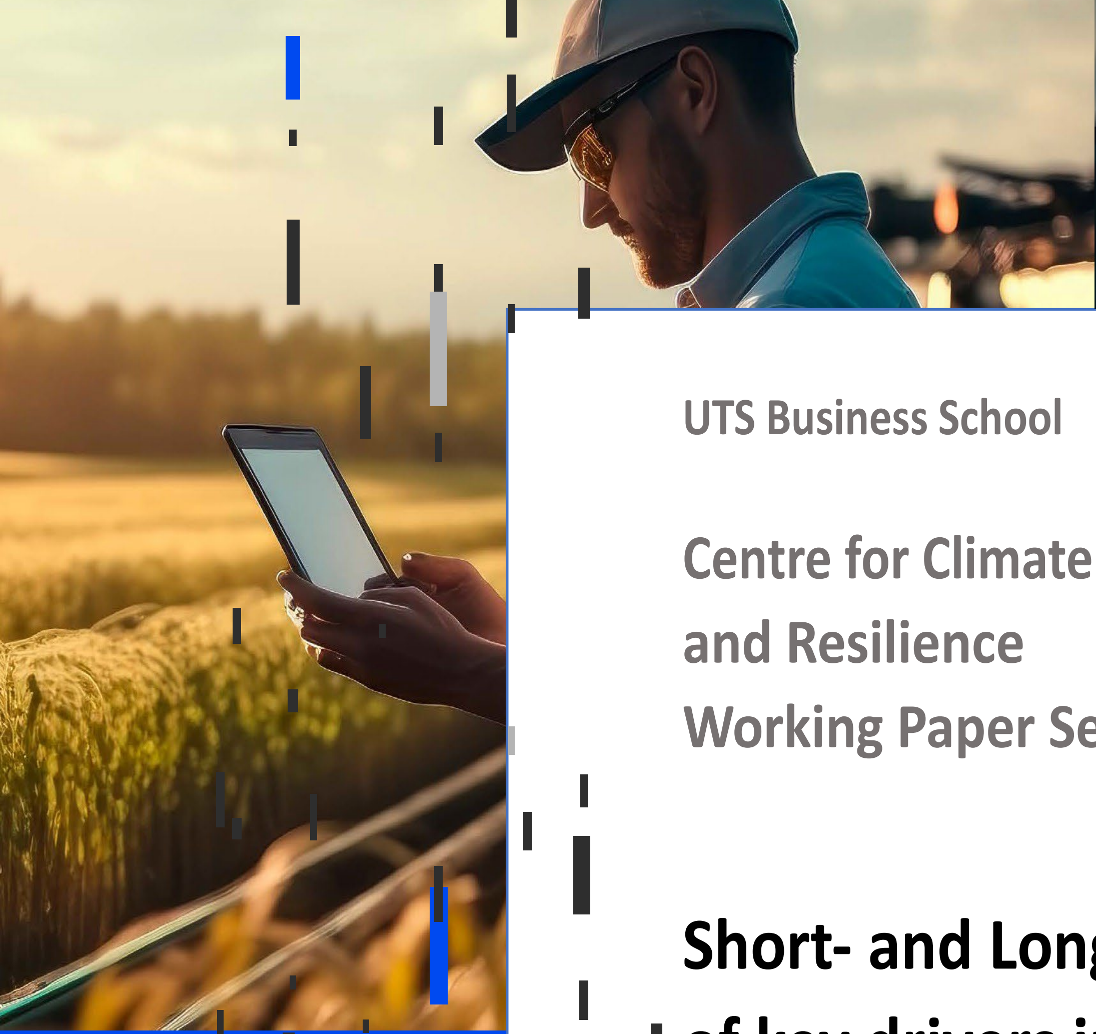


UTS Business School

Centre for Climate Risk
and Resilience
Working Paper Series

Short- and Long-term effects of key drivers in China's Natural Gas market

By Jinyong (Edward) Sung , Xunpeng (Roc) Shi, Sven Teske, Mengheng Li

Working Paper No. 2024/2

Series Editor: Martina Linnenluecke,
CCRR Director

Connect with CCRR
Web https://www.uts.edu.au/about/uts-business-school/research/centre-climate-risk-and-resilience
Email ccrr@uts.edu.au

Short- and Long-term effects of key drivers in China's Natural Gas market

Jinyong (Edward) Sung [a], Xunpeng (Roc) Shi [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

Abstract: Natural gas (NG) faces a dilemma in China between short-term development as a transitional fuel and long-term being phase-out, and the dynamics will significantly impact both global climate efforts and the world gas markets. China's carbon neutrality initiative, its NG market liberalisation, and the geopolitical developments, such as Russia's invasion of Ukraine and the Israel-Hamas war, further add uncertainty to the gas dynamics. This study uses a novel combination of machine learning and econometric techniques to identify the key driving factors influencing NG consumption in China and to analyse their short- and long-term impact. Through a long-run model that sequentially combines *Least Absolute Shrinkage and Selection Operator (LASSO)/Adaptive LASSO (ALASSO)* and *Error Correction Model (ECM)* techniques, we identify the key driving factors – temperature, thermal coal and HH gas prices, piped NG and LNG imports, NG market liberalisation, and NG infrastructure-related factors – and analyse their long-term effects. Subsequently, in a short-run model, utilising a combined methodology that simultaneously incorporates both ECM and LASSO/ALASSO, we initially investigate the preliminary short-term effects of these key driving factors and further explore more comprehensive short-run dynamics by identifying additional driving factors, such as ESG, global oil prices, coking coal prices, piped NG import prices, and economic indicators. Our findings provide valuable insights for policymakers aiming to develop effective energy policies towards carbon neutrality, considering short- and long-run effects.

Keywords: LASSO, ALASSO, ECM, China's Natural Gas Market, Carbon Neutrality, Energy Policy

1. Introduction

Natural gas (NG) as a transitional fuel faces a dilemma between the short and long run in China, and its dynamics will significantly impact both global climate efforts and the world NG markets. In September 2020, China announced to peak carbon emissions by 2030 and attain carbon neutrality by 2060. Owing to China's coal-dominant energy mix, NG has been consistently boosted in the past decade to reduce total carbon emissions and improve air quality. As part of the short-term goal towards carbon peaking, NG is gaining increasing significance as a transitional fuel within the 'coal-to-gas' transition strategy of the country's 14th Five-Year Plan (Hepburn et al., 2021; Stern & Xie, 2022). Furthermore, the Chinese government has promoted NG market liberalisation for a decade, not only to foster domestic NG consumption but also to influence both domestic and international NG production (T. Wang et al., 2020; Wei et al., 2023). However, since NG is a fossil fuel, China is expected to significantly phase down its NG consumption in the long run to achieve its net zero emissions goal (Meidan, 2020). Therefore, Chinese NG consumption will face a dilemma, and further studies of the future dynamics are necessary for China to fulfil its carbon-neutral commitments and for the global gas markets to prepare for uncertain futures.

The emerging dilemma and changing short- and long-term NG market conditions, compounded by various geopolitical factors, is a critical challenge for Chinese policymakers and industry players. The dynamics of the Chinese NG market are profoundly intricate, characterised by an extensive array of influencing factors that collectively shape its trajectory. Furthermore, while NG has emerged as a highly significant commodity in the global energy market due to Russia's invasion of Ukraine (Mbah & Wasum, 2022), China's NG market has been confronted with unprecedented challenges, including the global energy crisis caused by Russia's invasion of Ukraine and the sudden escalation in Middle Eastern geopolitical risk triggered by the Israel-Hamas war, amid the country's journey towards carbon neutrality and NG market reforms. To navigate through the dilemma, it is crucial to obtain a deeper understanding of key driving factors in the NG market from the short- and long-run perspectives. Given the country's prominent role in the global NG markets and LNG trade, deepening the comprehension of its NG market not only is important for China and the global climate efforts, but also affects the global NG market landscape.

Current studies lack a consistent view of these key driving factors, considering both short- and long-term effects. Various and numerous factors have been chosen as influencing factors in the literature, encompassing economic trends, energy policies, environmental considerations, technological advancements, and global trade dynamics. See Xu and Wang (2021), Bu et al. (2020), Jiang et al. (2020), and Wang and Li (2020) for recent examples. The sheer diversity and interrelatedness of these factors create a complex web of interactions and pose a formidable challenge to understanding the critical or relative importance of each in driving China's NG market. Furthermore, efforts to project China's NG demand often encounter significant hurdles due to numerous driving factors, and some of these factors may lack reliable data for accurate projection, yielding forecasts that can be misleading or imprecise. This inherent complexity underscores the need for a more sophisticated and holistic approach to analysing the Chinese NG market, one that can predict the future with minimal factors, integrate advanced modelling techniques, harness big data analytics, and account for both short-term fluctuations and long-term trends.

This research endeavours to employ a range of econometric, machine learning, and big data approaches for analysing drivers of China's NG market and their short- and long-term effects. The research first employs shrinkage methods, machine learning techniques of the *Least Absolute Shrinkage and Selection Operator (LASSO)* and *Adaptive LASSO (ALASSO),* to identify these key drivers. Subsequently, short- and long-run effects of the key drivers and additional short-term drivers are analysed using an econometric tool, the *Error Correction Model (ECM)*. By combining LASSO/ALASSO and ECM techniques, we identify key drivers (temperature, thermal coal and HH gas prices, piped NG and LNG imports, NG market liberalisation, and NG infrastructure-related factors), analyse their long-term effects, and examine short-run dynamics with additional factors (ESG, global oil prices, coking coal prices, piped NG import prices, and economic indicators). Our findings provide insights for policymakers to derive effective energy policies for carbon neutrality, considering both short- and long-run effects.

Our primary contribution to the literature is to introduce a novel framework that amalgamates machine learning and econometric methodologies into new energy policy analytics within the energy economics and policy field. Specifically, this research delves into the combined utilisation of LASSO-type variable selection and ECM techniques, considering the presence of time series variables. This innovative amalgamation can ultimately provide policymakers with selected short- and long-run factors for evidence-based policies balancing the short- and long-term effects. Secondly, to the best of our knowledge, this is the first time the ALASSO's extended capability has been used in the energy economics and policy field. Another marginal contribution is to introduce that, compared with existing LASSO studies in the energy economics and policy field, such as Shi et al. (2020), ALASSO has an extended capability to accurately select cointegrated variables, which signifies these variables' long-run relationship (Mendes, 2011), as well as its previously recognised capability to overcome LASSO's limitations, such as bias and inconsistency (Zou, 2006). Our final marginal contribution is that a long-run model, within the novel framework combining LASSO and ECM techniques, can explore the long-run relationship of key drivers without encountering a spurious regression problem or losing their inherent long-run relationship. Such an approach is imperative not only for enhancing our comprehension of this pivotal market but also for producing more accurate and reliable projections that can inform effective energy policies and investment strategies, both in China and across the global energy landscape.

The rest of this paper is organised as follows: Section 2 provides related literature pertaining to the influencing factors of China's NG market, as well as LASSO- and ECM-based approaches and applications. Sections 3 and 4 discuss data and methods, respectively. Section 5 presents the result, and the final section concludes the study. The supplementary materials include terms and abbreviations, tables, and figures used throughout this paper.

2. Literature review

In the dynamic context of China's NG market, comprehending the intricate web of factors that shape NG demand and the methodologies used to dissect them becomes crucial. This literature review focuses on two fundamental dimensions: the influencing factors of China's NG demand and the methods employed to uncover these driving factors and assess their short- and long-term consequences. Through this concise assessment, we can identify potential drivers for our forthcoming big data analysis. Additionally, the methodology review underlines the merits and advantages of the proposed research approach.

2.1 Influencing factors of China's NG demand

The research on influencing factors of Chinese NG demand has been conducted using various methods until recently. In existing literature, most influencing factors have been identified as a mixture of economic, energy, urbanisation, and demographic characteristics. Xu and Wang (2021) use the decomposition technique, Logarithmic Mean Divisia Index (LMDI), to analyse China's NG consumption based on economic growth, energy intensity, energy structure, and substitution effects. They find that as economic development increases, the energy substitution effect grows while the economic growth effect decreases. Bu et al. (2020) employ Social Network Analysis (SNA) 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) suggest that economic, urbanisation, and energy intensity factors are the major driving factors for NG demand in the long run. Wang and Li (2020) propose a grey model to examine the specific leading factors based on the main potential influencing factors on NG demand in various regions of China. Concluding that economic development (GDP), industry structure, environmental policy's qualitative index, urbanisation rate, population density, energy consumption intensity and energy consumption structure are the influencing factors, they indicate 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.

Besides, energy-related infrastructure, technological progress, environmental policy, and climate factors are additionally chosen as the main influencing factors in a few studies, along with the combination of economic, energy, urbanisation, and demographic factors. Gao and Dong (2018) use the decomposed technique LMDI to identify economic growth, urban spatial expansion, pipeline network density, pipeline length, urbanisation rate, population density, and NG substitution as influencing factors of NG demand in China. Liu et al. (2018) employ the General Least Squares Method to analyse the NG consumption of urban residents in China based on various influencing factors: NG price, household income, length of pipe, household size, energy substitution, temperature, and NG consumption intensity. Mu and Li (2018) develop a dynamic model of China's NG and demand system based on economic development, urbanisation, population growth, gas technological advancements, and environmental protection policy factors. They conclude that, in comparison to population growth, economic development, the impact of urbanisation, and gas technological advancements are the main influencing factors on China's NG consumption. In brief, 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 length of pipelines; energy-related technological progress; temperature; and environmental policy.

In addition to the influencing factors mentioned above, the distortion of NG prices due to Chinese government control has affected NG consumption by creating inefficiencies and distortions in the NG market. As a result, the country's NG market reforms are expected to resolve these inefficiencies and distortions. This NG market reform eventually affects the country's NG consumption. Shi et al. (2017) discuss one of the uncertainties in the Chinese gas market, stating that the 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 of China's regulatory price distortion indicates that energy price distortion negatively affects economic growth due to the inefficient use of and misallocation of energy caused by price distortion (Shi & Sun, 2017). Paltsev and Zhang (2015) propose that China's efforts to transition towards a market-based pricing system for NG address the demand and supply imbalance caused by price distortions.

ESG (Environmental, Social, and Governance) factors, which have recently attracted much attention, are also considered influencing factors of NG demand in China. Recent research on ESG factors highlights their increasing importance in the oil and gas sector by positively impacting the financial performance of oil and gas companies (Ramírez-Orellana et al., 2023). X. Wang et al. (2020) indicate that environmental regulation, a constituent of the Environmental pillar (E) within the ESG framework categorised by Puttachai et al. (2022), plays a crucial role in driving NG consumption in China.

From the literature, it is apparent that regulatory price distortion and ESG factors are recognised as influencing factors of NG consumption in China. Therefore, considering the main influencing factors sorted out above, all the potential influencing factors on NG consumption in China are classified by economic growth, energy indicators, urbanisation, demographics, infrastructure, environmental policy, climate, technological growth, regulatory price distortion, and ESG factors. Because of the various influencing factors, it is challenging to clearly understand the key drivers of China's NG market, and the analysis of the short- and long-term effects of these drivers is also limited. To address this, there is a growing need for new analytical approaches.

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

As for the new analytical approaches, three theoretical models have been explored: two LASSO-type variable selection models (LASSO and ALASSO) and one ECM. LASSO-type shrinkage and variable selection techniques have gained significant popularity in the machine learning and econometric literature. In the realm of time series analysis, ECM techniques have been widely used to separate the short- and long-run relationships among variables. Within the context of this research, the marriage of these techniques helps disentangle the short- and long-run effects of key drivers on China's NG consumption.

Recent advancements in energy economics and finance have witnessed a surge in machine learning analyses, particularly for big data-driven models. Regularisation techniques, with a focus on LASSO that shrinks near-zero effects to zero, have garnered substantial attention for their ability to control overfitting and identify crucial features in models (Albon, 2018). This growing interest in LASSO's distinctive capabilities has propelled its adoption in energy econometric inference. Notably, studies by Ghoddusi et al. (2019) highlight the effectiveness of machine learning analyses, including LASSO, in various aspects of energy research, such as classification, prediction, and policy analysis. Furthermore, in a specific application, Shi et al. (2020) successfully utilise LASSO to identify key driving factors affecting China's household carbon emissions (HCEs), thereby providing valuable insights for crafting low-carbon policies. This increased utilisation of LASSO underscores its importance in advancing energy economics and policy research, offering data-driven decision-making support and enhancing the understanding of intricate energy systems.

However, LASSO exhibits limitations, such as bias and inconsistency, in high-dimensional settings where the number of variables exceeds the sample size (Zou, 2006). To overcome these issues, ALASSO is introduced (Zou, 2006). Under certain conditions, ALASSO, as proposed by Zou, achieves the oracle property of model selection consistency by penalising different coefficients in the regularisation term of LASSO. As an extension of the research on ALASSO, Mendes (2011) utilises ALASSO as a model selection method to choose the most appropriate cointegrating regression with the accurate subset of variables and demonstrates its oracle property. Besides, Kock (2016) suggests that ALASSO exhibits oracle efficiency for consistent and conservative model selection in time series models, encompassing stationary and nonstationary autoregression. Recently, Duras et al. (2023) introduce ALASSO as a new variable selection method in the energy economics field. In various practical scenarios, they compare its variable selection performance with other methods, including LASSO.

The time series variables considered in the study include market, climate, and policy variables (see Section 3.1). Some of these variables exhibit temporal trends, indicating the presence of long-run or cointegration relationships. While long-run relationships between the variables are essential, the variables also influence short-term fluctuations in China's NG market driven by demand. To identify both long-term trends and short-term variations in the NG market, it is necessary to combine LASSO/ALASSO variable selection technologies with the ECM introduced by Engle and Granger (1987) to effectively disentangle the short-run effects from long-run ones; see also Shrestha and Bhatta (2018). This combination can offer a comprehensive perspective on the growth and fluctuations in China's NG market.

In a few studies, LASSO-type techniques for consistent and efficient model selection have been applied to ECMs; however, these studies have yet to focus on short- and long-run analysis. For example, Liao and Phillips (2015) use ALASSO within a vector ECM (VECM) to simultaneously determine the cointegration relationships and autoregressive order. Similarly, Liang and Schienle (2019) explore the cointegration rank and autoregressive lag order determination in high-dimensional settings using the LASSO-VECM approach. These studies have primarily centred around combined techniques involving LASSO with ECM rather than explicitly addressing the short- and long-run effects of variables.

Based on the above literature, a multitude of diverse influencing factors within China's NG market have been identified. However, the presence of these numerous diverse influencing factors often obscures a clear understanding of the critical or relative significance within the Chinese NG market. Such opacity can subsequently impede the development of coherent and efficacious policies. Furthermore, knowledge and analytical methods regarding the short- and long-term effects of the influencing factors are limited. This is significant when taking specific and effective policy measures towards achieving carbon neutrality. To fill these gaps, this study employs a novel framework that combines LASSO-type variable selection methods (LASSO/ALASSO) and ECM techniques to identify key drivers among influencing factors in China's NG market, as noted in prior studies, and to analyse short- and long-run effects of these key drivers.

3. Data

3.1 Data selection: 40 influencing factors

This study is conducted employing a set of 40 influencing factors as independent variables and considering NG consumption as the dependent variable. These 40 influencing factors are selected based on the literature review in the previous section. The 40 factors are categorised into six groups: (1) Economics & Energy (15 factors); (2) Global Energy Index (5 factors); (3) Urbanisation & Demographics (4 factors); (4) Infrastructure (4 factors); (5) Environment, Climate, and Technological growth (7 factors); (6) Price distortion (5 factors). All the factors consist of time series data on a monthly basis, spanning from July 2013 to December 2019 due to the availability of price distortion factors. Year or quarterly data has been converted into monthly data through interpolation. Economic factors are represented on a constant or real basis. Table 1 (see Appendix B) provides a detailed description of these factors along with their respective sources.

3.2 Data preparation

This study undertakes an analysis of both non-stationary and stationary time series variables within the context of examining the short- and long-term relationships among selected factors. A stationary time series maintains a constant mean and variance over time, while a non-stationary time series exhibits changing statistical properties. Two popular unit root tests, namely the Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests, are employed to identify whether a time series variable is stationary or not. For the factors U, AE, AF, and AG in Table 2 (see Appendix C), the existence of a unit root is further investigated using the Phillips-Perron test due to technical errors in the KPSS test. See Zivot and Wang (2006) for further information regarding these testing methods.

The regression analysis on non-stationary time series data relies on the utilisation of raw data. As per Shrestha and Bhatta (2018), non-stationary variables are converted into stationary ones using log transformation or differencing. If this is still insufficient, these variables are transformed by taking the logarithm and then the first or second difference. The "Stationary status" presented in Table 2 (see Appendix C) shows the transformation of variables into stationary ones – via log transformation or differencing – alongside the statistical significance assessment for each variable's stationarity.

Furthermore, we created lagged stationary variables from stationary independent and dependent variables. This process involved selecting the optimal lag order for each stationary variable based on the Vector Autoregression (VAR) model with the Akaike Information Criterion (AIC) (Stock & Watson, 2019). Table 2 shows that the total number of the lagged stationary variables is 117. These variables are subsequently used as inputs for LASSO/ALASSO modelling in conjunction with stationary time series analysis. The schematic outlining the preparation of time series data is visually depicted in Fig. 1.

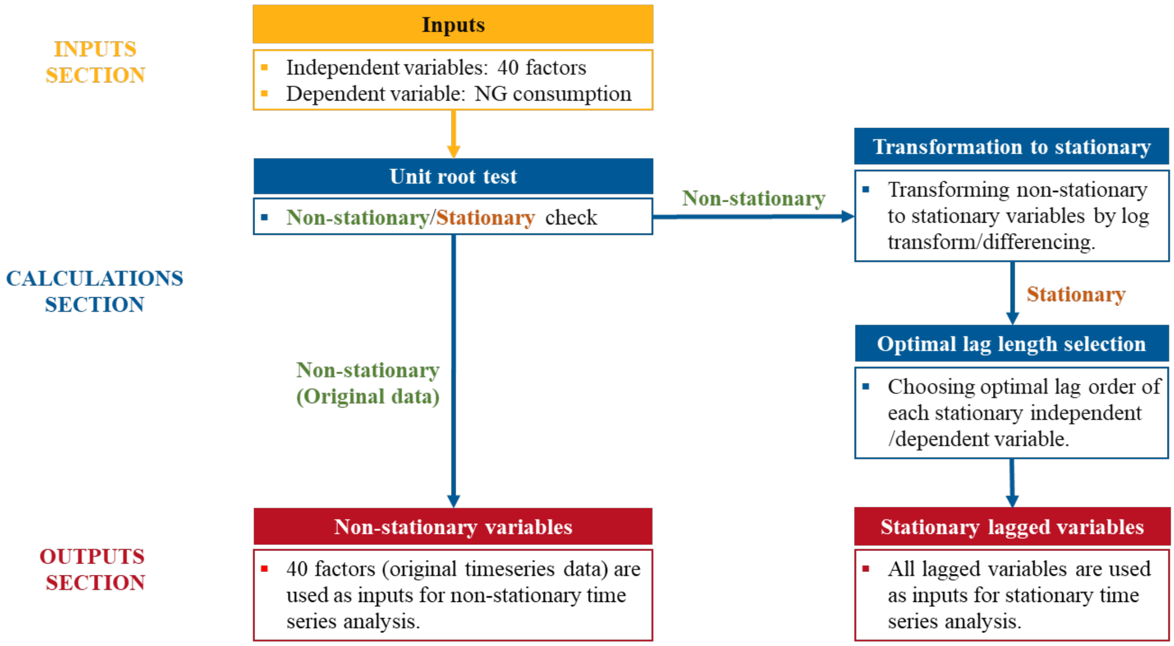


Fig. 1 Data preparation

4. Methods

4.1 Long-run model

This study utilises regression techniques that sequentially combine LASSO/ALASSO and ECM with original (non-stationary) time series datasets of 40 factors to identify the long-term effects of the key driving factors. Here, we face a dilemma regarding the use of non-stationary data. In terms of the general methodological framework for time series analysis, the regression model using non-stationary time series can cause a spurious regression, where linear regression shows a misleading relationship between variables (Shrestha & Bhatta, 2018). Therefore, for regression analysis, preference is given to time series datasets that are either inherently stationary or transformed to stationarity. On the other hand, the transformation from non-stationarity to stationarity introduces an additional challenge – the potential loss of the long-run relationship. Shrestha and Bhatta (2018) elucidate that when stationary time series datasets are modified to attain stationarity, they might forfeit their inherent long-run relationship or informative content.

We propose a long-run model using the combined techniques (LASSO/ALASSO and ECM) to explore the long-run relationship of key drivers without encountering a spurious regression problem or losing their inherent long-run relationship. As shown in Fig. 2, the long-run model progresses through the following stages. First, LASSO identifies the key driving factors from the original time series datasets of 40 factors without converting them into stationary series. Secondly, a residual-based cointegration test, which is the 1st step of the two-step procedure in Engle and Granger (1987), estimates the cointegration regression (long-run relationship) among these key driving factors, and then ALASSO verifies the robustness of these cointegrated factors. The ALASSO represents a cointegrated regression method – employed for estimating the long-run relationship – that incorporates the variable selection function. While individual non-stationary series might show spurious corrections or divergent behaviour, the cointegration tests help to determine if these series move together in the long term, suggesting a stable relationship even if they exhibit short-term fluctuations (Yu & Jin, 1992). Finally, ECM techniques (referred to as LASSO-ECM1 hereafter) are employed to quantify their long-run relationship. Further elaboration on each approach is provided in the subsequent sections.

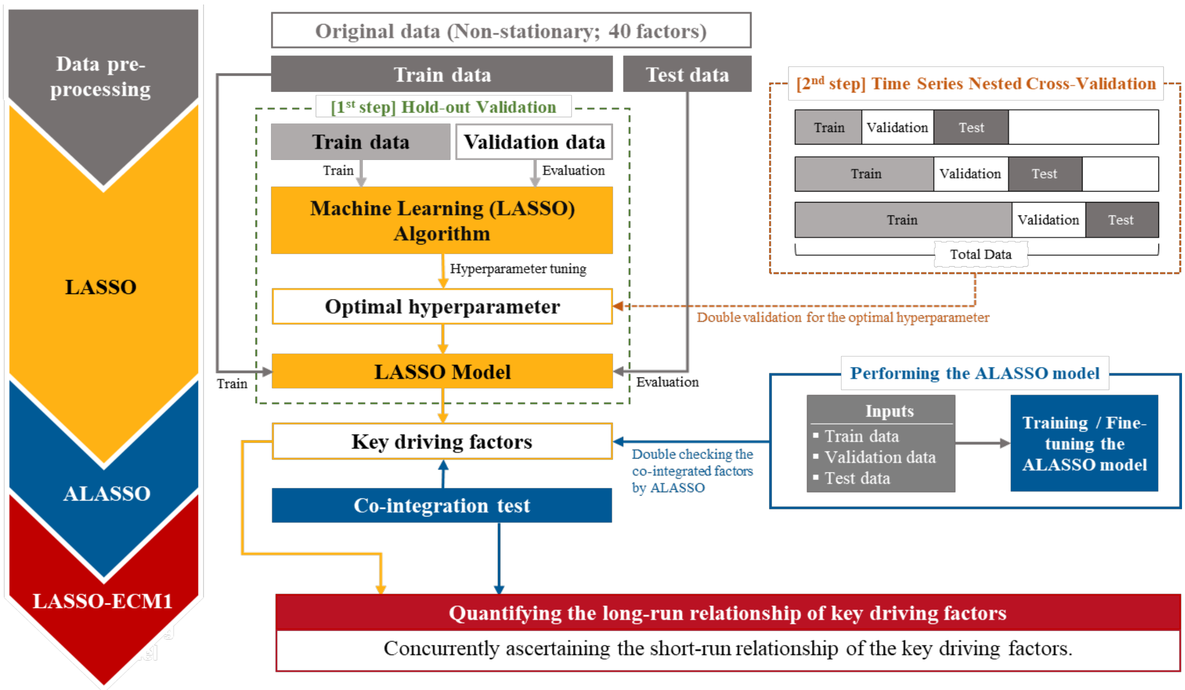


Fig. 2 Long-run model

4.1.1 LASSO and ALASSO

LASSO, including the advantages of machine learning, is a regression model that performs both feature selection and regularisation. LASSO is mathematically expressed as follows:

$$\boldsymbol{Minimize}_{\beta_0,\beta}\left(\frac{1}{2n}\sum_{i=1}^{n}(y_i-\beta_0-x_i^T\beta)^2+\alpha\sum_{j=1}^{m}|\beta_j|\right) \quad (1)$$

where n is the total number of observations; y_i is the dependent variable; m is the total number of independent variables $x_i = (x_{i1}, \dots, x_{im})^T$; β_0 is the intercept; β_j are the other parameters (coefficients); and α is the hyperparameter value.[1]

As seen in equation (1), LASSO applies a penalty to the cost function of the linear regression model. The penalty is the sum of the absolute values of the hyperparameter value α

[1] In machine learning algorithms, hyperparameter values typically control the learning process and assist in determining the appropriate model parameters (coefficients).

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 the penalty. Specifically, each coefficient in the penalty is reduced by increasing the hyperparameter value α. It can be observed that some coefficients are reduced enough to become zero. In this research, these unique properties of LASSO can lead to relatively choosing more essential variables, while eliminating less important variables by setting coefficients of less important variables to zero (Géron, 2022, pp. 158 - 161).

LASSO can inadvertently lead to overfitting by readily eliminating features with minor coefficients (Zou, 2006). However, ALASSO effectively tackles this concern by penalising larger coefficients, thereby facilitating the identification of the accurate set of true variables. This phenomenon is acknowledged as the “oracle properties” of ALASSO. These oracle properties empower ALASSO to correctly identify cointegrated variables, indicative of their long-run relationship (Mendes, 2011).

ALASSO represents an extension of the LASSO methodology using adaptive weights (Shortreed & Ertefaie, 2017). The mathematical formulation for ALASSO is presented as follows:

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

where $\hat{w}_j$ represents adaptive weights, calculated reciprocally from $|\hat{\beta}_j|$ values derived from LASSO; and the remaining parameters include n, y_i, $\hat{y}_i\,(= \beta_0 - x_i^T\beta)$, m, β_j, and α, mirroring the parameters found in LASSO. As illustrated 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 by Python, an open-source programming language.

4.1.2 LASSO-ECM1

The Error Correction Model (ECM), which constitutes the 2nd step of the Engle-Granger two-step procedure after a residual-based cointegration test (Engle & Granger, 1987), is an econometric approach aimed at estimating both short- and long-run effects of one-time series variable on another (Shrestha & Bhatta, 2018). The fundamental version of ECM is expressed mathematically as follows:

$$\boldsymbol{\Delta Y_t = \beta_0 + \beta_1\,\Delta X_t - \gamma\hat{\mu}_{t-1} + e_t} \quad \textbf{(3)}$$

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 residual – referred to the error correction term (ECT) – that is estimated through the linear regression model 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, when the ECT coefficient (γ) in equation (3) is negative and statistically significant, it indicates that deviations from the long-run equilibrium are subsequently corrected towards the equilibrium. The adjustment effect, as governed by the ECT coefficient (γ), is specifically observed in the two scenarios (Narayan & Smyth, 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.

Based on the ECT coefficient (γ), we can estimate the long-run relationship of the key drivers selected by LASSO. This analytical approach, carried out using EViews, a statistical software, is referred to as "LASSO-ECM1" in this study. Simultaneously, LASSO-ECM1 also estimates the short-run relationship among the key drivers. Further elaboration on the identification of short-run variables is provided in the next section.

4.2 Short-run model

The short-run model, comprising LASSO-ECM1 and LASSO-ECM2, utilises a combined methodology that simultaneously incorporates both ECM and LASSO/ALASSO to discern short-run variables. As depicted in Fig. 3, the short-run model commences with LASSO-ECM1, serving as a preliminary test, and then proceeds to LASSO-ECM2 for a more advanced test. On the one hand, LASSO-ECM1 yields relatively straightforward short-run effects of the key drivers chosen within the long-run model. This is because it relies on the fundamental ECM framework, as expressed by equation (3), without incorporating lagged variables. On the other hand, LASSO-ECM2 aims to uncover more comprehensive short-run dynamics by incorporating lagged stationary variables. In other words, the short-run model ultimately seeks to identify a more comprehensive set of short-run variables. Further elaboration on LASSO-ECM2 is provided in the subsequent sections.

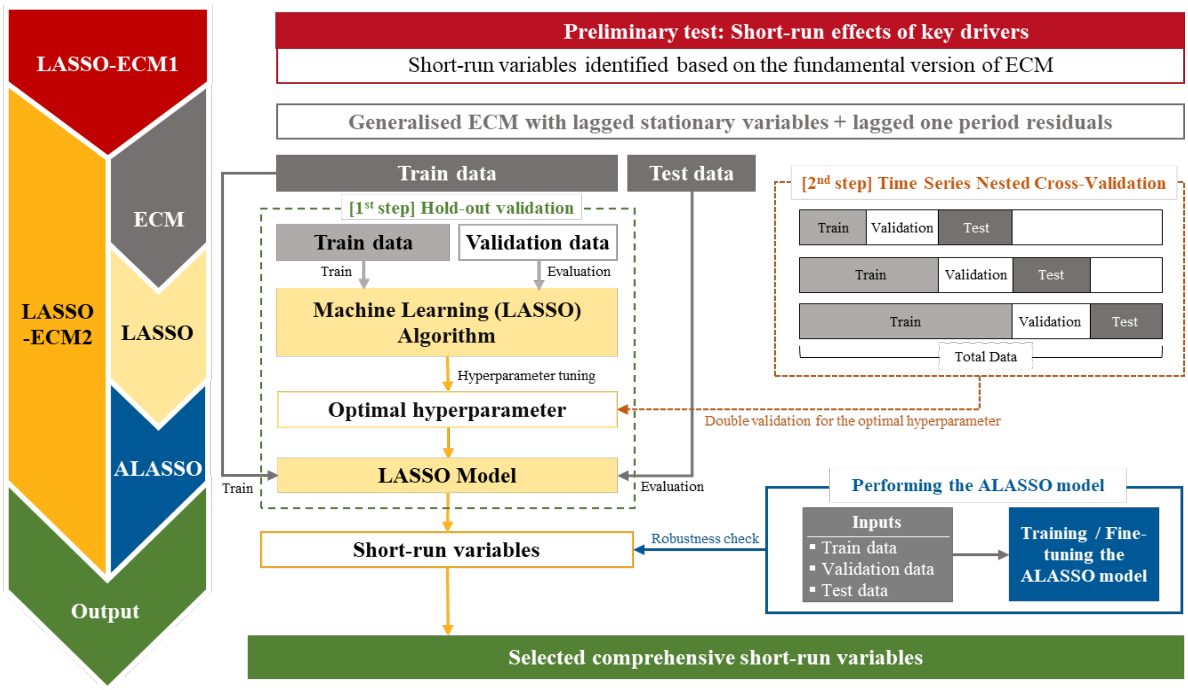


Fig. 3 Short-run model

4.2.1 LASSO-ECM2

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:

$$\Delta \boldsymbol{Y}_t = \boldsymbol{\beta}_0 + \sum_{i=0}^{n} \boldsymbol{\beta}_i \Delta \boldsymbol{X}_{t-i} + \sum_{j=1}^{m} \boldsymbol{\alpha}_j \Delta \boldsymbol{Y}_{t-j} - \boldsymbol{\gamma} \hat{\boldsymbol{\mu}}_{t-1} + \boldsymbol{e}_t \qquad \textbf{(4)}$$

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 linear regression model 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 on the system. The incorporation of lagged variables enables numerous economic researchers to evaluate the influence of past values on the current state of the system, thereby enhancing the understanding of its dynamics (Kondratieff, 1925). Therefore, to identify more comprehensive short-run variables, we initially formulate a generalised ECM with lagged stationary variables. Following this, in accordance with the general to simple (GETS) approach[2] suggested by Hendry (1985), we apply LASSO to simplify the model through data-driven variable elimination, eliminating less essential independent variables. Finally, ALASSO provides a robustness check for the LASSO outputs using its oracle properties. In summary, this analytical approach enables us to delve into more comprehensive short-run dynamics by introducing additional variables. In this study, we utilise EViews for this analytical approach and refer to it as "LASSO-ECM2".

5. Result

5.1 Long-run model results

5.1.1 Identifying key driving factors with LASSO

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, we split original time-series data into three sets – 60% as a training set, 20% as a validation set, and 20% as a test set – with a fixed random seed of 279. This seed, referred to as the "random_state" parameter in machine learning, represents the initial value used when shuffling data randomly and ensures consistent output when the same machine learning function is run multiple times (Müller & Guido, 2016, p. 18). For this long-run model, the same random_state parameter is constantly used. All input sets are standardised and scaled to unit variance.

[2] The GETS 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).

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 – the process of determining an optimal hyperparameter (α). The hyperparameter tuning for the LASSO model is processed in two steps:

- **Step 1**: We identify an optimal hyperparameter using a hold-out validation technique introduced by Géron (2022, pp. 34 - 35). This involves holding out a portion of the training set as a validation set, which is then used to assess several candidate models. These models are trained with different hyperparameters on the reduced training set (sub-train set). In this study, we select an optimal hyperparameter based on a graphical analysis of model evaluation metrics on the validation set: mean absolute error (MAE), root mean square error (RMSE), and R-squared (R^2). As shown in Fig. 4, the optimal hyperparameter chosen by considering the highest R^2, the lowest MAE, and RMSE is 0.1.

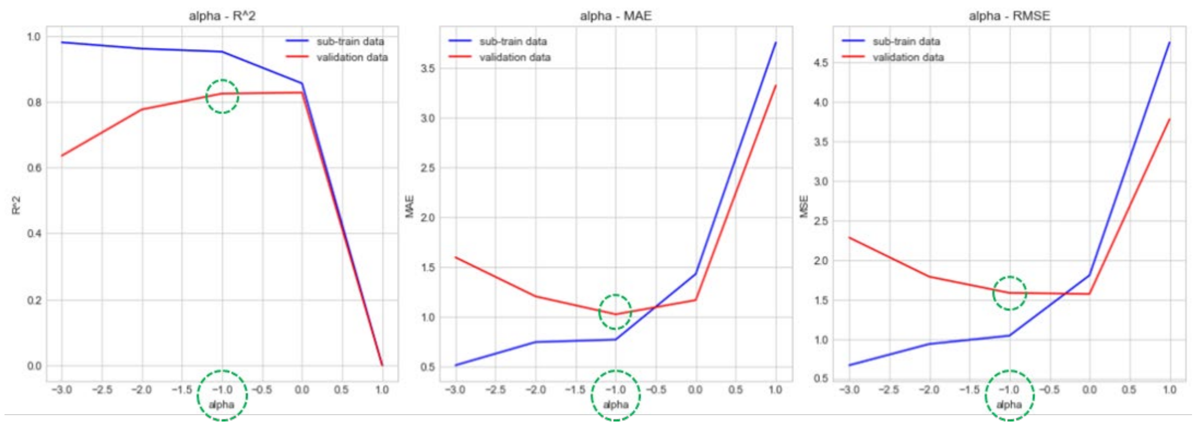


Fig. 4 The optimal hyperparameter (α) by hold-out validation

Notes: A base-10 log scale is utilised for the Y-axis ("α" values) of all the above graphs. The optimal α value is 0.1.

- **Step 2**: Following Kumar and Sarojamma (2017) for time series analysis, the optimal hyperparameter value (α) identified in the first step is validated using the Time Series Nested Cross-Validation (TSNCV). 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. The optimal hyperparameter is calculated based on RMSE. The optimal hyperparameter calculated through TSNCV also yields the same value, 0.1. The process of the TSNCV is illustrated in Fig. 5.

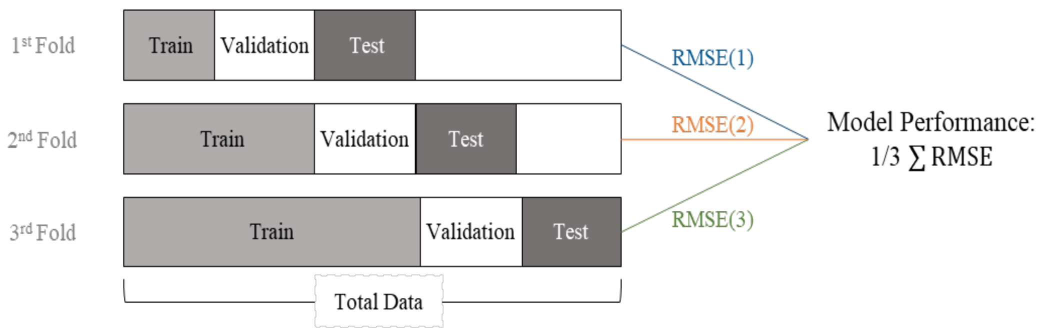
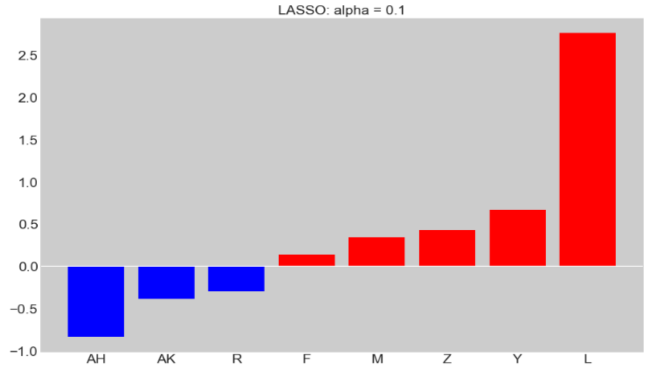


Fig. 5 The process of Time Series Nested Cross-Validation (TSNCV)

In the third phase, final performance and evaluation, the best model with the optimal hyperparameter, selected through hold-out validation and TSNCV, is trained on the full training set (including the validation set), using the previously fixed random seed. This model is considered the finalised LASSO model. Consequently, we employ the finalised LASSO model to evaluate the performance on 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 eight key driving factors, five factors exhibit a positive relationship with NG consumption: domestic thermal coal price, NG imports via pipelines, gas pipeline capacity, LNG terminal capacity, and LNG imports. Conversely, the remaining three factors have a negative relationship with NG consumption: temperature, price distortion, and HH gas price. The key driving factors and their relationships are illustrated in Fig. 6: blue and red bars signify negative and positive relationships, respectively.

Fig. 6 LASSO results (α = 0.1): Key driving factors



#		Negative/Positive relationship
		Negative relationship
1	AH	Average national temperature (°C, Celsius)
2	AK	Price distortion (City gate gas price / Piped gas price, %)
3	R	HH gas price (USD/MMBTU)
		Positive relationship
4	F	Domestic thermal coal price: Datong (5500 kcal/kg; USD/MT)
5	M	NG imports via pipelines (Metric tonne)
6	Z	Gas pipeline capacity (BCM)
7	Y	Accumulated LNG terminal capa. per year (10KT)
8	L	LNG imports (Metric tonne)

Notes: Fig. 6 displays only variables with non-zero coefficients.

5.1.2 Estimating the long-run (cointegration) relationship of the key driving factors

ALASSO is adopted to estimate further the long-run relationship between these key drivers and NG demand. ALASSO validates the robustness of these drivers by assessing the cointegration regression (long-run relationship) among them. ALASSO progresses through the following phases: (1) the cointegration test, (2) data pre-processing tailored for ALASSO, and (3) training and evaluating phases in ALASSO.

First of all, the cointegration test is executed based on the 1st step of the Engle-Granger two-step procedure (Engle & Granger, 1987) in EViews. This cointegration test determines the existence or absence of a long-run relationship between the key driving factors (selected by LASSO). For the cointegration test, a linear regression model (LRM) is constructed utilising the key driving factors through the ordinary least squares (OLS) method. Subsequently, the residuals of LRM are calculated, and two widely-used unit root tests, namely ADF and KPSS, are applied to these residuals, the independent variables (key driving factors), and the dependent variables (NG consumption). This assessment is carried out to determine whether these variables exhibit stationarity. As per Engle and Granger (1987), the results shown in Tables 3 and 4, where the residuals demonstrate I(0) and all variables manifest I(1), indicate that the key driving factors are cointegrated.

Secondly, in the data pre-processing phase of ALASSO, the original time-series data is divided into an 80% training dataset and a 20% test dataset, utilising the same fixed random seed employed in LASSO. All input sets are standardised and scaled to unit variance.

Table 3 LRM and Unit root tests 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 or I(0).
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 or I(0).

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 acronyms (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 or I(1). Based on the significance level of 0.1 and the p-value of ADF, an independent (L) variable is stationary at the first difference or I(1).

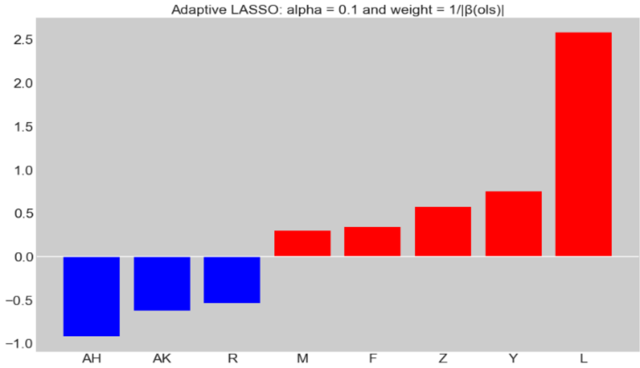
Finally, in the training and evaluating phases of ALASSO, as seen in equation (2), we implement adaptive weights ($\hat{w}_j$) within a LASSO model to construct an ALASSO model. As suggested by Zou (2006), we first solve a LASSO model. Subsequently, we compute the coefficients (selected by LASSO) using the OLS method, denoted as $\hat{\beta}_j(ols)$. Next, we calculate the weights as $\hat{w}_j = 1/|\hat{\beta}_j(ols)|$.[3] 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 shown in Fig. 7, the finalised ALASSO model attains the 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 remain consistent with LASSO. Since both LASSO and ALASSO models are designed to select more important factors, ALASSO results validate LASSO findings using

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

original time series datasets, without the need for transformation from non-stationarity to stationarity. Additionally, this implies that the key driving factors are in the long-run relationship, as supported by the cointegration test results.

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



#		Negative/Positive relationship
		Negative relationship
1	AH	Average national temperature (°C, Celsius)
2	AK	Price distortion (City gate gas price / Piped gas price, %)
3	R	HH gas price (USD/MMBTU)
		Positive relationship
4	M	NG imports via pipelines (Metric tonne)
5	F	Domestic thermal coal price: Datong (5500 kcal/kg; USD/MT)
6	Z	Gas pipeline capacity (BCM)
7	Y	Accumulated LNG terminal capa. per year (10KT)
8	L	LNG imports (Metric tonne)

Table 5 The 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 acronyms (F, L, M, R, Y, Z, AH, and AK) denoting factors and displays only variables with non-zero coefficients.

5.1.3 Quantifying the long-run relationship: LASSO-ECM1

Representing a fundamental version of ECM encompassing 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, LASSO-ECM1, carried out using EViews, reaffirms the long-run relationship of the selected key driving factors, as indicated by ALASSO.[4] As depicted in Table 6, the lagged Error Correction Term (ECT) in LASSO-ECM1 appears with a coefficient (γ) of -1.1646. This coefficient is negative, and its associated p-value is statistically significant at the 1 % level. According to Narayan and Smyth (2006), a γ value between -2 and -1 implies that the ECT converges towards the long-run equilibrium in a dampening manner. Consequently, we can deduce that the key driving factors have long-term effects on China's NG consumption.

Furthermore, the ECT coefficient (γ) enables LASSO-ECM1 to manifest short-term effects among the key driving factors. A more comprehensive exploration on this subject is presented in the subsequent section concerning the results of the short-run model.

5.2 Short-run model results

LASSO-ECM1 indicates that the selected key driving factors have short-term effects on China's NG consumption, serving as a preliminary test of the short-run model. However, as shown in Table 6, temperature, domestic thermal coal price, and piped NG imports among these selected factors are not statistically significant at the 10% level. Hence, these factors may be excluded from short-run variables in the LASSO-ECM1 outcomes. This could be attributed to the inherent nature of a fundamental ECM version that does not incorporate lagged variables.

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

In the event that 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) may exhibit short-term effects. Therefore, for a more comprehensive investigation of potential short-run variables, this study performs LASSO modelling with stationary lagged variables based on the generalised ECM, denoted as LASSO-ECM2. LASSO-ECM2, executed by Python, employs a combined approach that simultaneously applies both ECM and LASSO/ALASSO to identify a more comprehensive set of short-run variables.

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 = 1.88×10^{-6}). The series "(-1)" refers to the series lagged by one period. The lagged Error Correction Term (ECT) represents lagged one-period residuals.

5.2.1 Determining a more comprehensive set of short-run variables

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

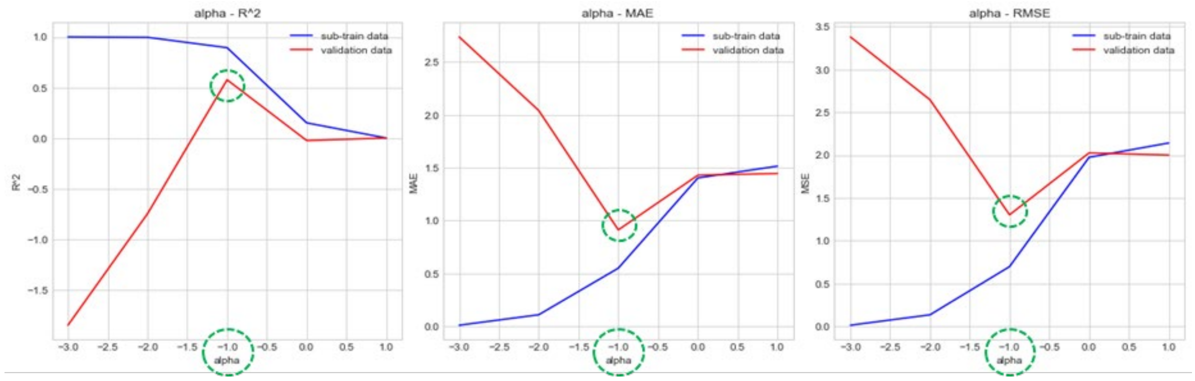


Fig. 8 The optimal hyperparameter (α) by hold-out validation

Notes: A base-10 log scale is utilised for the Y-axis ("α" values) of all the above graphs. The optimal α value is 0.1.

In the initial step, as detailed in Section 3.2 Data preparation, 117 lagged stationary variables are created from independent and dependent variables. The specifics of all the lagged stationary variables are shown in Table 2 (see Appendix C). Lagged one-period residuals are derived from the residuals of LRM, constructed using the selected key driving factors. These residuals exhibit stationarity, as indicated in Table 4. In total, 118 lagged stationary variables are utilised in the data pre-processing phase of the LASSO modelling – comprising 117 lagged stationary variables and one variable representing lagged one-period residuals. These 118 variables are structured according to the generalised ECM, as depicted in equation (4).

In the data pre-processing as a second step, the 118 variables are partitioned into training, validation, and test sets, mirroring the process used for LASSO in the long-run model.[5]

In the training, tuning, and evaluation as a third step, the training set is used to fit a LASSO model, while 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 model in the long-run model, hyperparameter tuning is performed based on hold-out validation and TSNCV. As shown in Fig. 8, the optimal hyperparameter is selected using overall model evaluation metrics that reflect the highest R^2, the lowest MAE, and RMSE scores on the validation set.

In the last modelling step of LASSO-ECM2, we use the full training set (including the validation set) to train the model with the fixed random seed used previously and the optimal hyperparameter selected through hold-out validation and TSNCV. This results in the finalised LASSO model. The finalised LASSO model selects a more comprehensive set of short-run variables from all the lagged variables by evaluating the model performance on the test set. Subsequently, ALASSO is employed to verify the robustness of these selected short-run variables, replicating the methodology employed for ALASSO in the long-run model.

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 (10KT)
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.

As shown in Table 7 and Fig. 9 (see Appendix D), the finalised LASSO model selects 21 comprehensive short-run variables from a pool of 118 lagged stationary variables with the high accuracy (R^2) of 87.96% on a train set and 83.47% on a test set. As depicted in Fig. 10 (see Appendix E), ALASSO also select the same short-run variables. Like the long-run model, while the coefficients of these variables in ALASSO are different from those in LASSO

[5] In order to produce consistent output, the pseudorandom number generator shuffling data is provided with a fixed random seed of 200.

because of adaptive weights, the (positive/negative) relationships between the variables and NG consumption in ALASSO align with those observed in LASSO.

The selected 21 variables are structured according to the generalised ECM and are also considered to have short-run informative content. As shown in Table 7, the lagged one-period residuals, RES(-1), are indicated with a negative value, thus confirming the alignment of the 21 variables with the generalised ECM structure. Besides, as time series variables transition from non-stationarity to stationarity, they may lose their inherent long-run relationship or information (Shrestha & Bhatta, 2018). Hence, we assert that the selected 21 variables carry short-run information. In other words, the selected 21 variables, forming the generalised ECM structure and containing short-run information, are identified as a more comprehensive set of short-run variables.

Through this more comprehensive estimation of short-term effects, the key drivers that do not exhibit statistical significance in the preliminary test of the short-run model can be identified as short-run variables. For example, the preliminary test does not demonstrate short-term effects between domestic thermal coal price and NG demand because, as shown in Table 6, the coefficient of the thermal coal price is not statistically significant. However, a more comprehensive estimation of short-term effects, as presented in Table 7, includes the thermal coal price as a short-run variable. A more in-depth discussion of these short-run dynamics is presented in the following section.

5.3 Further discussion of the short-run dynamics

The short-run model (LASSO-ECM2) captures more comprehensive short-run dynamic effects or relationships in the variables. The LASSO model of LASSO-ECM2, compared with the LASSO model of the long-run model, additionally selects ESG, coking coal import price, WTI oil price, piped NG import price, LNG average import price, trend restored CLI, and real GDP, apart from the key driving factors selected by the long-run model. Among the selected short-run variables, domestic thermal coal price, real GDP, environmental score, NG imports via pipelines, LNG terminal capacity, gas pipeline capacity, NG import price via pipeline, and LNG imports have a positive relationship with NG consumption. On the other hand, 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. In the positive relationship, the environment score lags by three periods, while the NG import price via pipelines has no lag. In the negative relationship, the environment score lags by one period, while the NG import price via pipelines lags by three periods. Further exploration provides a deeper understanding of these short-run dynamics as follows:

<u>*ESG factors*</u>

The short-run model selects two ESG factors – Environmental score and Country score for climate change – as additions to the long-run variables. A lagged three-period Environmental score shows a positive relationship with NG consumption, while a lagged one-period Environmental score and the Country score for climate change at zero lag exhibit negative relationships. These relationships are explained as follows:

- *The negative relationship between Country score for Climate change and NG consumption*: NG is a fossil fuel that generates carbon dioxide emissions. Although its carbon dioxide emission intensity is lower than that of coal, it still contributes to climate change. Therefore, an increase in the use of NG is negatively related to a reduction in the Country score for climate change.
- *The positive/negative relationship between Environmental score and NG consumption*: The positive relationship observed between NG consumption and the Environmental score lagged by three periods could be attributed to underlying factors or dynamics. For instance, there could be a delayed effect of changes in environmental practices or policies on NG consumption. On the other hand, the negative relationship between NG consumption and the Environmental score lagged by one period might indicate a more immediate impact. Theodori (2009) raises concerns about the environmental impact caused by NG development and production, which can explain the negative relationship for the Environmental score lagged by one period. Furthermore, since both the Environmental score and the Country score for climate change are considered environmental factors within ESG factors, these two scores exhibit the same (negative) relationship with NG consumption.

Chinese coking coal and domestic thermal coal price

The Chinese coking coal import price, as an additional short-run variable, exhibits a negative relationship with NG consumption. In the Chinese steelmaking industry, coking coal is typically used as a specialised fuel, and NG is not typically used as a substitute for coking coal (Wang et al., 2023).[6] In general, an increase in coking coal prices indicates higher steel prices, leading to a decrease in steel demand and, consequently, a reduction in steel production. In China, the steel industry is one of the high-electricity-consuming industries (Zhang et al., 2019). Therefore, a decrease in steel production means a reduction in electricity demand from steel mills, which in turn affects the overall demand for NG. This means less demand from gas-fired power plants.

In contrast, the domestic thermal coal price selected by the short-run model has a positive relationship with NG consumption, as similarly indicated by the long-run model. Typically, thermal coal and NG are common fuel sources for electricity generation in power plants. This positive relationship suggests that higher thermal coal prices could result in an increased demand for NG as a substitute for coal in electricity generation.

WTI oil price and HH gas price

WTI oil price, additionally chosen as a comprehensive short-run variable, shows a negative relationship with NG consumption. Likewise, HH gas price, demonstrating its short-term effects at the 10% significance level in the preliminary test of the short-run model, also displays the negative relationship. Villar and Joutz (2006) provide evidence of an interconnection between HH gas price and WTI oil price in both the short run and long run. Thus, we can infer a connection between the WTI oil price, chosen as an additional short-run variable, and the HH gas price, given its demonstrated short-term effects in the preliminary test. This suggests that

[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 & Tan, 2022); see also (Shen et al., 2021).

HH gas price exhibits characteristics of a short-run variable, even though, as shown in Table 7, it is not selected as a comprehensive short-run variable.

Other energy-related factors

The preliminary test in the short-run model uncovers the short-term effects of three energy-related factors - gas pipeline capacity, LNG terminal capacity, and LNG imports – at the 1% significance level. However, the short-run model identifies NG imports via pipelines, along with these three factors, as comprehensive short-run variables, all of which are integral components of the selected key driving factors. Besides, all four of these energy-related factors exhibit a negative relationship with NG consumption. This indicates that these four factors have short-term effects on NG consumption.

The short-run model additionally chooses two energy price factors: LNG average import price and NG import price via pipelines. LNG average import price lagged by one period exhibits a negative relationship with NG consumption. This negative relationship implies that changes in LNG prices in the previous period can influence NG consumer and industrial behaviour, leading to fluctuations in NG consumption in the current period. NG import price via pipelines, lagged by three periods, shows a negative relationship with NG consumption, while NG import price via pipelines at zero lag shows a positive relationship. This negative relationship indicates that changes in piped NG import prices take some time to influence NG consumption patterns. This lag can be attributed to factors such as term contracts, price adjustment mechanisms, inventory levels, economic decision-making processes, and seasonal variations in demand. On the other hand, this positive relationship might be attributed to the immediate impact of fluctuations in NG import price via pipelines on NG demand during the same period, such as increased use of NG as a peak-shaving tool in response to periods of high electricity demand when NG prices are elevated.

Economic factors

Two economic factors, trend restored Composite Leading Indicator (CLI) and real GDP, are selected as additional short-run variables, which are not identified in the preliminary test of the short-run model and long-run model. 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 allows for the possibility 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.[7]

In contrast, real GDP variables lagged by one period or five periods, representing a measure of China's economic output, have a positive relationship, suggesting that economic growth leads to higher NG consumption. These relationships suggest that, in the short term, changes in economic activity influence the demand for NG.

Temperature and price distortion

The short-run model chooses temperature and price distortion factors, similar to the long-run model. Among these factors, the preliminary test in the short-run model reveals only the short-term effects of a price distortion factor at the 10% significance level. However, as a more

[7] China's CLI components include the output of specific industries, such as chemicals, steel, construction, and motor vehicles, the diffusion index of 5000 Industrial Enterprises, and the Shanghai Stock Exchange turnover (OECD, 2022).

comprehensive set of short-run variables includes all these factors, we can conclude that these factors have short-term effects.

In a nutshell, a more comprehensive short-run dynamic analysis enables us to gain a detailed understanding of the dynamics of both the previously selected and additional short-run variables. Besides, statistically insignificant short-run variables in the preliminary test are demonstrated to exhibit short-run effects. Therefore, it can be observed that the selected key driving factors have both short- and long-run effects.

6. Conclusions

China's use of NG as a transitional fuel creates a dilemma. This dilemma arises from the opposing requirements for NG between periods of China's carbon peak and carbon neutrality. While China aims to expand its utilisation of NG as an alternative to coal, supported by a decade-long promotion of NG market reforms, in the short run to achieve its emissions peak, it will have to reduce its NG consumption to achieve net zero emissions in the long run. This dynamic has significant implications for both China and global climate efforts and the global NG markets. Understanding key drivers in China's NG market and their short- and long-term effects is essential for Chinese stakeholders in designing effective energy policies. In addition, given China's influential role in the global NG markets, a deeper understanding of these drivers is also vital for the global NG market dynamics. This understanding is even more critical in the context of the geopolitical developments, such as Russia's invasion of Ukraine and the Hamas-Israel conflict. The complexity of the task of identifying the drivers requires a more sophisticated approach that is absent in the existing literature.

This study's short- and long-term models aim to identify short- and long-term variables according to China's short-term strategy by 2030 and its long-term target by 2060 for achieving carbon neutrality. We integrate advanced modelling and big data analytics and propose a novel combination of LASSO/ALASSO and ECM techniques to find key drivers and assess their long-term effects through a long-run model. Subsequently, we demonstrate the short-run effects of these drivers and select additional short-run variables through a short-run model. Our main contributions include the introduction of this novel combination as new energy policy analytics in the energy economics and policy field, the initiation of ALASSO's extended capability for cointegrated variable selection in this field, and the analysis of long-run relationships among the drivers without encountering a spurious regression problem or disrupting their inherent long-run information.

The identified key drivers and their relationships with NG consumption provide not only crucial information for policy formulation but also valuable datasets for future research. In the long-run model, we first identify the key drivers – 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. We then find their long-run effects, which refer to the structural and lasting impacts of changes in these drivers on NG consumption. In the short-run model, we disclose that among the selected eight key drivers, three drivers (temperature, domestic thermal coal price, and piped NG imports), which the preliminary test of the short-run model may not confirm as having short-term effects, indeed exhibit such effects. Besides, we identify additional short-run variables, such as energy commodity prices, ESG, and economic factors. Hence, we conclude that the key drivers identified in the long-run model exert both short-term and long-term effects.

The results above offer policymakers insights into the key drivers, enabling the formulation of evidence-based policies that balance short- and long-term effects to achieve carbon neutrality. In this context, we propose four policy implications.

First, environmental costs should be internalised. Our findings indicate that thermal coal prices positively correlate with NG consumption in both the short and long term. In other words, adjusting the increase in domestic thermal coal prices in both the short and long term can promote the switch from coal to NG. This aligns with China's short-term carbon neutrality strategy, the 'coal-to-gas' transition strategy of the country's 14th Five-Year Plan (Hepburn et al., 2021; Stern & Xie, 2022). Therefore, the Chinese government can implement environmental pricing mechanisms, such as a carbon tax, in the short term to make thermal coal more expensive, resulting in increased NG consumption. This proposed policy can help achieve the country's short-term carbon neutrality strategy. From a long-term perspective (covering the period from China's 2030 carbon peak target to its 2060 carbon neutrality goal), the proposed policy should be expanded to include other fossil fuels. These fuels may become cheaper due to demand reductions along decarbonisation pathways. Therefore, to prevent their use as alternative fuels, the proposed policy needs to be sustained over the long term.

Second, price distortions in the NG market should be minimised. As observed, price distortions can significantly impact NG demand in both the short and long term. In the short run, these distortions in China's NG market enable end consumers to opt for cheaper energy sources, such as coal, over NG, leading to an imbalance of demand and supply in the NG market. Therefore, a market-based pricing system for NG can promote the adoption of NG instead of carbon-intensive and inexpensive energy sources. This aligns with the short-term carbon neutrality strategy of the coal-to-gas transition. In the long run, as China's NG prices align with international market values, the country's NG market is expected to foster collaboration with global gas markets and reflect environmental costs (Lu et al., 2024). This ultimately helps its NG market to meet the long-term global carbon neutrality goals, which seek to decrease the use of fossil fuels, including NG. Of course, this long-term policy should be implemented alongside the previously mentioned policy of internalising environmental costs.

Third, Environmental, Social, and Governance (ESG) criteria should be integrated into energy policy. Our study highlights the significance of ESG factors as a short-term variable influencing China's NG demand. Therefore, in alignment with the short-term carbon neutrality strategy, policymakers should consider integrating ESG criteria into energy-related decisions and regulations to promote sustainable and environmentally friendly energy practices. In the short run, this can help activate NG as a cleaner, affordable energy source during the transition decades to a low-emissions economy.

Lastly, infrastructure for NG imports should be enhanced. The infrastructure expanded to activate NG in alignment with the short-term carbon neutrality strategy can still be necessary for a long-term carbon-neutral future. According to our study, gas-related infrastructure is positively related to NG consumption in the short and long run. In line with short-term carbon neutrality strategies, the Chinese government can enhance the capability and flexibility of the NG supply chain by diversifying its supply sources through the development of new gas-related infrastructure, pipelines, and LNG terminals, which can result in increased NG demand. However, in the context of the country's long-term carbon neutrality target, the expanded NG infrastructure can be repurposed for hydrogen, which is emerging as an important energy source in a low-emission future (Ogden et al., 2018).

A potential extension of our ECM is to allow for time variation in the cointegration relationship following the methods of Bierens and Martins (2010) and Koop et al. (2011). The proposed ECM 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 will stay the same. We leave this to future research.

Another limitation of this study is that the short-run model (LASSO-ECM2), which has advantages for more comprehensive short-run dynamic analysis, has been restricted in investigating long-run effects. Underlying variables in LASSO-ECM2 may have a long-run relationship due to biased coefficients. LASSO artificially shrinks the coefficients of less important variables closer to zero, making the coefficients unable to accurately represent the true magnitude of the relationship between independent and dependent variables (Ranstam & Cook, 2018). For this reason, it is challenging to evaluate standard errors for biased coefficients through statistical tests or confidence intervals. This indicates that the coefficients estimated in LASSO-ECM2 might not be statistically significant. Moreover, LASSO-ECM2 is devised to understand more comprehensive short-run dynamics in China's NG market. Therefore, the ECM techniques to investigate the long-term effects are not employed in LASSO-ECM2. Additional research using an inference tool can offer a more in-depth exploration of the long-term effects.

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
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
GHG	Greenhouse gases
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 (1,000 tons)
LASSO	Least Absolute Shrinkage and Selection Operator
LNG	Liquefied Natural Gas
LRM	Linear Regression Model
MAE	Mean Absolute Error
MGMT	Management
MMBtu	Metric Million British Thermal Unit
MT	Metric Ton
NBS	National Bureau of Statistics of China
NG	Natural Gas
OECD	The Organization for Economic Cooperation and Development
OECD Stat	OECD Statistics; Database of OECD countries
PPP	Purchasing Power Party
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. Table 1 The description and relevant sources of the 40 factors

Factors	Description	Data source[8]
(1) Economics & Energy (15 factors)		
Real GDP by expenditure (CNY, Billions)	Quarterly → Monthly series	Bloomberg
CPI: Overall average	Monthly series	Refinitve; NBS
Trend restored CLI[9]	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[10]
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[11]	Refinitve
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 PPP)	Monthly series	World Bank Indicator
Industry structure[12]	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[13]
(3) Urbanisation & 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 (10KT)	Monthly series	KEEI
Gas pipeline capacity (BCM)	Monthly series	External source[14]
City public utilities: Length of gas pipelines (10KT)	Monthly series	NBS
Pipeline transportation: Length of gas pipelines (10KT)	Monthly series	CIEC[15]
(5) Environment, Climate, and Technological growth (7 factors)		
[ESG] Environmental score	Yearly → Monthly series	Refinitive
[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)[16]	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[17]
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	

[8] Bloomberg and Refinitive, 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.

[9] 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).

[10] Currently, China imports more coking coal than it domestically produces because coking coal, in contrast to thermal coal, is relatively scarce (Tu & Johnson-Reiser, 2012). Thus, the domestic coking coal price isn't factored into data collection, whereas both domestic and import prices of thermal coal are considered.

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

[12] 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 (2020).

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

[14] This reference is taken from Chen et al. (2021)'s research.

[15] CIEC is a global economic, industry, and financial data provider.

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

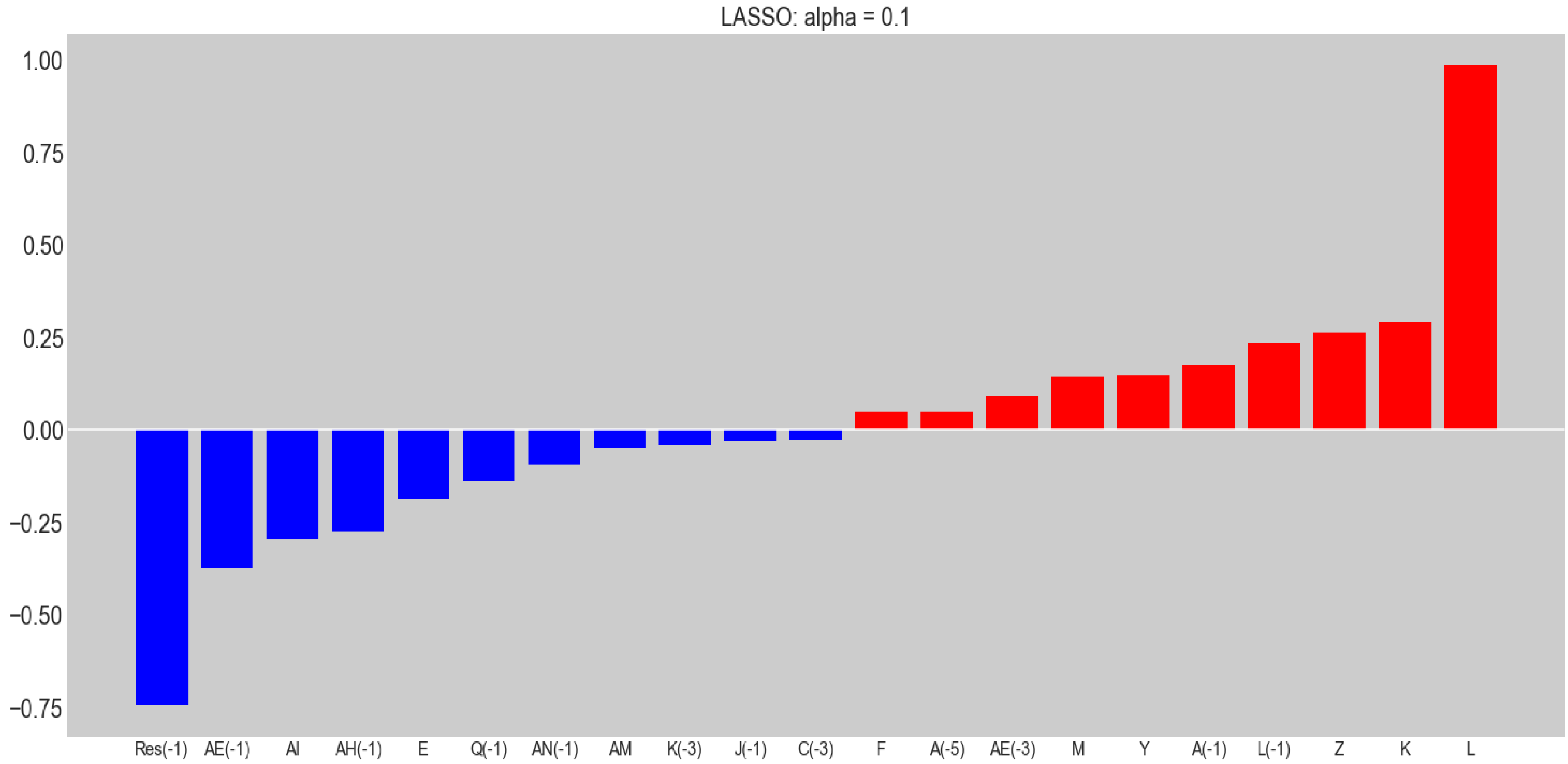
[17] Price distortion factors are calculated using the formula: "(City gate gas price / each energy price - 1) ×100".

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

D. Fig. 9 LASSO results (α = 0.1) in LASSO-ECM2: Selected comprehensive short-run variables

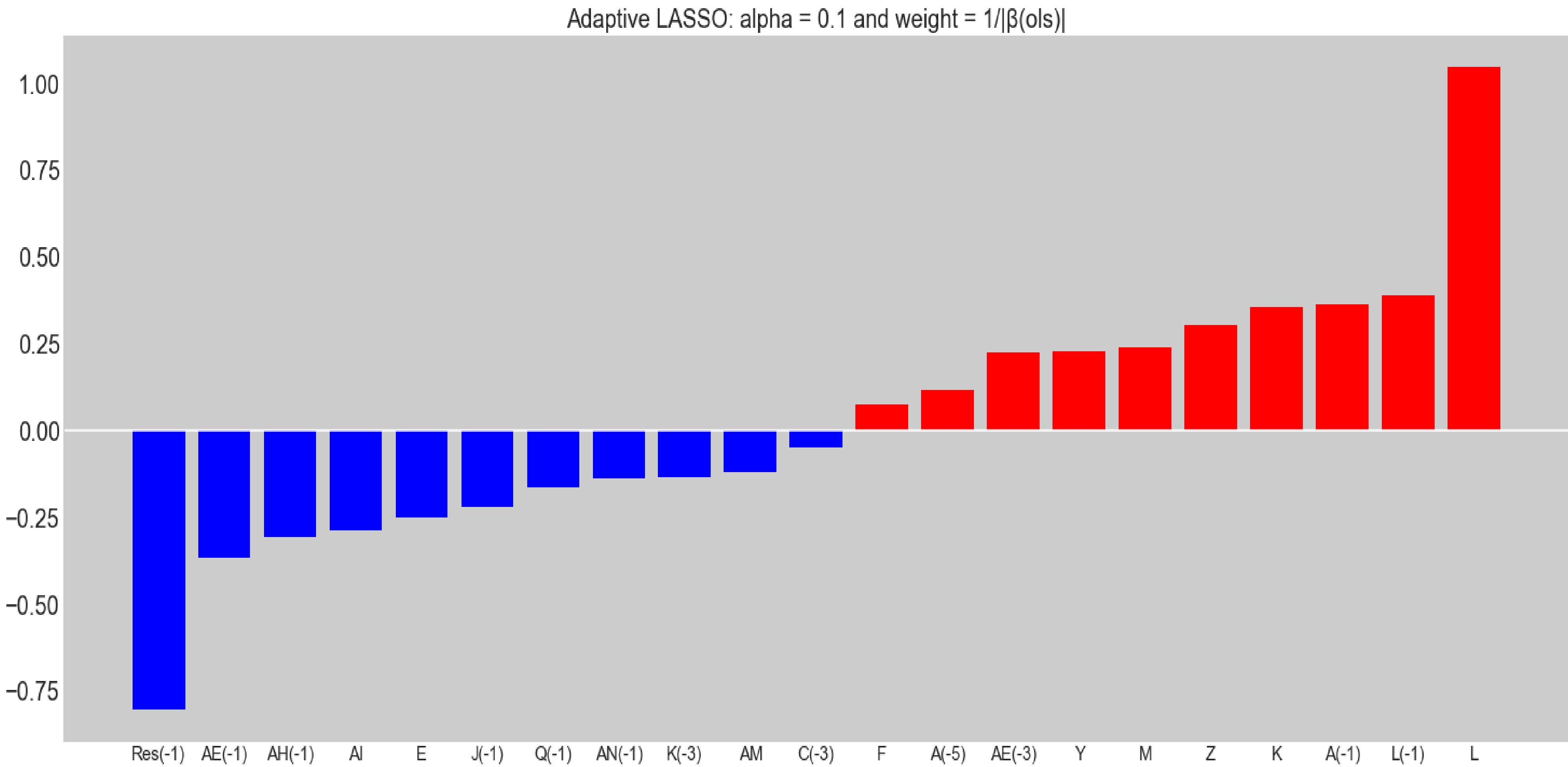


21 Coefficients selected by LASSO

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

Notes: Table 9 shows acronyms denoting relevant factors. These coefficients are rounded to four decimal places. Fig. 9 displays only variables with non-zero coefficients.

E. Fig. 10 ALASSO results ($\alpha = 0.1$ and $\hat{w}_j = 1/|\hat{\beta}_j(ols)$) in LASSO-ECM2: Selected comprehensive short-run variables



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

Notes: Table 9 shows acronyms denoting relevant factors. These coefficients are rounded to four decimal places. Fig. 10 displays only variables with non-zero coefficients

Reference

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

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

Bu, Y., Wang, E., Bai, J., & Shi, Q. (2020). Spatial pattern and driving factors for interprovincial natural gas consumption in China: Based on SNA and LMDI. *Journal of Cleaner Production*, *263*, 121392. https://doi.org/https://doi.org/10.1016/j.jclepro.2020.121392

Campos, J., Ericsson, N. R., & Hendry, D. F. (2005). General-to-specific Modeling: An Overview and Selected Bibliography. *International finance discussion papers*, *2005*(838), 1-94. https://doi.org/10.17016/IFDP.2005.838

Chen, Z., Wang, H., Li, T., & Si, I. (2021). Demand for Storage and Import of Natural Gas in China until 2060: Simulation with a Dynamic Model. *Sustainability*, *13*(15), 8674. https://www.mdpi.com/2071-1050/13/15/8674

Duras, T., Javed, F., Månsson, K., Sjölander, P., & Söderberg, M. (2023). Using machine learning to select variables in data envelopment analysis: Simulations and application using electricity distribution data. *Energy Economics*, *120*, 106621. https://doi.org/https://doi.org/10.1016/j.eneco.2023.106621

Engle, R. F., & Granger, C. W. J. (1987). Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica*, *55*(2), 251-276. https://doi.org/10.2307/1913236

Gao, J., & Dong, X. (2018). Stimulating factors of urban gas consumption in China [Article]. *Natural Gas Industry*, *38*(3), 130-137. https://doi.org/10.3787/j.issn.1000-0976.2018.03.016

Géron, A. l. (2022). *Hands-on machine learning with Scikit-Learn, Keras and TensorFlow : concepts, tools, and techniques to build intelligent systems* (Third edition. ed.). O'Reilly.

Ghoddusi, H., Creamer, G. G., & Rafizadeh, N. (2019). Machine learning in energy economics and finance: A review. *Energy Economics*, *81*, 709-727. https://doi.org/https://doi.org/10.1016/j.eneco.2019.05.006

Hendry, D. (1985). Econometric Methodology. *Paper presented to the Econometric Society Fifth World congress*.

Hepburn, C., Qi, Y., Stern, N., Ward, B., Xie, C., & Zenghelis, D. (2021). Towards carbon neutrality and China's 14th Five-Year Plan: Clean energy transition, sustainable urban development, and investment priorities. *Environmental Science and Ecotechnology*, *8*, 100130. https://doi.org/https://doi.org/10.1016/j.ese.2021.100130

Jiang, H., Dong, X., Jiang, Q., & Dong, K. (2020). What drives China's natural gas consumption? Analysis of national and regional estimates. *Energy Economics*, *87*, 104744. https://doi.org/https://doi.org/10.1016/j.eneco.2020.104744

Kock, A. B. (2016). CONSISTENT AND CONSERVATIVE MODEL SELECTION WITH THE ADAPTIVE LASSO IN STATIONARY AND NONSTATIONARY AUTOREGRESSIONS. *Econometric Theory*, *32*(1), 243-259. https://doi.org/10.1017/S0266466615000304

Kondratieff, N. D. (1925). The Static and the Dynamic View of Economics. *The Quarterly Journal of Economics*, *39*(4), 575-583. https://doi.org/10.2307/1883266

Koop, G., Leon-Gonzalez, R., & Strachan, R. W. (2011). Bayesian inference in a time varying cointegration model. *Journal of Econometrics*, *165*(2), 210-220.

Kumar, P. K., & Sarojamma, B. (2017). Time Series Cross-Validation Techniques for determining the order of the Autoregressive models. *International journal of advanced research in computer science*, *8*(3).

Liang, C., & Schienle, M. (2019). Determination of vector error correction models in high dimensions. *Journal of Econometrics*, *208*(2), 418-441. https://doi.org/https://doi.org/10.1016/j.jeconom.2018.09.018

Liao, Z., & Phillips, P. C. B. (2015). AUTOMATED ESTIMATION OF VECTOR ERROR CORRECTION MODELS. *Econometric Theory*, *31*(3), 581-646. https://doi.org/10.1017/S026646661500002X

Liu, G., Dong, X., Jiang, Q., Dong, C., & Li, J. (2018). Natural gas consumption of urban households in China and corresponding influencing factors. *Energy Policy*, *122*, 17-26. https://doi.org/https://doi.org/10.1016/j.enpol.2018.07.016

Lu, X., Xiao, J., Wang, X., Wen, L., & Peng, J. (2024). Government regulation and China's natural gas price distortion: A sectoral perspective. *Natural resources forum*. https://doi.org/10.1111/1477-8947.12410

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. *Advances in Social Sciences Research Journal*, *9*, 144-153. https://doi.org/10.14738/assrj.93.12005

Meidan, M. (2020). Unpacking China's 2060 carbon neutrality pledge. *OXFORD ENERGY COMMENT*(The Oxford Institute For Energy Studies.). https://a9w7k6q9.stackpathcdn.com/wpcms/wp-content/uploads/2020/12/Unpacking-Chinas-carbon-neutrality-pledge.pdf

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.

Mu, X., & Li, G. (2018). Study on natural gas consumption forecast and influence factors in China based on system dynamics model [Article]. *Journal of Engineering Studies*, *10*(1), 56-67. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85081687737&partnerID=40&md5=4aa67d5e377ce89547df19697bf7471f

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. *Contemporary economic policy*, *24*(2), 332-342. https://doi.org/10.1093/cep/byj019 (Contemporary Economic Policy)

OECD. (1998). *OECD Composite Leading Indicators: a Tool for Short‐Term Analysis*.

OECD. (2022). *OECD Composite Leading Indicators: Reference Turning Points and Component Series*.

https://www.oecd.org/sdd/leading-indicators/CLI-components-and-turning-points.pdf
Ogden, J., Jaffe, A. M., Scheitrum, D., McDonald, Z., & Miller, M. (2018). Natural gas as a bridge to hydrogen transportation fuel: Insights from the literature. *Energy Policy*, *115*, 317-329. https://doi.org/https://doi.org/10.1016/j.enpol.2017.12.049
Paltsev, S., & Zhang, D. (2015). Natural gas pricing reform in China: Getting closer to a market system? *Energy Policy*, *86*, 43-56. https://doi.org/https://doi.org/10.1016/j.enpol.2015.06.027
Puttachai, W., Phadkantha, R., & Yamaka, W. (2022). The threshold effects of ESG performance on the energy transitions: A country-level data. *Energy Reports*, *8*, 234-241. https://doi.org/https://doi.org/10.1016/j.egyr.2022.10.187
Ramírez-Orellana, A., Martínez-Victoria, M., García-Amate, A., & Rojo-Ramírez, A. A. (2023). Is the corporate financial strategy in the oil and gas sector affected by ESG dimensions? *Resources Policy*, *81*, 103303. https://doi.org/https://doi.org/10.1016/j.resourpol.2023.103303
Ranstam, J., & Cook, J. A. (2018). LASSO regression. *British Journal of Surgery*, *105*(10), 1348-1348. https://doi.org/10.1002/bjs.10895
Shen, J., Zhang, Q., Xu, L., Tian, S., & Wang, P. (2021). Future CO2 emission trends and radical decarbonization path of iron and steel industry in China. *Journal of Cleaner Production*, *326*, 129354. https://doi.org/https://doi.org/10.1016/j.jclepro.2021.129354
Shi, X., Hari Malamakkavu Padinjare, V., & Jacqueline, T. (2017). Global impact of uncertainties in China's gas market. *Energy Policy*, *104*, 382-394. https://doi.org/https://doi.org/10.1016/j.enpol.2017.02.015
Shi, X., & Sun, S. (2017). Energy price, regulatory price distortion and economic growth: A case study of China. *Energy Economics*, *63*, 261-271. https://doi.org/https://doi.org/10.1016/j.eneco.2017.02.006
Shi, X., Wang, K., Cheong, T. S., & Zhang, H. (2020). Prioritizing driving factors of household carbon emissions: An application of the LASSO model with survey data. *Energy Economics*, *92*, 104942. https://doi.org/https://doi.org/10.1016/j.eneco.2020.104942
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. *The Journal of finance and data science*, *4*(2), 71-89. https://doi.org/10.1016/j.jfds.2017.11.001
Stern, N., & Xie, C. (2022). China's new growth story: linking the 14th Five-Year Plan with the 2060 carbon neutrality pledge. *Journal of Chinese economic and business studies*, *ahead-of-print*(ahead-of-print), 1-21. https://doi.org/10.1080/14765284.2022.2073172
Stock, J. H., & Watson, M. W. (2019). *Introduction to econometrics*. Pearson Education.
Theodori, G. L. (2009). Paradoxical perceptions of problems associated with unconventional natural gas development. *Southern rural sociology*, *24*(3), 97-117.
Tu, K. J., & Johnson-Reiser, S. (2012). Understanding China's rising coal imports. *Policy outlook*(Carnegie Endowment for International Peace), 16. https://carnegieendowment.org/files/china_coal.pdf
Villar, J. A., & Joutz, F. L. (2006). The relationship between crude oil and natural gas prices. *Energy Information Administration, Office of Oil and Gas*, *1*, 1-43.
Wang, J., & Li, N. (2020). Influencing factors and future trends of natural gas demand in the eastern, central and western areas of China based on the grey model. *Natural Gas Industry B*, *7*(5), 473-483. https://doi.org/https://doi.org/10.1016/j.ngib.2020.09.005
Wang, T., Zhang, D., Ji, Q., & Shi, X. (2020). Market reforms and determinants of import natural gas prices in China. *Energy*, *196*, 117105. https://doi.org/https://doi.org/10.1016/j.energy.2020.117105
Wang, X., Lu, X., Zhou, N., Xiao, J., & Chen, J. (2020). Does Environmental Regulation Affect Natural Gas Consumption? Evidence from China with Spatial Insights. *Sustainability*, *12*(8), 3354. https://www.mdpi.com/2071-1050/12/8/3354
Wang, Y., Liu, J., Tang, X., Wang, Y., An, H., & Yi, H. (2023). Decarbonization pathways of China's iron and steel industry toward carbon neutrality. *Resources, Conservation and Recycling*, *194*, 106994. https://doi.org/https://doi.org/10.1016/j.resconrec.2023.106994
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. https://doi.org/https://doi.org/10.1016/j.energy.2023.128521
Xu, H., & Wang, Z. (2021). Analysis of driving factors of natural gas consumption in china: Based on regional integration and LMDI method [Article]. *Energy Engineering: Journal of the Association of Energy Engineering*, *118*(4), 1027-1043. https://doi.org/10.32604/EE.2021.015028
Yu, E. S. H., & Jin, J. C. (1992). Cointegration tests of energy consumption, income, and employment. *Resources and Energy*, *14*(3), 259-266. https://doi.org/https://doi.org/10.1016/0165-0572(92)90010-E
Yu, X., & Tan, C. (2022). China's pathway to carbon neutrality for the iron and steel industry. *Global Environmental Change*, *76*, 102574. https://doi.org/https://doi.org/10.1016/j.gloenvcha.2022.102574
Zhang, C., Su, B., Zhou, K., & Yang, S. (2019). Analysis of electricity consumption in China (1990–2016) using index decomposition and decoupling approach. *Journal of Cleaner Production*, *209*, 224-235. https://doi.org/https://doi.org/10.1016/j.jclepro.2018.10.246
Zivot, E., & Wang, J. (2006). *Modeling Financial Time Series with S-PLUS®* (2nd ed. 2006. ed.). Springer New York. https://doi.org/10.1007/978-0-387-32348-0
Zou, H. (2006). The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*, *101*(476), 1418-1429. https://doi.org/10.1198/016214506000000735