

Faculty of Engineering and Information Technology
University of Technology Sydney

High-dimensional Dependence Modelling for Cross-market Analysis

A thesis submitted in fulfillment of
the requirements for the degree of
Doctor of Philosophy

by

Jia Xu

February 2022

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, *Jia Xu* declare that this thesis, is submitted in fulfilment of the requirements for the award of *Doctor of Philosophy*, in the *Data Science Institute, 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 qualifications at any other academic institution. This research is supported by the Australian Government Research Training Program.

Signature: Production Note:
 Signature removed prior to publication.

Date: 17/02/2022

*To Charlotte,
my dearest daughter*

Acknowledgments

I would like to express my gratitude and appreciation to my supervisor Professor Longbing Cao. His knowledge, patience and encouragement brought great inspiration to me in the entire course of my PhD study, especially in the final phase of writing this thesis. He has set an example that will have a long-term influence on my attitude towards life and work.

I would also like to thank my co-supervisor Associate Professor Richard Xu for providing me with continuous support throughout my PhD study and research. Without his professional guidance and persistent help, this thesis would not have been possible.

My sincere thanks to my colleague Wei Wei for his valuable guidance and discussion. Also, I would like to thank my other colleagues in Advanced Analytics Institute for their enthusiastic help and support: Dr. Jinjiu Li, Dr. Zhigang Zheng, Dr. Yuming Ou, Dr. Allen Lin, Dr. Mu Li, Dr. Wei Cao, Dr. Liang Hu, Dr. Xuhui Fan, Dr. Junfu Yin, Dr. Guansong Pang, Dr. Shoujin Wang, and Dr. Chengzhang Zhu. I appreciate our intensive discussions, in-depth collaboration, and all the fun we have had in the last eight years.

Lastly, I would like to thank my parents and my uncle Dr. Qingping Cheng for their unconditional support and lovely encouragement over the past decades. Without their support, it would not be possible for me to conduct this research. In particular, I am very grateful for the tremendous support my wife Xue Xia have given me during my thesis. She was a great comfort to me whenever the writing caused stress. Retrospectively, I feel

ACKNOWLEDGMENTS

so blessed to have met the aforementioned people whose friendship, support and help will never fade in my memory.

Jia Xu

University of Technology, Sydney

November 2021

Contents

Certificate	i
Dedication	iii
Acknowledgment	v
List of Figures	xiii
List of Tables	xv
List of Publications	xvii
Abstract	xix
Chapter 1 Introduction	1
1.1 Background	1
1.2 Research Motivations and Challenges	2
1.3 Research Gaps in Cross-market Modeling	7
1.3.1 Essential Econometric Approaches	7
1.3.2 Traditional Machine Learning Models	8
1.3.3 Recurrent Neural Network Models	9
1.4 Research Contributions	10
1.5 Thesis Structure	12
Chapter 2 Preliminaries and Foundations	16
2.1 Copula	17
2.1.1 Copula Foundations	17
2.1.2 Parametric Copula Families	20
2.1.3 Dependence Measures	23
2.1.4 Bivariate Copula	26
2.1.5 Vine Copula and Partial Correlation	31

CONTENTS

2.2	Time Series	37
2.2.1	Autoregressive Integrated Moving Average Model . . .	39
2.2.2	Autoregressive Conditional Heteroskedasticity Model .	41
2.2.3	Recurrent Neural Network	43
2.2.4	Variational Autoencoder	46
2.3	Evaluation Metrics	49
2.3.1	In-Sample Tests	49
2.3.2	Out-of-sample Tests	53
Chapter 3 Literature Review		58
3.1	Multivariate Copula Modelling	59
3.1.1	Multivariate Copula Construction	59
3.1.2	Multivariate Elliptical Copulas	59
3.1.3	Nested Archimedean Copulas	60
3.1.4	Vine Copula Construction	66
3.1.5	Copula Model Fitting	71
3.1.6	Bayesian Vine Copula Generation	74
3.2	Multivariate Correlation and Dependence Modelling	75
3.3	Machine Learning of Market Couplings	78
3.4	Deep Sequential Modelling in Finance	79
Chapter 4 High Dimensional Cross-market Dependence Modelling via Partial D Vine Copula		83
4.1	Introduction	84
4.2	Weighted Partial D Vine Copula	88
4.2.1	Partial D Vine Structure Construction	89
4.2.2	Dependence Structure Truncation	95
4.2.3	Bivariate Copula Selection	95
4.2.4	Marginal Distribution Specification	97
4.3	Parameter Estimation	98
4.4	Value at Risk for Evaluating Business Performance	99
4.5	Experiments of Cross Markets	101

4.5.1	Data and Marginal Distributions Specification	101
4.5.2	Weighted Partial Vine Copula Tuning and Truncation .	102
4.5.3	In-Sample Performance Analysis	107
4.5.4	Out-of-Sample Performance Analysis	109
4.6	Summary	110

Chapter 5 Partial Regular Vine Copula for Portfolio and Risk

Management	113	
5.1	Introduction	114
5.2	Risk Averse Utility Maximisation Framework	119
5.3	Weighted Partial Regular Vine Copula	122
5.3.1	Weighted Partial Regular Vine Specification	122
5.3.2	Partial Regular Vine Construction	124
5.3.3	Bivariate Copula Family Specification	128
5.3.4	Vine Copula Optimisation	131
5.4	Model Fitting and Inference	131
5.5	Case Study for Portfolio Management	134
5.5.1	Data Description and Marginal Distributions Specifi- cation	134
5.5.2	Parameters Tuning and In Sample Test	137
5.5.3	Performance Analysis for Portfolios	141
5.6	Summary	143

Chapter 6 Partial Vine Copula based Asymmetry and Tail

Dependence Modelling	147	
6.1	Introduction	148
6.2	Partial Regular Vine Copula with Structural Asymmetric De- pendence	152
6.2.1	New Tail Dependencies Measures	153
6.2.2	Weighted Partial Regular Vine Copula Specification . .	157
6.2.3	Bivariate Copula Selection and Optimisation	162
6.2.4	Parameter Estimation for the Partial Regular Vine . .	164

6.3	Case Study	167
6.3.1	Data and Marginal Distribution Specification	167
6.3.2	Regular Vine Copula Structure Specification and Tail Dependence Analysis	169
6.3.3	Out-of-Sample Performance Analysis	172
6.4	Summary	175
Chapter 7 Copula Variational LSTM for High-dimensional Cross-market Dependence Modelling		178
7.1	Introduction	179
7.2	The WPVC-VLSTM Model	185
7.2.1	LSTM-Based Variational Autoencoder	185
7.2.2	Weighted Partial Regular Vine Construction	188
7.2.3	Vine Structure Optimisation and Copula Family Selec- tion	194
7.3	Variational Inference and Parameter Estimation	195
7.3.1	Inference of Copula Variational LSTM	195
7.3.2	Computational Complexity Analysis	197
7.4	Experiments	198
7.4.1	Date Sets	198
7.4.2	Model Specification	199
7.4.3	Baseline Models	199
7.4.4	Evaluation Metrics	200
7.4.5	Business Performance Assessment	201
7.4.6	Performance Evaluation	203
7.5	Summary	206
Chapter 8 Conclusions and Future Work		209
8.1	Conclusion	209
8.2	Discussion and Future Work	211
8.2.1	Research Gaps Discussion	211
8.2.2	Future work	212

Appendix A List of Abbreviations 215

Appendix B List of Notations 218

Bibliography 221

List of Figures

1.1	Illustration of cross-market interactions: homogeneous and heterogeneous couplings within and between stock and foreign currency markets in UK and US	3
1.2	Illustration of distinct multi-market conditions: heterogeneous distributions and asymmetric correlations between financial markets	5
1.3	Mapping between research problems and the contributions . .	12
1.4	The profile of work in this thesis	15
2.1	C vine with five variables	36
2.2	D vine with five variables	37
2.3	Elman's RNN architecture	44
2.4	LSTM cell illustration	46
2.5	GRU cell illustration	47
2.6	Illustration of a typical VAE structure	48
3.1	FNA copulas construction for $n = 4$	62
3.2	PNA copulas construction for $n = 4$	63
3.3	HNA copulas construction for $n = 8$	65
3.4	Example - D vine construction for four variables	68
3.5	VRNN operations demonstration	81
3.6	Relationship between the multivariate cross-market approaches and copula models	82
4.1	Illustration of cross markets dependence	84
4.2	Scatter plot of the same correlation applied on different copulas	85

LIST OF FIGURES

4.3	Partial D vine trees construction	93
4.4	Weight parameter tuning on coverage test	106
4.5	Performance analysis of truncated WPDC	107
4.6	Comparison of estimation time and log-likelihood	109
4.7	The VaR forecast of portfolio returns by using $WPDC_{0.05}$	111
5.1	Portfolio frontier by mean variance strategy	116
5.2	Illustration for stock dependence structure	117
5.3	Example - illustration of tree node generation	126
5.4	Parameter tuning for hyper parameters	139
5.5	Portfolio wealth against initiation for $\gamma = 3$	145
5.6	Portfolio wealth against initiation for $\gamma = 5$	145
5.7	Portfolio wealth against initiation for $\gamma = 7$	146
5.8	Portfolio wealth against initiation for $\gamma = 10$	146
6.1	Asymmetric dependencies across financial markets	148
6.2	Illustration of structural asymmetric dependence	150
6.3	Illustration of rotated Archimedean copulas with ($\tau = 0.05$) and $\tau = -0.05$	156
6.4	Example - Regular vine construction	161
6.5	Tree 1 - 3 for the specified regular vine	172
6.6	Tail dependence coefficient movement with different period	176
6.7	Tree 4 - 11 for the specified regular vine	177
7.1	Portfolio daily return predicted by LSTM	181
7.2	The vine copula-integrated variational LSTM: modelling both cross-market dependence degrees and structures	184
7.3	Illustrations of the learning process of VLSTM	187
7.4	Regular vine modelling the dependencies for latent variable	190
7.5	Truncation level tuning in WPVC-VLSTM	200
7.6	Performance evaluation of the Portfolio return forecasting	204
7.7	The VaR forecasting of portfolio returns by WPVC-VLSTM	206

List of Tables

2.1	Properties of bivariate elliptical copula families	28
2.2	Properties of bivariate Archimedean copula families	31
2.3	Clustering of exceedance	56
3.1	Example - degrees of freedom for bivariate student t copulas	67
4.1	Test of lag selection for ARMA models	103
4.2	Ljung-Box test for the fitted ARMA model	104
4.3	Log-likelihood performance of WPDC with Parameter m_0	104
4.4	The total number of edges against the truncation threshold	105
4.5	Vuong closeness test results	108
4.6	Backtesting results of VaR for portfolio	112
5.1	Statistics for experimental data set	134
5.2	Ljung-Box test for autocorrelation evaluation	137
5.3	AIC for WPRC weight parameter tuned from 0 to 1	138
5.4	Truncation parameter tuning	139
5.5	Bivariate copula families selected	140
5.6	Vuong test between the baselines and the $WPRC_{0.1}$	141
5.7	Performance of portfolio with different RRAs	144
6.1	The Tail Dependence of Copula Family	158
6.2	Ljung-Box test for autocorrelation evaluation	168
6.3	Log-likelihood of WPRC with different m_0	169
6.4	Truncation threshold parameter tuning	170
6.5	Tail dependence analysis on different measures	171

LIST OF TABLES

6.6	Backtesting results of VaR Forecasting	173
7.1	The performance of ARR and confusion matrix	205
7.2	Backtesting results of VaR in trading the portfolio	208

List of Publications

Papers Published

- **Jia Xu**, Wei Wei, Longbing Cao (2017), Copula-based high dimensional cross-market dependence modeling. *in* 'Proceedings of the 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA2017)', pp. 734-743, IEEE.
- **Jia Xu**, Longbing Cao (2018), Vine copula-based asymmetry and tail dependence modeling. *in* 'Proceedings of the 22th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD2018)', pp. 285-297.
- Qianqian Chen, Liang Hu, **Jia Xu**, Wei Liu, Longbing Cao (2015), Document similarity analysis via involving both explicit and implicit semantic couplings. *in* '2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA2015)', pp. 1-10, IEEE.

Papers Submitted/Under Review

- **Jia Xu**, Longbing Cao (2021), High-dimensional Cross-market Dependence Modeling by Copula Variational LSTM, submitted to the IEEE transactions on neural networks and learning systems (TNNLS)

Abstract

In the age of the information explosion, data mining and machine learning techniques have become heavily involved in financial market modelling. Many scholars have demonstrated the significance of dependence modelling across the multiple financial markets, especially during the catastrophic global financial crisis (GFC) in 2008. Sharp fluctuations across different markets demonstrate that the dependence is high-dimensional, contains various hierarchical and horizontal relationships, and often presents complicated dependence structures and characteristics such as an asymmetrical structure and tail dependence. Thus, a strong understanding of cross-market dependence is critical in cross-market applications such as portfolio management and risk management.

Unfortunately, modelling the dependence across multiple financial markets is highly challenging for the following reasons: (i) the cross-market dependence structure is often embedded with strong and complicated coupling relationships of high dimensionality as with any complex behavioural and social system; (ii) financial variables such as daily return have been demonstrated to follow non-Gaussian distributions, which means that dependence models should cover a wide range of dependencies to capture the asymmetric dependencies; and (iii) various tail dependencies such as lower and upper tail dependence are ubiquitous in financial markets.

Typical approaches such as Markov models, probabilistic graphical models, and neural network models could have advantages by building a conditional dependence structure between random variables to resolve the high-

dimensional problem. However, these models always impose unrealistic assumptions (e.g., Gaussian or mixtures of it), which leads to failure in capturing the complex dependence structures in the real world. In addition, copulas have been demonstrated to be effective in presenting dependence between variables in the statistics and finance communities. By splitting the joint distribution into dependence between variables and independent marginal distributions, copulas provide a flexible mechanism for investigating the specifications of the dependence across the market and the marginal distributions independently. Nevertheless, building effective dependence structures to address the aforementioned complexities is still a significant challenge for existing copula approaches. Over the last decade, various mixed models (e.g., tree-structured copula models and copula Bayesian networks) have been developed by utilising the advantages of both traditional copula models and probabilistic graphical models; however, assumptions and restrictions on the dependence structure have still not been avoided.

Based on the aforementioned research limitations and challenges, this thesis proposes weighted partial D vine copula, weighted partial regular vine copula and weighted regular vine variational long short-term memory (LSTM) models for high-dimensional cross-market modelling.

In Chapter 4, a novel bottom-to-top approach with no prior dependence assumptions called a weighted partial D vine copula is presented to capture the nonlinear and asymmetric dependence structures of cross-market data. By releasing these restrictions regarding the Gaussian assumptions, the new model is able to capture more sophisticated dependence structures between variables. The new modelling outcomes are applied to stock and exchange markets data as a case study, the extensive experimental results demonstrate that this model and its intrinsic design significantly outperform typical models and industry baselines.

Chapter 5 extends the approach presented in Chapter 4 to a more general structure, namely weighted partial regular vine. The new model can capture the nonlinear and asymmetric dependence structures with more flexible way

by utilising the advantages of the regular vine structure and non-restricted bivariate copula families. Then, the application of model for asset allocation by optimising the utility of a portfolio is presented as a case study. Compared with the general approaches, such as minimum variance, the optimised utility function using the new weighted partial regular vine can avoid the Gaussian assumption, which is always implied in these models due to computational issues.

Chapter 6 discusses the benefits of taking flexible bivariate copulas with different tail dependencies using the new regular vine copula model. Experiments are conducted to implement the new model on the exchange market to analyse the dynamic movement of the tail dependencies, consequently demonstrating better performance.

Chapter 7 introduces a novel vine copula-based variational autoencoder (VAE) to generate randomness for LSTM to model the cross-market data. Current VAE models usually apply mean-field assumption to simplify the calculation process; however, such an assumption may lead to posterior collapse by removing the dependencies between variables. Our new model provides a two-step parameter estimation process to be incorporated with variational LSTM for capturing the complex dependencies on latent variables. By maintaining such dependency relationships over latent variables, the empirical results demonstrate a dramatic improvement for cross-market data modelling and avoid posterior collapse.

All of the aforementioned approaches and frameworks for high-dimensional cross-market dependence modelling are applied in accordance with business applications, such as value at risk. These models not only provide insightful knowledge for investors to control and reduce the aggregation risk of the portfolio, but also show promising potential for further exploration and development.

