Resources Policy 119 (2026) 105943

Contents lists available at ScienceDirect

Resources Policy

journal homepage: www.elsevier.com/locate/resourpol

Identifying key natural gas drivers towards China's carbon neutrality through a novel temporal shrinkage framework

Jinyong Sung (Edward)[a,*], Xunpeng Shi (Roc)[b], Sven Teske[a], Mengheng Li[c]

[a] Institute for Sustainable Futures, University of Technology Sydney (UTS), Sydney, Australia
[b] Australia-China Relations Institute, University of Technology Sydney (UTS), Sydney, Australia
[c] Economics Discipline Group, UTS Business School, Sydney, Australia

ARTICLE INFO

Dataset link: Identifying key natural gas drivers towards China's carbon neutrality through a novel temporal shrinkage framework (Original data)

Keywords:
Carbon neutrality
LASSO
ALASSO
ECM
Natural gas
China

ABSTRACT

Natural gas (NG) plays a critical short-term role but must be gradually phased out in the long term to achieve China's ambitious dual carbon goals (DCGs), making it essential to identify key drivers that reconcile these divergent temporal demands. This study employs a novel temporal shrinkage framework that integrates big data analytics (LASSO/ALASSO) with advanced time-series modelling (ECM), proposing the long-run model (sequentially combining LASSO/ALASSO and ECM) and the short-run model (simultaneously incorporating both ECM and LASSO/ALASSO). The long-run model first identifies the key drivers in the NG market—temperature, thermal coal and HH gas prices, piped NG and LNG imports, NG market liberalisation, and NG infrastructure-related factors—and analyses their long-term effects. The short-run model then preliminarily assesses the short-term effects of these drivers while further exploring more comprehensive short-run dynamics by identifying additional drivers, such as ESG, global oil price, coking coal price, LNG and PNG import prices, and economic indicators. Our findings offer policymakers insights into these drivers, enabling the formulation of initiatives that balance short- and long-term effects to advance DCGs for China's sustainable development.

1. Introduction

In the pursuit of China's *dual carbon goals (*DCGs)—carbon peak by 2030 and carbon neutrality by 2060—natural gas (NG) occupies a paradoxical position, presenting dilemmas on two distinct fronts. In the short term, NG serves as a bridge to a low-carbon future—most notably within the ‘coal-to-gas’ strategy—offering a pragmatic solution to reduce carbon emissions by 2030 (Stern et al., 2022; Hepburn et al., 2021). In contrast, looking ahead to 2060, long-term reliance on NG threatens carbon neutrality efforts. The technology is fraught with challenges, including methane leakage (Howarth, 2024; Qin et al., 2018), the crowd-out effect that may undermine cleaner energy adoption (Gürsan et al., 2021), the risk of carbon lock-in (Gürsan et al., 2021; Unruh, 2000), and the emergence of stranded assets (Santillán Vera et al., 2023). To formulate effective initiatives for achieving DCGs, it is therefore crucial to understand the key drivers in the NG market from both short- and long-term perspectives. A core challenge for the industry lies in reconciling the burgeoning NG demand for an immediate energy transition with NG's gradual phasing out required for long-term carbon neutrality. This tension underscores the necessity for a rigorous, temporally informed analysis of the underlying NG drivers of these conflicting imperatives.

Adding to the inherent contradiction, China's NG market is shaped by dual layers of uncertainty. In the short term, market instability is exacerbated by rapid structural changes stemming from domestic market liberalisation (Wang et al., 2020a; Wei et al., 2023) and volatile global developments, such as the prolonged Russo–Ukrainian war (Mbah et al., 2022). In the long term, uncertainty increases further due to the planned fossil fuel phase-out (Meidan, 2020) and fluctuations in electricity demand—driven by the accelerated electrification necessary for achieving China's carbon neutrality goals (Li et al., 2022). Under these uncertainties, current studies find it difficult to identify the key drivers, especially to reconcile the contradictory needs in the short and long term. These studies have highlighted various factors as primary NG drivers, including economic trends, energy policies, environmental considerations, technological advancements, and global trade dynamics (See Xu et al. (2021), Bu et al. (2020), Jiang et al. (2020), Wang et al. (2020b) for recent examples). However, these studies lack consistency in the drivers' selection criteria and fail to account for the NG market's short- and long-term uncertainties. Accordingly, a more nuanced

---

* Corresponding author. University of Technology Sydney, Bldg 10, 235 Jones St, Ultimo NSW, 2007, Australia.
E-mail address: jinyong.sung@uts.edu.au (J. Sung).

https://doi.org/10.1016/j.resourpol.2026.105943
Received 22 March 2025; Received in revised form 10 May 2026; Accepted 10 May 2026

understanding of the underlying dynamics is crucial for formulating robust policies and industry strategies that balance short-term energy transition demands with long-term carbon neutrality goals, thereby supporting the effective realisation of China's DCGs and broader sustainable development agenda.

This study aims to fill these gaps by employing a range of machine learning (ML), big data, and econometric approaches to analyse the key drivers of China's NG market. This study first employs ML-based shrinkage techniques—the *Least Absolute Shrinkage and Selection Operator (LASSO)* and *Adaptive LASSO (ALASSO)*—to identify these drivers from time-series data on 40 factors related to China's NG market. These methods streamline variable selection while improving prediction accuracy, model consistency, and interpretability in high-dimensional settings. Subsequently, the selected drivers' short- and long-run effects are examined using the *Error Correction Model (ECM)*, an econometric tool that simultaneously analyses both short-term fluctuations and long-term equilibrium dynamics of variables. This novel combined LASSO-ALASSO-ECM framework identifies key drivers—temperature, thermal coal and HH gas prices, piped NG and LNG imports, NG market liberalisation, and NG infrastructure-related factors—and analyses their temporal effects. These drivers link short-term energy transition needs with long-term carbon neutrality goals by capturing both structural shifts from market liberalisation and decarbonisation—areas overlooked in previous studies. The framework also examines additional short-run variables—ESG, global oil and coal prices, LNG/PNG import prices, and economic indicators—that remain underexplored in prior studies to provide a more comprehensive view of China's NG market dynamics. Our findings provide insights for policymakers to develop effective energy initiatives for carbon neutrality, considering both short- and long-run effects.

Our primary contribution to the literature lies in the introduction of a novel temporal shrinkage framework that integrates ML and econometric methodologies into new energy policy analytics. Specifically, it explores the combined utilisation of LASSO/ALASSO-driven variable selection and ECM-based temporal analysis for identifying key short- and long-run NG drivers. To the best of our knowledge, this marks the first application of ALASSO's extended capability in the energy economics and policy field, contributing to a deeper analysis of the long-term effects. Compared to its previously recognised capability to overcome LASSO's limitations, such as bias and inconsistency (Zou, 2006), it has an extended capability to accurately select cointegrated (long-run) variables, thus identifying their long-run relationship (Mendes, 2011). This capability allows for analysis using the inherent long-term information within the data without altering it or encountering spurious regression (Shrestha et al., 2018). Another key contribution is the identification of a robust set of NG drivers capable of reconciling the contradictory imperatives of short-term energy transition and long-term carbon neutrality. This integrated approach directly addresses the tension highlighted at the beginning of the paper, providing policymakers with these drivers for developing evidence-based energy policies that reconcile short-term transition dynamics with the pursuit of long-term carbon neutrality. Furthermore, the drivers can be used as a foundational input for future forecasting research aimed at designing investment strategies for energy industry stakeholders navigating the short- and long-term complexities of the carbon-neutral transition.

The remainder of this paper is organised as follows: Section 2 outlines the background of China's NG market and reviews related literature, including influencing factors of the NG market and LASSO- and ECM-based approaches and applications. Sections 3 and 4 describe data and methods, respectively. Section 5 presents results, Section 6 provides discussion and policy implications, and the final section concludes the study. Appendix include terms and abbreviations, tables, and figures used throughout this paper.

2. Background and literature review

At the centre of China's NG market lie dual temporal uncertainties and the contrasting roles NG plays under the DCGs. In September 2020, China announced to peak carbon emissions by 2030 and achieve carbon neutrality by 2060. In the short term, NG is positioned as a key transition fuel under the coal-to-gas initiative outlined in the 14th Five-Year Plan, intended to help meet carbon dioxide (CO_2) reduction targets and improve air quality (Stern et al., 2022; Hepburn et al., 2021). Accordingly, China's short-term NG consumption is expected to rise (Stern et al., 2022). However, the uncertainties in NG consumption could also be expected to rise in the short term due to domestic structural changes resulting from market liberalisation (Wang et al., 2020a; Jia et al., 2023) and international NG price fluctuations caused by the prolonged Russo-Ukrainian war and Middle Eastern crisis (Mbah et al., 2022; Lee et al., 2023).

Conversely, China's carbon-neutral policies aim to decrease all fossil fuel consumptions, including NG, in the long term (Meidan, 2020). This long-term role of NG creates continuous uncertainties in the NG market, largely due to the absence of a clearly defined direction for phasing out NG entirely (Tuyu Zhou et al., 2022), which conflicts with more ambitious carbon neutrality goals suggested by NG's emerging critical environmental issues (Jia et al., 2022; Patel, 2024). For instance, methane emissions from the NG supply chain—approximately 80 times more potent than CO_2 as a greenhouse gas (Howarth, 2024)— undermine its climate benefits in mitigating CO_2 emissions compared to coal (Qin et al., 2018). Additionally, NG's role in potentially crowding out renewables (Gürsan et al., 2021) could result in carbon lock-in (Gürsan et al., 2021; Unruh, 2000) and stranded assets in the long term (Santillán Vera et al., 2023). These long-term uncertainties could be exacerbated by China's primary carbon neutrality strategies—aggressive renewable expansion and electrification (Liu et al., 2022)—which intensify electricity demand fluctuations, ultimately raising the volatility in the NG consumption, including the use of NG to complement intermittent renewables (Li et al., 2022).

To analyse the shifting dynamics and temporal uncertainties of China's NG demand under DCGs, it is critical to understand both the intricate web of factors shaping NG demand and the methods used to recognise them. This literature review first examines the influencing factors behind China's NG demand and, subsequently, the methods used to identify key drivers among these factors and assess their short- and long-term effects.

2.1. *Influencing factors of China's NG demand*

Research has explored the factors influencing Chinese NG demand using diverse methodological frameworks. The literature generally identifies the most influencing factors as a mixture of economic, energy, urbanisation, and demographic characteristics. Xu et al. (2021) apply the logarithmic mean Divisia index (LMDI) to analyse China's NG consumption by decomposing economic growth, energy intensity, energy structure, and substitution effects, highlighting that as economic development increases, the energy substitution effect grows while the economic growth effect decreases. Bu et al. (2020) employ social network analysis and LMDI to identify not only the economic effect reflecting per capita GDP, fossil energy structure, and energy intensity effect calculated as an inverse index of energy efficiency as main influencing factors of China's NG consumption at the national level, but also energy efficiency and population density as the main influencing factors at the provincial level. Jiang et al. (2020) find that long-run NG demand is primarily driven by economic growth, urbanisation, and energy intensity. Wang et al. (2020b) propose a grey model to determine leading NG demand factors across regions of China, identifying economic development (GDP), industry structure, environmental policy's qualitative index, urbanisation rate, population density, energy consumption intensity and energy consumption structure as the influencing factors.

Notably, they observe that the leading factors in eastern and central China are energy consumption structure, GDP, and urbanisation rate, whereas, in western China, the leading factors are urbanisation rate, industry structure, and population density.

Additionally, several studies consider energy-related infrastructure, technological progress, environmental policy, and climate factors as influencing factors, along with the combination of economic, energy, urbanisation, and demographic factors.Gao et al. (2018) use LMDI to identify economic growth, urban spatial expansion, pipeline network density and length, urbanisation rate, population density, and NG substitution as NG demand drivers in China. Liu et al. (2018) apply generalised least squares to analyse the NG consumption of urban residents in China based on NG price, household income, pipeline length, household size, energy substitution, temperature, and NG consumption intensity. Mu et al. (2018) develop a system dynamics model of China's NG supply–demand structure and conclude that economic development, urbanisation impact, and gas technological advancements are the main influencing factors on China's NG consumption rather than population growth. Overall, the main influencing factors of NG consumption in China can be categorised as follows: economic growth, including GDP; energy indicators, including NG/LNG prices and consumption and NG substitution (oil and coal); urbanisation rate; demographics, including population growth and density; NG infrastructure, such as pipeline length; energy-related technological progress; temperature; and environmental policy.

In addition to these factors, NG price distortions caused by the Chinese government's control have affected NG demand. Shi et al. (2017a) argue that government intervention in the domestic NG prices could affect NG demand by hindering the final consumers from benefiting from the low import prices. A case study on China's regulatory price distortion indicates that such interventions negatively affect economic growth due to the inefficient use of and misallocation of energy (Shi et al., 2017b). Paltsev et al. (2015) suggest that China's efforts to transition towards a market-based NG pricing system could address the supply and demand imbalance caused by price distortion.

Furthermore, environmental, social, and governance (ESG) factors are increasingly recognised as influencing the NG demand. Research underscores ESG's rising influence in the oil and gas sector by positively impacting the value of oil and gas companies and their financial condition (Ramírez-Orellana et al., 2023). Wang et al. (2020c) highlight environmental regulation—a key constituent of the environmental pillar (E) within the ESG framework (Puttachai et al., 2022)—as a major driver of NG consumption in China.

By sorting out the literature thus far, the factors influencing the NG demand are classified by economic growth, energy indicators, urbanisation, demographics, environmental policy, infrastructure, climate, technological growth, price distortion, and ESG factors. However, identifying the key drivers of China's NG market from the literature is challenging due to the varying criteria for selecting the factors. Furthermore, the analysis of their short- and long-term effects is limited. Therefore, there is a growing need for a new data-driven decomposition approach involving time-series analysis.

2.2. Methods to reveal drivers and their short- and long-term effects

To address this need, this study explores three theoretical models: two LASSO-type variable selection models (LASSO and ALASSO) and one ECM. LASSO-type shrinkage and variable selection techniques have gained prominence in the ML and econometric literature. In the realm of time-series analysis, ECM techniques have been widely used to differentiate between short- and long-run relationships among variables. This study combines these techniques to disentangle the short- and long-run effects of key drivers on China's NG consumption.

Traditional econometric models, such as decomposition techniques including time-series analysis, have limitations in big data contexts. For instance, Adow et al. (2025) examine Gulf Cooperation Council countries from 1990 to 2022 using a cross-sectional dependence (CSD)-robust panel auto-regressive distributed lag (ARDL) model within the environmental Kuznets curve framework, observing that financial development reduces CO_2 emissions in the long run, whereas energy use and urbanisation increase them. Similarly, Zhou et al. (2025) apply the CSD-ARDL model to 41 high-emission countries from 1990 to 2021 and demonstrate that energy intensity and non-renewable energy consumption increase emissions, whereas renewable energy consumption decreases them. Unlike ML techniques, these traditional models lack automated variable selection capabilities from big data, which limits the big data-driven analyses of this study.

Recent advancements in energy economics and finance have increased ML analyses, particularly for big data-driven models. Regularisation techniques, such as LASSO, which shrinks the coefficients of specific variables to zero, have garnered substantial attention for their ability to control overfitting and identify key variable characteristics (Albon, 2018), thus propelling LASSO's adoption in the energy sector. Ghoddusi et al. (2019) highlight the effectiveness of ML analyses, including LASSO, in energy-related research, covering classification, prediction, and policy analysis. Shi et al. (2020) apply LASSO to identify key driving factors affecting China's household carbon emissions, thereby providing valuable insights for formulating low-carbon policies. Accordingly, LASSO has emerged as a valuable tool in energy economics and policy research for data-driven decision-making support and understanding of complex energy systems.

However, LASSO exhibits limitations, including bias and inconsistency, in high-dimensional settings where the number of variables exceeds the sample size (Zou, 2006). ALASSO, which addresses these limitations, has been studied in various fields, including energy. Zou (2006) introduces ALASSO, which achieves model selection consistency (the oracle property) by penalising different coefficients in LASSO's regularisation term. Mendes (2011) extends ALASSO to cointegrated regressions, selecting the correct variable subset and demonstrating its oracle property. Additionally, Kock (2016) demonstrates ALASSO's oracle property through consistent and conservative model selection in time series, including stationary and nonstationary autoregressions. Duras et al. (2023) introduce ALASSO as a new variable selection method in energy economics, comparing its variable selection performance with other methods, including LASSO.

Building on these, this study aims to integrate LASSO/ALASSO's variable selection technologies with the ECM to identify NG demand drivers under DCGs and disentangle their short-run effects from long-run ones (Shrestha et al., 2018; Engle et al., 1987). Although prior research has applied LASSO-type techniques to ECMs, few have explicitly addressed temporal uncertainties. For example, Liao et al. (2015) use ALASSO within a vector ECM (VECM) to simultaneously determine the cointegration relationships and autoregressive order. Similarly, Liang et al. (2019) use the LASSO-VECM approach to explore the cointegration rank and autoregressive lag order determination in high-dimensional settings. However, these studies focus on the dynamic relationships between variables and the structural characteristics of the models rather than the short- and long-run effects of variables.

In summary, while previous literature identifies a multitude of diverse influencing factors within China's NG market, divergent selection criteria of these factors along with limited temporal analyses could obscure a clear understanding of the critical drivers and their dynamic effects. Such opacity and the lack of temporal effect analysis can impede the development of coherent and time-sensitive policies towards achieving DCGs. To address this, this study employs a novel temporal shrinkage framework integrating LASSO/ALASSO with ECM to identify key drivers in China's NG market and analyse their short- and long-run effects.

3. Data

This study examines 40 factors—identified through the literature

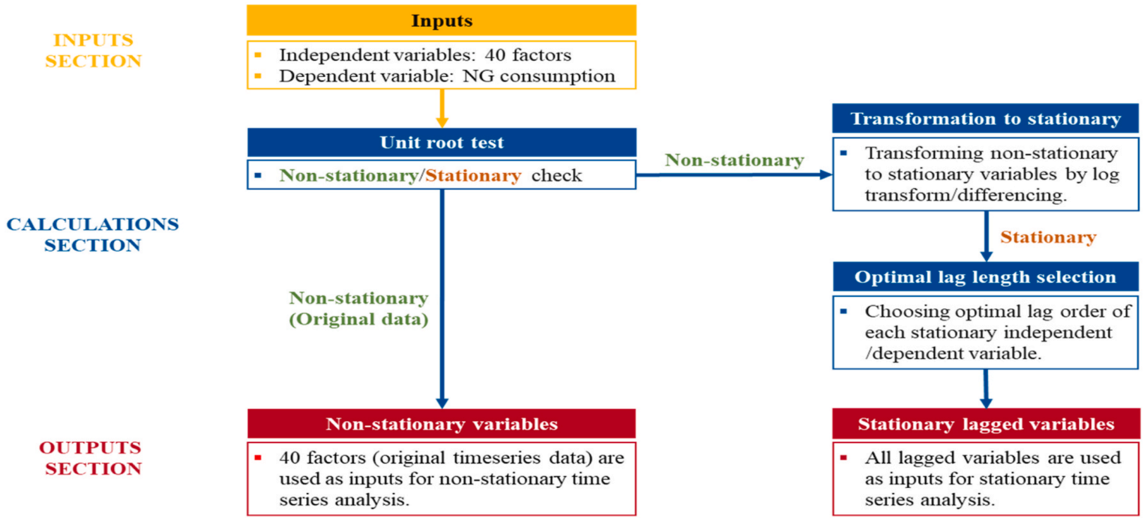


Fig. 1. Data preparation.

review—as independent variables influencing NG consumption, which serves as the dependent variable. These factors are categorised into six groups: (1) Economics and Energy (15 factors); (2) Global Energy Index (5 factors); (3) Urbanisation and Demographics (4 factors); (4) Infrastructure (4 factors); (5) Environment, Climate, and Technological Growth (7 factors); and (6) Price Distortion (5 factors).

These factors consist of monthly time-series data collected from July 2013 to December 2019, based on the availability of price distortion data. Annual or quarterly data have been converted to monthly data through interpolation. Economic factors are expressed in constant or real values. Table 1 provides a detailed description of these factors and their respective sources (see Appendix B).

3.1. Data preparation

Stationary and non-stationary time-series analyses are conducted to prepare the input data for LASSO/ALASSO and ECM modelling.[1] First, to determine whether a time-series variable is non-stationary or stationary, two main unit root tests—augmented Dickey–Fuller (ADF) and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests—are employed. For the U, AE, AF, and AG factors listed in Table 2 (see Appendix C), the Phillips–Perron test is used to further investigate the presence of a unit root, given the known finite-sample limitations in the KPSS test (see Zivot et al., 2006). Second, the choice of whether the analysis focuses on non-stationary or stationary time series determines whether to use raw data or apply transformations. Non-stationary time-series analysis uses raw data, whereas stationary time-series analysis involves transforming non-stationary variables into stationary ones through log transformation or differencing (Shrestha et al., 2018). Table 2 presents the transformation of variables into stationary ones—via log transformation or differencing—alongside their statistical significance assessment. Finally, as lagged stationary variables are used in the stationary analysis, their optimal lag order is selected using the vector autoregression (VAR) model with the Akaike information criterion (AIC) (Stock et al., 2019). As presented in Table 2, the total number of the lagged stationary variables is 117. Fig. 1 presents the schematic outlining the preparation of time-series data.

4. Methods

This section first outlines the hybrid regression techniques that sequentially integrate LASSO/ALASSO and ECM (hereafter, *long-run model*), and then presents the combined methodology that simultaneously incorporates both ECM and LASSO/ALASSO (hereafter, *short-run model*).

4.1. Long-run model

This study utilises original (non-stationary) time-series datasets of 40 factors to analyse their key drivers and temporal effects. Here, a dilemma arises regarding non-stationary data. In general, the regression model using non-stationary time series can cause a spurious regression, where linear regression shows a misleading relationship between variables (Shrestha et al., 2018). Therefore, for regression analysis, preference is given to time-series datasets that are either inherently stationary or transformed to stationarity. Moreover, transforming non-stationary time series to stationary can lead to the loss of their inherent long-run information (Shrestha et al., 2018).

The proposed long-run model identifies key drivers in China's NG market and explores their long-run effects without encountering spurious regression or losing long-run information. As shown in Fig. 2, the long-run model progresses through the following stages. First, LASSO is applied directly to the original time-series datasets of 40 factors without converting them into stationary series to identify the key driving factors. Second, a residual-based cointegration test—the 1st step of the two-step procedure in Engle et al. (1987)—estimates the cointegrated regression (long-run relationship) among these factors. ALASSO then verifies the robustness of these cointegrated factors, indicating their stable relationship even if they exhibit short-term fluctuations (Yu et al., 1992). Finally, ECM (hereafter, *LASSO-ECM1*) is used to quantify their long-run relationship. Further technical details of each approach of the long-run model—LASSO, ALASSO, and LASSO-ECM1—are provided in Appendix D.

4.2. Short-run model

The short-run model comprises LASSO-ECM1 and ECM-LASSO-ALASSO (hereafter, *LASSO-ECM2*), a simultaneously integrated approach for identifying a broader set of short-run variables. As depicted in Fig. 3, the short-run model begins with LASSO-ECM1, a preliminary test, and then proceeds to LASSO-ECM2, a more advanced test. LASSO-ECM1 yields relatively straightforward short-run effects of the key

---

[1] A stationary time series has a constant mean and variance over time, while a non-stationary time series shows changing statistical properties.

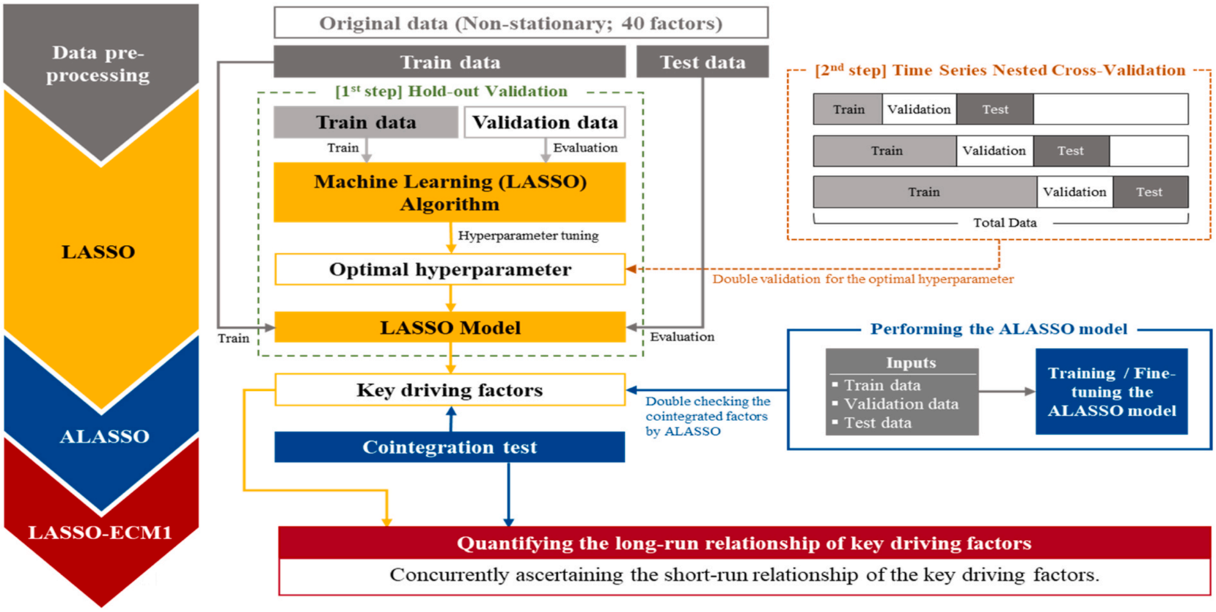


Fig. 2. Long-run model.

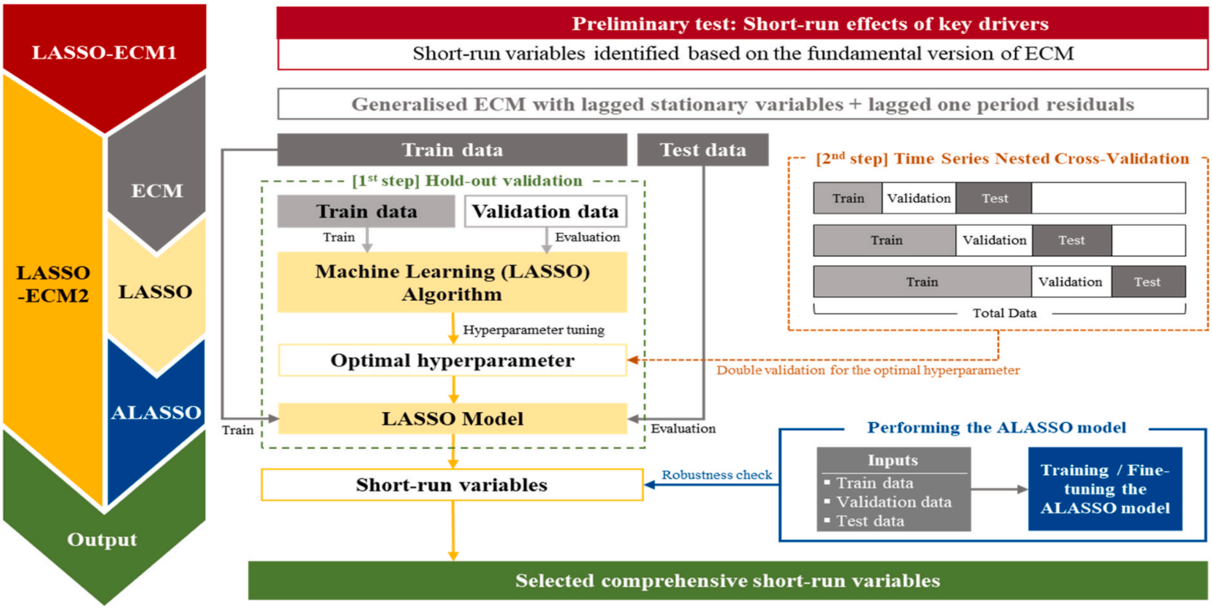


Fig. 3. Short-run model.

drivers selected within the long-run model, as it relies on the fundamental ECM framework described in Equation (3), without incorporating lagged variables. Alternatively, LASSO-ECM2 aims to capture more comprehensive short-run dynamics by incorporating lagged stationary variables. In essence, the short-run model ultimately seeks to identify a more comprehensive set of short-run variables. Further technical details of LASSO-ECM2 is explained in Appendix E.

5. Results

This section first presents the results of the long-run model, followed by the results of the short-run model.

5.1. *Long-run model results*

Long-run model results comprise three components: (1)

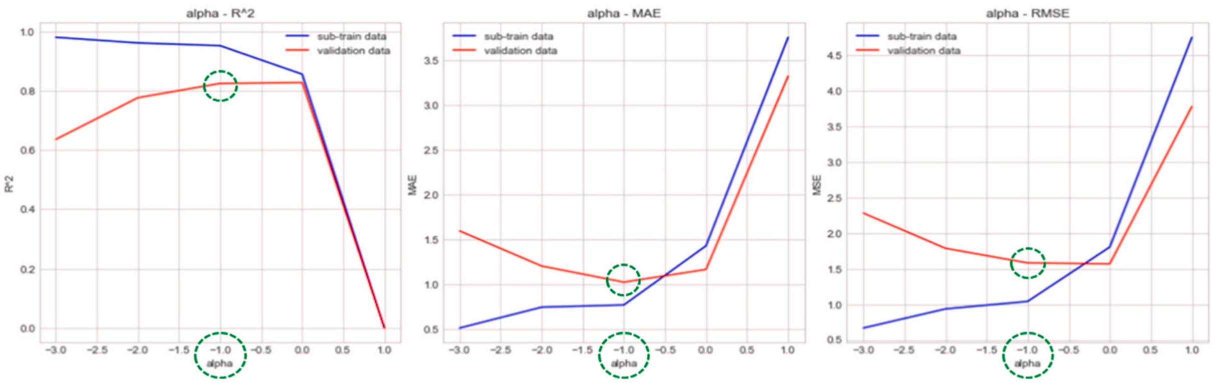


Fig. 4. Optimal hyperparameter (**α**) by hold-out validation.
Notes: A base-10 log scale is utilised for the Y-axis ('α' values) of all the above graphs. The α value is 0.1.

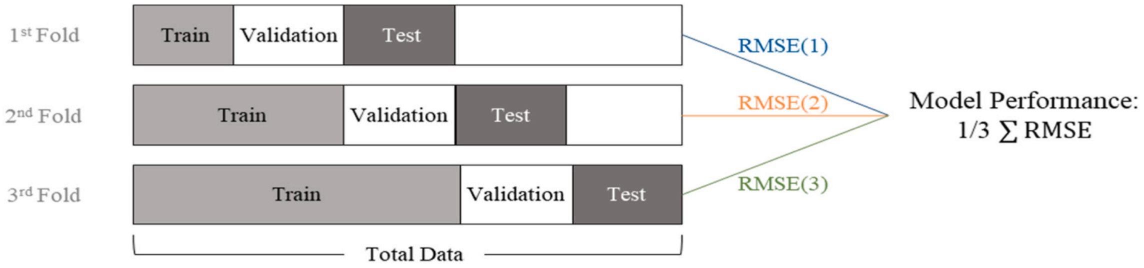


Fig. 5. Process of time series nested cross-validation.

identification of the key driving factors by LASSO; (2) verification of their long-run relationship through the residual-based cointegration test and validated by ALASSO; and (3) quantification of the relationship using LASSO-ECM1.

5.1.1. Key driving factors

Three phases are involved in LASSO modelling for selecting key driving factors: (1) data pre-processing; (2) training, tuning, and evaluation; and (3) final performance and evaluation.

In the first phase, data pre-processing, the original time-series data is split into a training set (60%), validation set (20%), and test set (20%) using a fixed random seed of 279. This seed, referred to as the *random_state* parameter in ML, represents the initial value used when shuffling data randomly, ensuring consistent output when the same ML function is run multiple times (Müller et al., 2016). The same random_state parameter is constantly used thereafter. All input sets are standardised and scaled to unit variance.

In the second phase—training, tuning, and evaluation—the training set is used to fit a LASSO model, while the validation set is utilised to evaluate the model for hyperparameter tuning, that is, determining an optimal hyperparameter (α). The hyperparameter tuning is processed in two steps:

- **Step 1**: α is identified using hold-out validation (Géron, 2022), which entails holding out a portion of the training set as a validation set, which is then used to assess several candidate models. First, these models are trained with different hyperparameters on the reduced training set. Then, α is selected through a graphical analysis of metrics, such as mean absolute error (MAE), root mean square error (RMSE), and R-squared (R^2) on the validation set. Fig. 4 indicates that α selected for the highest R^2 and lowest MAE and RMSE is 0.1.
- **Step 2**: To validate the α chosen in the first step, time series nested cross-validation (TSNCV) is employed (Kumar et al., 2017). TSNCV involves performing the train/validation/test splits of the original data in a time-ordered manner. A robust estimate of the model error (RMSE) is computed by training and tuning the model on the train/validation set of each fold and averaging the errors on the test sets. α is calculated based on RMSE. α calculated through TSNCV also yields the same value, 0.1. Fig. 5 illustrates TSNCV process.

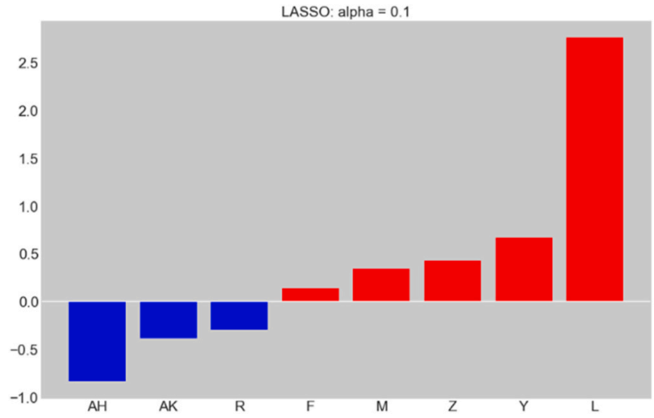


Fig. 6. LASSO results (**α** = 0.1): Key driving factors.
Note: Fig. 6 displays only variables with non-zero coefficients.

In the third phase—final performance and evaluation—the best

Table 3
LRM and Unit root test results.

LRM results

Coefficients of independent variables (key driving factors)

F	L	M	R	Y	Z	AH	AK	C
0.0327	1.625e-06	4.809e-07	−0.8487	0.0011	0.0366	−0.1039	−2.7543	7.0185
Dependent Variables: NG consumption			Method: OLS			R^2 (Adjusted R^2) : 0.944 (0.938)		

Unit root test results (ADF/KPSS) of the residuals

Unit root test		Description
ADF	P-value: 0.00 (ADF statistic: −7.90)	Based on the significance level of 0.05 and the p-value of ADF, H_0 can be rejected. Hence, the residuals are stationary at level.
KPSS	P-value: 0.10 (KPSS statistic: 0.03)	Based on the significance level of 0.05 and the p-value of KPSS, H_0 cannot be rejected. Hence, the residuals are stationary at level.

Notes: In the ADF test, H_0 (null hypothesis) represents that the series has a unit root (non-stationarity), while H_1 (alternative hypothesis) represents that the series has no unit root (stationarity). In the KPSS test, H_0 denotes that the series is stationary, while H_1 denotes that the series is not stationary. Table 4 shows notations (F, L, M, R, Y, Z, AH, and AK) denoting relevant factors. 'e−0N' means ten to the minus 'N' power (i.e. 1.625e-06 = 1.625 × 10^{-6}). 'C' of OLS regression results represents a constant.

Table 4
ADF test result of independent and dependent variables.

Factor		t-Statistic				P-value
		Test critical values			ADF statistic	
		1% level	5% level	10% level		
D(F)***	Domestic thermal coal price	−4.085092	−3.470851	−3.162458	−6.252749	0.0000
D(L)*	LNG imports	−2.601024	−1.945903	−1.613543	−1.932583	0.0515
D(M)***	NG imports via pipelines	−4.083355	−3.470032	−3.161982	−16.28665	0.0001
D(R)***	HH gas price	−4.083355	−3.470032	−3.161982	−10.05880	0.0000
D(Y)***	Accum. LNG terminal capa. per year	−4.083355	−3.470032	−3.161982	−7.152722	0.0000
D(Z)***	Gas pipeline capacity	−4.083355	−3.470032	−3.161982	−4.748410	0.0013
D(AH)***	Average national temperature	−4.103198	−3.479367	−3.167404	−9.971515	0.0000
D(AK)***	Price distortion	−4.083355	−3.470032	−3.161982	−9.983380	0.0000
D(NG)***	NG consumption	−3.533204	−2.906210	−2.590628	−7.644665	0.0000

Notes: ***, **, and * indicate ADF statistical significance at the 1%, 5%, and 10% levels, respectively. D ('factor') denotes the first difference of the factor. Based on the significance level of 0.01 and the p-value of ADF, independent (F, M, R, Y, Z, AH, and AK) and dependent (NG consumption) variables are stationary at the first difference. Based on the significance level of 0.1 and the p-value of ADF, an independent (L) variable is stationary at the first difference.

model with the optimal hyperparameter (α), selected through hold-out validation and TSNCV, is trained on the full training set (including the validation set), using the fixed random seed. This model is considered the finalised LASSO model, which is then used to evaluate the performance of the test set.

The finalised LASSO model, achieving the high accuracy (R^2) of 93.84% (on a train set) and 93.53% (on a test set), selects eight key driving factors among the 40 factors. Among these five factors—domestic thermal coal price, NG imports via pipelines, gas pipeline capacity, LNG terminal capacity, and LNG imports—are positively associated with NG consumption. The remaining three factors—temperature, price distortion, and HH gas price—exhibit a negative relationship with NG consumption. These results are illustrated in Fig. 6, where red and blue bars represent positive and negative relationships, respectively.

5.1.2. *Long-run relationship*

The long-run relationship between the selected factors and NG demand is estimated using a cointegration test and ALASSO. The cointegration test preliminarily investigates whether a long-run relationship exists, whereas ALASSO validates the robustness of the cointegration test. This process involves three phases: (1) cointegration test, (2) ALASSO-specific data pre-processing, and (3) training and evaluating phases in ALASSO.

First, the cointegration test is executed based on the 1st step of the Engle–Granger two-step procedure (Engle et al., 1987). An LRM is constructed from the key driving factors using the ordinary least squares (OLS) method. Residuals from the LRM are then calculated and two widely-used unit root tests—ADF and KPSS—are applied to these residuals, the independent variables (key driving factors), and the dependent variables (NG consumption) to determine whether these variables exhibit stationarity. As shown in Tables 3 and 4, the residuals are stationary at level and all variables are stationary at the first difference, indicating cointegration and thus a long-run relationship among the key driving factors (Engle et al., 1987).

Second, in the ALASSO-specific data pre-processing phase, the original time-series data is divided into an 80% training set and a 20%

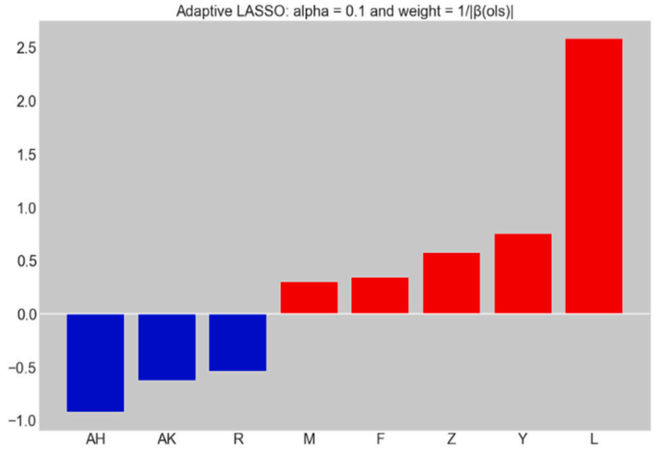


Fig. 7. ALASSO results ($\alpha = 0.1$; $\hat{w}_j = 1/|\hat{\beta}_j(ols)|$): Key driving factors.

Table 5
Coefficients by LASSO vs ALASSO.

Model	Coefficients							
	AH	AK	R	F	M	Z	Y	L
LASSO	−0.829377	−0.381393	−0.291971	0.139179	0.345636	0.430412	0.672438	2.758738
ALASSO	−0.916028	−0.619827	−0.538088	0.338705	0.293858	0.573565	0.74688	2.577915

Notes: Fig. 7 shows notations (F, L, M, R, Y, Z, AH, and AK) denoting factors and displays only variables with non-zero coefficients.

Table 6
LASSO-ECM1 results: Short- and long-term effects of key driving factors.

Variables		LASSO-ECM1 outputs			
		Coefficient	Std. Error	t-Statistic	P-value
D(AH)	Average national temperature	−0.05959	0.033334	−0.178754	0.8587
D(AK)*	Price distortion	−2.955824	1.592273	−1.856355	0.0679
D(F)	Domestic thermal coal price	0.0039266	0.035823	1.096104	0.2771
D(L)***	LNG imports	1.88e-06	2.20e-07	8.541529	0.0000
D(M)	NG imports via pipelines	3.38e-07	3.06e-07	1.104004	0.2737
D(R)*	HH gas price	−0.533474	0.308705	−1.728105	0.0887
D(Y)***	Accumulated LNG terminal capa. per year	0.002841	0.000994	2.857464	0.0057
D(Z)***	Gas pipeline capacity	0.079203	0.024306	3.258604	0.0018
D(NG)(-1)	Lagged one-period NG consumption	0.115895	0.087708	1.321375	0.1910
ECT***	Lagged error correction term	−1.164600	0.159314	−7.310104	0.0000
C	Constant	−0.133114	0.130626	−1.019050	0.3120

Notes: ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. D (‘factor’) denotes the first difference of the factor. ‘e−0N’ means ten to the minus ‘N’ power (i.e. 1.88e-06 = 1.88 $\times 10^{-6}$). The series ‘(-1)’ refers to the series lagged by one period. The lagged ECT represents lagged one-period residuals.

test set, utilising the same fixed random seed employed in LASSO. All input sets are standardised and scaled to unit variance.

Finally, in the training and evaluating phases of ALASSO, adaptive weights ($\widehat{w}_j$) are incorporated into a LASSO model to construct the ALASSO model, as per Equation (2). Following Zou (2006), a LASSO model is initially solved, then OLS is used to compute the coefficients (selected by LASSO) denoted as $\widehat{\beta}_j(ols)$. The adaptive weights are then calculated as the inverse of the absolute value of $\widehat{\beta}_j(ols)$: $\widehat{w}_j = 1/|\widehat{\beta}_j(ols)|$.[2] The ALASSO model is fitted using the training sets, these weights, and the optimal hyperparameter from LASSO. The performance of the finalised ALASSO model is then evaluated on the test set.

As Fig. 7 illustrates, the finalised ALASSO model attains a high accuracy (R^2) of 94.20% on a train set and 94.41% on a test set while selecting the same eight key driving factors as LASSO. Notably, as indicated in Table 5, the coefficients of the key driving factors in ALASSO differ from those in LASSO due to the adaptive weights. Nonetheless, the (positive/negative) relationships between the key driving factors and NG consumption in ALASSO align with LASSO. This implies that ALASSO results validate those of LASSO. Additionally, ALASSO confirms these selected factors’ long-run relationship, as preliminarily indicated by the cointegration test (Mendes, 2011).

5.1.3. *Quantification of long-run relationship*

LASSO-ECM1 represents a fundamental version of ECM, using the first-differenced key driving factors and lagged first-differenced NG consumption as independent variables, and the first-differenced NG consumption as the dependent variable. This model reaffirms the long-run relationship of the selected factors, as indicated by ALASSO.[3] As shown in Table 6, the lagged ECT in LASSO-ECM1 has a coefficient (γ) of −1.1646, which is negative and statistically significant at the 1 % level. According to Narayan and Smyth (Narayan et al., 2006), a γ value between −2 and −1 suggests ECT convergence towards the long-run equilibrium in a dampening manner. This confirms that the selected factors have long-term effects on China's NG consumption. The γ value also reveals their short-term effects (Narayan et al., 2006), further discussed in the following section.

5.2. *Short-run model results*

LASSO-ECM1 serves as a preliminary test of the short-run model, indicating that the key driving factors have short-term effects on China's NG consumption. However, as Table 6 shows, temperature, domestic thermal coal price, and NG imports via pipelines among these selected factors are not statistically significant at the 10% level. These factors may be excluded from short-run variables in this model. This limitation

---

[2] Very small numbers (1.00e-20), close to zero, are used as substitutes for zero coefficients, which appear as denominators in the weight calculation.

[3] LASSO-ECM1 can be conducted once the cointegration test results are satisfied.

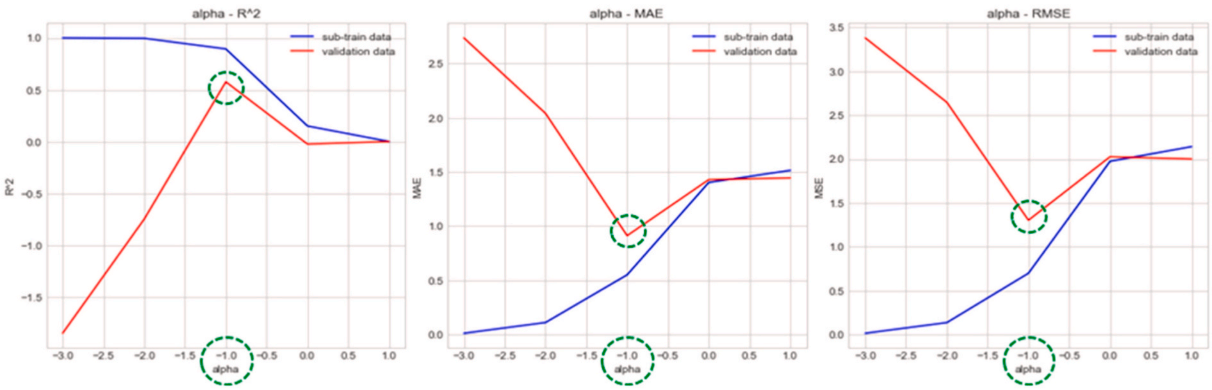


Fig. 8. Optimal hyperparameter ($\boldsymbol{\alpha}$) by hold-out validation.
Notes: A base-10 log scale is utilised for the Y-axis ('α' values) of all the above graphs. The optimal α value is 0.1.

Table 7
LASSO-ECM2 results: Selected comprehensive short-run variables.

#	Negative/Positive relationship	
		Negative relationship
1	Res(-1)	Lagged one-period residuals
2	AE(-1)	[ESG] Environmental score
3	AI	[ESG] Country score for climate change (%)
4	AH(-1)	Average national temperature (°C, Celsius)
5	E	Chinese coking coal import price (USD/MT)
6	Q(-1)	WTI oil price (USD/bbl)
7	AN(-1)	Price distortion (City gate gas price/HH, %)
8	AM	Price distortion (City gate gas price/JCC, %)
9	K(-3)	NG import price via pipelines (USD/MMBtu)
10	J(-1)	LNG average import price (USD/MMBtu)
11	C(-3)	Trend restored CLI
		Positive relationship
12	F	Domestic thermal coal price: Datong (5500 kcal/kg; USE/MT)
13	A(-5)	Real GDP by expenditure (CNY, Billions)
14	AE(-3)	[ESG] Environmental score
15	M	NG imports via pipelines (Metric tonne)
16	Y	Accumulated LNG terminal capa. per year (10 KT)
17	A(-1)	Real GDP by expenditure (CNY, Billions)
18	L(-1)	LNG imports (Metric tonne)
19	Z	Gas pipeline capacity (BCM)
20	K	NG import price via pipelines (USD/MMBtu)
21	L	LNG imports (Metric tonne)

Notes: The series ('- Number') denotes the series lagged by '#' period. Table 1 describes the calculation of price distortion factors.

may stem from the inherent nature of a fundamental ECM version, which excludes lagged variables. In case a more extensive set of lagged variables is utilised to analyse the short-run variables, it is possible that the variables excluded or not identified in the preliminary test (LASSO-ECM1) could exhibit short-term effects (Kondratieff, 1925; Reed, 2015). Therefore, for a more comprehensive examination of potential short-run variables, this study simultaneously combines LASSO/ALASSO modelling with stationary lagged variables based on the generalised ECM (LASSO-ECM2).

5.2.1. *Implementation of the model*

LASSO-ECM2 involves four main steps: (1) creating all lagged stationary variables based on the generalised ECM; (2) LASSO-specific data pre-processing; (3) training, tuning, and evaluation phases within LASSO; and (4) final performance and evaluation of LASSO and robustness check of ALASSO.

In the initial step, 118 variables are prepared for data pre-processing. This comprises 117 lagged stationary variables derived from both independent and dependent variables (as detailed in Section 3.1 and Table 2; see Appendix C), along with one lagged one-period residual variable from the LRM based on the key driving factors (as specified in Table 3). Given their stationarity, as indicated in Table 2, these variables are organised in line with the generalised ECM shown in Equation (4).

In the second step, the 118 variables are split into training, validation, and test sets, mirroring the process used for LASSO in the long-run model.[4]

In the third step, the training set is used to fit a LASSO model, whereas the validation set is used to evaluate the LASSO model for hyperparameter tuning—the process of determining an optimal hyperparameter (α). Similar to the LASSO in the long-run model, hyperparameter tuning employs hold-out validation and TSNCV. As shown in Fig. 8, α is selected based on the highest R^2 and lowest MAE and RMSE on the validation set.

In the last step, the model is trained on the full training set, including the validation set, with the fixed random seed used previously and α selected through hold-out validation and TSNCV. This results in the finalised LASSO model that selects a more comprehensive set of short-run variables on the test set. ALASSO then verifies their robustness, replicating the long-run model's approach.

As presented in Table 7 and Fig. 9 (see Appendix F), the finalised LASSO model selects 21 comprehensive short-run variables from 118 lagged stationary variables, achieving a high accuracy (R^2) of 87.96% on a training set and 83.47% on a test set. As depicted in Fig. 10 (see Appendix G), ALASSO selects the same short-run variables. Like the long-run model, while ALASSO's coefficients differ from those in LASSO due to adaptive weights, the (positive/negative) relationships between the variables and NG consumption in ALASSO align with those observed in LASSO.

5.2.2. *A more comprehensive set of short-run variables*

The short-run model (LASSO-ECM2) selects 21 variables based on the generalised ECM structure and considers them to have short-run informative content. As shown in Table 7, the lagged one-period residuals, RES(-1), are negative, confirming the alignment of these 21 variables with the generalised ECM structure. As time-series variables transition from non-stationarity to stationarity, they may lose their inherent long-run relationship or information (Shrestha et al., 2018). Hence, it can be asserted that the 21 variables carry short-run information.

Similar to the long-run model, the short-run model also selects similar factors related to NG consumption. For example, the price distortion factor selected in the long-run model is also considered in the short-run model. The preliminary test results of the short-run model show that the price distortion factor exhibits a short-run effect at the 10% significance level, and this factor is also included in the 21

---

[4] To produce consistent output, the pseudorandom number generator for data shuffling is provided with a fixed random seed of 200.

variables. As such, the findings of the short-run model are broadly consistent with the eight key drivers identified in the long-run model.

However, the short-run model aims to identify a more comprehensive set of variables capturing short-run dynamics. More specifically, this model additionally selects ESG, coking coal import price, WTI oil price, PNG import price, LNG average import price, trend restored CLI, and real GDP. Among all the short-run variables identified by the model, domestic thermal coal price, real GDP, environmental score, NG imports via pipelines, LNG terminal capacity, gas pipeline capacity, NG import price via pipelines, and LNG imports have a positive relationship with NG consumption. In contrast, ESG, temperature, coking coal import price, WTI oil price, price distortion, NG import price via pipelines, LNG average import price, and trend restored CLI exhibit a negative relationship with NG consumption.

Moreover, the short-run model includes climate factors, energy market price indicators, and macroeconomic indicators whose short-run effects have not been sufficiently highlighted in previous studies, thereby reflecting a more comprehensive set of short-run dynamics of China's NG demand. In this context, the short-run dynamics are further unpacked from three analytical perspectives.

First, this study shows that three key drivers identified in the long-run model—domestic thermal coal price, temperature, and PNG imports—can also be treated as short-run variables. The preliminary test does not reveal statistically significant short-term effects of these variables on NG demand, as shown in Table 6. However, a more comprehensive estimation of short-run effects, presented in Table 7, includes them as short-run variables, implying the presence of their short-run dynamics.

Second, the HH gas price can also be interpreted as a short-run variable. While the HH gas price exhibits a short-run effect at the 10% significance level in the preliminary test, it is not selected as part of the comprehensive set of short-run variables, as shown in Table 7. By contrast, the WTI oil price is selected as an additional short-run variable, and notably, the HH gas price and the WTI oil price exhibit interconnections in both the short run and the long run (Villar et al., 2006). Accordingly, an association between the WTI oil price and the HH gas price can be inferred, suggesting that the latter exhibits short-run characteristics.

Finally, two additional economic variables—trend restored CLI and real GDP—which are not identified in the preliminary test and long-run model can be included as part of a more comprehensive set of short-run variables. The trend restored CLI variable lagged by one period, representing an economic activity indicator, has a negative relationship with NG consumption, indicating that an economic downturn in the short-term business cycles precedes increased NG consumption. The trend restored CLI suggests that low-frequency movements in CLI are driven by the output of specific industries and financial and market indicators, which can affect the short-run dynamics of NG demand (Gyomai et al.).[5] Conversely, real GDP variables lagged by one period or five periods, representing a measure of China's economic output, exhibit a positive relationship, suggesting that economic growth leads to higher NG consumption (Li et al., 2019). These relationships suggest that changes in economic activity influence NG demand in the short term.

In brief, a comprehensive set of variables capturing short-run dynamics that are not identified in the long-run model or the preliminary test has been examined. The following section examines these dynamics in greater detail and discusses their policy implications in the context of short- and long-run relationships.

6. Discussion and policy implications

The NG key drivers identified in this study reflect both the structural changes resulting from market liberalisation and the pursuit of carbon neutrality—aspects that have received limited attention in previous studies—offering insight into how short-term transition imperatives interact with longer-term decarbonisation goals. At the same time, these drivers reveal the structural dilemma confronting China's energy transition policy. Taken together, their temporal dynamics indicate inherent policy tensions surrounding the role of NG along China's decarbonisation pathway.

In the short term, NG functions as an essential transitional fuel by enabling coal substitution and supporting the achievement of China's 2030 carbon peaking target. However, the same factors that underpin this transitional role risk reinforcing fossil-fuel-based path dependency over the longer term, thereby complicating the pursuit of carbon neutrality by 2060. This inherent duality suggests that NG policy must move beyond a narrow emphasis on supply expansion and instead adopt a more sophisticated and coherent policy framework that clearly delineates the temporal scope and functional boundaries of NG within the broader energy transition. Accordingly, this study derives five policy implications—concerning coal reduction, NG pricing mechanisms, ESG considerations, the mitigation of price distortions, and adjustments to NG infrastructure—which are discussed sequentially in the subsequent sections. This is followed by a concluding synthesis that situates these implications within a broader framework of China's energy transition and the evolving structural role of NG.

6.1. Coal-specific staged environmental cost policy

While previous studies suggest that carbon pricing on coal has yielded measurable effects (Paltsev et al., 2015), this study highlights that the internalisation of environmental costs needs to be designed more precisely by accounting for coal type and the timing of policy implementation. Domestic thermal coal prices exhibit a positive relationship with NG consumption not only in the short-run model but also in the long-run model. This positive relationship suggests that higher thermal coal prices could result in an increased demand for NG as an energy-generating substitute in both the short and long run.

Contrastingly, China's coking coal import price, selected as an additional short-run variable, exhibits a negative relationship with NG consumption. Coking coal is a specialised input in China's steel production process, and NG is not generally used as a direct substitute for coking coal (Wang et al., 2023).[6] When coking coal prices rise, steel prices increase accordingly, leading to a contraction in steel demand and, consequently, a reduction in steel production. As the steel industry is one of the most electricity-intensive sectors in China (Zhang et al., 2019), a decline in steel production reduces electricity demand at steel plants. This, in turn, lowers overall electricity demand, which subsequently affects NG consumption. As a result, NG demand from gas-fired power plants declines.

These findings suggest that environmental costs associated with coal should be internalised in a staged manner: prioritising thermal coal in the short term, while extending the scope to coking coal in the long term. Raising prices for thermal coal promotes fuel switching toward NG, whereas higher prices for coking coal reduce NG demand through impacts on industrial activity and electricity consumption. Taken together, coal pricing policy embodies a policy tension by generating divergent effects on NG demand. Specifically, empirical results from both the short- and long-run models indicate that thermal coal prices are positively correlated with NG consumption. Accordingly, raising domestic

---

[5] China's CLI components include outputs from key industries—such as chemicals, steel, construction, and motor vehicles—the diffusion index of 5000 industrial enterprises, and the turnover of the Shanghai Stock Exchange (OECD).

[6] In recent research, the substitution of coking coal in steel production is most likely associated with the application of net-zero breakthrough technologies, such as CCS and green hydrogen (Yu et al., 2022; Shen et al., 2021).

thermal coal prices in the short term (by 2030) through environmental pricing mechanisms can induce fuel switching towards NG, thereby supporting China's 2030 carbon peaking strategy—the coal-to-gas switching in the 14th Five-Year Plan (Stern et al., 2022; Hepburn et al., 2021). In the long term (covering the period from the 2030 carbon peak to the 2060 carbon neutrality target), extending environmental charges to coking coal, alongside the continued application of such charges to thermal coal, could support the achievement of long-term carbon neutrality, including a gradual reduction in NG consumption (Zhang et al., 2016).

6.2. *Environmental pricing via gas price factors*

NG price factors—LNG average import price and NG import price via pipelines—can be used as effective policy instruments for internalising environmental costs, as they exert immediate influences on NG consumption. The short-run model indicates that these price factors exhibit comprehensive short-run characteristics. Specifically, LNG average import price lagged by one period exhibits a negative relationship with NG consumption, indicating that LNG price changes in the previous period can influence NG consumer and industrial behaviour, causing fluctuations in NG consumption in the current period (Maxwell et al., 2011).

Additionally, NG import price via pipelines lagged by three periods shows a negative relationship with NG consumption, whereas NG import price via pipelines at zero lag shows a positive relationship. The positive relationship reflects the immediate response of NG demand to higher PNG import prices when NG is used as a peak-shaving fuel during periods of high electricity demand (Gu et al., 2016). By contrast, the negative relationship reflects a short-run delayed effect, whereby changes in PNG import prices influence NG consumption patterns with a certain time lag due to factors such as term contracts and price adjustment mechanisms (Wang et al., 2020a).

These findings suggest that imposing environmental costs on LNG and PNG import prices can have an immediate impact on consumption decisions and effectively internalise environmental externalities. When combined with environmental costs already applied to coal, such pricing measures can promote coal-to-gas fuel substitution and facilitate adjustments in NG demand during the energy transition period. While such coal-to-gas substitution is designed as a transitional instrument up to the 2030 carbon peak, its unintended extension into the post-peak period risks reinforcing NG dependence, highlighting a policy tension between the short-term promotion of NG and its long-term climate constraints. Avoiding such carbon lock-in therefore requires clearly sequenced NG policies. In this context, extending environmental pricing to NG in a gradual manner after 2030 becomes a necessary component of a long-term decarbonisation strategy, enabling the progressive reduction of NG consumption in line with carbon neutrality objectives (Zhang et al., 2016).

6.3. *ESG factors as short-run gas policy instruments*

This study finds that ESG factors constitute an important component of China's short-term pathway toward carbon neutrality. The short-run model selects two ESG factors—Environmental score and Country score for climate change—as additional short-run variables. A lagged three-period Environmental score shows a positive relationship with NG consumption, whereas a lagged one-period Environmental score and the Country score for climate change at zero lag exhibit negative relationships.

These relationships are explained through two empirical findings. First, the negative relationship between Country score for climate change and NG consumption reflects the fact that NG is a fossil fuel that generates CO_2 emissions. Although its CO_2 emission intensity is lower than that of coal, it contributes to climate change. Therefore, increasing NG consumption is negatively related to reducing the Country score for climate change. Second, the Environmental score exhibits both positive and negative relationships with NG consumption at different lags, highlighting short-run responses. The positive relationship between NG consumption and the Environmental score lagged by three periods could be attributed to underlying factors or dynamics. There could be a delayed short-run effect of changes in environmental practices or policies on NG consumption. Alternatively, the negative relationship between NG consumption and the Environmental score lagged by one period could indicate a more immediate impact. Methane leakage from NG consumption currently raises environmental concerns (Howarth, 2024; Qin et al., 2018), explaining the negative relationship for the Environmental score lagged by one period. Both the Environmental score and the Country score for climate change are considered environmental factors within ESG factors, hence these two scores exhibit the same (negative) relationship with NG consumption.

These results suggest a policy tension resulting from the coexistence of efforts to promote NG as a short-term coal substitute and ESG-driven decarbonisation objectives, underscoring that ESG performance already operates as a meaningful short-run governance factor shaping China's natural gas demand rather than merely a long-term normative framework. In particular, the negative relationships between NG consumption and both the Environmental score and the Country score for climate change at short lags indicate that ESG factors are already embedded in short-term NG demand decision-making processes.

ESG factors should be integrated into NG policy to support China's short-term low-carbon pathway. ESG factors are generally viewed as long-term determinants in energy investment and policy (Fu et al., 2024). However, this study suggests that ESG factors can also function as effective instruments in managing the short-term NG demand. Meanwhile, although regulations requiring large publicly listed companies to disclose ESG information are expected to be introduced by 2026, ESG reporting is not yet mandatory in China (Interesse, 2024). Additionally, progress toward certain carbon neutrality targets has remained limited (Myllyvirta et al., 2024; Cushing, 2024). Against this background, incorporating ESG factors into NG policy and regulatory frameworks is essential to advance China's short-term low-carbon transition.

6.4. *Phased gas pricing reform*

This study underscores the importance of minimising price distortions in the NG market by distinguishing short-from long-term price structures. The study demonstrates that price distortions negatively impact NG demand in both the short and long term, with a large LASSO coefficient, smaller than temperature but still impactful. More specifically, the study shows that NG consumption responds negatively to price distortions linked to international energy prices—HH gas and JCC—in the short run, and to PNG import prices in the long run.

These findings suggest that a phased pricing reform, designed to minimise NG price distortions based on short- and long-term price structures, could serve as a key policy instrument in supporting the DCGs. In the short term (carbon peak by 2030), the introduction of a market-based NG pricing reform linked to international energy prices could enhance the relative competitiveness of NG vis-à-vis coal. Such a reform could promote NG use over carbon-intensive and inexpensive energy sources (Paltsev et al., 2015), aligning with China's short-term strategy for carbon neutrality through the coal-to-gas transition.

In the long term, in line with the 2060 carbon neutrality target—including a gradual reduction in NG consumption—under the DCGs, adopting a market-based pricing reform anchored to PNG import prices, which are typically governed by long-term contracts referenced to global energy price benchmarks and thus offer greater supply security and lower price volatility (Pye et al., 2025), could enable a more rational linkage between domestic gas prices and underlying market fundamentals (Lu et al., 2025). Such a more stable market-based pricing reform could align domestic gas markets with global pricing mechanisms, allowing for the internalisation of fossil-fuel externalities, while exposing NG prices to market-based competition from increasingly cost-competitive renewables. This could reduce NG competitiveness against renewables (Yang et al., 2020; Wesseh et al., 2019) and eventually facilitate a gradual decline in NG consumption along the long-term decarbonisation pathway (Green et al., 2024).

Taken together, and viewed more broadly, this temporal interaction reveals a structural policy tension: market-based pricing reform enhances efficiency and transparency in the short run, yet may ultimately erode the competitiveness of NG relative to renewables, accelerating its decline within a carbon-neutral energy system.

6.5. *Strategic optimisation of NG infrastructure*

This study demonstrates that NG infrastructure factors—LNG terminal capacity and gas pipeline capacity—positively affect NG consumption in both the short and long run. More specifically, the long-run model shows that these infrastructure factors, alongside LNG imports, exhibit large LASSO coefficients and maintain a positive long-run relationship with NG consumption, suggesting that they are major determinants in NG consumption changes. Meanwhile, the preliminary test of the short-run model reveals that these factors exert statistically significant short-term effects at the 1% significance level. The short-run model also identifies them as comprehensive short-run variables. In addition, imports associated with this NG infrastructure—LNG and PNG imports—exhibit a positive short-term relationship with NG consumption. Taken together, these results indicate that the NG infrastructure factors, together with their associated import factors, exert positive effects on NG consumption in both the short and long run.

These findings highlight a central policy trade-off with direct implications for energy security, reflecting the broader policy tension between short-term transition priorities and long-term decarbonisation objectives. While expanded LNG terminals and pipeline infrastructure can enhance short-term supply security and system flexibility, they may, over time, undermine long-term energy security by entrenching import dependence and increasing exposure to external supply and geopolitical risks. Moreover, large-scale and long-lived gas infrastructure investments can raise the political and economic costs of a rapid transition away from NG, thereby reinforcing carbon lock-in. Accordingly, NG import infrastructure needs to be strategically optimised by explicitly differentiating its short- and long-term functional roles. In this context, infrastructure developed to support NG in line with China's 2030 carbon peak target should gradually transition toward configurations that reduce NG consumption to align with the 2060 carbon neutrality objective. In the short term, China could enhance the NG supply chain's capability and flexibility by diversifying supply sources and developing new gas infrastructure, such as pipelines and LNG terminals, to support NG demand. In the long term, this infrastructure can be repurposed for hydrogen, which is emerging as an essential energy source towards carbon neutrality (Ogden et al., 2018).

6.6. *The role of natural gas in China's energy transition*

Overall, this study demonstrates that NG plays neither a purely bridging nor a purely obstructive role in China's energy transition. Rather, its role is structurally contingent upon policy design, market reform, and infrastructure choices. In the short term, NG can support decarbonisation by enabling coal substitution, enhancing system flexibility, and responding swiftly to price and ESG signals. In the longer term, however, these same mechanisms—if not carefully coordinated—risk reinforcing import dependence, infrastructure lock-in, and delayed substitution by zero-carbon energy sources. This highlights the need to govern NG not as an end-state fuel, but as an explicitly time-bound and policy-conditioned component of China's pathway towards a zero-carbon energy system.

7. Conclusions

China's DCGs place NG in a paradoxical position: it is indispensable as a short-term transitional fuel but poses a long-term risk to carbon neutrality. This tension is compounded by short-term market volatility and long-term uncertainties, underscoring the need to identify key drivers in China's NG market that reconcile these conflicting temporal demands. Leveraging advanced analytical methods, this study offers a systematic framework for isolating the most relevant market drivers. The insights derived from this framework lay the groundwork for policies and strategies that effectively balance immediate energy needs with long-term sustainability.

This study employs a novel temporal shrinkage framework integrating big data analytics (LASSO/ALASSO) with advanced time-series modelling (ECM), through two models: the long-run model (sequentially combining LASSO/ALASSO and ECM) and the short-run model (simultaneously incorporating both ECM and LASSO/ALASSO). The long-run model initially identifies key drivers in the NG market—average national temperature, price distortion, HH gas price, NG imports via pipelines, domestic thermal coal price, gas pipeline capacity, LNG terminal capacity per year, and LNG imports—and assesses their long-term effects. Subsequently, the short-run model shows that three of these drivers (domestic thermal coal price, temperature, and PNG imports), although not confirmed as having short-term effects in the model's preliminary test, indeed exhibit such effects. Additionally, it identifies additional short-run variables, such as energy commodity prices, ESG, and economic factors. Together, these models clarify the key drivers' short- and long-term effects.

The study not only introduces a novel temporal shrinkage framework for energy policy analytics but also demonstrates ALASSO's extended capability in accurately selecting cointegrated variables. This framework leverages the inherent long-term information within the data without disrupting it or encountering spurious regression (Shrestha et al., 2018). Furthermore, the study results offer policymakers insights into the key drivers, enabling the formulation of initiatives that balance short- and long-term effects to advance DCGs for China's sustainable development. Based on these findings, this study derives five policy implications—spanning coal, gas pricing, ESG, price distortion, and NG infrastructure—which provide a foundation for supporting China's ambitious implementation of carbon neutrality and clarify the role of NG as an explicitly time-bound and policy-conditioned component of the transition.

Furthermore, in parallel with these policy implications, sustained, government-supported investments in net-zero breakthrough technologies—including renewables and hydrogen—can boost their competitiveness over fossil fuels and complement their phase-out (Stern et al., 2021). This may enable an ambitious carbon-neutral pathway—the complete phase-out of fossil fuels through the technologies (Yu et al., 2022)—similar to Denmark's case, which now sources 89% of its electricity from renewables (IEA, 2023a) and aims to phase out NG entirely by 2030 (IEA, 2023b). This pathway could serve as a novel feasible guideline for achieving China's carbon neutrality.

One limitation of this study is its temporal focus on China's 2030 carbon peaking (short-term) and 2060 carbon neutrality (long-term) goals. However, some forecasts suggest that China's emissions peak may be delayed until around 2040, corresponding to a medium-term horizon (Fang et al., 2019). Although it may be feasible to design medium-term policies based on long-term determinants, such an approach entails methodological limitations. The proposed ECM framework primarily models short-term adjustments towards the long-term equilibrium of the variables and is thus not suitable for quantifying medium-term effects. On the transition path towards the medium-term goal, one cannot rule out the possibility that the long-term equilibrium could remain constant. As such, a potential extension of the ECM framework would be to allow for time variation in the cointegration relationship following Bierens et al. (2010) and Koop et al. (2011). Such extension warrants further investigation.

Another limitation is that LASSO-ECM2, while effective for short-term analysis, is restricted in examining long-term effects. Although the selected negative RES(-1) in Table 5 may suggest a long-run relationship, LASSO-ECM2 can distort the proper relationship between independent and dependent variables by shrinking less important coefficients towards zero, resulting in a biased coefficient for RES(-1) (Ranstam et al., 2018). Additionally, LASSO-ECM2 cannot evaluate standard errors for the biased coefficient, indicating that the coefficient might not be statistically significant. Furthermore, LASSO-ECM2 does not use the ECM techniques for long-term effects. Future research should employ additional statistical tools to investigate these effects more comprehensively.

CRediT authorship contribution statement

Jinyong Sung: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Xunpeng Shi:** Resources, Supervision, Writing – review & editing. **Sven Teske:** Supervision, Writing – review & editing. **Mengheng Li:** Methodology, Validation.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Jinyong (Edward) SUNG reports financial support and article publishing charges were provided by University of Technology Sydney Institute for Sustainable Futures. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research is supported by an Australian Government Research Training Program Stipend (UTS Research Excellence) and Fee Offset Scholarship. The views expressed in this paper are solely those of the authors and do not necessarily reflect the views of the supporting agencies and authors' affiliations.

Appendix

A. Terms and abbreviations

Term	Description
ACRI	Australia-China Relations Institute at UTS
ALASSO	Adaptive LASSO
Bbl	Barrel
BCM	Billion cubic metres
Brent	A global oil benchmark
CLI	Composite Leading Indicator; Early signals of turning points in business cycles
CNY	Chinese yuan renminbi; The official currency of the People's Republic of China
CPI	Consumer Price Index
DCGs	Dual Carbon Goals
ECM	Error Correction Model
ECT	Error Correction Term
Envir.	Environment
EIA	U.S. Energy Information Administration.
ESG	Environmental, Social, and Governance
GDP	Gross Domestic Product
HH	Henry Hub; A distribution hub on the NG pipeline system in Louisiana in the U.S.
JCC	Japan Crude Cocktail; In general, JCC, the average price of customs-cleared crude oil imports into Japan, is used by LNG traders in APAC to price LNG contracts.
KEEI	Korea Energy Economics Institute
KT	Kiloton (1000 tons)
LASSO	Least Absolute Shrinkage and Selection Operator
LNG	Liquefied Natural Gas
LRM	Linear Regression Model
MAE	Mean Absolute Error
MGMT	Management
ML	Machine Learning
MMBtu	Metric Million British Thermal Unit
MT	Metric Ton
NBS	National Bureau of Statistics of China
OECD Stat	Organization for Economic Cooperation and Development Statistics
PNG	Piped Natural Gas
RMSE	Root Mean Squared Error
USD	US Dollar; the currency of the United States
UTS	University of Technology Sydney
WTI	West Texas Intermediate; A global oil benchmark

B. Description of data: 40 factors

Table 1
Description and relevant sources of the 40 factors

Factors	Description	Data source[a]
(1) Economics and Energy (15 factors)		
Real GDP by expenditure (CNY, Billions)	Quarterly → Monthly series	Bloomberg
CPI: Overall average	Monthly series	Refinitiv; NBS
Trend restored CLI[b]	Monthly series	OECD Stat
Average city gate gas price (USD/MMBtu)	Monthly series	Bloomberg; Internal source
Chinese coking coal import price (USD/MT)	Monthly series	Bloomberg
Domestic thermal coal price: Datong (5500 kcal/kg; USD/MT)	Monthly series	Sxcoal; Internal source[c]
Domestic thermal coal price: Datong (5800 kcal/kg; USD/MT)	Monthly series	
Newcastle coal index: (6000 kcal/kg; USD/MT)	Monthly series	
Crude petroleum oil import price (USD/bbl)	Average monthly price[d]	Refinitiv
LNG average import price (USD/MMBtu)	Monthly series	Bloomberg
NG import price via pipelines (USD/MMBtu)	Monthly series	
LNG imports (Metric tonne)	Monthly series	
NG imports via pipelines (Metric tonne)	Monthly series	
Energy intensity level of primary energy (MJ/GDP measured in constant 2017 USD at Purchasing Power Party)	Monthly series	World Bank Indicator
Industry structure[e]	Yearly → Monthly series	Bloomberg
(2) Global Energy Index (5 factors)		
Brent oil price (USD/bbl)	Monthly series	Bloomberg
LNG JCC import price (USD/MMBtu)	Monthly series	
WTI oil price (USD/bbl)	Monthly series	EIA
HH gas price (USD/MMBtu)	Monthly series	
Int'l spot thermal coal price benchmark (USD/MT)	Monthly series	CCTD[f]
(3) Urbanisation and Demographics (4 factors)		
Urbanisation rate (%, Urban population of total population)	Monthly series	World Bank Indicator
Per capita income (constant 2015 USD)	Monthly series	
Population density (people per sq. km of land area)	Monthly series	
Population (Millions)	Monthly series	Bloomberg
(4) Infrastructure (4 factors)		
Accumulated LNG terminal capa. per year (10 KT)	Monthly series	KEEI
Gas pipeline capacity (BCM)	Monthly series	External source[g]
City public utilities: Length of gas pipelines (10 KT)	Monthly series	NBS
Pipeline transportation: Length of gas pipelines (10 KT)	Monthly series	CEIC[h]
(5) Environment, Climate, and Technological Growth (7 factors)		
[ESG] Environmental score	Yearly → Monthly series	Refinitiv
[ESG] Social score	Yearly → Monthly series	
[ESG] Governance score	Yearly → Monthly series	
[ESG] Country score for climate change (%)	Yearly → Monthly series	
Investment % of Chg Y/Y: Water conservancy, Envir. & Public utilities mgmt.	Monthly series	
Average national temperature (°C, Celsius)[i]	Monthly series	NBS
Technology and Innovation: Envir.-related patents of total patents (%)	Yearly → Monthly series	OECD Stats
(6) Price Distortion (5 factors)		
Price distortion (City gate gas price/LNG import price, %)	Monthly series	Bloomberg, Internal source, and estimation[j]
Price distortion (City gate gas price/Piped gas price, %)	Monthly series	
Price distortion (City gate gas price/Average of LNG import price and Piped gas price, %)	Monthly series	
Price distortion (City gate gas price/JCC, %)	Monthly series	
Price distortion (City gate gas price/HH, %)	Monthly series	

[a] Bloomberg and Refinitiv, global economic and financial data providers, are accessed through UTS Business School; The internal source is provided by UTS ACRI; Sxcoal, a Chinese coal data provider, is referenced from http://www.sxcoal.com/site/about-us/en.

[b] The trend restored CLI provides early signals of turning points in business cycles, reflecting economic trends such as GDP. This index is calculated by multiplying the raw CLI, which varies around 100, by the trend of the reference series such as GDP (OECD, 1998).

[c] Currently, China imports more coking coal than it domestically produces because coking coal, in contrast to thermal coal, is relatively scarce (Tu et al., 2012). Thus, the domestic coking coal price is not factored into data collection, whereas both domestic and import prices of thermal coal are considered.

[d] Crude Petroleum Oil Import Price is the average monthly price calculated as Total monthly price/Total monthly volume.

[e] Industry structure is calculated as Secondary Industry's real GDP/Total real GDP. Secondary industry includes construction and manufacturing, both of which consume a significant amount of energy. This calculation is suggested by Wang and Li (Wang et al., 2020b).

[f] CCTD is a Chinese coal data provider, referenced from http://www.cctdcoal.com/.

[g] This reference comes from Chen et al. (2021).

[h] CEIC is a global economic, industry, and financial data provider.

[i] This factor is calculated as the average temperature of all provinces per month.

[j] Price distortion factors are calculated using the formula: ‘(City gate gas price/each energy price - 1) ×100’.

C. Lagged stationary variables

Table 2
Lagged stationary variables of 40 factors

Factors			Stationary status	Lag order by AIC	Num. of lagged variables
1	A	Real GDP	Log transform & 2nd difference***	6	7
2	B	CPI	1st difference*	1	2
3	C	Trend restored CLI	Log transform & 2nd difference***	3	4
4	D	Average city gate gas price	1st difference***	0	1
5	E	Chinese coking coal import price	1st difference***	1	2
6	F	Domestic thermal coal price: Datong - 5500 kcal/kg	1st difference***	6	7
7	G	Domestic thermal coal price: Datong - 5800 kcal/kg	1st difference***	3	4
8	H	Newcastle coal index: 6600 kcal/kg	1st difference***	1	2
9	I	Crude petroleum oil import	1st difference***	1	2
10	J	LNG average import price	1st difference***	0	1
11	K	NG import price via pipelines	1st difference***	3	4
12	L	LNG imports	1st difference***	4	5
13	M	NG imports via pipelines	1st difference***	2	3
14	N	Industry structure	2nd difference***	5	6
15	O	Energy intensity level of primary energy	2nd difference***	0	1
16	P	Brent oil price	1st difference***	1	2
17	Q	WTI oil price	1st difference***	1	2
18	R	HH gas price	1st difference***	1	2
19	S	Int'l spot thermal coal price benchmark	1st difference***	1	2
20	T	LNG JCC import price	1st difference***	1	2
21	U	Urbanisation rate	Non-difference**	2	3
22	V	Per capita income	Log transform & 2nd difference***	0	1
23	W	Population	Log transform & 2nd difference***	0	1
24	X	Population density	Log transform & 2nd difference***	0	1
25	Y	Accumulated LNG terminal capa. per year	1st difference***	0	1
26	Z	Gas pipeline capacity	1st difference***	0	1
27	AA	City public utilities: Length of gas pipelines	Log transform & 2nd difference***	0	1
28	AB	Pipeline transportation: Length of gas pipeline	Log transform & 2nd difference***	0	1
29	AC	Water conservancy, Envir. & Public utilities mgmt.	2nd difference***	6	7
30	AD	Envir.-related patents of total patents	1st difference***	0	1
31	AE	[ESG] Environmental score	2nd difference***	5	6
32	AF	[ESG] Social score	2nd difference***	5	6
33	AG	[ESG] Governance score	2nd difference***	5	6
34	AH	Average national temperature	1st difference***	6	7
35	AI	[ESG] Country score for climate change	1st difference*	0	1
36	AJ	Price distortion (City gate gas price/LNG import price)	1st difference***	0	1
37	AK	Price distortion (City gate gas price/Piped gas price)	1st difference***	1	2
38	AL	Price distortion (City gate gas price/Average of LNG import price and Piped gas price)	1st difference***	5	2
39	AM	Price distortion (City gate gas price/JCC)	Non-difference*	2	3
40	AN	Price distortion (City gate gas price/HH)	1st difference***	1	2
41	[Dependent variable] NG consumption		1st difference***	2	2
Total number of lagged stationary variables					**117**

Notes: ***, **, and * indicate statistical significance of unit root tests at the 1%, 5%, and 10% levels, respectively. The actual optimal lag order of the factor (AL) is 5, but a lag order of 1 is practically use due to a limit of available data. In this study, an alphabetical assignment is attributed to each factor.

D. Technical details of long-run model

This appendix outlines the technical details of the long-run model, which is structured as a sequential framework integrating LASSO/ALASSO and LASSO-ECM1. The technical exposition is organised into two stages. First, the LASSO and ALASSO methodologies employed to identify key drivers are described. Second, the ECM-based approach—LASSO-ECM1—to characterise the long-run relationships among these drivers is presented.

D.1 LASSO and ALASSO

LASSO, an ML's technique, performs both feature selection and regularisation and is mathematically expressed as follows:

$$\boldsymbol{Minimize}_{\beta_0,\beta}\left(\frac{\mathbf{1}}{\mathbf{2}\boldsymbol{n}}\sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{n}}\left(\boldsymbol{y}_{\boldsymbol{i}}-\boldsymbol{\beta}_{\mathbf{0}}-\boldsymbol{x}_{\boldsymbol{i}}^{\boldsymbol{T}}\boldsymbol{\beta}\right)^{\mathbf{2}}+\boldsymbol{\alpha}\sum_{\boldsymbol{j}=\mathbf{1}}^{\boldsymbol{m}}\left|\boldsymbol{\beta}_{\boldsymbol{j}}\right|\right) \tag{1}$$

where n is the total number of observations; y_i is the dependent variable; m is the total number of independent variables $x_i = (x_{i1}, \ldots, x_{im})^T$; β_0 is the intercept; β_j represents the other parameters (coefficients); and α is the hyperparameter value.[7]

Equation (1) shows that LASSO penalises the cost function of the linear regression model (LRM). The penalty is the sum of the absolute values of the hyperparameter value α multiplied by coefficients, which is called *L1 Regularisation*. The cost function is the sum of the difference between y_i (actual result) and $\hat{y}_i = \beta_0 - x_i^T\beta$ (estimated result).

LASSO aims to find β_0 and β that minimise the cost function and penalty. Specifically, each coefficient in the penalty is reduced by increasing the hyperparameter value α. This process reduces some coefficients to zero. These unique properties of LASSO can lead to relatively choosing more essential variables, while excluding less important variables by setting their coefficients to zero (Géron, 2022).

However, LASSO can inadvertently eliminate features with small coefficients, causing overfitting (Zou, 2006). To overcome this limitation, ALASSO penalises larger coefficients, thereby identifying an accurate set of true variables. This phenomenon—known as the *oracle property*—enables ALASSO to select cointegrated variables, indicating their long-run relationship (Mendes, 2011).

ALASSO extends LASSO using adaptive weights (Shortreed et al., 2017). The mathematical formulation for ALASSO is as follows:

$$\boldsymbol{Minimize}_{\beta_0,\beta}\left(\frac{1}{2n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2+\alpha\sum_{j=1}^{m}\hat{w}_j|\beta_j|\right) \tag{2}$$

where $\hat{w}_j$ represents adaptive weights, calculated reciprocally from $|\hat{\beta}_j|$ values derived from LASSO. The remaining parameters include n, y_i, $\hat{y}_i$ $\left(= \beta_0 - x_i^T\beta\right)$, m, β_j, and α, mirroring LASSO's parameters. As demonstrated in Equation (2), ALASSO introduces adaptive weights to impose penalties on different coefficients within L1 Regularisation. In this study, both LASSO and ALASSO models are developed using Python, an open-source programming language.

D.2 LASSO-ECM1

The ECM constitutes the 2nd step of the Engle–Granger two-step procedure after a residual-based cointegration test (Engle et al., 1987), designed to estimate both short- and long-run effects of one time-series variable on another (Shrestha et al., 2018). The fundamental version of ECM is mathematically expressed as follows:

$$\Delta Y_t = \beta_0 + \beta_1\,\Delta X_t - \gamma\hat{\mu}_{t-1} + e_t \tag{3}$$

where β_1 serves as the short-run coefficient, signifying the immediate impact of a change in X_t on a change in Y_t; $\hat{\mu}_{t-1}$ represents the lagged one-period residuals—referred to by the error correction term (ECT)—that are estimated through the LRM of two different series X_t and Y_t; γ stands as the coefficient of the ECT—also known as the adjustment effect or error correction coefficient—dictating the speed of adjustment towards the long-run equilibrium; e_t accounts for the white noise error term; β_0 denotes the constant term; and Δ represents the first difference operator.

In general, as presented in Equation (3), a statistically significant and negative ECT coefficient (γ) indicates that deviations from the long-run equilibrium are subsequently corrected towards the equilibrium. The adjustment effect, as governed by γ, is specifically observed in the two cases (Narayan et al., 2006): (1) When γ falls between -1 and 0, the adjustment monotonically converges to the long-run equilibrium; and (2) when γ ranges between -2 and -1, the adjustment oscillates around the long-run equilibrium with a damping effect. The long-run relationship of the LASSO-selected key drivers within these cases can be estimated, but not others. This analytical approach, carried out using the statistical software EViews, corresponds to LASSO-ECM1 as introduced earlier in this study. Simultaneously, LASSO-ECM1 also estimates the short-run relationship of these drivers.

E. Technical details of short-run model: LASSO-ECM2

LASSO-ECM2 represents a simultaneously integrated ECM–LASSO–ALASSO approach in the short-run model. More specifically, LASSO-ECM2 provides estimates for identifying more comprehensive short-run variables by applying LASSO/ALASSO techniques to lagged stationary variables, formulated in the structure of a generalised ECM. The generalised ECM is mathematically expressed as follows:

$$\Delta Y_t = \beta_0 + \sum_{i=0}^{n}\beta_i\Delta X_{t-i} + \sum_{j=1}^{m}\alpha_j\Delta Y_{t-j} - \gamma\hat{\mu}_{t-1} + e_t \tag{4}$$

where α_j and β_i are short-run coefficients; n and m are lags of variables; $\hat{\mu}_{t-1}$ is the lagged one-period residuals estimated from the LRM of two different series X_t and Y_t; γ is the coefficient of ECT; e_t is the white noise error term; β_0 is the constant term; and Δ represents the first difference operator.

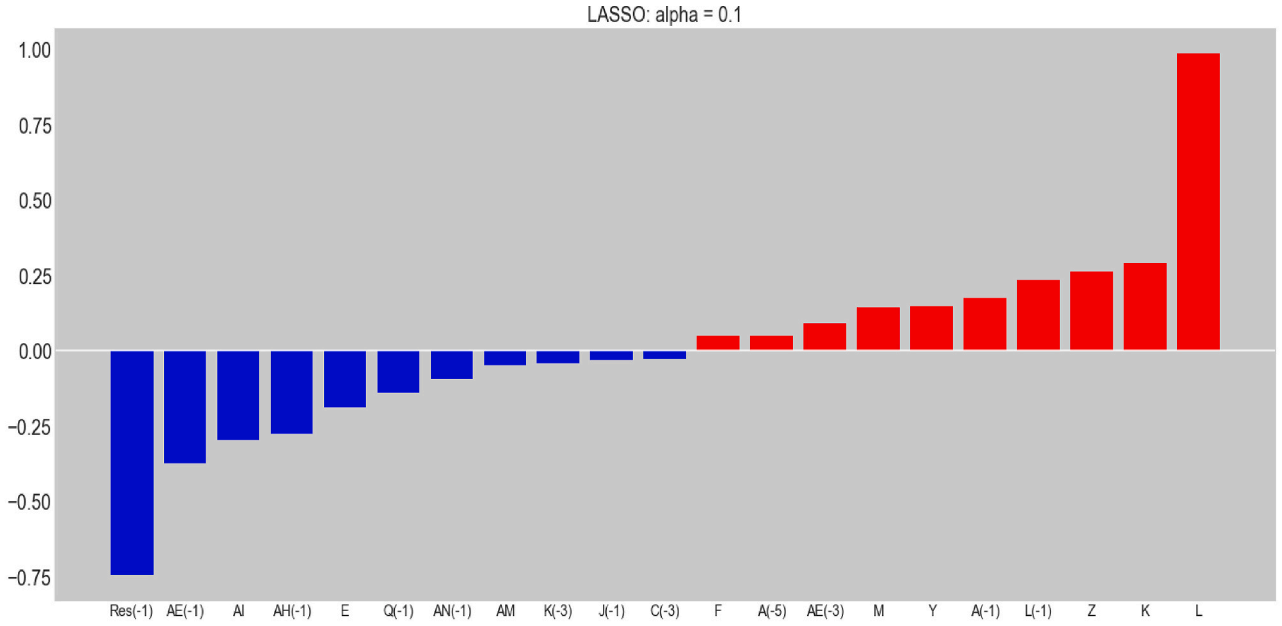
In general, while raw time-series variables (at zero lag) might not impact a system in the short term, lagged variables can exert such effects. Incorporating lagged variables enables researchers to evaluate how past values impact the current system, improving their understanding of its dynamics (Kondratieff, 1925; Reed, 2015). Hence, to identify more comprehensive short-run variables, a generalised ECM with lagged stationary variables is first formulated. Subsequently, following the general-to-specific approach,[8] LASSO is applied to eliminate less essential independent variables in a data-driven manner. Finally, ALASSO provides a robustness check for the LASSO outputs using its oracle properties. In summary, this analytical approach—LASSO-ECM2—enables us to delve into more comprehensive short-run dynamics by introducing additional variables. This analysis is conducted using Python.

---

[7] In ML algorithms, hyperparameters control the learning process and assist in determining the appropriate model parameters (coefficients).

[8] The general-to-specific approach is an econometric methodology designed to simplify a general model that initially includes all potentially relevant variables. This simplification is achieved by systematically eliminating insignificant or irrelevant variables (Campos et al., 2005).

F. LASSO results in LASSO-ECM2

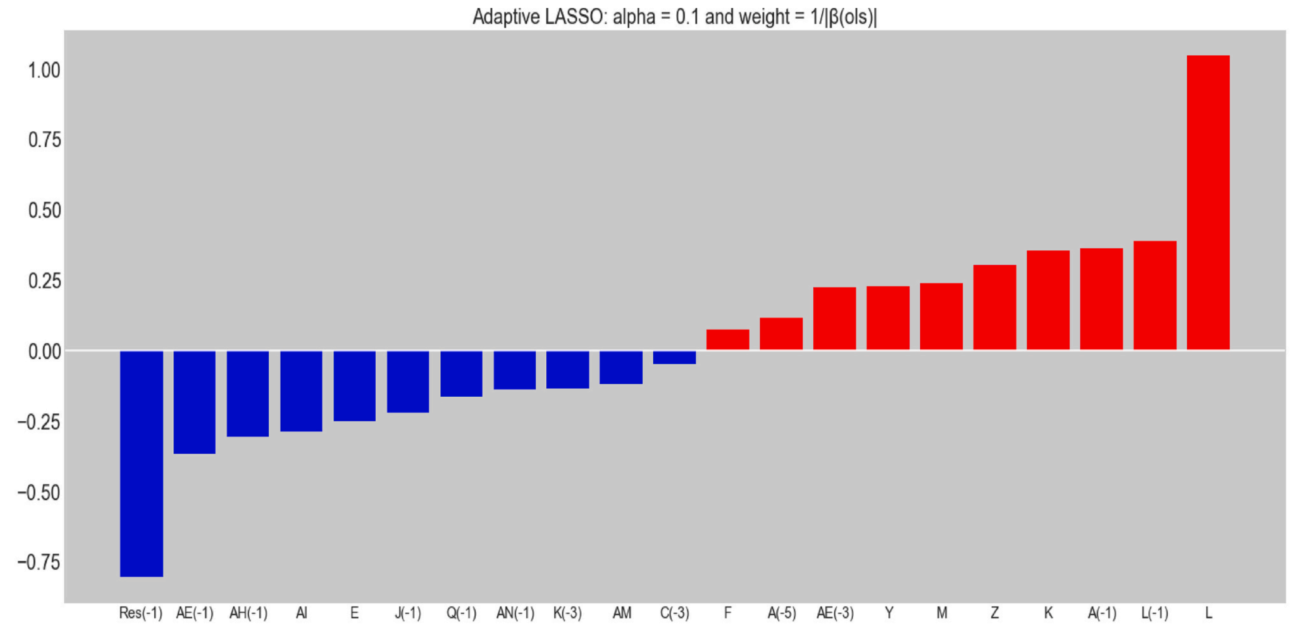


21 Coefficients selected by LASSO

Res(-1)	AE(-1)	AI	AH(-1)	E	Q(-1)	AN(-1)	AM	K(-3)	J(-1)	C(-3)	F	A(-5)	AE(-3)	M	Y	A(-1)	L(-1)	Z	K	L
- 0.7424	- 0.3733	- 0.2958	-0.2738	- 0.1889	- 0.1395	- 0.0925	- 0.0467	- 0.0414	- 0.0314	- 0.0286	0.0492	0.0498	0.0906	0.143	0.1487	0.1742	0.2363	0.2614	0.2905	0.9852

Fig. 9. LASSO results ($\alpha = 0.1$) in LASSO-ECM2: Selected comprehensive short-run variables
Notes: Table 2 shows notations denoting relevant factors. These coefficients are rounded to four decimal places. Fig. 9 displays only variables with non-zero coefficients.

G. ALASSO results in LASSO-ECM2



21 Coefficients selected by ALASSO

Res(-1)	AE(-1)	AH(-1)	AI	E	J(-1)	Q(-1)	AN(-1)	K(-3)	AM	C(-3)	F	A(-5)	AE(-3)	Y	M	Z	K	A(-1)	L(-1)	L
-0.8028	-0.3673	-0.3060	-0.2874	-0.2492	-0.2209	-0.1664	-0.1396	-0.1339	-0.1217	-0.0489	0.0731	0.1146	0.2230	0.2288	0.2383	0.3039	0.3565	0.3627	0.3891	1.0455

Fig. 10. ALASSO results ($\alpha = 0.1$ and $\widehat{w}_j = 1/|\widehat{\beta}_j(ols)|$) in LASSO-ECM2: Selected comprehensive short-run variables
Notes: Table 2 shows notations denoting relevant factors. These coefficients are rounded to four decimal places. Fig. 10 displays only variables with non-zero coefficients.

Data availability

I have shared the link to my data/code at the Attach File step.

Identifying key natural gas drivers towards China's carbon neutrality through a novel temporal shrinkage framework (Original data) (Mendeley Data)

References

Adow, A.H., Mahmood, H., Furqan, M., 2025. Exploring the effects of energy consumption, financial market development, and urbanization on CO2 emissions in GCC countries: Cross-Sectional dependence analysis. Int. J. Energy Econ. Pol. 15 (3), 146.

Albon, C., 2018. Machine Learning with Python Cookbook: Practical Solutions from Preprocessing to Deep Learning. O'Reilly Media, Incorporated, Sebastopol.

Bierens, H.J., Martins, L.F., 2010. Time-varying cointegration. Econom. Theory 26 (5), 1453–1490.

Bu, Y., et al., 2020. Spatial pattern and driving factors for interprovincial natural gas consumption in China: based on SNA and LMDI. J. Clean. Prod. 263, 121392.

Campos, J., Ericsson, N.R., Hendry, D.F., 2005. General-to-specific modeling: an overview and selected bibliography. Int.Finance Discussion Papers 2005 (838), 1–94.

Chen, Z., et al., 2021. Demand for storage and import of natural gas in China until 2060: simulation with a dynamic model. Sustainability 13 (15), 8674.

Cushing, D.S.C., 2024. China off track on emissions goals as energy demand offsets renewables push. Researchers Say.

Duras, T., et al., 2023. Using machine learning to select variables in data envelopment analysis: simulations and application using electricity distribution data. Energy Econ. 120, 106621.

Engle, R.F., Granger, C.W.J., 1987. Co-Integration and error correction: representation, estimation, and testing. Econometrica 55 (2), 251–276.

Fang, K., et al., 2019. Will China peak its energy-related carbon emissions by 2030? Lessons from 30 Chinese provinces. Appl. Energy 255, 113852.

Fu, Y., et al., 2024. Short-term impacts vs. long-term contributions: the role of clean energy and ESG investments in China. Renew. Energy 233, 121131.

Gao, J., Dong, X., 2018. Stimulating factors of urban gas consumption in China. Nat. Gas. Ind. 38 (3), 130–137.

Ghoddusi, H., Creamer, G.G., Rafizadeh, N., 2019. Machine learning in energy economics and finance: a review. Energy Econ. 81, 709–727.

Green, F., et al., 2024. No new fossil fuel projects: the norm we need. Sci.(American Association for the Advancement of Science) 384 (6699), 954–957.

Géron, A.l., 2022. Hands-on Machine Learning with Scikit-Learn, Keras and Tensorflow : Concepts, Tools, and Techniques to Build Intelligent Systems, third ed. O'Reilly, Sebastopol.

Gürsan, C., de Gooyert, V., 2021. The systemic impact of a transition fuel: does natural gas help or hinder the energy transition? Renew. Sustain. Energy Rev. 138, 110552.

Gu, Y., et al., 2016. Overall review of peak shaving for coal-fired power units in China. Renew. Sustain. Energy Rev. 54, 723–731.

Gyomai, G. and E. Guidetti, International Seminar on Early Warning and Business Cycle Indicators 14 to 16 December 2009 Scheveningen, The Netherlands.

Hepburn, C., et al., 2021. Towards carbon neutrality and China's 14th five-year Plan: clean energy transition, sustainable urban development, and investment priorities. Environ. Sci. Ecotechnol. 8, 100130.

Howarth, R.W., 2024. The greenhouse gas footprint of liquefied natural gas (LNG) exported from the United States. Energy Sci. Eng. 12 (11), 4843–4859.

IEA, 2023a. *Denmark* (Energy Mix). https://www.iea.org/countries/denmark/energy-mix.

IEA, 2023b. Denmark 2023. IEA, Paris. Available from: https://www.iea.org/reports/denmark-2023.

Interesse, G., 2024. China Unveils Its First Set of Basic Standards for Corporate Sustainability (ESG) Disclosure.

Jia, W., et al., 2023. Potential changes of regional natural gas market in China amidst liberalization: a mixed complementarity equilibrium simulation in 2030. Energy 284, 129254.

Jia, X., et al., 2022. Multi-objective energy planning for China's dual carbon goals. Sustain. Prod. Consum. 34, 552–564.

Jiang, H., et al., 2020. What drives China's natural gas consumption? Analysis of national and regional estimates. Energy Econ. 87, 104744.

Kock, A.B., 2016. Consistent and conservative model selection with the adaptive LASSO IN stationary and nonstationary AUTOREGRESSIONS. Econom. Theory 32 (1), 243–259.

Kondratieff, N.D., 1925. The static and the dynamic view of economics. Q. J. Econ. 39 (4), 575–583.

Koop, G., Leon-Gonzalez, R., Strachan, R.W., 2011. Bayesian inference in a time varying cointegration model. J. Econom. 165 (2), 210–220.

Kumar, P.K., Sarojamma, B., 2017. Time Series Cross-Validation Techniques for determining the order of the Autoregressive models. Int. J. Adv. Res. Comput. Sci. 8 (3).

Lee, S., Lee, M., 2023. The Israel-Hamas War, oil price volatility, and anticipated impacts: implications for Korean industries. Korea Institute for Industrial Economics and Trade Research Paper No, vol. 23.

Li, J., et al., 2022. China's flexibility challenge in achieving carbon neutrality by 2060. Renew. Sustain. Energy Rev. 158, 112112.

Li, Z.-G., Cheng, H., Gu, T.-Y., 2019. Research on dynamic relationship between natural gas consumption and economic growth in China. Struct. Change Econ. Dynam. 49, 334–339.

Liang, C., Schienle, M., 2019. Determination of vector error correction models in high dimensions. J. Econom. 208 (2), 418–441.

Liao, Z., Phillips, P.C.B., 2015. Automated estimation of vector error correction models. Econom. Theory 31 (3), 581–646.

Liu, G., et al., 2018. Natural gas consumption of urban households in China and corresponding influencing factors. Energy Policy 122, 17–26.

Liu, Z., et al., 2022. Challenges and opportunities for carbon neutrality in China. Nat. Rev. Earth Environ. 3 (2), 141–155.

Lu, X., et al., 2025. Government regulation and China's natural gas price distortion: a sectoral perspective. Nat. Resour. Forum.

Maxwell, D., Zhu, Z., 2011. Natural gas prices, LNG transport costs, and the dynamics of LNG imports. Energy Econ. 33 (2), 217–226.

Mbah, R.E., Wasum, F., 2022. Russian-Ukraine 2022 War: a review of the economic impact of Russian-Ukraine crisis on the USA, UK, Canada, and Europe. Adv. Soc. Sci. Res. J. 9, 144–153.

Meidan, M., 2020. *Unpacking China's 2060 Carbon Neutrality Pledge.* Oxford Energy Comment. The Oxford Institute For Energy Studies.

Mendes, E.F., 2011. Model selection consistency for cointegrating regressions. arXiv: Methodology.

Müller, A.C., Guido, S., 2016. Introduction to Machine Learning with Python: a Guide for Data Scientists. O'Reilly Media, Incorporated, Sebastopol.

Mu, X., Li, G., 2018. Study on natural gas consumption forecast and influence factors in China based on system dynamics model. J.Eng. Stud. 10 (1), 56–67.

Myllyvirta, Lauri, Qi, Qin, Shen, Xinyi, Qiu, Chengcheng, Schäpe, Belinda, Xunpeng Shi, M.Y., Zhang, Ying, 2024. China's Climate Transition: Outlook 2024. Centre for Research on Energy and Clean Air (CREA). Available from: https://energyandcleanair.org/wp/wp-content/uploads/2024/11/Report_-CREA_Chinas-Climate-Transition-Outlook-2024_EN-1.pdf.

Narayan, P.K., Smyth, R., 2006. What determines MIGRATION FLOWS from low-income to high-income countries? AN empirical investigation of Fiji-U.S. Migration 1972-2001. Contemp. Econ. Policy 24 (2), 332–342.

OECD. OECD Composite Leading Indicators dataset. https://data-explorer.oecd.org/vis?fs[0]=Topic%2C1%7CEconomy%23ECO%23%7CLeading%20indicators%23ECO_LEA%23&pg=0&fc=Topic&bp=true&snb=1&vw=tb&df[ds]=dsDisseminateFinalDMZ&df[id]=DSD_STES%40DF_CLI&df[ag]=OECD.SDD.STES&df[vs]=&dq=.M.LI … AA … H&lom=LASTNPERIODS&lo=10&to[TIME_PERIOD]=false.

OECD, 1998. OECD Composite Leading Indicators: a Tool for Short-Term Analysis.

Ogden, J., et al., 2018. Natural gas as a bridge to hydrogen transportation fuel: insights from the literature. Energy Policy 115, 317–329.

Paltsev, S., Zhang, D., 2015. Natural gas pricing reform in China: getting closer to a market system? Energy Policy 86, 43–56.

Patel, A., 2024. Experts: what to Expect in China's Climate Pledge for 2035.

Puttachai, W., Phadkantha, R., Yamaka, W., 2022. The threshold effects of ESG performance on the energy transitions: a country-level data. Energy Rep. 8, 234–241.

Pye, S., et al., 2025. The global implications of a Russian gas pivot to Asia. Nat. Commun. 16 (1), 386.

Qin, Y., et al., 2018. Challenges of using natural gas as a carbon mitigation option in China. Energy Policy 117, 457–462.

Ramírez-Orellana, A., et al., 2023. Is the corporate financial strategy in the oil and gas sector affected by ESG dimensions? Resour. Policy 81, 103303.

Ranstam, J., Cook, J.A., 2018. LASSO regression. Br. J. Surg. 105 (10), pp. 1348–1348.

Reed, W.R., 2015. On the practice of lagging variables to avoid simultaneity. Oxf. Bull. Econ. Stat. 77 (6), 897–905.

Santillán Vera, M., et al., 2023. Drivers of electricity GHG emissions and the role of natural gas in mexican energy transition. Energy Policy 173, 113316.

Shen, J., et al., 2021. Future CO2 emission trends and radical decarbonization path of iron and steel industry in China. J. Clean. Prod. 326, 129354.

Shi, X., et al., 2020. Prioritizing driving factors of household carbon emissions: an application of the LASSO model with survey data. Energy Econ. 92, 104942.

Shi, X., Hari Malamakkavu Padinjare, V., Jacqueline, T., 2017a. Global impact of uncertainties in China's gas market. Energy Policy 104, 382–394.

Shi, X., Sun, S., 2017b. Energy price, regulatory price distortion and economic growth: a case study of China. Energy Econ. 63, 261–271.

Shortreed, S.M., Ertefaie, A., 2017. Outcome-adaptive lasso: variable selection for causal inference. Biometrics 73 (4), 1111–1122.

Shrestha, M.B., Bhatta, G.R., 2018. Selecting appropriate methodological framework for time series data analysis. J. Financ. Data Sci. 4 (2), 71–89.

Stern, N., Valero, A., 2021. Innovation, growth and the transition to net-zero emissions. Res. Pol. 50 (9), 104293.

Stern, N., Xie, C., 2022. China's new growth story: linking the 14th five-year Plan with the 2060 carbon neutrality pledge. J. Chin. Econ. Bus. Stud. 1–21. **ahead-of-print** (ahead-of-print).

Stock, J.H., Watson, M.W., 2019. Introduction to Econometrics. the Pearson Series in Economics. Pearson Education.

Tu, K.J., Johnson-Reiser, S., 2012. Understanding China's Rising Coal Imports. Policy Outlook. Carnegie Endowment for International Peace, p. 16.

Tuyu Zhou, H.X., Gosens, Jorrit, Jotzo, Frank, 2022. China's net-zero plans. In: Policy Brief on near-term Policy Challenges & Australia-China Links in Decarbonization. Crawford School of Public Policy. https://iceds.anu.edu.au/files/ANU%20policy%20brief%20-%20Chinese%20climate%20and%20energy%20policy%20trends_0.pdf.

Unruh, G.C., 2000. Understanding carbon lock-in. Energy Policy 28 (12), 817–830.

Villar, J.A., Joutz, F.L., 2006. The relationship between crude oil and natural gas prices. In: Energy Information Administration, Office of Oil and Gas, 1, pp. 1–43.

Wang, J., Li, N., 2020b. Influencing factors and future trends of natural gas demand in the eastern, central and western areas of China based on the grey model. Nat. Gas Ind. B 7 (5), 473–483.

Wang, T., et al., 2020a. Market reforms and determinants of import natural gas prices in China. Energy 196, 117105.

Wang, X., et al., 2020c. Does environmental regulation affect natural gas consumption? Evidence from China with spatial insights. Sustainability 12 (8), 3354.

Wang, Y., et al., 2023. Decarbonization pathways of China's iron and steel industry toward carbon neutrality. Resour. Conserv. Recycl. 194, 106994.

Wei, Q., Zhou, P., Shi, X., 2023. The congestion cost of pipeline networks under third-party access in China's natural gas market. Energy 284, 128521.

Wesseh, P.K., Lin, B., 2019. Environmental policy and ‘double dividend’ in a transitional economy. Energy Policy 134, 110947.

Xu, H., Wang, Z., 2021. Analysis of driving factors of natural gas consumption in China: based on regional integration and LMDI method. Energy Eng. J. Assoc. Energy Eng.: J. Assoc.Energy Eng. 118 (4), 1027–1043.

Yang, X., et al., 2020. How does China's green institutional environment affect renewable energy investments? The nonlinear perspective. Sci. Total Environ. 727, 138689.

Yu, E.S.H., Jin, J.C., 1992. Cointegration tests of energy consumption, income, and employment. Resour. Energy 14 (3), 259–266.

Yu, X., Tan, C., 2022. China's pathway to carbon neutrality for the iron and steel industry. Glob. Environ. Change 76, 102574.

Zhang, C., et al., 2019. Analysis of electricity consumption in China (1990–2016) using index decomposition and decoupling approach. J. Clean. Prod. 209, 224–235.

Zhang, X., et al., 2016. Carbon emissions in China: how far can new efforts bend the curve? Energy Econ. 54, 388–395.

Zhou, Z.Z., et al., 2025. From pollution to sustainable environment: unlocking the role of energy intensity and renewable energy in highly pollutant nations. Environ. Res. Commun.

Zivot, E., Wang, J., 2006. Modeling Financial Time Series with S-PLUS®, second ed. Springer New York, New York, NY. 2006. ed.

Zou, H., 2006. The adaptive lasso and its oracle properties. J. Am. Stat. Assoc. 101 (476), 1418–1429.