang, W., Wei, Y., Wei, H., Chen, Y., Huang, G., Li, X., Li, R., Yao, N., Wang, X., Gu, X., et al. (2023). Survey on explainable ai: From approaches, limitations and applications aspects. *Human-Centric Intelligent Systems*, 3(3):161–188.

[158] Yu, J. and Mallory, M. L. (2014). Exchange rate effect on carbon credit price via energy markets. *Journal of International Money and Finance*, 47:145–161.

[159] Yu, L., Li, J., Tang, L., and Wang, S. (2015). Linear and nonlinear granger causality investigation between carbon market and crude oil market: A multi-scale approach. *Energy Economics*, 51:300–311.

[160] Zeng, Q., Liu, W., Cao, R., Shen, X., and Guo, K. (2017). Price determinants in China's pilot carbon markets. In *2017 4th International Conference on Industrial Economics System and Industrial Security Engineering (IEIS)*, pages 1–6. IEEE.

[161] Zeng, S., Fu, Q., Yang, D., Tian, Y., and Yu, Y. (2023). The influencing factors of the carbon trading price: A case of China against a double carbon background. *Sustainability*, 15(3).

[162] Zhang, C., Wu, Y., and Yang, Y. (2018a). The Influencing Factors of sCER Price Dynamics Under the Clean Development Mechanism: Theory and Econometric Analysis. *Journal of Systems Science and Complexity*, 31(5):1244–1272.

[163] Zhang, C., Wu, Y., and Yang, Y. (2018b). The influencing factors of sCER price dynamics under the clean development mechanism: Theory and econometric analysis. *Journal of Systems Science and Complexity*, 31:1244–1272.

[164] Zhang, G. and Qi, M. (2005). Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research*, 160(2):501–514. Decision Support Systems in the Internet Age.

[165] Zhang, S., Ji, H., Tian, M., and Wang, B. (2022). High-dimensional nonlinear dependence and risk spillovers analysis between China's carbon market and its major influence factors. *Annals of Operations Research*.

[166] Zhao, L., Gong, J., Huang, D., and Hu, C. (2021a). Fault feature selection method of gearbox based on Fisher Score and maximum information coefficient. *Control and Decision*, 36(9):2234–2240.

[167] Zhao, L.-T., Miao, J., Qu, S., and Chen, X.-H. (2021b). A multi-factor integrated model for carbon price forecasting: Market interaction promoting carbon emission reduction. *Science of The Total Environment*, 796:149110.

[168] Zheng, A. and Casari, A. (2018). *Feature engineering for machine learning: principles and techniques for data scientists*. " O'Reilly Media, Inc.".

[169] Zheng, G., Li, K., Yue, X., and Zhang, Y. (2024). A multifactor hybrid model for carbon price interval prediction based on decomposition-integration framework. *Journal of Environmental Management*, 363:121273.

[170] Zheng, L. (2020). Short-term parking demand prediction method based on variable prediction interval. *IEEE Access*, 8:58594–58602.

[171] Zhou, J. and Wang, S. (2021). A carbon price prediction model based on the secondary decomposition algorithm and influencing factors. *Energies*, 14(5):1328.

[172] Zhou, J., Xu, Z., and Wang, S. (2022). A novel hybrid learning paradigm with feature extraction for carbon price prediction based on bi-directional long short-term memory network optimized by an improved sparrow search algorithm. *Environmental Science and Pollution Research*, 29(43):65585–65598.

[173] Zhou, K. and Li, Y. (2019). Influencing factors and fluctuation characteristics of China's carbon emission trading price. *Physica A: Statistical Mechanics and its Applications*, 524:459–474.

[174] Zhu, B., Wang, P., Chevallier, J., and Wei, Y. (2015). Carbon price analysis using empirical mode decomposition. *Computational Economics*, 45:195–206.

[175] Zhu, B., Ye, S., Han, D., Wang, P., He, K., Wei, Y.-M., and Xie, R. (2019). A multiscale analysis for carbon price drivers. *Energy Economics*, 78:202–216.

[176] Zhu, Y., Chen, Y., Hua, Q., Wang, J., Guo, Y., Li, Z., Ma, J., and Wei, Q. (2024). A hybrid model for carbon price forecasting based on improved feature extraction and non-linear integration. *Mathematics*, 12(10).

[177] Zotero (2020). Zotero [computer software]. Originally published 2006. Corporation for Digital Scholarship.