European Gas Markets: Market Integration & Market Efficiency: A Network Perspective

WORONIUK, DAVID,JAMES

European Gas Markets: Market Integration & Market Efficiency - A Network Perspective

David Woroniuk

Department of Economics and Finance
Durham University Business School
Durham University

This thesis is submitted for the degree of
Doctor of Philosophy

June 2023

Three Essays on European Oil and Gas Markets

Abstract

In this thesis, I study the market integration of European Natural Gas markets through two papers, whilst the third paper considers the impact of News Sentiment on the pricing and trading of Clean Energy and Traditional Energy stocks. Specifically, the first paper studies the level of harmonisation of European Natural Gas prices, characterised by 12 European gas hubs. The key finding is that, under normal market conditions, European Natural Gas markets are becoming increasingly integrated, with few physical barriers to increased market integration. Conversely, the detection of non-physical barriers to trade suggests that the liberalisation and development of certain national gas markets is yet to be fully achieved, inferring that improvements in technical arrangements are required.

The second paper provides a framework for forecasting the short term presence of physical barriers to market integration of European Natural Gas markets. The identification of infrastructure congestion is an important prerequisite in enforcing price competition, and the implementation of an internal European gas market. In order to address this challenge, the underlying infrastructure network is learnt as a graph, and a deep learning framework, Graph Convolutional Long Short-Term Memory Neural Network (GC-LSTM), based on the topology of the infrastructure network, is applied to learn the interactions between different pipelines, and forecast gas flows throughout the network. Empirical results show that the GC-LSTM outperforms baseline methods in predicting gas pipeline flows.

The third paper studies the impact of News sentiment on pricing and trading for European Clean Energy companies and Traditional Energy companies. Using daily news extracted from Bloomberg, we estimate Vector Autoregressive (VAR) models and evaluate the dynamic spillover effects between News sentiment, stock returns and trading volumes. We find that European Clean Energy firms and Traditional Energy firms share the same patterns; that News sentiment positively affects both stock returns and trading volumes, and in return, stock returns and trading volumes have a limited impact on News sentiment. Nevertheless, the spillovers are relatively moderate and asymmetric.

Declaration

The work in this thesis is based on research carried out at the Department of Economics and Finance in Durham University Business School, England. No part of this thesis has been submitted elsewhere for any other degree or qualification and it is my own work unless referenced to the contrary in the text, or where it is otherwise clearly indicated that it is the work of others.

The first paper (Chapter 2) contained in this thesis has been published by the Energy Policy Research Group (EPRG), University of Cambridge, as part of the working papers series. The relevant publication is listed below:

Publication: *David Woroniuk, Arzé Karam and Tooraj Jamasb (2019). European Gas Markets, Trading Hubs, and Price Formation: A Network Perspective. Energy Policy Research Group, University of Cambridge.*

Acknowledgements

There are many people to whom I would like to express my sincere gratitude for their help, guidance and support both academically and personally during my PhD study:

First and foremost, I am immensely grateful to my supervisor, Prof. Tatiana Damjanovic. She has provided me with so much support, advice and encouragement in both my PhD studies and my wider academic experience. I am very grateful that she agreed to accept me as her student, and I will always be thankful.

Next, I would like to thank my family for their unwavering love and support. Their sacrifices, tremendous understanding and encouragement, enabled me to complete my studies. My family were the inspiration to begin my PhD journey, and have provided immense support throughout it's completion.

I would like to thank my fiancé, Emma (Wenyi), who's perseverance, energy and positivity has helped me through difficult times during my studies. Finally, I would like to thank my friends and former colleagues, for their kind support and help.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis aims to explore two main topics in Energy Economics: market integration and market efficiency. To study the determinants of market integration, I focus on two perspectives, which are price transmission efficiency, determined through dynamic network density terms, and isolation of physical network constraints. To contribute to the understanding of market integration, I emphasise that day ahead gas markets in Europe are developing, with high variability in the rates of development, and integration into European Gas pricing. Moreover, the findings in Chapter 2 suggest a limited number of physical barriers to market integration under normal market operating conditions. Chapter 3 provides a mechanism for the monitoring, and prediction of short term physical barriers to market integration. A comprehensive review of the existing literature and empirical analysis of European Natural Gas market integration are presented in Chapters 2 & 3, respectively. A brief outline of the motivation for the thesis, research background, notable research gaps, and key findings of three individual papers are presented as follows.

The issue of whether European gas markets are integrated has been explored by numerous authors (Asche et al. 2000; Asche et al. 2002; Neumann et al. 2006; Renou-Maissant 2012), using a multitude of methodological approaches, such as cointegration, causality and state space modelling, achieving contrasting results. Whilst some authors accept the hypothesis of a harmonised European market (Asche et al. 2000; Asche et al. 2002), others find evidence to reject this (Neumann et al. 2006; Renou-Maissant 2012). Although the application of complex network theory is in its infancy, the literature has previously exploited symmetric (correlation-based) measures to formulate undirected networks, with

Minimum Spanning Trees (MST) introduced to identify hierarchical market power within crude oil markets (Ji and Fan 2016) and gas markets (Geng et al. 2014).Whilst this methodology permits dynamic assessment of interconnectivity within a network, it reduces the network to its most primitive of structures, containing only the most intuitive information exhibited by the network. Through use of nonlinear extensions of Granger causality (Geweke 1982), Granger causal networks can be developed, which can elucidate which hubs act as 'price setters' and 'price takers', permitting observation of the interactive evolution of market dynamics.

The results presented in Chapter 2 confirm that day ahead gas markets in Europe are developing, however each hub holds unique characteristics, providing different rates of development and integration. The low number of physical barriers to price harmonisation within the European day ahead gas market implies that the Third Energy Package's focus on national gas market integration through cross-border mechanisms has been broadly successful, reducing pipeline capacity constraints throughout the European network. Conversely, the detection of non-physical barriers to trade suggests that the liberalisation and development of certain national gas markets is yet to be fully achieved, inferring that improvements in technical arrangements are required. It is imperative that system operators throughout Europe engage in providing full integration of day ahead gas markets through a number of mediums, primarily market concentration, market design, regulation and security of supply, all of which are crucial to the development of a single European market. The aim of elimination of physical barriers to integration, in conjunction with improved legislative integration, implies the convergence of day ahead gas prices toward a single European price (in the absence of transmission costs), however this also holds the additional benefit of reducing pipeline congestion, increasing usage efficiency, and reducing the market power of actors within national gas markets.

Chapter 3 builds on the findings presented in Chapter 2, developing an early warning system for the detection of physical barriers to price harmonisation of European Gas prices. Whilst Chapter 2 finds a limited number of persistent physical barriers to market integration, identifying infrastructure congestion issues is an important prerequisite in

enforcing competition, implementing an internal European gas market and increasing the security of gas supply of EU members.

Many of the prior infrastructure forecasting studies employing statistical methods (Dieckhöner et al. 2013; Lochner and Bothe 2007; Monforti and Szikszai 2010) were developed when the size of datasets was limited. Due to the limited capability of statistical models to handle high dimensional time-series, infrastructure constraints have historically been aggregated to the country level. Utilising the benefits of increased computational power and data availability, I forecast infrastructure constraints at a pipeline specific level. Moreover, the spatiotemporal characteristics of prior studies (Dieckhöner et al. 2013; Monforti and Szikszai 2010) are often inappropriately specified, with the impact of upstream congestion issues not passed to downstream assets. This is corrected through the inclusion of a normal operations reachability matrix (NORM), based on the ability of each 'parcel' of gas to reach different points within the network.

The model's ability to adapt to unplanned outages and maintenance cycles verifies the applicability of the modelling technique to pipeline utilization forecasting, and subsequently, gas flow state forecasting. The prediction of pipeline utilization rates throughout the European Gas system relative to these events provides valuable information pertaining to how the impact of exogenous shocks propagate around the European Gas Infrastructure Network. Additionally, this provides important information relating to how pipeline utilization rates, and subsequent gas flows may react to exogenous shocks.

Chapter 4 departs from the topic of market integration, instead considering the impact of News Sentiment on the price formation and trading volumes of "Traditional Energy" and "Clean Energy" stocks. Using daily news extracted from Bloomberg, we estimate VAR models and evaluate the dynamic spillover effects between News sentiment, stock returns and trading volumes. We find that European Clean Energy firms and Traditional Energy firms share the same patterns; that News sentiment positively affects both stock returns and trading volumes, and in return, stock returns and trading volumes have a limited impact on News sentiment. Nevertheless, the spillovers are relatively moderate and asymmetric.

The remainder of the thesis is organized as follows. Chapter 2 presents the full details of the first paper, titled *European Gas Markets, Trading Hubs, and Price Formation: A Network Perspective*. Chapter 3 demonstrates the full content of my second paper, titled *Congestion Learning and Forecasting within Gas Markets: A Deep Learning Approach*. Chapter 4 details my third paper, titled *Clean Energy, Brown Energy & The Impact of News Sentiment: A Comparative Analysis*. Chapter 5 concludes and provides implications for future research.

Chapter 2

European Gas Markets, Trading Hubs, and Price Formation: A Network Perspective

2.1. Introduction

The creation of a single European Energy Market has been a long-standing European objective, initially proposed through the Treaties of Rome in 1957. Through the Treaties of Rome, the European Union (EU) established a customs union, agreeing to the progressive reduction of customs duties, and the establishment of a single market for goods, labour and services. Following this, the European Commission (EC) has issued a number of liberalisation directives for natural gas markets, notably the First Energy Package (1998), Second Energy Package (2003) and Third Energy Package (2009), which aim to increase market competition and consumer protection throughout Europe.

The Third Energy Package, enacted on 3^{rd} September 2009 (Directive 2009/73/EC)[1], was intended to further develop the internal European natural gas market. The directive aimed to address issues surrounding market access, transparency and consumer protection through the development of a competitive, integrated market. However, obstacles to an integrated[2], single European market for natural gas remain. These are not only struc-

[1] Directive 2009/73/EC of the European Parliament and of the Council of 13 July 2009 concerning common rules for the internal market in natural gas and repealing Directive 2003/55/EC

[2] Throughout this paper, the terms 'Market Integration' and 'Market Harmonization' are used interchangeably, inferring the development of a unified, competitive natural gas market within Europe, characterised by convergence of price returns toward a singular value, in the absence of transportation costs.

tural, with the relative level of transmission liberalisation holding price impact, but also contractual, with producers, who hold a high degree of market power, constraining market liberalisation through long term 'take-or-pay' contracts. The European Commission has taken the above measures to enforce competition and increase cross border trade, offering all consumers a choice of supply, through development of short-term trade in gas markets. As such, it is possible that, although the European gas network is imperfectly connected, with substantial pipeline bottlenecks and regions of spare capacity, and considering the geographic disparity of market pairs, price signalling may still propagate throughout the short-term gas markets, tacitly inferring that markets may be competitive.

The issue of whether European gas markets are integrated has been explored by numerous authors (Asche et al. 2000; Asche et al. 2002; Neumann et al. 2006; Renou-Maissant 2012), using a multitude of methodological approaches, such as cointegration, causality and state space modelling, achieving contrasting results. Whilst some authors accept the hypothesis of an integrated European market (Asche et al. 2000; Asche et al. 2002), others find evidence to reject this (Neumann et al. 2006; Renou-Maissant 2012). Many such studies in economics and financial time series employ Granger Causality (Granger 1969). This method investigates pairs of price returns which exhibit an equilibrium relationship. Discovering an equilibrium relationship between two variables, X and Y, implies that variable X "causes" movements in Y when the predictive capacity of Y is improved through the addition of lagged values of X. Previous studies have provided evidence of long-run convergence (Renou-Maissant 2012; Growitsch et al. 2015; Neumann and Cullmann 2012) through the measurement of long-run cointegration or equilibrium relationships. Through the analysis of short-run, causal interactions amongst price returns, this paper improves the resolution of analysis of European gas market dynamics.

The aim of the study is to assess the European gas network's integration and evolution. We investigate the time-varying interactions amongst European gas prices by measuring the Granger-Geweke (Geweke 1982) causality between price return dynamics during the period 2016-2018. This enables dynamic observation of which hubs[3] act as 'price setters'

[3]Hubs can be both physical or virtual points on the gas transmission system where ownership rights of gas can be transferred, logistically supported through the provision of market services by an impartial

and 'price takers', which can be extracted to draw inferences relating to the development of the Single European market for gas. We also seek to examine whether abnormal price behaviour within European gas markets can be explained by historical phenomena. Further to this, the origin and propagation of events which disrupt normal market functioning is elucidated. In addition, we analyse the efficacy of the Third European Gas Directive through the generation of a European network density time-series. In the case of an informationally efficient market, short term price returns should not be related to the lagged values of other variables, hence a Granger-Geweke test should not detect a high degree of causality. However, given the stated European objective of a unified, competitive natural gas market, price returns, in the absence of transportation limitations, should converge toward a singular value across all European hubs, quantified through the network density term.

As European gas injection is predominantly by pipeline, with gas transited between hubs by an interconnected pipeline network, application of network theory, which can measure the degree of interdependence and direction of causality between hubs, represents an appropriate methodology for assessment of the European gas market.

We use a novel approach based on graph-theory to model the dynamic interactions amongst European gas prices. The interactions between daily day-ahead gas prices are modelled as a connectivity network, in such instances, nodes represent different European trading hubs and edges between them denote the strength and direction of statistically significant influences between relative price variations. The dynamic propagation patterns between the different hubs return series provides valuable information on the degree of market integration.

Although the application of complex network theory is in its infancy, literature has previously exploited symmetric (correlation-based) measures to formulate undirected networks, with Minimum Spanning Trees (MST) introduced to identify hierarchical market power within crude oil markets (Ji and Fan 2016) and gas markets (Geng et al. 2014).Whilst this methodology permits dynamic assessment of interconnectivity within a

entity.

network, it reduces the network to its most primitive of structures, containing only the most intuitive information exhibited by the network. Through use of nonlinear extensions of Granger causality (Geweke 1982), Granger causal networks can be developed, which can elucidate which hubs act as 'price setters' and 'price takers', permitting observation of the interactive evolution of market dynamics. As such, this paper holds novelty through the use of a network approach which specifies multivariate, directed weighted networks in order to understand the multivariate interactions between European gas prices, rather than static investigation into univariate profiles.

We show that the connectivity of the European natural gas markets remained at a low mean throughout the sample period. A number of spikes were observed throughout the sample period, however abnormally positive and negative changes in connectivity were similar in number and magnitude. We infer that the path toward attainment of an sustained, high degree of gas market integration, which characterises a Single European Market for natural gas, appears to be long.

This paper is organised as follows: Section 2.2 discusses the background information related to this study, whilst Section 2.3 outlines the methodologies employed throughout the paper. Section 2.4 specifies the dataset, with Section 2.5 reporting the empirical findings, which are subsequently discussed in Section 2.6. Section 2.7 concludes.

2.2. The European Natural Gas Market

The price of gas depends on a range of economic and non-economic supply and demand parameters, including weather conditions, proximity to production, transmission constraints, geopolitical factors and import diversification. This section provides a brief overview of the European spot gas markets, then providing a critical analysis of previous studies on European natural gas market integration.

2.2.1. Background

European gas companies were largely considered to be conventionally regulated national monopolies until the EU embarked upon its natural gas reformation programme, aiming to correct distortions and increase competitive pricing through legislating three Energy Directives (1998, 2002 and 2009). The market framework underwent substantial change due to new unbundling rules, capacity allocation rules and the removal of destination clauses. Critically, two barriers to a singular, internal European market for gas remain; non-physical barriers, characterised by contract inflexibility and a lack of market liberalisation, and physical barriers, namely trade constriction through a lack of available transmission capacity. Although the three directives have increased competition within European gas markets, a dual-tier pricing system has developed, with long term supply contracts and hub-based mechanisms emerging. Whilst hub-based pricing is short-term in nature, the inherent limitations in transmission capacity and contracted volume flexibility provide gas markets with relatively unique characteristics, causing the spot market to act as a one-day forward market.

European day ahead trading responds to the prevailing domestic supply and demand fundamentals, with Brown and Yucel (2008) finding that crude oil prices, weather, seasonality and storage play ancillary roles in price discovery. Meanwhile long-term contracts are typically negotiated for a period of 10-30 years, with natural gas pricing indexed to the pricing of oil products. Whilst long-term contracts still account for 70% (Heather 2015) of gas procurement in Europe, a transition from oil product indexation to short-term hub-based pricing is underway, with consumer dissatisfaction with inflexible, long-term contracts driving an increased volume of day ahead market trade. The relatively low "swing" volume and destination inflexibility inherent in long-term contracts, combined with the growing volume of day ahead trade, motivate this study's investigation into day ahead pricing.

Gas markets hold a number of stylised facts which reduce the incumbent's ability to exercise time or space arbitrage. Two important examples of this are pipeline transmission constraints and the national interests of each hub pertaining to wholesale natural gas

markets. That said, due to the EU objective of a single, common pricing structure, capacity decisions and pricing strategies are becoming increasingly more simultaneous amongst European natural gas markets, based on increases in shared information and cooperation amongst regulatory bodies.

The European gas network is broadly considered to be imperfectly connected, with substantial pipeline bottlenecks and regions with spare capacity, with ENTSOG issuing biennial Ten-Year Network Development Plans (TYNDP) in order to correct these distortions. However, even in an imperfectly connected network of geographically disparate market pairs, price signalling may still propagate throughout European gas markets, tacitly inferring that markets may be competitive.

2.2.2. Previous Literature

The liberalisation of natural gas markets has driven a growing academic interest in the process of market integration. A number of studies (De Vany and Walls 1993; Walls 1994; King and Cuc 1996; Serletis 1997; Cuddington and Wang 2006) have focused on North American gas market integration following the Federal Energy Regulatory Commission's (FERC) regulatory developments throughout the 1980s. European natural gas market literature is both narrower in scope and more current, owing to more recent regulatory developments. As such, there is a growing interest in ascertaining whether natural gas markets in Europe are converging toward the EU goal of a single, competitive internal market, often measuring the degree to which price coupling and integration is being practically achieved. Since the early 2000's, a broad array of methodologies have provided a series of contradictory findings, with some studies displaying evidence of increasing integration amongst natural gas pricing, and others finding the opposite. Notably, none of these studies have assessed the dynamic relationships between every European hub in order to assess the time-varying nature of European market integration.

For example, based on a cointegration analysis, (Asche et al. 2000; Asche et al. 2002) accept the market integration hypothesis for French and German import pricing, finding that the "Law of One Price" holds. This finding was updated by the findings of Bourbon-

nais and Geoffron (2007), who found an increasing degree of integration between Belgian, French, German, Italian, UK and Spanish natural gas prices from 1999-2005. Conversely, state space model based studies such as Neumann et al. (2006), suggest that convergence between Zeebrugge-UK increased following the addition of an interconnector, whilst market integration amongst continental European hubs was not apparent. Taken in unison, the contrasting findings of these studies highlight the importance of the geographic proximity of hubs studied, which increases the probability of market integration. Furthermore, the assessed time period will impact the empirical findings of the study, emphasising the requirement of a dynamic model specification, which continually reassesses the degree of market integration. Robinson (2007) augments Neumann et al. (2006) study through the application of the Nahar and Inder (2002) methodology to annual retail pricing of Finland, France, Ireland, Netherlands, Spain and the UK, finding only Dutch pricing to converge throughout the observed period.

More recently, Renou-Maissant (2012) found strong integration amongst continental European gas markets, with the exception of the Belgian market. Furthermore, Growitsch et al. (2015) employed state space modelling with the extension of an error correction model to demonstrate the increasing integration of hubs within North-West Europe. Whilst the measured hubs exhibited an increased degree of integration since the introduction of the 'entry-exit' system, information efficiency between the hubs also increased. Moreover, Neumann and Cullmann (2012), who extended Growitsch et al. (2015)'s dataset, found geographical proximity and available transmission capacity to play a role in price integration, with Danish consumers paying a premium due to network externality, and Austrian pricing diverging from other measured hubs. These findings are unsurprising, as Austria serves as a thoroughfare for Russian gas into Southern and South-East Europe (Baumgarten), with hubs in these regions disregarded in the data specification.

The objective of a unified, competitive European gas market should, in the absence of transportation capacity limitations, determine the convergence of natural gas prices toward a singular price at all European trading hubs. If this assertion holds, the integration

of European gas markets, characterised by the network density term, can be expected to increase throughout the sample period.

2.3. Methodology

This section explains the methodologies employed throughout the paper. Firstly, we discuss construction of directed, multivariate weighted networks, developed through a rolling-window computation of Granger causality (Geweke 1982) between price return series of European natural gas hubs. The topology of the network is then quantified through the measurement of node weights and network density.

In order to investigate the evolution of the European natural gas pricing network, it is important to measure not only the degree of connectedness between European gas hubs, but also the directionality of each of these relationships. Consequently, we propose employing Granger-Geweke causality (Geweke 1982), a statistical notion of causality based on the relative forecasting power of multiple time series. Through the use of a VAR structure, we apply the Granger-Geweke network theory to the European gas network. The mathematical framework supplied by directed networks characterises interdependent systems well, enabling assessment of network evolution over time, which provides valuable information on how prices interact with one another dynamically.

Furthermore, 70% of European gas is landed by pipelines, with gas transited between hubs by a large, interconnected pipeline network. As such, the application of network theory, which can measure the degree of interdependence and direction of causality between hubs, appears to be a logical step in the assessment of the European gas market. If the European gas market is informationally efficient, short term price returns should not be related to the lagged values of other variables, hence a Granger-Geweke test should not detect a high degree of causality. However, given the stated European objective of a unified, competitive gas market throughout Europe, price returns, in the absence of transportation limitations, should converge toward a singular value across all European hubs, quantified through the network density term.

This section outlines the construction methodology of the multivariate, directed weighted

network used to assess the interdependencies of the European natural gas network, initially developed by Seth (2010). As we are employing a multivariate (MVAR) system, Granger-Geweke causality (Geweke 1982), which is simply Granger-causality adjusted for multivariate systems, is employed. The individual autoregressive representation of each return series is displayed in 2.1 & 2.2:

$$x(t) = \sum_{K=1}^{p} a_{1K} x(t-k) + \epsilon_1(t), \qquad \text{var}(\epsilon_1(t)) = \Sigma_1 \tag{2.1}$$

$$y(t) = \sum_{K=1}^{p} d_{1K} y(t-k) + \eta_1(t), \qquad \text{var}(\eta_1(t)) = R_1 \tag{2.2}$$

where a_{1K} and d_{1K} are autoregressive coefficients, $\epsilon_1(t)$ and $\eta_1(t)$ are noise terms and $K = 1, \ldots, p$. The joint description of the bivariate series $[x(t), y(t)]$ is given by the p^{th} order AR 2.3 and 2.4, where noise terms are not correlated over time and the covariance matrix is expressed as 2.5.

$$x(t) = \sum_{K=1}^{p} a_{2K} x(t-k) + \sum_{K=1}^{p} b_{2K} y(t-k) + \epsilon_2(t) \tag{2.3}$$

$$y(t) = \sum_{K=1}^{p} c_{2K} x(t-k) + \sum_{K=1}^{p} d_{2K} y(t-k) + \eta_2(t) \tag{2.4}$$

$$\Sigma = \begin{pmatrix} \Sigma_2 & r_2 \\ r_2 & R_2 \end{pmatrix} \tag{2.5}$$

where $\Sigma_2 = \text{var}(\epsilon_2(t)), R_2 = \text{var}(\eta_2(t))$ and $r_2 = \text{cov}(\epsilon_2(t), \eta_2(t))$. Whereas Σ_1 measures the ability of previous values of $x(t)$ to predict the present value of $x(t)$, Σ_2 represents the predictive capacity of $x(t)$ and $y(t)$'s previous values. If $x(t)$ and $y(t)$ are independent, then b_{2K} and c_{2K} are zero. Following Granger (1969), causality is defined as 2.6, with an MVAR approach applied to create the network 2.7, where each matrix, A_m, is determined by elements a_{ij} describing the linear relationship between $y_j(t-m)$ and y_i, p is the number of lags, and $E(t)$ is the vector of error terms.

$$GC_{y\to x} = w_{yx} = \log\left(\frac{\Sigma_1}{\Sigma_2}\right)$$
$$GC_{x\to y} = w_{xy} = \log\left(\frac{R_1}{R_2}\right) \tag{2.6}$$

$$Y(t) = \sum_{m=1}^{p} A_m Y(t-m) + E(t) \tag{2.7}$$

As the MVAR model described in 2.7 provides a time-invariant representation of the country Y's price as a function of the 11 other input values and their lags, a sliding window estimation is employed, facilitating time variation. A time varying model, employing daily data allows a time-series of network connectivity strength to be generated, facilitating the study of gas market integration over time.

The MVAR model shall be estimated using the Burg algorithm (Burg 1967), which recursively calculates the solution to an equation containing a Toeplitz matrix. Schlögl and Supp (2006) show that when the Burg algorithm (Burg 1967) is extended for multivariate Autoregressive models through the Nuttall-Strand method, the MVAR Burg algorithm provides the most accurate estimates. Furthermore, the Burg algorithm is considered more appropriate than non-parametric methods due to three distinct properties; windows are not applied to data, with the assumption that autocorrelation series outside the window $\neq 0$ relaxed, minimisation of backward and forward prediction errors in the least squares sense and the consistent yield of a stable Autoregressive model.

The MVAR model is considered to be theoretically more appropriate than a Spatial Vector Autoregression (sp-VAR) model (Beenstock et al. 2019), owing to the network topology of the physical gas infrastructure network. Whilst the sp-VAR model specifies the physical distance between different market zones well, it fails to consider the underlying network infrastructure through which gas is transported. As such, use of a sp-VAR model may lead to spurious conclusions concerning the level of physical market connection.

The model order q can be selected according to the Akaike criterion:

$$AIC(p) = 2log[det(\Sigma)] + 2pM^2/n \tag{2.8}$$

where Σ is the estimated noise covariance matrix of the bivariate Autoregressive model, n denotes the length of the data window, and M is the number of time series employed by the model. $2log[det(\Sigma)]$ holds an inverse relationship to p, whereas $2pM^2/n$ punishes models with a high order. The Akaike information criterion optimises q in order to minimise the cost function, which is defined by balancing the variance of the Autoregressive model against the volume of coefficients estimated.

2.3.1. Node Strength

The mathematical representation of a network is the adjacency matrix, from which the node strength can be determined. The node strength represents the sum of the weights of the total edges extending to other nodes within the system, which in turn, is divided between in-strength d_{in}, and out-strength d_{out}, due to the directional nature of granger causality, hence the relationships:

$$d_{\text{in}}(i) = \sum_{j \in V} A_{i,j} \qquad d_{\text{out}}(i) = \sum_{j \in V} A_{j,i} \tag{2.9}$$

$d_{in}(i)$ represents the total strength of incoming edges for vertex i, where V is the number of nodes, $A_{i,j}$ is the causal direction from node j to i, with causality weight between -1 and 1 determined through equation 2.7. Any interactions which do not achieve statistical significance shall be set to 0. $d_{\text{out}}(i) = \sum_{j \in V} A_{j,i}$ represents the total strength of outgoing edges for vertex i, where V is the number of nodes.

2.3.2. Network Connection Density

In order to assess overall market integration at a given time period, the global connection density, D, shall be computed:

$$D = \frac{1}{N(N-1)} \sum_{i,j \in V} A_{i,j} \tag{2.10}$$

where V is the set of available nodes and N is the total number of nodes. This facilitates measurement of the total amount of causal activity throughout the sample period, with the value determining the coordination of markets measured.

2.4. Data Specification

In order to analyse hub price convergence and by extension, European gas market integration, a definition of a gas hub must first be derived. Neumann et al. (2006) provides an appropriate hub definition, proffering that a hub can be both "physical (local) or virtual (notional) - on the gas transmission system where the transfer of natural gas can take place, logistically supported by a body offering the follow-up of the transfer of ownership, standardised contracts for trade at freely negotiated prices and other services".

Since Order 1775 was enacted in 2005, the European gas network has operated an entry-exit system, in which 'entry' gas is transported to either a physical or virtual trading hub within the network, from which it can be transported to an 'exit' point. Although operator published entry and exit pricing is location specific, volume-based price discrimination cannot be enacted amongst network users. As such, zones, based on proximity to entry-exit points, are defined within the transmission network, within which a singular price for gas is quoted, giving rise to 'virtual' trading hubs (VTP).

In order to accurately capture the dynamics of the European natural gas network, a hub, either physical or virtual, is specified within each 'trading zone'. As such, daily time-series data of a sample of 12 wholesale gas day ahead prices, covering the period 2016 to 2018 are considered. In order to fully represent the dynamics of the European gas market, the sample period begins upon commencement of day ahead trading at VTP Gaz System (Poland), the most junior trading hub within the network. The markets considered are: CEGH (Austria), VTP Gaz System (Poland), VHP-Gaspool (Germany), VHP-NCG (Germany), PSV (Italy), PVB (Spain), PEG Nord (France), PEG TRS (France), Zeebrugge (Belgium), TTF (Netherlands), ETF (Denmark) and NBP (UK). Midpoint day ahead prices were obtained from Bloomberg with time sampling of one-day.

Whilst prior studies reference an array of financial instruments within the delivery

curve, day ahead pricing was considered the most theoretically appropriate metric for this study, as day ahead pricing represented the most liquid instrument within the gas delivery curve within our study. Owing to its comparably high liquidity, day ahead pricing can be considered to be most representative of local supply-demand imbalances within the European Gas network.[4]

In order to negate any lag-structure requirements arising from time-zone differences, daily periodicity was chosen, with midpoint pricing best reflecting the trading prices of a given day. As such, this is in line with the pan-European analysis we intend to develop. All prices not quoted in €/MWh are converted by the daily spot midpoint price in the case of currency, and standard unit conversions (1 UK Therm = 0.02931MWh) in the case of unit. Each market's time series is comprised of 498 observations, as only 'gas days' are employed, due to this providing the most accurate portrayal of European gas markets functioning under normal conditions, in which all market participants are actively engaged with the market. Inclusion of non 'gas day' data may result in misleading conclusions on the state of integration of the European gas network, as market participants in certain regions may not actively participate in price discovery. As 'gas day' misalignment due to national holiday schedules occurs throughout the sample, it is negated through synchronizing the data through addition of the last available information pertaining to a missing observation. This has limited impact on the autocorrelation structure of the return series, with Granger and Ramanathan (1984) showing that the power of the augmented Dickey-Fuller test is increased through substituting missing observations with the previous observation.

2.4.1. Hub Pricing

The 12 daily day ahead natural gas prices (in €/MWh) and hub characteristics are described in Table 2.1. The highest gas price present in the sample is at the VTP Gaz System, Poland (€88.63/MWh) on 01/03/2018. This coincides with the highest price

[4]The body of literature focusing on European natural gas market integration, such as Renou-Maissant (2012), Neumann and Cullmann (2012) and Growitsch et al. (2015) evaluate day ahead price market integration, owing to the comparatively high liquidity of the day ahead market. Additionally, day ahead gas markets are considered to be representative of local supply-demand imbalances for the next market day.

observed at 10 of 12 hubs measured, which can be attributed to a demand shock following an unseasonal pan-European 'cold snap'. The pan-European 'cold snap' observed was referred to as 'The Beast from the East', with the meteorological name of 'Storm Emma', and precipitated a UK Formal Deficit warning due to a lack of gas availability and low storage levels. The high price observed in Poland is likely due to a number of concomitant factors, such as Polish gas infrastructure acting as the first European landfall for a large proportion of Russian swing capacity, extremely low temperatures in Eastern Europe and Poland's physical network externality.

Table 2.1. Descriptive Statistics

This table reports the mean and standard deviation of wholesale day-ahead natural gas prices between 2016-2018. The normality distribution (Jarque-Bera) test, including skewness and kurtosis values, are also reported.

Hub	**Mean**	**Std. Dev.**	**Skewness**	**Kurtosis**	**JB Test**
PVB (Spain)	22.54	5.06	1.22	2.31	≤0.001
PSV (Italy)	21.86	4.93	4.28	35.08	≤0.001
PEG-TRS (France)	21.45	5.41	3.65	31.83	≤0.001
VTP Gaz System (Poland)	21.18	4.84	6.07	76.65	≤0.001
CEGH (Austria)	20.19	3.89	3.03	21.83	≤0.001
NBP (UK)	20.00	5.03	5.46	68.89	≤0.001
NCG (Germany)	19.93	4.85	5.82	68.38	≤0.001
PEG-Nord (France)	19.78	4.43	3.65	31.83	≤0.001
Gaspool (Germany)	19.68	4.14	2.75	20.27	≤0.001
TTF (Netherlands)	19.68	4.51	4.62	49.90	≤0.001
Zeebrugge (Belgium)	19.51	3.90	1.40	4.82	≤0.001
ETF (Denmark)	18.86	3.89	1.91	9.24	≤0.001

Physical network externality can be a pertinent factor, as NBP (UK), which also exhibits physical network externality, registered the second highest price in the sample. The lowest price is measured at NBP (UK) on 15/06/2017, perhaps due to a supply surplus and lack of interconnector activity between the NBP (UK) and both Zeebrugge (Belgium) and TTF (Netherlands). All hubs observed exhibit a leptokurtic price distribution, with the dataset also exhibiting a high degree of right-hand skewness. CEGH (Austria) displays a mean gas price (€20.19/MWh) which is closest to the mean of the sample (€20.39/MWh), with Danish gas pricing holding the lowest mean (€18.86/MWh), likely due to indigenous production and physical proximity to a number of large gas fields in Norway and

The Netherlands. Finally, PVB (Spain) exhibits the highest mean (€22.54/MWh), likely due to its network externality, relatively high LNG usage and physical distance from sizable gas production. Due to the characteristics of Granger-causality (Granger and Ramanathan 1984), prices aren't normalised, as raw data should provide more insight into network topology at each point in time.

2.4.2. Traded Physical Volumes

Day ahead traded volume provides an important metric pertaining to market activity and development, as it provides a concise estimation of market liquidity and information critical to the development of market liquidity. Day ahead traded volumes are an important component in the determination of churn rate, with large absolute volumes in the day ahead market typically associated with a high churn rate, indicative of a well-developed market with a large number of market participants. Table 2.2 displays the mean day ahead volumes at each hub measured and coefficient of variation of day ahead volumes, which depicts variance of traded volume at each hub. Zeebrugge (Belgium), Gaspool (Germany), NCG (Germany), PSV (Italy) and PEG-Nord (France) display have the largest day ahead traded volumes, each holding a mean day ahead trading volume above 100MCM/d. CEGH (Austria), TTF (Netherlands) and VTP Gaz System (Poland) hold relative large mean day ahead trading volumes, with TTF (Netherlands) exhibiting a large coefficient of variation which can be associated with production at the Groningen field. PEG-TRS (France) and NBP (UK) display low traded volumes in the day ahead market, with NBP (UK) exhibiting the second largest coefficient of variation of the sample, implying that physical trade between the UK and continental Europe is largely dependent on domestic supply and demand balances in the UK. Finally, PVB (Spain) and ETF (Denmark) have a fraction of day ahead traded volume of the most liquid hubs, trading 7.2% and 2.3% of the mean volume traded at Zeebrugge (Belgium) respectively.

Table 2.2. Traded Physical Volumes

The mean day ahead volume in MCM/d and coefficient of variation, depicting the relative standard deviation in day ahead trading volumes and the number of physical pipeline connections (Connect.) of each hub.

Hub	Avg. Volume	Coeff. of Variation	Connect.
Zeebrugge (Belgium)	121.08	0.43	5
Gaspool (Germany)	116.27	0.19	5
NCG (Germany)	111.45	0.26	8
PSV (Italy)	109.47	0.24	3
PEG-Nord (France)	104.35	0.24	4
TTF (Netherlands)	90.73	0.61	4
CEGH (Austria)	90.48	0.27	2
VTP Gaz System (Poland)	75.60	0.17	1
PEG-TRS (France)	37.57	0.22	2
NBP (UK)	34.62	0.73	2
PVB (Spain)	8.66	0.52	1
ETF (Denmark)	2.76	0.87	2

2.4.3. Churn Ratio

Heather (2017) proffer that churn ratio is amongst the most important measures of each gas hub's commercial success, as it considers day ahead and total volume traded at each hub, whilst tacitly considering the number of market participants engaged in trading activities at each hub. The churn ratio, total traded volume divided by physical volume, provides an accurate indicator, based on the number of times each physical 'parcel' of gas is traded at the hub, of hub liquidity and commercial success. Gas markets are typically considered to be commercially successful, mature, hubs when the churn ratio is larger than ten, with many market participants reluctant to engage in activity at hubs with churn ratios below this threshold.

Table 2.3 shows that of the European gas hubs measured, TTF (Netherlands) is by far the most liquid and commercially successful hub, with a continued development of 7.35% in 2017. NBP (UK) remains the second most liquid trading hub within the sample, although it exhibited a decline in churn ratio in 2017, likely due to a combination of political uncertainty within the UK, cost of currency hedging for European market participants and the declining production of North Sea gas fields. It is clear that TTF (Netherlands)

and NBP (UK) are the most commercially relevant hubs within the sample, as they are the only hubs which exhibit churn ratios above ten, with total volume at NBP (UK) and TTF (Netherlands) exceeding that of the ten other hubs combined.

Table 2.3. Churn Ratio

The churn ratios of each hub for both 2016 (Heather 2017) and 2017, with the percentage change presented on the right.

Hub	2016 Churn Ratio	2017 Churn Ratio	% Change
TTF (Netherlands)	57.10	61.30	7.35%
NBP (UK)	22.10	21.81	-1.31%
CEGH (Austria)	5.70	3.51	-38.44%
Zeebrugge (Belgium)	4.10	2.95	-28.17%
NCG (Germany)	4.00	4.50	12.44%
Gaspool (Gemany)	2.50	3.05	21.90%
PEG-Nord (France)	1.70	2.19	29.11%
PSV (Italy)	1.20	3.15	162.26%
ETF (Denmark)	1.20	2.03	69.37%
VTP Gaz System (Poland)	0.80	1.09	36.23%
PEG-TRS (France)	0.60	1.10	83.14%
PVB (Spain)	0.10	0.61	513.94%

Interestingly, CEGH (Austria) and Zeebrugge (Belgium) both experienced a substantial decline in churn ratio in 2017, with the decline at Zeebrugge (Belgium) linked to a reduced output of North Sea fields and decreasing usage of Norwegian gas in favour of cheaper supply from Russia.

Notably, PSV (Italy) experienced a substantial growth in churn rate (162.26%), with Snam SpA attempting to leverage the multiple sources of supply, established national grid and well-developed infrastructure in order to increase the liquidity, and ultimately the importance of the Italian gas market. ETF (Denmark) also exhibited a substantial growth in churn ratio (69.37%), however this phenomenon is likely due to the loss of over 90% of Denmark's indigenous production at the Tyra field, which has been closed due to required infrastructure updates. Other hubs, such as NCG (Germany), Gaspool (Germany), PEG-Nord (France) and VTP Gaz System (Poland) have increased in churn ratio, indicating an overall increase in the liquidity of the European gas market. However, none of the aforementioned hubs exhibit a churn ratio above five, indicating a low degree

of hub development. Finally, hubs such as PEG-TRS (France) and PVB (Spain) have exhibited large percentage growth in churn ratios, although this is from a very low base, and the hubs are still considered illiquid and have marginal commercial importance.

2.4.4. Market Participants

The number of market participants engaged at each trading hub is indicative of the development and commercial success of a market, as it provides information pertaining to the barriers to market entry of new participants, whilst also providing information on ease of interaction for incumbent market participants. In an ideal scenario, the type of each active market participant would also be recorded, however, due to data availability, the total number of market participants are recorded at each hub. As a general rule, a larger number of active market participants increases market competition, lowering the bid-offer spread, increasing market depth and ultimately reducing susceptibility to market manipulation.

Both TTF (Netherlands) and NBP (UK) were considered to be substantially further developed in terms of market participation in 2014, which is confirmed by the large churn ratios and mean traded volumes observed in Table 2.2 and Table 2.3. Whilst TTF (Netherlands) and NBP (UK) have been usurped in the number of registered participants by NCG (Germany) and Gaspool (Germany), Neumann and Cullmann (2012) notes that the large number of market participants in Germany is a function of the vast number of subordinated network operators.

PSV (Italy) and PEG-Nord (France) have seen substantial development in the number of market participants, with other markets, such as VTP Gaz System (Poland), PEG-TRS (France), ETF (Denmark), and PVB (Spain) displaying large growth rates in market participation, albeit from a low base. These results suggest that although VTP Gaz System (Poland), ETF (Denmark), PEG-TRS (France) and PVB (Spain) have exhibited growth, they cannot yet be considered as deep, transparent and liquid natural gas hubs.

Table 2.4. Market Participants

Number of registered market participants at each measured hub in both 2014 (Heather 2017) and 2017. This includes companies both registered to trade and registered as shippers, rounded to the nearest five participants. The data was collated from a range of sources, with no standardised calculation methodology.

Hub	Participants (2014)	Participants (2017)
NCG (Germany)	90	502
Gaspool (Germany)	105	230
PSV (Italy)	118	190
NBP (UK)	200	190
TTF (Netherlands)	130	140
PEG-Nord (France)	55	120
Zeebrugge (Belgium)	82	115
PVB (Spain)	70	105
VTP Gaz System (Poland)	58	80
PEG-TRS (France)	37	65
ETF (Denmark)	43	60
CEGH (Austria)	53	30

2.5. Results

The following section reports the main results. Firstly, the in-strength estimations are presented, followed by out-strength estimations. The net-strength estimations are presented, followed by the results relative to global connectivity. Finally, the network representations are discussed.

2.5.1. In-Strength Estimations

As per the Methodology, in-strength estimations detail the degree to which a given hub's price returns are impacted by price signals from other hubs within the network. Exhibition of a high mean in-strength indicates that hub pricing is highly impacted by other European gas prices. Conversely, a low mean in-strength indicates that hub pricing is marginally impacted by price formation in other markets, instead, it is determined by local factors.

Upon estimation, the day-ahead market which was least impacted by other European natural gas prices, as characterised by the lowest mean in-strength causality (0.179) was

PEG-TRS (France). PEG-TRS (France) displays the characteristics typical of an emerging hub, with a Churn Ratio of 1.10 (Table 2.3) and a mean traded volume of 37.57MCM/d (Table 2.2). The low levels of liquidity and physical interconnection of PEG-TRS are probable explanations for susceptibility to local supply and demand imbalances.

Table 2.5. In-Strength Estimations

The mean in-strength causality, mean price (€/MWh) and number of physical connections (Connect.) of each gas trading hub measured in the sample.

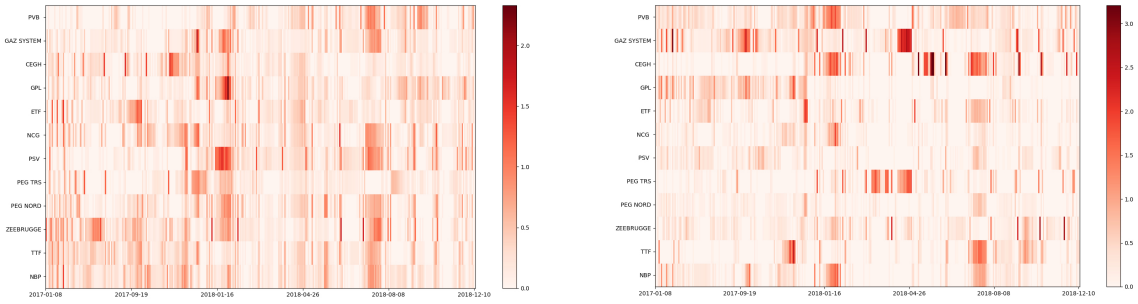
Hub	Avg. In-Strength	Avg. Price	Connect.
Zeebrugge (Belgium)	0.284	19.51	5
PSV (Italy)	0.268	21.86	3
NBP (UK)	0.263	20.00	2
NCG (Germany)	0.262	19.93	8
TTF (Netherlands)	0.244	19.68	4
Gaspool (Germany)	0.239	19.68	5
PEG-Nord (France)	0.236	19.78	4
ETF (Denmark)	0.219	18.86	2
CEGH (Austria)	0.209	20.19	2
VTP Gaz System (Poland)	0.205	21.18	1
PVB (Spain)	0.190	22.54	1
PEG-TRS (France)	0.179	21.45	2

The second lowest mean in-strength (0.190) was recorded at PVB (Spain), which is physically connected to PEG-TRS (France) at VIP-Pirineos. The low mean in-strength causality exhibited at PVB (Spain) can be attributed to a number of concomitant factors, including physical network externality, competitive pipeline import pricing from North Africa and substantial regasification capacity. This is further discussed in Section 2.6.1.

The largest mean in-strength causality (0.284) was recorded at Zeebrugge (Belgium), indicating that, on average, it is mostly Granger caused by other gas prices. Further to this, the in-strength variance was lower than the sample average, indicating that day ahead pricing at Zeebrugge (Belgium) is consistently influenced by other European gas prices. This is intuitive, as Zeebrugge (Belgium) is physically connected to five other large, liquid trading hubs.

It is possible to observe a number of periods in which individual node in-strength values increase within the sample period (Figure 2.1a). However no clear patterns emerge

with regard to average node in-strength throughout the period sampled. This implies that European gas price fluctuations held a higher degree of dependence upon variations in other European prices at certain periods within the sample, characterised by the intense colours depicted in Figure 2.1a. Meanwhile, there is no trend indicating an increasing degree of price harmonisation across the period. This phenomenon is further investigated through the utilization of a network density term (Section 2.5.4).



(a) In-strengths **(b)** Out-strengths

Figure 2.1. European Price In-Strength and Out-Strength values

Behaviour of the 12 European price in-strength and out-strength values between 2016-2018. Darker colours indicate larger values, as indicated on the colour bar. Each unit on the y-axis represents one of the 12 markets.

2.5.2. Out-Strength Estimations

Out-strength estimations detail the degree to which a given market's pricing impacts price signals at other markets within the network. Exhibition of a high mean out-strength indicates that the market holds a high degree of importance in price formation of gas prices at other European hubs. Conversely, a low mean out-strength indicates that the market holds a limited capacity to impact price formation at other hubs.

The hub price with both the highest mean out-strength (0.412) and standard deviation (0.594) was VTP Gaz System (Poland). This indicates that Polish gas pricing showed a large potential of influencing other European gas prices under certain market conditions, specifically in periods of high day ahead gas demand within continental Europe (Figure 2.1). This is unsurprising, as Poland holds one of the main transit routes for European gas imported from Russia (Yamal-EuRoPoL), carrying one-fifth of all Russian gas imports to

Europe.

Table 2.6. Out-Strength Estimations

The mean out-strength causality, mean price (€/MWh) and number of physical connections (Connect.) of each gas trading hub measured in the sample.

Hub	Avg. Out-Strength	Avg. Price	Connect.
VTP Gaz System (Poland)	0.412	21.18	1
PVB (Spain)	0.347	22.54	1
CEGH (Austria)	0.343	20.19	2
Zeebrugge (Belgium)	0.227	19.51	5
Gaspool (Germany)	0.221	19.68	5
PEG-TRS (France)	0.215	21.45	2
NBP (UK)	0.213	20.00	2
TTF (Netherlands)	0.212	19.68	4
ETF (Denmark)	0.207	18.86	2
PSV (Italy)	0.156	21.86	3
NCG (Germany)	0.138	19.93	8
PEG-Nord (France)	0.112	19.78	4

PVB (Spain) also exhibits a high mean out-strength causality, indicating that Spanish gas pricing shows potential to influence other European gas prices under specific market conditions, albeit these conditions are considerably different (Figure 2.1b, Figure 2.2) to those exhibited by VTP Gaz System (Poland), with the relative out-strengths holding a correlation of -0.24. When considering the fundamental drivers of European gas pricing, flexible oil product-indexed pipeline gas from Russia or Norway and Liquified Natural Gas (LNG), the global context in which these markets operate must also be considered. The ability of pipeline gas or LNG to influence European pricing, characterised by out-strength values, is largely dependent on Asian hub pricing, as a substantial premium at Asian hub prices tends to draw spot LNG cargoes away from delivery to Europe, leaving Russian or Norwegian swing capacity to dictate European day ahead pricing. Conversely, weak demand in Asia allows for increased spot LNG delivery to Europe, displacing Russian or Norwegian gas as European price setters (Figure 2.2).

Conversely, the hub which recorded the lowest mean out-strength (0.112), therefore the lowest ability to impact European gas prices, was PEG-Nord (France). Further to this, PEG-Nord also displayed the lowest out-strength variance of the sample, indicating

that PEG-Nord consistently had a marginal capacity to influence other European day ahead gas prices. This is in contrast to PEG-TRS (France), which held a considerably larger ability to influence European gas pricing throughout the sample, recording a mean out-strength of 0.215. The bifurcation of the French gas market is further discussed in Section 2.6.

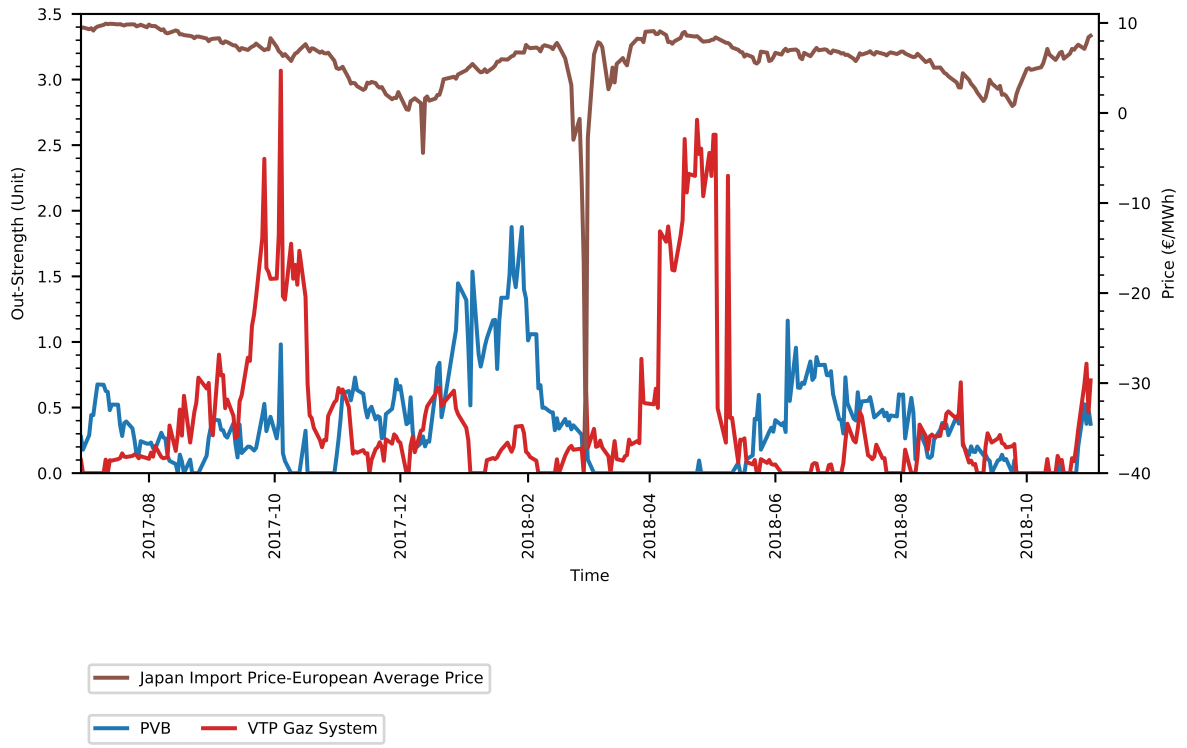


Figure 2.2. Japan-Europe Premium

A visual representation of the systematic premium between Japanese Import Price and the average European gas price (€/MWh), PVB (Spain) out-strength causality and VTP Gaz System (Poland) out-strength causality. When the Japan-Europe premium is large, LNG cargoes are attracted to Asia, allowing VTP Gaz System (Poland) to act as a price setter in European gas, characterised by high out-strength causality values (Correlation = 0.141). Conversely, when the Japan-Europe premium is low, LNG cargoes are attracted to Spain's large regasification capacity, increasing PVB's out-strength causality. (Correlation = -0.058)

2.5.3. Net-strength Estimations

Following estimations of both in-strengths and out-strengths, the net-strengths are estimated. A large, positive net-strength indicates that the trading hub shows a high potential for influencing prices at other hubs, acting as a 'price setter' within European

day-ahead gas pricing. Conversely, hubs which exhibit negative net-strengths indicates that pricing at the hub is mostly Granger caused by pricing at other hubs within the sample, acting as a 'price taker'.

Table 2.7. Net-strength Estimations

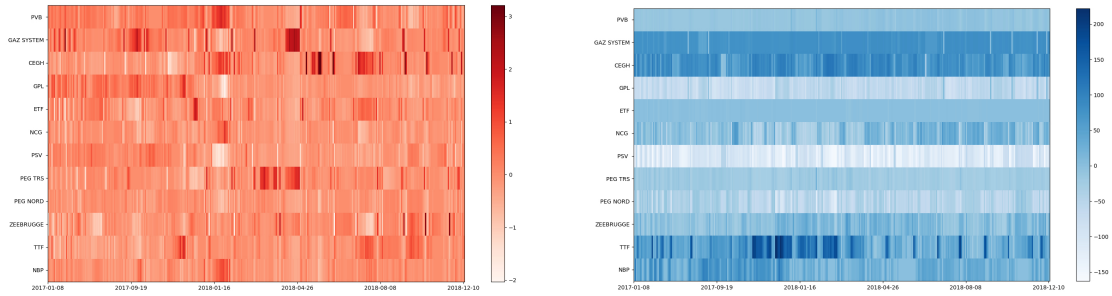
The mean net-strength causality, variance of net-strength causality and mean price (€/MWh) of each gas trading hub measured in the sample.

Hub	Avg. Net-Strength	Variance	Avg. Price
VTP Gaz System (Poland)	0.206	0.663	21.18
PVB (Spain)	0.158	0.449	22.54
CEGH (Austria)	0.134	0.699	20.19
PEG-TRS (France)	0.036	0.510	21.45
ETF (Denmark)	-0.012	0.450	18.86
Gaspool (Germany)	-0.018	0.429	19.68
TTF (Netherlands)	-0.032	0.394	19.68
NBP (UK)	-0.050	0.286	20.00
Zeebrugge (Belgium)	-0.057	0.513	19.51
PSV (Italy)	-0.113	0.394	21.86
NCG (Germany)	-0.124	0.302	19.93
PEG-Nord (France)	-0.124	0.253	19.78

The hub which recorded the largest mean net-strength (0.206) is VTP Gaz System (Poland), followed by 0.158 recorded at PVB (Spain) and 0.134 recorded at CEGH (Austria). This indicates that the aforementioned hubs exhibit the highest degree of 'price setter' behaviour within European gas markets. When considered in isolation, each hub has a number of unique characteristics, however the aforementioned hubs hold a commonality of physical proximity to gas injection into the European network. Whether it is LNG regasification capacity at PVB (Spain), or major pipeline (Yamal-EuRoPoL,Soyuz) terminals from production fields in Russia (VTP Gaz System, CEGH), the hubs which display the positive net-strengths acted as net exporters to other hubs, with the exception of PVB (Spain).

VTP Gaz System (Poland), which acts as the first European landfall of one-fifth of Russian exports transited by the Yamal-EuRoPoL pipeline, had a mean net export of 75.60MCM/d exclusively to Gaspool (Germany). As such, VTP Gaz System's prominence as a 'price setter' within the study is unsurprising. This hypothesis is further investigated

through the analysis of adjacency matrices in Section 2.6.5.



(a) Net-strengths **(b)** Net export volumes

Figure 2.3. European Price Net-Strength Values

Left: The 12 European price net-strength values between 2016 and 2018. Darker colours indicate larger influential values, whereas lighter colours indicate more prominent values. The units of the y-axis represent the 12 markets. Right: Net export volumes from each trading hub between 2016-2018, expressed in million cubic meters per day (MCM/d). Darker colours indicate large net export volumes, whilst lighter colours indicate large net import volumes, as per the colour bar.

Extended periods of disconnection from the European pricing system, further discussed in Section 2.6.4, were observed at PVB (Spain), culminating in a mean net-strength value of 0.158. Critically, PVB (Spain) also displayed the largest mean price of the sample, indicating that an improved degree of integration into the European pricing system could be beneficial for domestic consumers.

CEGH (Austria) recorded the third largest mean Net-Strength (0.134), indicating CEGH's importance as a 'price setter' within European gas markets. Given Baumgarten's importance in servicing the Southern Gas Corridor and the central European market, namely Germany (NCG and Gaspool) and Italy (PSV), this is unsurprising. CEGH (Austria) also held the largest mean net-export value (Figure 2.3) of 77.84MCM/d, with a mean daily export of 81.36MCM/d to PSV (Italy). That said, pricing at CEGH (Austria) only achieved statistical significance in influencing pricing at PSV (Italy) for 18 days of the sample (5.19%).

From this, a number of conclusions can be drawn; Firstly, hubs which have access to substantial physical injection into the European network hold influence over European natural gas pricing. When interpreting Figure 2.3a, it becomes apparent that a determin-

ing characteristic in the pricing of European gas is exogenous to the study, with Asian hub pricing most likely playing a role in determining whether LNG or Russian and Norwegian swing capacity acts as the price setter in the European day ahead market. This is intuitive, as high prices at Asian hubs act to draw LNG away from European regasification plants, allowing Russian and Norwegian swing capacity to act as the European day ahead price setter, evidenced by the large net-strengths exhibited at VTP Gaz System (Poland) and CEGH (Austria).

2.5.4. Network Density of European Gas Prices

Network Density, as defined in Section 2.3.2, remained around the relatively low mean value of 0.141 throughout the sample time frame, with a standard deviation of 0.074. The maximal value of network density is displayed as the largest peak in (Figure 2.4), which displays the behaviour of day ahead European gas pricing throughout the period 2016-2018. Through observing the causal interactions between EU natural gas prices, the model provides a dynamic quantification of European gas market price integration, the system's network density, which exhibits stochastic behaviour (Figure 2.4).

Subsequently, the distribution of the Network Density term was analysed, with the z-score calculated in order to understand at which observations of time the network density was 3 standard deviations above the mean ($Z>1.96$, $p<0.05$). Values lying outside the confidence bounds imply abnormal market behaviour at the observation, displaying an unusually large network density term.

The abnormally large network density terms coincided with factors such as pipeline capacity reductions, seismic activity and uncharacteristically cold weather. Periods of lower than normal connectivity occurred as frequently as those of greater than normal connectivity, indicating that long run gas market integration is not substantially pronounced. Results relative to global connection density display the occurrence of a large spike in January 2018, reaching a magnitude of 0.45 on 29/01/2018. This can be attributed to the pipeline disruption which occurred between Oude Statenzijl (Netherlands) and Bunde (Germany) between 27/01/2018 and 28/01/2018. Further to this, an initial interconnec-

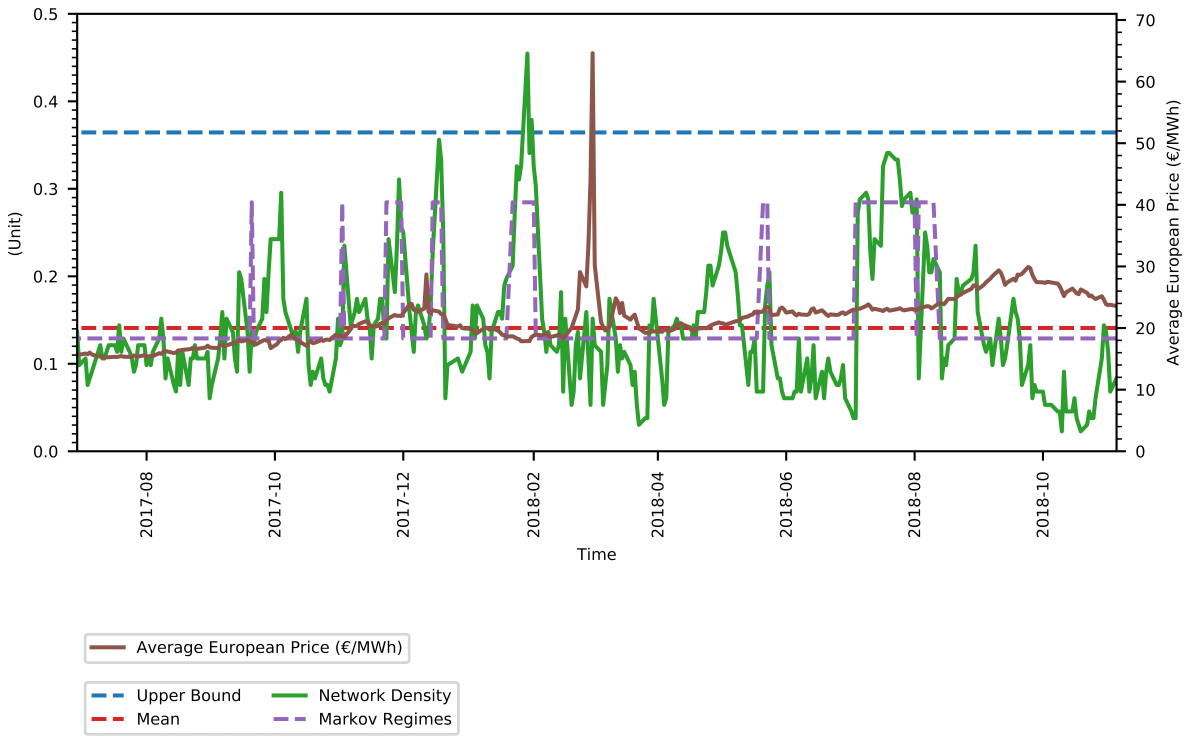


Figure 2.4. Network Density

Network density of 12 European day ahead gas prices between 2016-2018. The dotted horizontal lines represent the mean and upper confidence bounds (Z>1.96). Time is on the x-axis, with the network density value displayed on the y axis. The two markov regimes are depicted by the blue dotted line, with the respective coefficients of 0.129 and 0.285.

tivity peak of c.ca 0.30 can be observed within the first 69 model shifts (04/10/2017).

The following peaks reached 0.31 and 0.36 respectively, with the mean level of day ahead gas market network density remaining c.ca 0.15 throughout the sample. The longest peak in global density was recorded during the period 05/07/2018 to 02/08/2018, with an over basal increase of 194%. When referring to Figure 2.4, there is a noticeable lack of correlation between peaks in the average European day ahead gas price and the network density term, indicating that abnormal price phenomena did not produce the network density peaks observed.

The Markov regime switching model was subsequently applied to the connection density time series to detect the exact points in which connection density peaks occur. Observed jumps in global connectivity were considered as changes to another regime, which occurred seven times throughout the sample, displayed by (Figure 2.4).

An ARIMA (1,0,0) model was subsequently applied to the network density term in order to investigate whether the network density, and by extension, market integration of European day ahead gas markets, represents a stochastic process. Results show that a random walk process could have produced the network density time series, therefore an AR(1) model is an appropriate representation of European network density. From this, it can be surmised that the most accurate representation of the network density term at time t is network density at time $t-1$.

2.5.5. Network Representation of European Gas Prices

The system of European natural gas price interactions, summarised throughout the global connection density series by computing the degree of Granger causality between gas price returns, produces a dynamic network. Two examples of the network estimation are displayed in (Figure 2.5), with the adjacency matrix and network graph shown at the points at which the largest and smallest network density terms were observed. From this visualisation, it is clear that a substantial difference in the intensity and number of Granger causal interactions were observed throughout the sample, evidenced by the number of arrows displayed within the network graphs, and intensity displayed by the adjacency matrices.

2.6. Discussion

The following section discusses the findings reported in Section 2.5. Firstly, the in-strength estimations are discussed, followed by a discussion of the out-strength estimations. This is followed by discussion of the net-strength estimations, culminating in a discussion of the physical and non-physical reasons for market decouplings. Finally, Network Density is discussed.

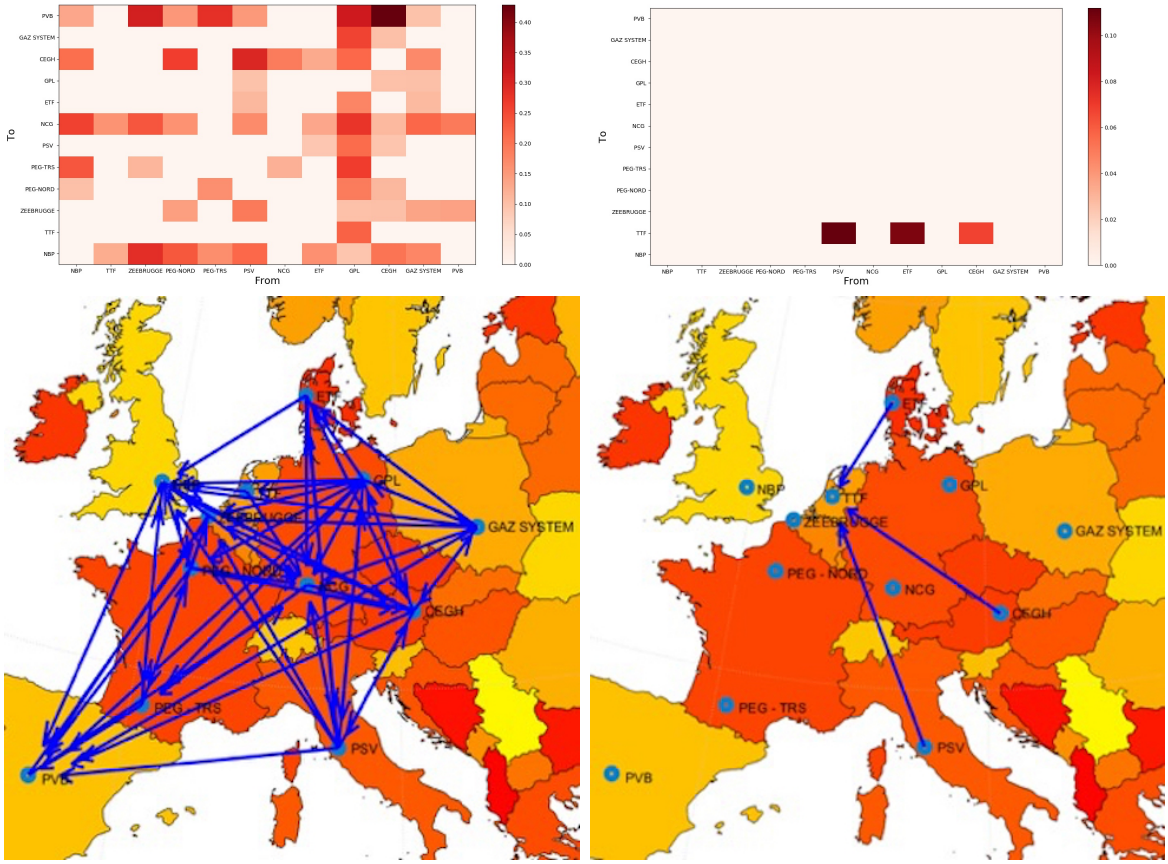


Figure 2.5. Network Graphs and Adjacency Matrices

The network graphs (lower panel) and adjacency matrices (upper panel) relative to the highest and lowest observations of network density are presented. The highest network density term was observed on 29/01/2018 (left panel), with the lowest network density term observed on 10/10/2018 (right panel). The direction of the arrow corresponds to the direction of the Granger causality between the variations of European gas prices, whilst the adjacency matrices (upper panel) depict the intensity of Granger causal interactions between markets. Each matrix entry displays the intensity of Granger-causality between the price variations observed in two sample countries, depicted by the adjacent colour scale.

2.6.1. In-Strengths

The mean in-strength causality of 0.190 exhibited at PVB (Spain) (Figure 2.5) can be attributed to a number of concomitant factors, including physical network externality, competitive pipeline import pricing from North Africa and substantial re-gasification capacity. The low number and physical capacity of interconnections to the European market (16.89MCM/d) makes Spain reliant on pipeline imports from North Africa to meet its domestic demand, with Algeria supplying 56.8% of total import demand in 2017 (IEA,

2018). Further to this, Spain currently holds 39% of total European regasification capacity, enabling LNG imports to constitute a substantial proportion of the domestic gas mixture, with term contracts in place with Norway, Nigeria and Qatar. As such, PVB (Spain) exhibits a low mean in-strength, as day ahead gas pricing is determined by a combination of LNG imports and North African pipeline imports to the Iberian Peninsula. This indicates that the Spanish gas market is largely driven by domestic supply and demand conditions, as opposed to pan-European supply and demand balances.

When considering the Iberian Peninsula's physical network externality, the systematic disparity of €1.08/MWh between the adjacent trading regions of PEG-TRS (France) and PVB (Spain) is comprehensible, since PEG-TRS (France) has a low reliance on gas for electricity generation, as a consequence of the high French nuclear baseload capacity. As the aforementioned trading regions are connected by 16.89 MCM/d of pipeline capacity at VIP Pirineos, which is rarely utilized at maximal capacity (Figure 2.6), the premia paid at PVB (Spain) appears to be a consequence of a lack of pipeline contract flexibility. According to Heather and Petrovich (2017), a lack of readily available transmission capacity between PEG-TRS (France) and PVB (Spain) for day ahead market participants who aren't engaged in term contracts between the markets has inhibited reduction of the premium paid at PVB (Spain), and subsequently the Iberian peninsula's integration into the European natural gas network.

Both PEG-TRS (France) and PVB (Spain) display the characteristics typical of emerging hubs, detailed in Table 2.2 and Table 2.3. As such, low levels of liquidity may adversely impact the degree of cross border trade, characterised by the pipeline utilization rates displayed at VIP Pirineos (Figure 2.6). This has the resultant impact of decreasing arbitrage flows, with arbitrage in the day-ahead market only occurring when the PVB (Spain) - PEG-TRS (France) premium becomes uncharacteristically large (Figure 2.6), reducing integration of the Iberian Peninsula into the European gas system. This is confirmed through analysis of the adjacency matrices, as PEG-TRS (France) Granger caused pricing at PVB (Spain) (above a 95% confidence) in 136 of 347 (39.19%) of sampled days.

Broadly, when physical network positioning of different hubs is considered, hubs which

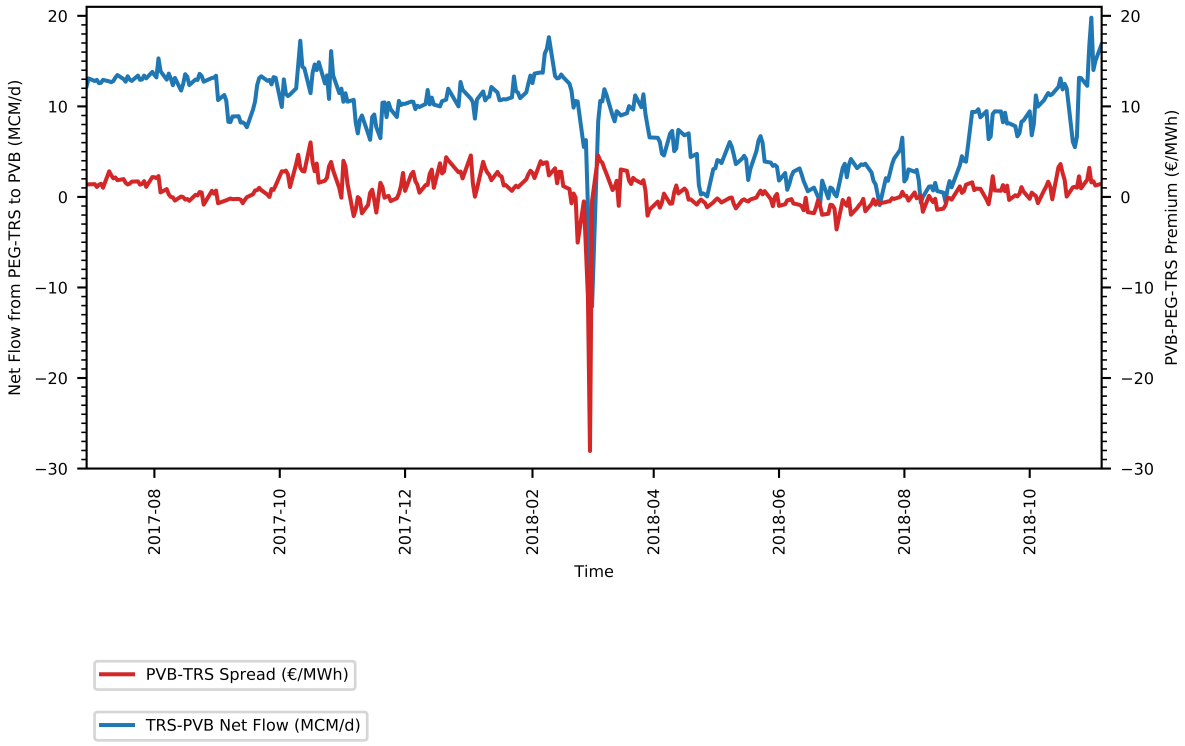


Figure 2.6. Systematic Price Difference

A visual representation of the systematic price difference between PVB (Spain) and PEG TRS (France), combined with the net flow in million cubic meters per day (MCM/d) through the VIP Pirineos pipeline, connecting the trading regions.

exhibit physical network externality (Figure 2.5, such as PVB (Spain), PEG-TRS (France) and VTP Gaz System (Poland), display a substantially larger mean day ahead price than hubs which exhibit network centrality (Table 2.5). Furthermore, these hubs are typically characterised by a low mean in-strength causality, indicating that the day-ahead pricing is marginally impacted by pricing at other European hubs.

A notable exception is Gaspool (Germany), which holds a relatively central network position, directly connected to five other trading hubs by pipeline, yet it exhibits a low degree of prominence, combined with a high variance of in-strength. This can be accounted for by the concentration of Russian and Norwegian gas imports at Gaspool (Germany), with day ahead price returns at Gaspool (Germany) showing a lower dependence on other European markets, as the relationship with Russian and Norwegian day ahead pricing supersedes this. The high degree of in-strength variance confirms this, as under specific market conditions, pricing at Gaspool (Germany) can be highly dependent on other Eu-

ropean day ahead prices, which coincides with a high network density term, however under normal market functioning, variables beyond the scope of the study, Russian and Norwegian day ahead pricing, have a large impact on day ahead pricing at Gaspool (Germany).

The day-ahead gas price which displayed the largest mean in-strength causality was Zeebrugge (Belgium), indicating that, on average, it is mostly Granger caused by other gas prices. Further to this, the in-strength variance was relatively low, indicating that day ahead pricing at Zeebrugge (Belgium) is consistently influenced by other European natural gas prices. This result is intuitive, as Zeebrugge (Belgium) holds a relatively central network position, directly connected to five other trading hubs by pipeline. Furthermore, Zeebrugge's status as the oldest trading hub within continental Europe, physical proximity to other large, liquid trading hubs (TTF, PEG-Nord, NBP) and status as a net-exporter of gas throughout the sample compound the characteristic of network centrality, increasing the degree to which Zeebrugge acts as a price taker within the day ahead market.

2.6.2. Out-Strengths

The largest mean out-strength of 0.412 was recorded at VTP Gaz System (Poland). This is intuitive, as Poland holds one of the main transit routes for European bound gas from Russia, with the Yamal-EuRoPoL pipeline carrying one-fifth of all Russian gas imports to Europe. Therefore, Poland is one of the first points at which Russian day ahead capacity is priced into the sample. As such, in periods of market stress, characterised by an uncharacteristically large demand, VTP Gaz System (Poland) is the market which, due to the physical location of the gas fields which serve European day ahead supply (Russia), exerts a high degree of influence on European pricing, characterised by a high mean and variance of out-strength causality.

Whilst Gaz System (Poland) holds 112.39 MCM/d of interconnection capacity to Gaspool (Germany), day ahead gas pricing within Poland holds a systematic premium to pricing at Gaspool (Germany), with 18.44% of adjacency matrices showing day ahead pricing at Gaspool (Germany) is Granger caused by VTP Gaz System (Poland) at the 5% significance level. Meanwhile, pricing at Gaspool (Germany) influenced pricing at VTP

Gaz System (Poland) for 24.78% of sampled days.

Surprisingly, throughout the sample, despite the systematic €1.50/MWh premia at VTP Gaz System (Poland), negligible arbitrage flows occurred from Gaspool (Germany) to VTP Gaz System (Poland). Conversely, Gaz System (Poland) acted as a net exporter to Gaspool (Germany), exporting an average of 75.60 MCM/d (67.27% pipeline utilization) throughout the sample. This supports Heather and Petrovich (2017)'s assertion that capacity constraints are not the primary factor in Poland's systematic price premium in the day ahead market, with pricing determined by local supply and demand imbalances, inferring that the country's gas system is not fully liberalised, or integrated into the European gas network.

PVB (Spain) also exhibits a high mean out-strength causality (0.347, Table 2.6), indicating that Spanish gas pricing shows potential to influence other European gas prices under specific conditions. That said, these conditions are considerably different (Figure 2.1b) to those exhibited by VTP Gaz System (Poland), with the relative out-strengths holding a correlation of -0.24. When considering the fundamental drivers of European natural gas pricing, flexible oil product-indexed pipeline gas from Russia and Norway or Liquified Natural Gas (LNG), the global context in which these markets operate must also be considered. The ability of pipeline gas or LNG to influence European pricing, characterised by out-strength values, is largely dependent on Asian hub pricing, as a substantial premium at Asian hub prices tends to draw spot LNG cargoes away from delivery to Europe, leaving Russian or Norwegian swing capacity to dictate European day ahead pricing. Conversely, weak demand in Asia allows for increased spot LNG delivery to Europe, displacing Russian or Norwegian gas as European price setters (Figure 2.2).

Considering the above, the out-strengths displayed in (Figure 2.1b) are intuitive, as Russian day ahead capacity, which is initially priced into the sample at VTP Gaz System (Poland) or CEGH (Austria) holds a negative correlation (-0.24 and -0.37 respectively) with countries with large LNG regasification facilities, such as PVB (Spain). As the Iberian Peninsula accounts for 44% of European LNG regasification capacity, combined with landfall of pipelines from Algeria (Medgaz) and Morocco (Maghreb Europe Gas),

PVB (Spain) has considerable influence on European day ahead pricing when Asian LNG demand is weak.

Conversely, the hub which recorded the lowest mean out-strength (0.112, Table 2.6), therefore the least ability to impact European gas prices, was PEG-Nord (France). Further to this, PEG-Nord also displayed the lowest out-strength variance of the sample, indicating that PEG-Nord consistently had a limited capacity to influence other European day ahead gas prices.

This can be attributed to France's 70% dependence on Nuclear capacity for energy generation, lack of indigenous production following the closure of the Lacq field in 2013, and net importer status throughout the sample. When considering the bifurcation of the French gas market, a difference in the origin of gas mixture at the two hubs provides a plausible explanation for the disparity in price dynamics. As LNG is essential to meet demand within PEG-TRS (France), constituting 39% of volume (ENTSOG, 2017), PEG-TRS (France) is subject to the same global LNG dynamics as PVB (Spain), however PEG-Nord (France), which holds marginal LNG volume (3%, ENTSOG, 2017) within the import portfolio, is subject to day ahead pricing from Russia, Norway and the Netherlands.

However, when considering the North-South (PEG-Nord, PEG-TRS) division of the French gas market, an interesting disparity between the characteristics and pricing of the two trading hubs can be observed, as PEG-TRS (France) displays the sixth largest mean out-strength, whilst PEG-Nord (France) displays the lowest. Firstly, PEG-Nord's physical interconnection to Zeebrugge (Belgium), NCG (Germany) and PEG-TRS (France) provides it with a substantially larger degree of network centrality than PEG-TRS (France), which holds comparatively small interconnection capacity to PVB (Spain) and PEG-Nord (France). Furthermore, PEG Nord (France) also had a substantial average net import volume throughout the sample, (46.7MCM/d), whereas PEG-TRS (France) had low net import volumes (21.1MCM/d), most of which was delivered from PEG Nord (France). The high mean utilisation of the Liason Nord-Sud pipeline (96%) begins to explain the systematic price disparity between the two French markets, as the premium at PEG-TRS

increases as the Liason Nord-Sud pipeline utilisation rate increases beyond 95%. Furthermore, when adjacency matrices are considered, PEG-Nord Granger caused pricing at PEG-TRS on 11.82% of days sampled, whilst PEG-TRS only Granger caused PEG-Nord on 10.09% of days sampled.

On November 1^{st} 2018, the French gas market merged its two virtual trading points (VTPs), PEG-Nord and PEG-TRS, in order to form a single VTP called Point d'échange de Gaz (PEG), within a single trading region, Trading Region France (TRF). As a consequence, the systematic price difference (Figure 2.1) between PEG-Nord (France) and PEG-TRS (France) is expected to be reduced, as a single national wholesale gas market is established, with an increase in liquidity, competition, and ultimately market integration anticipated. Prior to the merger, PEG-Nord (France) was well supplied through direct pipeline linkages to Zeebrugge (Belgium) and NetConnect Germany (NCG), and LNG regasification capacity at Dunkerque and Montoir de Bretagne. Conversely, PEG-TRS held limited access to other injection sources, with LNG capacity at the Fos terminals supplying a substantial proportion of demand in the PEG-TRS (France) region.

Following the merger, there was a palpable change in dynamics of the French gas market, with a decrease in mean out-strength from 0.217 to 0.031 exhibited at PEG-TRS (France) (Figure 2.1b). This indicates that following the merger, price formation characteristics in France were similar to those exhibited at PEG-Nord (France) prior to the merger, as opposed to PEG-TRS (France). It is anticipated that these dynamics shall continue to evolve due to the expected completion of the Val de Saone and Gascogne-Midi pipeline projects, which aim to equalise gas pricing within the French domestic market.

2.6.3. Net-Strengths

Following discussion of both in-strength and out-strengths, the net-strengths, which constitutes the degree to which each hub acts as a 'price setter' or 'price taker' are discussed.

VTP Gaz System (Poland) exhibits the largest net-strength of 0.206 (Table 2.7) of the sample, which is intuitive, given that Poland acts as the first European landfall of one-fifth

of Russian export capacity. VTP Gaz System's (Poland) ability to act as a 'price setter' within central European gas pricing is confirmed through analysis of adjacency matrices. Pricing at Gaspool (Germany), NCG (Germany) and ETF (Denmark) was influenced by VTP Gaz System (Poland) in 46, 64 and 40 instances respectively.

CEGH (Austria) recorded the third largest net-strength of 0.134 (Table 2.7), exhibiting price dynamics which were largely similar to those observed at VTP Gaz System (Poland). This is unsurprising given that Baumgarten, within the CEGH (Austria) trading region, is one of the major European terminals of Russian export flows, acting as a thoroughfare for day ahead capacity to South-Eastern Europe and the central European market, namely Germany (NCG and Gaspool) and Italy (PSV). CEGH (Austria) also recorded the largest mean net export value of 77.84MCM/d, with net exports of 81.36MCM/d to PSV (Italy). However, pricing at CEGH (Austria) only achieved statistical significance in influencing pricing at PSV (Italy) 5.19% of the time.

As discussed in Section 2.5.3, PVB (Spain) exhibited the second largest Net-Strength of 0.158 (Table 2.7) due to extended periods of market disconnection. This is precipitated by a multitude of reasons, addressed in Section 2.6.4 and Section 2.6. Critically, the systematic premium paid for day-ahead gas at PVB (Spain) is not derived from infrastructure constraints, as average pipeline utilization at VIP Pirineos remained at 50.3% throughout the sample (Figure 2.6), with LNG regasification infrastructure utilization c.ca 40.0%. Of the concomitant factors restricting PVB (Spain) integrating into the European gas network, the most pertinent appears to be the lack of availability of transmission capacity for day ahead market participants who are not engaged in term contracts (Heather and Petrovich 2017). As PVB (Spain) matures as a trading hub, increasing in liquidity and churn rate, and PEG-TRS (France) is integrated into a singular French market, it is anticipated that PVB (Spain) shall become more integrated into a singular European gas market.

2.6.4. Physical and Non-Physical Reasons for Market Decouplings

As indicated in Sections 2.6.1, 2.6.2, and 2.6.3, the dislocation of certain markets from the European network, and resulting reduction in network density term, may be the result of delinkages precipitated by one of two core reasons; physical and non-physical.

When considering the physical reasons for a lack of market integration, pipeline congestion is amongst the most pertinent factors. The €1.67/MWh mean differential displayed between PEG-Nord (France) and PEG-TRS (France) (Table 2.7) may be the result of a lack of available capacity on the liason Nord-Sud pipeline which connects the two trading zones (OIES,2017), as mean capacity utilization remained at 96% throughout the sample. Whilst other factors, such as the disparity in LNG composition of the gas mixture (39% at PEG-TRS (France), 3% at PEG-Nord (France)), relative consumption volumes and network centrality may have impacted the differential, the pipeline capacity constraints on the liason Nord-Sud pipeline appears to be the most critical determinant of the differential magnitude between the two trading zones. As the two zones were merged on November 1^{st} 2018, an equilibrium value between the two hub prices is anticipated, increasing the degree of price integration within European gas pricing.

However, given the absence of capacity constraints surrounding PVB (Spain) and VTP Gaz System (Poland), non-physical factors appear to be pertinent in the lack of price integration, and by extension market integration within European day ahead gas markets. These hubs share similarity in a number of characteristics, namely a lack of indigenous gas production, the capacity to inject or transit large volumes of gas into the European market, and physical network externality.

Extended periods of disconnection from the European pricing system were observed at PVB (Spain), leading Platts to declare the Iberian Peninsula as a "gas island". Whilst pipeline utilization remained at 50.3% throughout the sample, a persistent premium (Figure 2.6) over PEG-TRS (France) was paid at PVB (Spain), indicating that arbitrage forces are not fully operating between the two hubs, with pricing at PVB (Spain) determined by

domestic supply and demand conditions. Heather and Petrovich (2017) provide a plausible explanation, arguing that a lack of available transmission capacity for day ahead market participants who are not engaged in term contracts has an adverse impact on capacity utilization, and consequently arbitrage flows. Further to this, low levels of liquidity, displayed in Tables 2.2 and 2.3, may discourage market participants from entering into space arbitrage between the two hubs.

It is anticipated that following the merger of PEG-TRS (France) and PEG-Nord (France) into a singular French gas market, PEG, liquidity shall increase, and increased arbitrage flows shall decrease the premium paid at PVB (Spain), increasing its integration into the European gas market. Further to this, the expected completion of the Val de Saone and Gascogne-Midi pipeline projects should alleviate capacity restrictions on the Liaison Nord-Sud pipeline and effectively eliminate the premium paid within the day ahead market in the region formerly identified as PEG-TRS (France).

Furthermore, VTP Gaz System's (Poland) systematic price premium over Gaspool (Germany) stipulates that arbitrage forces should be in effect, however VTP Gaz System (Poland) had a mean net export of 75.60 MCM/d to Gaspool (Germany) throughout the sample, indicative of a non-physical barrier to market integration and price integration. Heather and Petrovich (2017) suggest that a low degree of Polish internal market liberalisation is the primary factor in the lack of integration into the European gas market, with pricing largely determined by local supply and demand balance. Notably, of the 12 hubs sampled, VTP Gaz System (Poland) is the only hub which is denominated in an emerging market currency, Polish Zloty. It is plausible that market participants impound foreign exchange risk into pricing at VTP Gaz System (Poland), which is less of a consideration amongst Euro denominated hubs within continental Europe or Sterling denominated hubs (NBP and TTF).

2.6.5. Network Density and The Third Energy Package

To what degree has the Third Energy Package impacted gas markets? The European day ahead gas market network density term remained at a relatively low mean level

of 0.141 throughout the sample, indicating a low degree of price integration within the European market. Further to this, the results of applying an ARIMA (1,0,0) model to the network density term indicate that network density exhibits a random walk process, indicating that the expected value of change in network density is best characterised by white noise. As such, the best estimate of network density at time t is represented by the value produced at time $t-1$. Additionally, the number and magnitudes of both abnormally positive and negative network density behaviours implies a lack of increasing, sustained market integration amongst day ahead prices displayed at European natural gas trading hubs. Taken in unison, these findings provide a clear indication that price integration within the European day ahead gas market did not exhibit a substantial, sustained increase between 2016 and 2018.

The regime model applied to the network density term provides evidence of two distinct regimes within the European day ahead gas market, with the peaks in network density of the European market coinciding with pipeline disruptions, hub disruptions and seismic activity. On 12/12/2017, an explosion at Baumgarten, Austria forced the operator to close the facility and physical trading to cease at CEGH (Austria), leading Italy to declare a national emergency regarding gas supplies. Therefore, it is likely that the day ahead market responded to this outage through efficiently impounding the loss of capacity into the market, characterised by a spike in network density until 19/12/2017. Additional evidence for this hypothesis is obtained through the analysis of out-strengths (Figure 2.1b) and net-strengths (Figure 2.3), where CEGH (Austria) uncharacteristically held no capacity to influence pricing at other European hubs (sample mean net-strength of 0.134), acting as a strong 'price taker' at this time. Furthermore, PSV (Italy), which has a heavy import reliance on CEGH (Austria) and subsequently declared a national state of emergency, temporarily transitioned to 'price setter' behaviour following the outage, returning to 'price taking' behaviour on 20/12/2017.

A possible cause of the large peak in January 2018 could possibly be attributed to the day-ahead demand impact of extreme temperatures throughout Europe, combined with a pipeline disruption between Oude Statenzijl (Netherlands) and Bunde (Germany)

between 27/01/2018 and 28/01/2018. An increase in both in-strength and out-strength was observed across different markets in late January 2018, however, through analysis of the relationship between the average European heating/cooling days, a correlation value of 0.102 was not able to explain the spike in connection density, eliminating the possibility of the impact of the low temperatures characteristic of Storm Emma and the ensuing formal gas deficit warning issued within the UK. Therefore, it is more probable that the pipeline disruption observed between the Netherlands and Germany impacted trade (Figure 2.4), integrating pricing and producing a spike in the network density term.

The average in-strength exhibited stochastic behaviour throughout the period sampled, indicating that price returns at any given hub were not subject to a larger amount of influence from other European natural gas markets at the end of the time period as opposed to the beginning, suggesting no sustained increase in price integration amongst European day ahead gas prices. This is confirmed by the network density term, which exhibited stochastic behaviour throughout the sample period. Taking these findings in unison, it becomes clear that, although periods of high network integration did occur, there was no sustained increase in price integration amongst European day ahead gas prices.

2.7. Conclusions

This study applies graph theory in order to model the interactions between 12 European day ahead gas markets during the period 2016-2018. The interrelations between the observed European gas markets is measured through the system's network density term (3.3), which characterises the quantity of causal interactions within the system at a given point in time. The novelty of this work lies in the application of network theory and Granger-Geweke causality to gas markets, disentangling the dynamic relationships between markets and measuring market integration.

The methodology is verified through the identification of historical exogeneous events and subsequently observing the dynamics of connectivity measures, such as in-strengths or out-strengths relative to these occurrences. In the cases of the Baumgarten terminal

explosion on 12/12/2017 and Oude Statenzijl-Bunde disruption on the 27/01/2018, the market dynamics display the anticipated dynamics, with CEGH (Austria) uncharacteristically becoming a strong price taker and a reduction in TTF's (Netherlands) ability to act as a price setter over this period.

Analysis of the network density term resulted in the detection of a two regime Markov model, with an abnormally large spike of c.ca 0.45 observed on 29/01/2018, which possibly reflects the impact of a number of factors, including storm Emma, the UK formal gas deficit warning, the decision to reduce production at the Groningen field due to seismic activity and pipeline capacity reduction between Oude Statenzijl (Netherlands) and Bunde (Germany).

Aside from the two regimes detected by the Markov model, abnormal behaviour in connectivity was evenly distributed, with abnormal positive and negative changes in connectivity essentially similar in number and magnitude. Through application of an ARIMA (1,0,0) model to the network density term, it becomes apparent that network density is a random walk process, indicating that the expected change in network density is best characterised by white noise. From this, the assertion that the best estimate of network density at time t is represented by the value observed at time $t-1$ can be drawn. Taken in unison, these findings imply a lack of sustained increase in market integration amongst European day ahead gas prices, indicating that attainment of a high degree of gas market integration within Europe appears to be some distance away.

Our results confirm that day ahead gas markets in Europe are developing, however each hub holds unique characteristics, providing different rates of development and integration. The low number of physical barriers to price integration within the European day ahead gas market implies that the Third Energy Package's focus on national gas market integration through cross-border mechanisms has been broadly successful, reducing pipeline capacity constraints throughout the European network. Conversely, the detection of non-physical barriers to trade suggests that the liberalisation and development of certain national gas markets (VTP Gaz System, PVB) is yet to be fully achieved, inferring that improvements in technical arrangements are required. It is imperative that

system operators throughout Europe engage in providing full integration of day ahead gas markets through a number of mediums, primarily market concentration, market design, regulation and security of supply, all of which are crucial to the development of a single European market. The aim of elimination of physical barriers to integration, in conjunction with improved legislative integration, implies the convergence of day ahead gas prices toward a single European price (in the absence of transmission costs), however this also holds the additional benefit of reducing pipeline congestion, increasing usage efficiency, and reducing the market power of actors within national gas markets.

Further to this, the methodology elucidates the importance of employing a dynamic model, which is capable of monitoring the time-varying interactions within a network structure, as each market has a continually evolving ability to influence (out-strength) and be influenced by (in-strength) other markets. The model, which is able to observe the dynamic nature of interactions, is also able to detect underlying changes in market integration, which can be driven by either exogenous (Baumgarten) or endogenous events (Oude-Stanzijl-Bunde). As such, it can be considered a suitable tool for measurement of the both market dynamics and the degree of market integration.

Chapter 3

Congestion Learning and Forecasting within Gas Markets: A Deep Learning Approach

3.1. Introduction

The creation of a single European energy market is a long-standing European objective, initially proposed through the Treaties of Rome in 1957, and progressively implemented through a number of liberalisation directives; the First Energy Package (1998), Second Energy Package (2003) and Third Energy Package (2009).

The First Energy Package (Directive 98/30/E)[1] aimed to reform the monopolistic market structure, characterised by pre-defined concession areas and the proprietary use of pipelines, through unbundling national monopolies and permitting third parties to obtain non-discriminatory access to the gas transmission network. Although the First Energy Package increased competition within European gas markets, the Directive failed to achieve the impact anticipated by the European Commission.

Consequently, the Second Energy Package (Directive 2003/55/EC)[2], was legislated, introducing strengthened provisions stipulating the separation of transmission and distribution activities, aiming to improve cross-border competition, increase the security of supply and generate a single, integrated European market for natural gas.

[1] Directive 98/30/EC of the European Parliament and of the Council of 22 June 1998 concerning common rules for the internal market in natural gas

[2] Directive 2003/55/EC of the European Parliament and of the Council of 26 June 2003 concerning common rules for the internal market in natural gas and repealing Directive 98/30/EC

As a consequence of the shortcomings of the Second Energy Package, a third legislative package, the Third Energy Package (Directive 2009/73/EC)[3] was adopted in September 2009, stipulating the effective separation of supply and production activities from the operation of transmission and distribution systems, and increased cross-border regulation through the establishment of a community level regulatory body, the Agency for Cooperation of Energy Regulators (ACER).

Notably, many of the liberalisation directives' goals, such as consumer protection, market access and increased transparency, are underpinned by the stimulation of cross-border competition within natural gas markets. Given the homogeneous nature of natural gas, if cross-border competition is being practically achieved, European natural gas markets should be highly integrated, with the price differentials between markets reflecting transaction costs. This "Law Of One Price" (Mankiw 2020) can be ensured by spatial arbitrages, as profit maximising market participants can exploit price differentials between locations, enforcing price convergence between different markets.

The stimulation of transparent, cross-border competition plays a pivotal role in reducing market concentration, and by extension, the market power of incumbent participants. If cross-border competition is being practically achieved, the subsequent reduction of market concentration and increase in market integration acts to improve consumer protection, achieving the stated aims of the European Commission's liberalisation directives (Directive 2009/73/EC)[4]

However, several stylised facts must be considered when applying the "Law Of One Price" to natural gas markets. Unlike most commodities, which can be transported between markets by a multitude of mechanisms (i.e., barge, road, rail, air), natural gas' physical characteristics stipulate that it must be transported through pipelines, or liquefied (Liquified Natural Gas/LNG), and regasified following transportation. As liquefaction of gas is only cost competitive over long distances (Ritz 2019), a market participant's ability to exploit price differentials between locations is restricted by the available pipeline

[3] Directive 2009/73/EC of the European Parliament and of the Council of 13 July 2009 concerning common rules for the internal market in natural gas and repealing Directive 2003/55/EC

[4] Directive 2009/73/EC of the European Parliament and of the Council of 13 July 2009 concerning common rules for the internal market in natural gas and repealing Directive 2003/55/EC

capacity between those locations. As such, identifying congestion issues within the gas infrastructure network is crucial to maintaining a market participant's ability to exercise spatial arbitrage, and developing an internally competitive European gas market.

In this context, the European gas market is confronted by substantial challenges over the coming decade; within the European Union (EU), natural gas production is projected to decline due to limited reserves, maturing gas fields and production caps. This is anticipated to be particularly apparent within the largest gas producing countries of the EU, especially the Netherlands, where Groningen field production is anticipated to be completely curtailed by 2022 (Reuters, 2019)[5]. Whilst gas consumption within the EU is anticipated to gradually decline as renewable energy technologies become increasingly competitive, production decline is expected to outpace demand decline, generating a domestic supply deficit, and an increasing dependence on imported gas (Figure 3.1).

Due to the declining production of EU member states and subsequent reliance on imported gas, increasingly large volumes of gas shall be arriving at the borders of the EU. In order to support these increased import volumes – which are projected to be 300 billion cubic meters per year (bcm/y) by 2020[6] – an increase in import capacity is required. More importantly, infrastructure utilisation patterns within the EU shall also evolve, as pipeline gas or LNG (Liquefied Natural Gas) shall require transportation from the EU border to end consumers.

Furthermore, as indigenous production declines and imports rise, transportation distances of natural gas shall increase, with the consequence of increased network usage, and a higher probability of network congestion issues.

Another potential challenge for the European gas market is the danger of short-term supply disruptions, as observed on 12 December 2017 (Bros 2018), when an explosion at the Baumgarten gas terminal forced the operator to close the facility, severely reducing cross-border gas transmission from Austria to Italy. The Baumgarten crisis, and Italy's subsequent declaration of a state of national emergency (Bros 2018) highlighted the dependence of some EU member states on specific gas transit routes, indicating po-

[5] Netherlands to halt Groningen gas production by 2022

[6] Own elaboration based on BP Statistical Review of World Energy 2019.

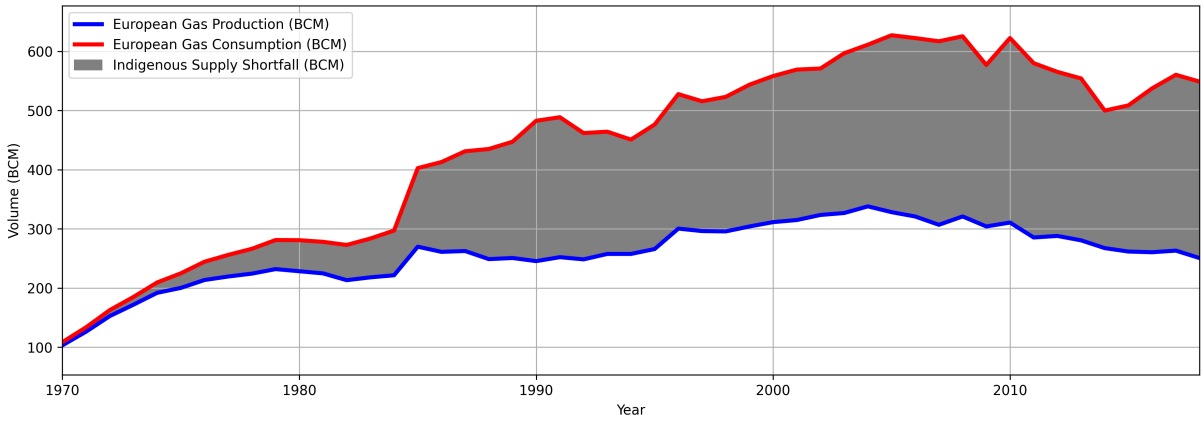


Figure 3.1. EU Domestic Supply–Demand Balance

The historical domestic supply–demand balance of European Union member countries. This illustrates an increasing dependence on gas imports.

tential weaknesses in supply security. As such, the Baumgarten crisis emphasised the interrelationship between security of gas supply and the flexibility of the underlying gas transportation infrastructure.

Taken together, identifying infrastructure congestion issues is an important prerequisite in enforcing competition, implementing an internal European gas market and increasing the security of gas supply of EU members. The progressive liberalisation of European natural gas markets has generated a substantial body of both academic (Neumann et al. 2006; Renou-Maissant 2012; Bastianin et al. 2019) and policy (ACER and CEER 2017; ACER 2018) literature investigating the attainment of a single, internal market for natural gas. Existing literature on European natural gas markets can be roughly categorized into two main strands; price integration studies, and infrastructure simulation studies. This paper departs from earlier studies in a number of ways.

Firstly, most of the early literature (Asche et al. 2000; Asche et al. 2002; Neumann et al. 2006) investigates the law of one price, providing an *ex-post* assessment of market integration. Although more recent studies address this deficiency through the use of a quantitative, time-varying approach (Renou-Maissant 2012; Bastianin et al. 2019; Woroniuk et al. 2019), to the best of our knowledge, this is the first study to forecast short-term infrastructure congestion issues, providing an *ex-ante* assessment of physical barriers to market integration.

Secondly, many of the infrastructure forecasting studies using statistical methods (Dieckhöner et al. 2013; Lochner and Bothe 2007; Monforti and Szikszai 2010) were developed when the size of datasets was limited. As statistical models' capability of handling high dimensional time-series is restricted, the forecasts generated typically aggregate infrastructure constraints to the country level. With recent developments in computational power and increasing data availability, infrastructure forecasting can be undertaken at a pipeline specific level, with machine learning methods applied to capture complex non-linear relationships. Consequently, this study augments the extant literature through increasing forecasting accuracy, output resolution and interpretability of output.

Thirdly, many infrastructure forecasting studies (Dieckhöner et al. 2013; Monforti and Szikszai 2010) inappropriately specify the dispatch order or spatiotemporal characteristics of the underlying infrastructure network. These models also fail to explicitly consider the impact of upstream congestion issues on downstream gas flow states, which can lead to spurious conclusions. Furthermore, the models inappropriately conjecture that gas flow states of two geographically disparate locations can contemporaneously influence each other. To mitigate this, a normal operations reachability matrix (NORM) based on the ability of each 'parcel' of gas to reach different points within the network is applied to a graph convolutional operator, enabling learning of the localised 'neighbourhood' of each node within the transmission network.

In this study, we learn the gas infrastructure network as a graph and conduct a convolution on the gas infrastructure network graph. In order to extract localised features and incorporate the physical characteristics of the network, we employ a graph convolution operator. Based on this operator, we propose a Graph Convolutional LSTM (GC-LSTM) to model the short-term dynamics of pipeline flows and capture spatio-temporal interdependencies between pipelines.

Evaluation results show that the proposed GC-LSTM model outperforms advanced forecasting baseline models. More importantly, through extraction of the model's weights, the proposed model is capable of identifying the most important pipelines within the gas infrastructure network. As such, it can be considered a suitable tool for short-term conges-

tion forecasting within the gas infrastructure network, whilst also highlighting pipelines which are important to the smooth functioning of a competitive, internal European gas market.

The main contributions of this paper are as follows:

1. We propose a graph convolutional operator which is capable of accommodating the physical specialities of gas infrastructure networks and extract comprehensive features.

2. A Graph Convolutional LSTM recurrent neural network is proposed to learn the complex spatial and dynamic temporal interdependencies presented by gas flows.

3. The graph convolution weights can be extracted and interpreted, identifying the most important pipelines within the European gas system.

This paper is organised as follows: Section 3.2 reviews the extant literature, whilst Section 3.3 outlines the methodologies employed throughout the paper. Section 3.4 specifies the dataset, with Section 3.5 discussing the empirical findings. Section 3.6 concludes.

3.2. Literature Review

Due to the recent increase in data availability and computational power, increasingly complex predictive models are applied to learn spatiotemporal relationships within Energy market data. Whilst statistical approaches such as Autoregressive Integrated Moving Average (ARIMA) have been applied to extract temporal patterns within energy market data (Contreras et al. 2003), the ARIMA model is incapable of extracting the spatial interdependencies which characterise network data. Conversely, latent space modelling or k-Nearest Neighbour (k-NN) models can be applied to capture spatial correlations within energy networks. Whilst these approaches show promise in extracting spatial dependencies, application to spatiotemporal issues in energy markets revealed limited efficacy (Grundmann et al. 2016).

Recently, neural networks have been applied to a range of spatiotemporal forecasting problems, typically employing Convolutional Neural Networks (CNNs) to extract spatial

features. Whilst CNNs have shown good forecasting results, the extension of the CNN architecture to graph-structures through development of Graph Convolutional Networks (GCNs) (Kipf and Welling 2016) has increased the applicability of CNNs to network data. In order to compress the spatial data into 2 dimensions (2D), the neighbouring local information is combined into one grid of image by a localized filter. Following this, feature maps are extracted by convolutional layers, which consider the relationships between adjacent nodes.

As the features extracted by GCNs are time invariant, Long Short Term Memory recurrent neural networks (LSTM) (Hochreiter and Schmidhuber 1997) are often applied to capture autoregressive sequential dependencies and long-term dependencies within the data. Given the highly interconnected nature of the European Gas Infrastructure Network, the ability to capture sequential dependencies between nodes is particularly useful when learning and predicting congestion issues, as we are able to assess how congestion issues propagate throughout the system. Additionally, the use of long-term history corresponding to each node improves forecasting performance, as gas demand often exhibits temporal recurrence, including hourly, daily, weekly and monthly patterns (Hulshof et al. 2016).

3.3. Methodology

3.3.1. Gas Flow Forecasting

The forecasting of Gas flow states refers to predicting future flow states given previously observed gas flow states within an infrastructure network consisting of N sensor locations, where links connect the sensor locations. In the context of gas markets, the sensor locations are border points, where gas flows are measured, and the links between border points are represented by pipelines.

The topological relationship between the sensor locations, is characterised well by an undirected graph, G, where $G = (\mathcal{V}, \varepsilon, A)$, with N nodes representing the sensor locations, $v_i \in \mathcal{V}$ and edges $(v_i, v_j) \in \varepsilon$ representing the pipeline connections between them. The

network can be expressed as an adjacency matrix, $A \in \mathbb{R}^{N \times N}$, in which each element $A_{i,j} = 1$ if a pipeline exists between the nodes i and j, and $A_{i,j} = 0$ otherwise. Based on the adjacency matrix, $A \in \mathbb{R}^{N \times N}$, a link counting function $d(v_i, v_j)$ can be defined as counting the minimum number of links traversed between node i and node j.

The European gas infrastructure network graph has a number of properties which differentiate it from other graph-based systems such as molecule graphs, document citation graphs and social network graphs. Principally, the European Gas Infrastructure graph contains no isolated nodes or edges, as pipelines are inherently designed to carry a given commodity between two locations. Subsequently, each point within the graph is accessible from any other location within the graph by traversing a minimum of d links (pipelines).

Moreover, whilst other graph-based systems such as molecule graphs or document citation graphs are typically static, the dynamic nature of gas flows stipulates that the flow state recorded at each sensor location (node) varies with time.

Critically, the links (edges) within the European gas infrastructure network have meaningful physical characteristics, namely the length, direction and maximal technical capacity of the associated pipeline. In order to account for these characteristics, a distance adjacency matrix $D \in \mathbb{R}^{N \times N}$ can be defined, where each element, $D_{i,j}$ represents the pipeline distance between sensor locations (nodes) i and j.

Subsequently the graph signals recorded at each sensor location (node), at time t, can be expressed as $X_t \in \mathbb{R}^{N \times P}$, where P is the number of features associated with each node (sensor). As this study exclusively considers gas flows, measured as a percentage of firm technical capacity, at each node, $P = 1$.

In summary, the gas flow forecasting model aims to learn a function $F(\cdot)$ to map T' historical graph signals to the subsequent T time step of graph signals:

$$F([X_{(t-(T'+1))}, \ldots, X_t]; G(\mathcal{V}, \varepsilon, A, D)) = [X_{t+1}, \ldots, X_{t+T}] \tag{3.1}$$

Furthermore, the forecasting process enables an improved understanding of the complex interdependencies between sensor locations (nodes), and the impact of any congestion between two nodes on the broader system.

3.3.2. Infrastructure Graph Convolution

In order to develop a Graph Convolution on the Infrastructure Network, we must firstly define the r-hop neighbourhood of each node i. The r-hop matrix can be defined as $NB_i = v_i \in \mathcal{V} jd(v_i, v_j) \leq r$, where $r = 1$ is exactly equal to the adjacency matrix for graph G.

Consequently, the r-hop adjacency matrix can be computed through calculating the r^{th} product of A, with the r^{th} order adjacency matrix defined as:

$$\tilde{A}^r = B_i(\prod_{i=1}^{r} A + I) = B_i(A^r + I) \tag{3.2}$$

Where B_i constrains the values of all elements of $(A^r + I)$ to a maximum of 1,with $A^r + I \in \{0, 1\}^{N \times N}$. The Identity Matrix (I) makes each node within the network self accessible, with an example $\tilde{A}^r$ with respect to the square red node is displayed by Figure 3.2.

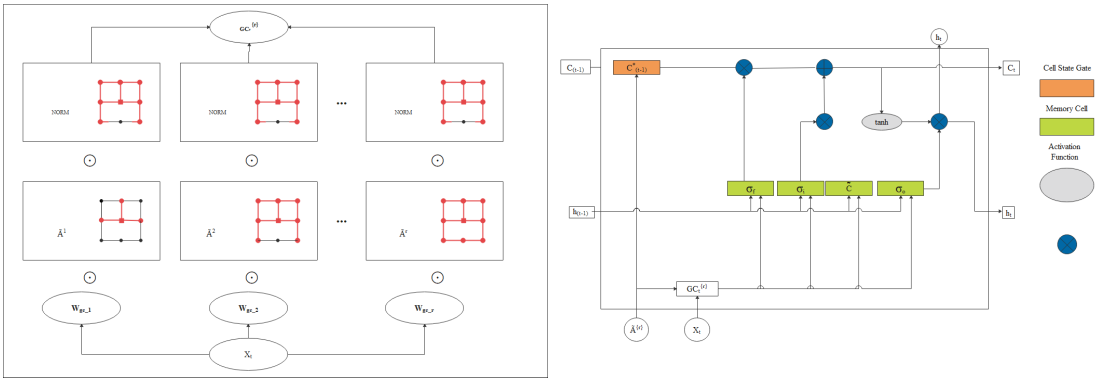


Figure 3.2. The Graph Convolution Component of the Model

The graph convolution component of the model is displayed on the left side of the figure, detailing the unfolding of the convolution at time t, in which $\tilde{A}^r$ and $NORM$ are displayed with respect to the square red node. The architecture of the Graph Convolutional LSTM model is detailed on the right side of the figure.

The r-hop graph convolution can be defined as follows:

$$GC^r = (W_{gc_r} \odot \tilde{A}^r)X_t \tag{3.3}$$

Where $\odot$ is a multiplication operator for each element held within the matrix, W_{gc_r} is the

r-hop weight matrix for the r^{th} order adjacency matrix ($\tilde{A}^r$), and $X_t \in \mathbb{R}^{N\times 1}$ is the gas flow state at time t.

This study augments the convolutional operator by assessing the operational dynamics of the underlying gas infrastructure network, and considering the mechanisms by which congestion propagates throughout an interconnected system.

Firstly, congestion which occurs at a given node, i, has the potential to propagate both upstream and downstream in the network. Intuitively, congestion in the prevailing direction of travel is likely to impact flow states at nodes upstream, as gas flows have to be directed away from transmission through this node, in order to comply with the technical capacity[7] and linepack[8] flexibility rules of each individual pipeline.

Secondly, due to the nature of the European Gas Infrastructure network, there are a finite number of paths by which the network can be traversed, with each node pair having a most efficient path, defined by counting the minimum number of links traversed between node i and node j. In a commercial environment, the most efficient path for the transmission of gas between node i and node j is typically the most cost effective path for gas shippers, as it requires less pipeline usage, incurring lower flow-based charges[9]. Thus, for the European Gas Infrastructure network graph, the impact transmission between non-adjacent nodes cannot bypass intermediate nodes. Consequently, the impact of any congestion between adjacent and nearby node pairs requires specification within the graph convolution.

To account for this, a Normal Operations Reachability Matrix (NORM), $NORM \in \mathbb{R}^{N\times N}$, which considers the accessibility of adjacent nodes from a given node, i, is defined:

$$NORM_{i,j} = \begin{cases} 1, & S_{i,j}m \vartriangle t - D_{i,j} \geq 0 \\ 0, & \text{otherwise} \end{cases}, \forall v_i, v_j \in \mathcal{V} \tag{3.4}$$

[7]Technical capacity is defined as the maximum firm capacity that the transmission system operator can offer to the network users, taking account of system integrity and the operational requirements of the transmission network (Regulation2015/703).

[8]Linepack refers to the storage of gas by compression in gas transmission and distribution systems, but not including facilities reserved for transmission system operators carrying out their functions (Regulation2015/703)

[9]A flow-based charge covers costs associated with the quantity and distance of gas flows. This charge is uniform for all entry and exit points associated with each balancing zone.

where $S_{i,j}$ is the average flow rate between nodes i and j assuming no congestion, $\triangle t$ represents time duration, and m is a value counting how many time-intervals are required to travel between nodes i and j. Subsequently, each element $NORM_{i,j}$ is exactly equal to one if gas can traverse between nodes i and j within m time steps at the average flow rate $S_{i,j}$, and zero otherwise. All $NORM$ diagonal values are set to one, indicating self access for each node. The Graph Convolution is consequently updated to incorporate the Normal Operations Reachability Matrix ($NORM$):

$$GC^r = (W_{gc_r} \odot \tilde{A}^r \odot NORM)X_t \tag{3.5}$$

where the r-hop adjacency matrix and $NORM$ are multiplied element-wise. For a specific network graph, when r is increased, the $\tilde{A}^r \odot NORM$ term will eventually converge to the point that $\tilde{A}^r \odot NORM = NORM$. A comparison between the convolution presented in equations (3.5) and (3.3) is presented in Figure 3.2. As such, a maximum of r hops of features need to be extracted from the data, X_t. The features extracted by the graph convolution at time, t, are subsequently concatenated together:

$$\mathbf{GC}_t^r = [GC_t^1, GC_t^2, \ldots, GC_t^r] \tag{3.6}$$

where $\mathbf{GC}^r \in \mathbb{R}^{N \times r}$ is the r^{th} order graph convolution features, which can be fed into the LSTM model described in the following subsection.

3.3.3. Graph Convolutional LSTM

In order to forecast short-term congestion within the Gas Infrastructure Network, we propose a Graph Convolutional Long Short Term Memory (LSTM) recurrent neural network, with the architecture detailed on the right side of Figure 3.2. The Graph Convolutional LSTM model combines the Graph Convolution's ability to learn complex spatial dependencies with the LSTM's ability to learn the dynamic temporal dependencies presented within gas flow data.

The LSTM model is a type of recurrent neural network designed to overcome the

issues associated with basic recurrent neural networks, principally vanishing and exploding gradients (Hochreiter and Schmidhuber 1997). The ability to cope with the vanishing and exploding gradient issue makes LSTM recurrent neural networks capable of learning long-term dependencies, which is particularly useful when working with seasonal data, such as natural gas pipeline flows.

In this model, the gates structure of Hochreiter and Schmidhuber's (1997) LSTM and hidden state are unchanged, but the input unit is replaced by the features extracted by the Graph Convolution, which are reshaped into a vector $\mathbf{GC}^{\{r\}} \in \mathbb{R}^{rN}$. Following (Hochreiter and Schmidhuber 1997), the forget gate f_t, input gate i_t, output gate o_t and input cell state $\tilde{C}_t$ at time, t are defined as:

$$f_t = \sigma_g(W_f \cdot \mathbf{GC}_t^{\{r\}} + U_f \cdot h_{t-1} + b_f) \tag{3.7}$$

$$i_t = \sigma_g(W_i \cdot \mathbf{GC}_t^{\{r\}} + U_i \cdot h_{t-1} + b_i) \tag{3.8}$$

$$o_t = \sigma_g(W_o \cdot \mathbf{GC}_t^{\{r\}} + U_o \cdot h_{t-1} + b_o) \tag{3.9}$$

$$\tilde{C}_t = tanh(W_c \cdot \mathbf{GC}_t^{\{r\}} + U_c \cdot h_{t-1} + b_c) \tag{3.10}$$

where $\cdot$ represents a matrix multiplication operator, W_f, W_i, W_o and $W_c \in \mathbb{R}^{rN \times N}$ represent the weight matrices, which map the input to the three respective gates and input cell state. U_f, U_i, U_o and $U_c \in \mathbb{R}^{N \times N}$ represent the weight matrices of the preceding hidden state and b_f, b_i, b_o and $b_c \in \mathbb{R}^N$ represent four bias vectors. The σ_g is the gate activation function, whilst $tanh$ represents the hyperbolic tangent function.

As each node within the Gas Infrastructure Network is influenced by it's own previous states, and the previous states of adjacent nodes, the LSTM cell state of each node should also be impacted by innovations in the cell states of adjacent nodes. Following Gers et al. (2002), an additional cell state gate is incorporated into the LSTM, which is defined as follows:

$$C_{t-1}^* = W_N \odot (\tilde{A}^r \odot NORM) \cdot C_{t-1} \tag{3.11}$$

where W_N is a weight matrix which measures the contributions of neighbouring cell states.

W_N is limited through multiplication of a $NORM$ based r-hop adjacency matrix $\tilde{A}^r \odot NORM$. Through the addition of a cell state gate, the influence of the adjacent cells states is considered when the cell state is recurrently input to the following time step. As such, the final cell state and hidden state are specified:

$$C_t = f_t \odot C^*_{t-1} + i_t \odot \tilde{C}_t \tag{3.12}$$

$$h_t = o_t \odot tanh(C_t) \tag{3.13}$$

3.3.4. Loss Function and Regularization

Regularization of Graph Convolution Weights

As the weights of the Graph Convolution are not constrained as non-negative and the extracted features of each node are influenced by the features of adjacent nodes, the weights of the Graph Convolution can vary substantially whilst undergoing training.

As the convolution weights hold information detailing the relationship between nodes within the network, visualisation and interpretation of the convolution weights is informative for better understanding the complex relationships between different nodes within the network. Critically, the interpretability of non-regularized convolution weights is low, as large and small weights appear randomly, which, when combined, negate each other. When combined, these weights can illuminate important relationships within the network structure, but don't provide an accurate representation of inter-nodal relationships. Following Kipf and Welling (2016), an L1-norm regularization term is added to the graph convolution weight matrices to make these matrices as sparse as possible:

$$R^{\{1\}} = ||\mathbf{W}_{gc}||_1 = \sum_{i=1}^{r} |W_{gc_i}| \tag{3.14}$$

Through addition of the L1 regularization term, the graph convolution weight is stable and sparse, hence interpretation of the relative importance of adjacent nodes is more intuitive.

Regularization of Graph Convolution Features

Given that the impact from an influencing node must be transmitted through all intermediary nodes to a node of interest, the features extracted from different hops of the graph convolution do not exhibit a high degree of variance. In order to restrict the variance between features extracted by different hops of the graph convolution, an L2-norm feature regularization term is included in the loss function:

$$R^{\{2\}} = ||\mathbf{GC}_t^r||_2 = \sqrt{\sum_{i=1}^{r-1}(GC_t^i - GC_t^{i+1})^2} \tag{3.15}$$

Through the application of the L2-norm regularization term, the features extracted from adjoining hops of the graph convolution have limited variance, more accurately reflecting the realities of the underlying relationships present within the European Gas Infrastructure network.

Loss Function

As with most supervised regression models, the GC-LSTM output at time t is a predicted value h_t, which can also be denoted as $\hat{Y}_t$ and corresponds to a known value (label), Y_t. Consequently, the loss can be defined as:

$$L_t = LossFunction(\hat{Y}_t - Y_t) \tag{3.16}$$

where $LossFunction(\cdot)$ is a function to calculate the difference between the predicted value $\hat{Y}_t$ and the true value (label) Y_t. Following Wallach and Goffinet (1989), $LossFunction(\cdot)$ is a Mean Squared Error (MSE) function, which has been shown to evaluate the prediction accuracy of continuous values well.

For a sequential model, the true value (label) of time step t is equal to the input of the next step in the sequence $(t+1)$. As such, Y_t can also be expressed as X_{t+1}, with the

loss function expressed as:

$$L_t = LossFunction(\hat{Y}_t - Y_t) = LossFunction(h_t - X_{t+1}) \tag{3.17}$$

Through addition of the L1-norm of graph convolution weight matrices outlined in 3.4.1. and the L2-norm feature regularization term outlined in 3.4.2., the total loss function at time t is defined as:

$$L_t = LossFunction(h_t - X_{t+1}) + \lambda_1 R^{\{1\}} + \lambda_2 R_t^{\{2\}} \tag{3.18}$$

3.4. Data

In order to learn the spatial and temporal interdependencies within the European Gas Infrastructure Network, this study constructs a graph G, where $G = (\mathcal{V}, \varepsilon, A)$, consisting of N nodes and edges $(v_i, v_j) \in \varepsilon$. In the context of European gas markets, the nodes (sensors) are the measurement points of gas flows, which correspond to border points between balancing zones within the European gas network, whilst edges between the nodes represent the maximum technical capacity of pipelines between the border points. In the instance that multiple pipelines exist between two border points, the maximal technical capacity is combined.

To capture the dynamics of the European Gas Infrastructure network, both 'virtualised' and 'physical' border points are specified. As such, hourly time series data of 79 border points are considered (Figure 3.3), covering the period 2016 to 2019.

The time series is of sufficient length to highlight potential congestion issues within the European Gas Infrastructure Network, whilst also identifying the pipelines most important to maintaining an integrated European Gas market. This data was obtained from the ENTSOG Transparency platform with time sampling of one-hour. To negate lag structure requirements arising from time-zone differences, all data was standardised to European Central time, resulting in 2,584,248 observations.

As this study is primarily concerned with the detection of short-term congestion within

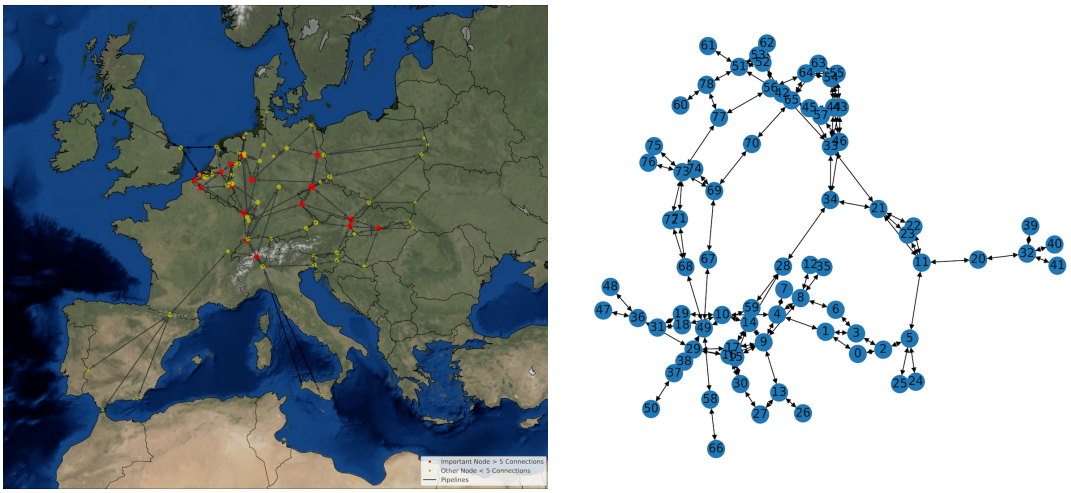


Figure 3.3. European Gas Infrastructure Network

Each point on the map corresponds to a border point within the Dataset. The Data was retrieved from ENTSOG Transparency Platform with a time sampling of one hour.

the European gas system, gas flows are measured as a percentage of the maximal technical capacity between nodes i and j (Equation 3.19) . This method has three primary advantages when compared with measurement of total gas flows.

Firstly, measuring gas flows as a percentage of maximum technical capacity can be considered as a method of feature scaling. Feature scaling is a common practice in data pre-processing, with the explicit purpose of scaling features of different orders of magnitude to increase the speed of training and prevent the gradient descent algorithm from incorrectly specifying a local optima as the global optima (Han et al. 2011).

Secondly, in order to measure or predict congestion, the maximum technical capacity and gas flow through a given pipeline are required. Notably, the maximum technical capacity of a given pipeline is dynamic, as it is subject to ancillary factors such as air temperature changes, pressure reductions and planned or unplanned maintenance cycles. In order to account for these stylised facts, this paper employs the below calculation method:

$$Observations = \frac{Gas\ Flow\ Recorded}{Maximum\ Technical\ Capacity} \tag{3.19}$$

Thirdly, use of the metric outlined in Equation 3.19 provides a highly interpretable model output, with a pipeline utilisation value predicted for each border point at time, t. This enables the easy identification of potential congestion issues at different confidence

intervals.

3.5. Empirical Findings

This section first discusses the Baseline models to which GC-LSTM is compared, followed by the modelling assumptions applied, and performance metrics by which the models are assessed. Subsequently, the experimental results, training efficiency and model outputs are reported and discussed.

3.5.1. Baseline Models

In order to evaluate the performance of the GC-LSTM model, we compare its performance to a number of baseline models. The baseline models considered are the Long Short Term Memory recurrent neural network (Hochreiter and Schmidhuber 1997) and Defferrard et al. (2016)'s localised spectral graph convolution LSTM model (LSGC-LSTM).

3.5.2. Modelling Assumptions

The neural network hidden state dimensions are set as the number of nodes within the Gas Infrastructure Network, in this instance 79. Whilst the size of hops in the graph convolution can vary, for the purpose of model evaluation, r is set to 3, indicating that $NORM$ is calculated based on 3 time steps.

The initial learning rate for all models is set as 5×10^{-6}, with the batch size of all models set to 30. As per Section 3.3.4, the model is trained by minimising the Mean Squared Error (MSE) using RMSProp (Tieleman and Hinton 2012). Finally, the regularization term learning rates are set as 0.01.

3.5.3. Performance Metrics

In this study, the performance of the proposed model and the baseline models outlined in Section 3.5.1 are evaluated through the use of multiple performance metrics. This

subsection presents each performance metric employed within the study, whilst providing an overview of the strengths and weaknesses associated with each error metric.

Mean Absolute Error (MAE) is calculated as the arithmetic average of forecast error values, where all of the forecast values are made absolute and hold equal weight. As such, MAE measures the average magnitude of forecast errors without considering the direction associated with each error. MAE is a scale-dependent accuracy measure, meaning that the output units are synonymous with input units, and comparison across different scales can lead to inappropriate conclusions. MAE is defined as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|(Y_t - \hat{Y}_t)| \tag{3.20}$$

where $\hat{Y}_t$ is the predicted value, Y_t is the true value (label) and n is the sample size.

Mean Absolute Percentage Error (MAPE) removes the scale dependence of MAE, enabling easy comparison across different datasets. Although this characteristic is advantageous, MAPE exhibits a number of weaknesses, most importantly, the handling of any 0 values corresponding to the label (Y_t). Division by 0 is undefined, hence any 0 label pairs ($Y_t = 0$) are dropped within this study, reducing the efficacy of this evaluation metric. MAPE is defined as:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| \times 100\% \tag{3.21}$$

Mean Squared Error (MSE) represents the mean of the squared difference between true values (Y_t) and predicted values ($\hat{Y}_t$) . Whilst MAE assigns equal weights to each forecast error, MSE and RMSE penalise variance as they both assign errors with large absolute values higher weight than errors with small absolute values. MSE is defined as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_t - \hat{Y}_t)^2 \tag{3.22}$$

Notably, MSE is measured in squared units corresponding to Y_t and $\hat{Y}_t$, reducing the interpretabilty of the output. To rectify this, the Root Mean Squared Error (RMSE), which represents the square root of the second sample moment of forecast errors, is spec-

ified. RMSE is defined as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_t - \hat{Y}_t)^2} \tag{3.23}$$

All of the above metrics range between 0 and ∞ and are negatively orientated scores, such that a lower value indicates a better forecasting accuracy. Whilst all of the above performance metrics are indifferent to the direction of forecasting error, inclusion of an additional metric which is sensitive to error direction provides additional information pertaining to any systematic overestimation (underestimation) of pipeline utilization.

As such, the Mean Bias Error (MBE) performance metric is included. MBE shares a number of characteristics with MAE, however, MBE specifies that forecast error values should not be absolute. This uncovers potential systematic overestimation (underestimation) issues within the model. MBE is defined as:

$$MBE = \frac{1}{n}\sum_{i=1}^{n}(Y_t - \hat{Y}_t) \tag{3.24}$$

3.5.4. Experimental Results

Table 3.1 provides a number of performance metrics pertaining to the GC-LSTM model and two baseline models. Through analysis of Table 3.1, the GC-LSTM model can be identified as the best performing model, indicating its ability to identify complex spatiotemporal dependencies within the data.

Notably, Defferrard et al. (2016)'s one-layer localised spectral graph convolution LSTM (LSGC-LSTM) fails to outperform the LSTM model in this task, likely due to the parameters lacking sufficient breadth to appropriately represent the underlying features of the network.

The GC-LSTM model, which both identifies features inherent to the graph structure, and also incorporates the physical characteristics of the gas infrastructure network, outperforms the baseline models with respect to all performance metrics (Table 3.1).

From Table 3.1, we note that the GC-LSTM and baseline models consistently gener-

ated a negative MBE value, indicative of an underestimation of pipeline utilization rates throughout the European gas system. However, given the consistently low magnitude of MBE values across each model and the continually evolving maximum technical capacity outlined in Section 4.3, underestimation errors of this magnitude are not a cause for substantial concern.

Table 3.1 displays that the GC-LSTM model produced the smallest MBE value, indicating the lowest underestimation of gas flows. As this study focuses on the detection of congestion within the European Gas Infrastructure Network, defined as Equation 3.19 $\geq 95\%$, this suggests that the GC-LSTM model is the most appropriate tool to detect congestion within the European Gas Market.

Table 3.1. Performance Metrics

A comparison of different approaches and performance metrics. As per Section 3.5.2, r is set to 3.

Metric	LSTM	LSGC-LSTM	GC-LSTM
Mean Bias Error (MBE)	-0.72%	-0.67%	-0.39%
Mean Absolute Error (MAE)	3.80%	5.17%	2.33%
Mean Absolute Percentage Error (MAPE)	12.96%	13.82%	6.43%
Mean Squared Error (MSE)	0.29%	0.57%	0.29%
Root Mean Squared Error (RMSE)	5.43%	7.52%	5.41%

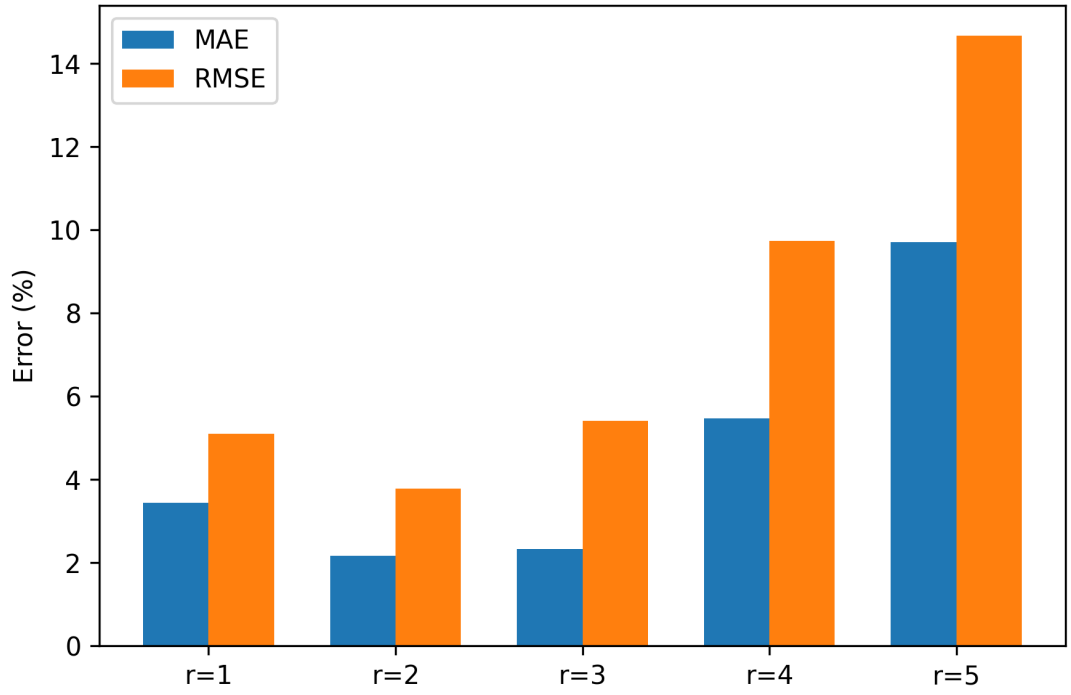
Whilst the MBE value considers the direction of forecasting errors (overestimation or underestimation), MAE and MAPE consider the absolute forecasting errors. The GC-LSTM model substantially outperforms the baseline models with respect to absolute errors, indicating that it is the most appropriate model for learning and detecting congestion issues within the European Gas network.

However, when analysing the model MSE and RMSE values, the GC-LSTM model marginally outperforms the LSTM model. Given that MAPE and MAE assign equal weighting to each forecast error, these metrics do not penalise large forecasting errors as rigorously as MSE and RMSE. These results, combined with MAPE, tacitly infer that the GC-LSTM model may be prone to the generation of larger forecasting errors.

Figure 3.4 displays the relative performance of different hops (r-values) of graph convolution in the GC-LSTM model. The model performance, as defined by MAE and RMSE,

Figure 3.4. MAE and RMSE

A histogram comparing the MAE and RMSE of different hops (r-values) of graph convolution of the GC-LSTM model.



improves as r is increased from 1 to 2, with performance marginally decreasing as r is increased to 3. Following this, a pronounced decrease in model performance is observed (Figure 3.4) as the value of r is increased. This can be attributed to the highly interconnected nature of the European Gas Infrastructure Network, with large r-values limiting the localisation of the graph convolution.

3.5.5. Training Efficiency

This subsection compares the training efficiency of the GC-LSTM model with the baseline models outlined in Section 3.5.1. Figure 3.6 displays the validation loss curves of each model compared with the number of training epochs completed by the model. As we specify an early stopping patience of 10, the number of training epochs differs between the models presented in Figure 3.6. Figure 3.7 displays that the GC-LSTM model requires fewer epochs than the LSTM and LSGC-LSTM model, whilst also exhibiting the fastest decline in Validation Loss.

However, on a per Epoch basis, the GC-LSTM model is the most expensive, followed by the LSGC-LSTM model (Figure 3.7). Whilst the GC-LSTM model is the most expensive per Epoch, it is the only model which achieves early stopping (Figure 3.6), indicating that it is less expensive when considering total training time.

Figure 3.5. Training Efficiency

A comparison of training efficiency when applying different hops (r-values) of graph convolution of the GC-LSTM model.

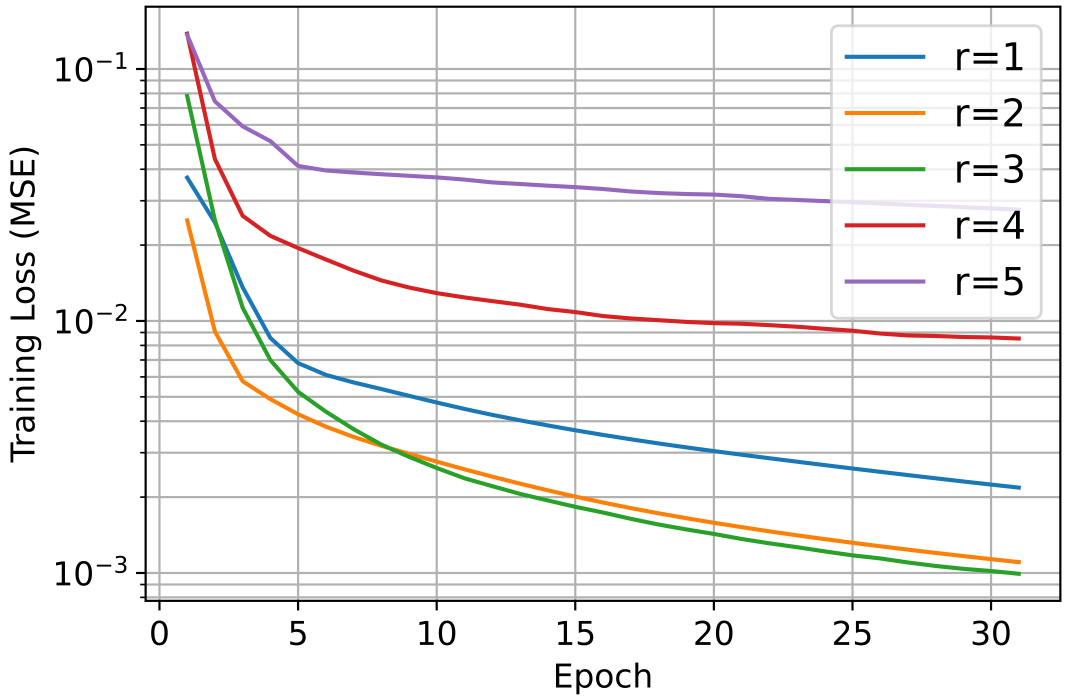


Figure 3.5 displays the training losses of the GC-LSTM model when incorporating different graph convolution components. The speed of convergence is fastest when using $r = 2$ or $r = 3$, with the speed of convergence decreasing as the value of r is increased. The training results indicated by the speed of convergence (Figure 3.5) are consistent with the validation results indicated by the performance metrics (Figure 3.4). As such, we consider $r = 2$ or $r = 3$ appropriate localisations when conducting graph convolutions on the European Gas Infrastructure Network.

Figure 3.6. Validation Loss

Validation loss versus the number of completed training epochs with a Batch Size of 30, Early stopping patience of 10.

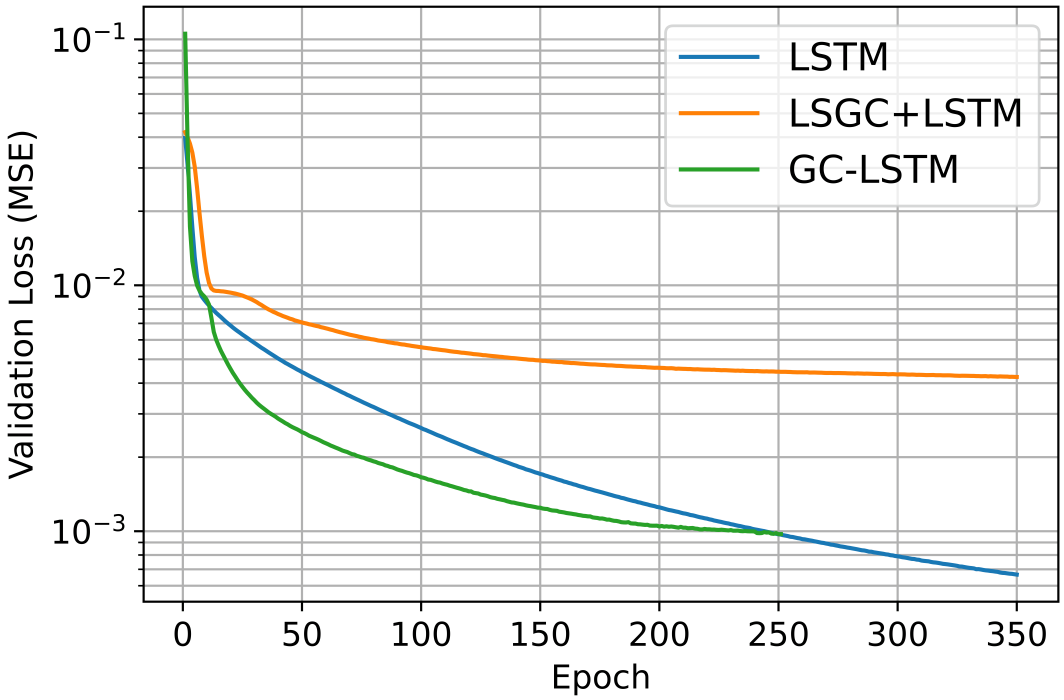
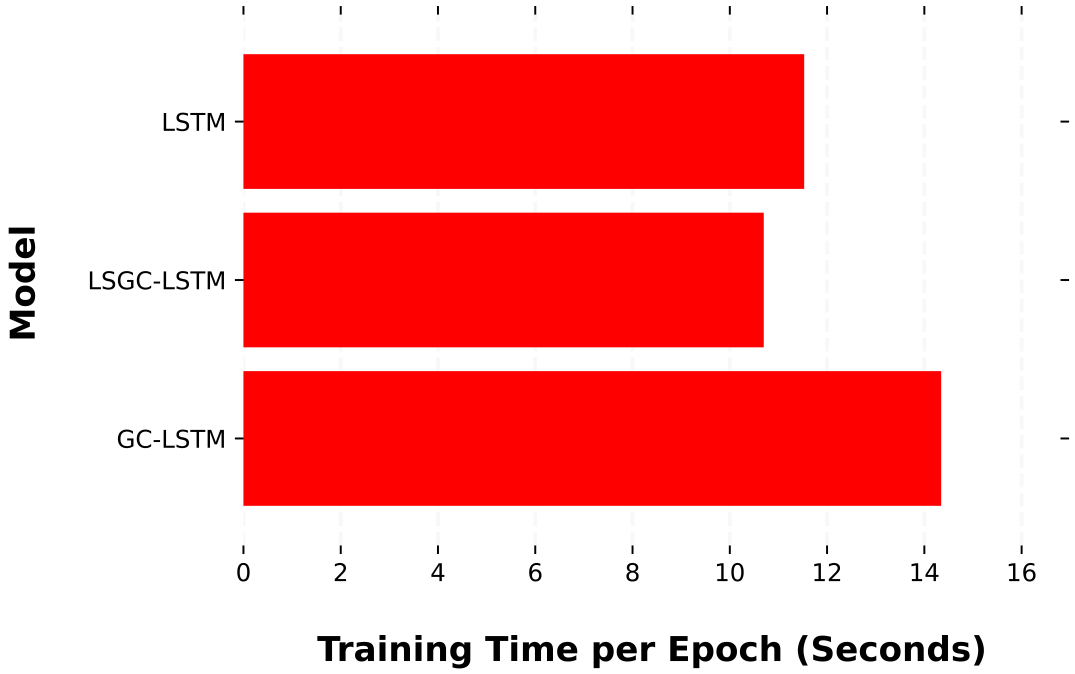


Figure 3.7. Training Time per Epoch

The training times per epoch for the GC-LSTM model and baseline models.



3.5.6. Model Outputs

This subsection provides a detailed analysis of performance metrics corresponding to each border point outlined by Figure 3.3. Whilst performance metrics are an important tool when assessing the efficacy of different models, in order for the prediction error to hold meaningful economic information, the scaling factor outlined in Equation 3.19 is reapplied to the metrics, expressing the forecast error in the context of pipeline associated gas flow values.

Figure 3.8. Pipeline Utilization Forecasting - Low MAE

Pipeline Utilization (%) forecasting visualisation for the Nordlohne, Bocholtz-Vetshau and Etzel border points, which correspond to the lowest, second lowest and third lowest MAE values recorded.

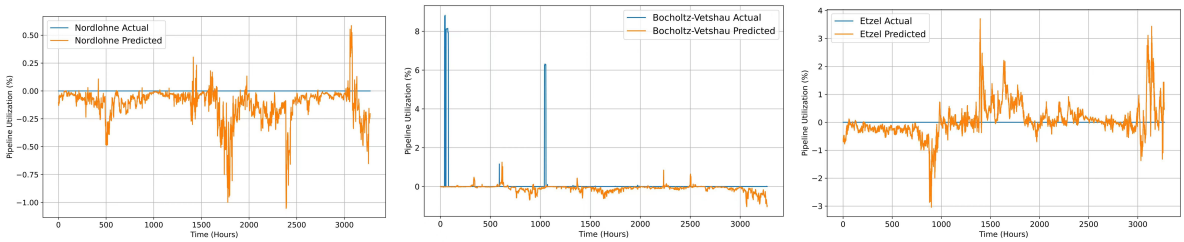


Figure 3.8 presents the border points at which the lowest MAE were recorded. Notably, all three locations which recorded the lowest MAE are situated on the border of the Gaspool balancing zone, and typically record very low percentage utilization values. As such, the low MAE value recorded at these locations can be attributed to the associated low pipeline utilization rates, as opposed to good forecasting accuracy. To apply an economic context to the low MAE scores recorded at these locations, a representative maximal technical capacity of the border points must also be considered (800 MWh/h, 1000 MWh/h and 4,612.5 MWh/h respectively). [10]. Given the representative technical capacity of the aforementioned border points, it is intuitive that these points do not transit a large proportion of European natural gas, and have limited contributions to the model weights.

Consequently, the three largest capacity border points, Uzhgorod, Baumgarten and

[10]The maximal technical capacity of each border point is subject to fluctuations, as outlined in Section 4.3

Griefswald are considered. Notably, representative maximum technical capacity values corresponding to Uzhgorod, Baumgarten and Griefswald constitute 18.90% of a representative maximal technical capacity of the sample, indicating their potential importance in the European Natural Gas mixture, and subsequent importance to security of supply.

Figure 3.9. Pipeline Utilization Forecasting - MTC

Pipeline Utilization (%) forecasting visualisation for the Uzhgorod, Baumgarten and Griefswald border points, which correspond to the three largest technical capacity pipelines in the sample.

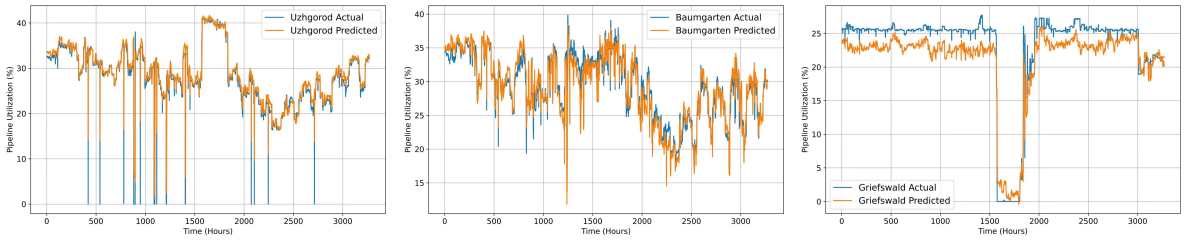


Figure 3.9 visualises the forecast pipeline utilization rates and the recorded values at Uzhgorod, Baumgarten and Griefswald. Though the patterns recorded at each border point are substantially different, the visualisation demonstrates that the pipeline utilization is forecast well under a range of scenarios, demonstrating the model's ability to adjust to maintenance cycles and unscheduled outages at each border point. In order to qualify the MAE values in an economic context, the error values recorded at Uzhgorod, Baumgarten and Griefswald correspond to 918 MWh/h, 769 MWh/h and 1479 MWh/h, which, when considering representative maximum capacities of 86,667 MWh/h, 67,167 MWh/h and 65,429 MWh/h, is an acceptable margin of error.

3.6. Conclusions

We learn the underlying geographic relationships between 78 nodes of the European Gas Infrastructure Network during the period 2016-2019. In order to extract spatial features from the European Gas Infrastructure Network, a graph convolution operator is defined. Consequently, a Graph Convolutional Long Short Term Memory (GC-LSTM) model is proposed, which is capable of incorporating the complex spatial dependencies obtained from the Graph Convolution and the temporal relationships determined by the

LSTM model. As such, the GC-LSTM is appropriate for forecasting interdependent spatiotemporal data. Additionally, the regularization terms applied to the GC-LSTM's loss function (Equation 3.18) enable a more stable model, with the additional benefit of increased interpretability.

The GC-LSTM model substantially outperforms the baseline models outlined in Section 3.5.1, displaying consistently superior performance metrics (Table 3.1).

The model's ability to adapt to unplanned outages and maintenance cycles verifies the applicability of the modelling technique to pipeline utilization forecasting, and subsequently, gas flow state forecasting. The prediction of pipeline utilization rates throughout the European Gas system relative to these events provides valuable information pertaining to how the impact of exogenous shocks propagate around the European Gas Infrastructure Network. Additionally, this provides important information relating to how pipeline utilization rates, and subsequent gas flows may react to exogenous shocks.

The forecasting results displayed in Figures 3.8 and 3.9 display the model's ability to accurately predict pipeline utilization patterns at a range of locations and under a range of different conditions. The progressive increase in pipeline utilization rates throughout the sample indicates the growing importance of identification of congestion within the European Gas Infrastructure Network. However, the low number of congestion occurrences identified within the sample implies that the Third Energy Package's (Directive 2009/73/EC) focus on enforcing competition by enabling cross-border mechanisms has been broadly successful.

Whilst limited instances of congestion in the European Gas Infrastructure Network were recorded, the novelty of this work lies in the provision of a model capable of learning complex spatiotemporal interdependencies and accurately forecasting gas pipeline utilization rates. As European gas production continues to decline and imports rise, it is anticipated that transportation distances of gas shall increase, with the consequence of increased pipeline utilization. Consequently, a model which is capable of identifying and forecasting infrastructure utilization patterns is of considerable value in enforcing cross-border competition, maintaining an internal European gas market and increasing the

security of gas supply to EU member states.

Chapter 4

Clean Energy, Brown Energy & The Impact of News Sentiment: A Comparative Analysis.

4.1. Introduction

Owing to an increasing awareness of Sustainable Development initiatives, the Clean Energy industry, and it's growing importance in financial markets has attracted growing attention from policymakers and investors alike. Increased investor awareness of climate change has given rise to three core innovations; promotion of companies which exhibit awareness of Climate Change through inclusion into Environmental Social Governance (ESG) or Climate Change indices, punishment of extractive or pollutive companies through exclusion from ESG and Climate Change indices, and increased borrowing costs for extractive or pollutive companies.

In light of this, most investors are interested in considering Clean Energy stocks within their asset allocations, both as a source of diversification (Kuang 2021), and due to increasingly punitive investment mandates for highly extractive or pollutive (Hunt and Weber 2019) industry sectors, such as Oil and Gas. Consequently, understanding the factors which impact price formation of Clean Energy and Traditional Energy stocks is imperative for informing portfolio allocation and risk management frameworks.

The relationship between the stock price of energy companies and energy prices is multifaceted and enforced through multiple channels (Hamilton 1983). The inherent linkage between energy prices and the associated cash flows of energy companies is well

understood (Huang et al. 1996), with the downstream impact on stock valuation, and subsequently stock price, well documented (Reboredo and Ugolini 2018). As the stock prices of energy companies are understood to be influenced by prevailing energy prices (Jones and Kaul 1996), a body of literature has documented the various channels through which this relationship is enforced.

Firstly, due to the high degree of interconnection amongst financial markets (Geng et al. 2021) and increasing financialisation of the energy sector (Creti and Nguyen 2015), energy market risk can quickly propagate to equity markets (Arouri et al. 2011). Secondly, due to energy's status as an essential good (Fabra et al. 2020), energy markets are susceptible to geopolitical risk, with market participants often paying a premium to increase their security of supply. This can lead to positive externalities for energy companies, whilst creating negative externalities for energy intense industries, such as manufacturing.

Whilst the relationship between energy price and energy companies stock prices' is well understood, the relationship between investor sentiment and energy companies stock prices' is comparatively less explored. As stock markets are composed of a multitude of agents, who are operating at different time horizons and with different objectives, price formation of stock is likely to be driven by investor sentiment. When investors display a positive (negative) sentiment, they impound their optimism (pessimism) toward an asset into the price. Some investors may rely on the beliefs of others (Ghosh and Bouri 2022), while others are influenced by positive or negative news (Baker and Wurgler 2007; Tetlock 2007), creating a divergence in investor sentiment.

A large volume of extant literature examines the impact of investor sentiment on financial assets, with a myriad of data sources and sentiment measurement techniques employed. Whilst much of the earlier work (Tetlock 2007; Chen et al. 2011) focused on newspapers such as the Wall Street Journal, the increasing adoption of social media precipitated academic interest in the impact of different social media channels, such as Google Search (Da et al. 2011), Facebook (Karabulut 2013) and Twitter (Bollen et al. 2011; Mao et al. 2011) on stock returns. Da et al. 2011 find increased Google search volume to reliably predict higher stock prices over the following 2 weeks, with the exception of

Initial Public Offerings (IPOs), whereby high Google search volume predicts long-term under-performance. Through adoption of the Facebook Gross National Happiness index (FGNHI), Karabulut 2013 demonstrates the impact of retail investor sentiment on the returns of the aggregate U.S. stock market. Similarly, Bollen et al. 2011 through the extraction of investor sentiment from 10 million tweets, find investor sentiment to Granger cause returns of the Dow Jones Industrial Average. Mao et al. 2011 provide additional support for the findings of Bollen et al. 2011 and Da et al. 2011, showing news sentiment, twitter sentiment and Google search to hold a significantly positive relationship with stock returns.

This study aims to identify differences in the impact of Investor sentiment on the Clean Energy and Traditional Energy sectors. From a behavioural finance perspective, companies in non-cyclical sectors, which undergo corporate restructurings and dividend modifications less frequently than cyclical sectors, have clearer earnings expectations and theoretical valuations, ceteris paribus. Consequently, we conjecture that the News sentiment index will hold a stronger relationship with companies in the Clean Energy sector than the Traditional Energy sector, as theoretical valuations within the Clean Energy sector are considerably more opaque than the Traditional Energy sector. This hypothesis is underpinned by the earlier findings of Khan et al. 2020, who report their sentiment index to have a stronger correlation with the financials, technology, health care and consumer discretionary sectors, than other sectors in a sample of the S&P 500.

Moreover, Baker and Wurgler 2006; Baker and Wurgler 2007 and Lemmon and Portniaguina 2006 emphasise the merit of a cross sectional approach when considering the impact of Investor sentiment on stocks, demonstrating that firm characteristics can determine the magnitude of Investor sentiment impact on stock prices. Notably, Baker and Wurgler 2006; Baker and Wurgler 2007 demonstrate that companies which have a more subjective valuation, such as those that are small, young, volatile, do not pay a dividend or exhibit extreme valuation ratios are more exposed to Investor sentiment than companies which do not exhibit these characteristics.

Taken in unison, the findings of Baker and Wurgler 2006; Baker and Wurgler 2007 and

Lemmon and Portniaguina 2006 indicate that firm characteristics play a role in determining the impact of investor sentiment on stock returns. Given the comparative maturity of the Traditional Energy sector, as characterised by mean size, mean age and mean dividend payment, we anticipate Investor sentiment to have a larger impact on the returns and volumes of Clean Energy stocks.

A large amount of literature provides evidence that investor sentiment plays an important role in the formation of stock prices. However, the body of literature discussing the influence of investor sentiment on clean energy stocks is limited, with few studies drawing comparative analyses between the impact of investor sentiment on clean energy and traditional energy stocks.

The energy market literature is comprised of two distinct strands; investor sentiment and oil prices, and investor sentiment and stock prices. Guo and Ji 2013 find a long-term relationship between oil prices and investor attention, as measured by Google search volumes (GSV). Notably, Guo and Ji 2013 document that increased investor attention has a direct impact on oil price volatility. Ji et al. 2019 confirm this through an analysis of spillovers between WTI returns and investor sentiment indices, finding an asymmetric impact of investor sentiment on WTI returns, with sentiment playing a more prominent role in negative returns. Similarly, Han et al. 2017 show GSV to be significant when forecasting day-ahead and week-ahead oil prices. Moreover, Gupta and Banerjee 2019 who studied the relationship between OPEC news announcements and U.S. stock returns, find a significantly negative relationship.

A limited body of literature isolates the relationship between investor sentiment and energy companies' stock returns. Reboredo and Ugolini 2018 examine the interactions between investor sentiment, as characterised by tweets, returns, volatility and volume of 17 clean energy stocks, finding investor sentiment to have a limited impact on returns, volatility, or trading volumes. Using GSV, Song et al. 2019 find investor sentiment to have a weak impact on the returns of renewable energy securities, corroborating the earlier findings of Reboredo and Ugolini 2018.

The literature exploring the interaction between investor sentiment and energy stocks

is insufficient, with prior studies using limited sample sizes (Reboredo and Ugolini 2018), or measures of investor attention (Song et al. 2019) to determine the strength of the relationship between investor sentiment and energy stocks. Through construction of the News Sentiment Index, this study closes three distinct gaps within the literature.

The first gap concerns the construction of investor sentiment measures. Prior studies on the interaction between investor sentiment and energy stock returns proxy investor sentiment through the use of publicly available data sources, such as Google Search Volume (GSV), or Tweets. Construction of the News Sentiment Index captures information available to professional investors, which may otherwise be behind a paywall. This ensures that we appropriately characterise investor sentiment as the information available to professional investors, who dominate stock ownership within our sample.

The second gap concerns the geographies studied by the extant literature. Reboredo and Ugolini (2018) study U.S. securities contained by the Wilder Hill Clean Energy Index and Song et al. (2019) study the Wilder Hill Clean Energy Index and S&P Clean Energy Index. Whilst the Wilder Hill Clean Energy Index is exclusively comprised of U.S. domiciled securities, the S&P Clean Eneextrgy Index is geographically disparate. Through the use of the STOXX Europe Total Market Index, we close a literature gap through the addition of empirical evidence from Europe.

The third literature gap relates to the lack of context of the empirical findings. Whilst Reboredo and Ugolini (2018) find investor sentiment to have a limited impact on the returns of clean energy companies, no context is provided on the impact of investor sentiment on other industry sectors. We close this gap in the literature by presenting a bifurcated sample of "Traditional Energy" companies, and "Clean Energy" companies, and draw comparative analyses between the impact of investor sentiment on the segments.

To fill the gap in literature, this paper examines the role of investor sentiment in pricing and trading Clean Energy and Traditional Energy stocks, using a sample of 35 Clean Energy and 50 Traditional Energy companies from the STOXXt Europe Total Market Index (BKXP). Prior studies focus on information from social media, and find that investor sentiment aggregated from social media affects stock returns (Bollen et al.

2011; Mao et al. 2011; Siganos et al. 2014). However, Nofer and Hinz (2015) and Reboredo and Ugolini (2018) pinpoint that investor sentiment from Twitter has no sizeable impact on returns, volatility or trading volumes, and indicate that the wisdom of the Twitter crowd is not substantial in shaping prices and trading for renewable energy companies. Unlike social media, such as Twitter and Facebook, Bloomberg serves as a news platform for both institutional investors and professional investors, who comprise the largest portion of shareholders within our sample [1], and are the dominant shareholders within the energy industry (Ritchie, Dowlatabadi, et al. 2015) and renewable energy industry (Kaminker and Stewart 2012).

Since Bloomberg is commonly used by professional investors to inform investment decisions, we use information from the Bloomberg News Platform to measure investor sentiment, as characterised by the sentiment embedded within Bloomberg News articles.

Following the dynamic spillover model developed by Diebold and Yilmaz (2009), Diebold and Yilmaz (2012), and Diebold and Yılmaz (2014), we test the spillover effects from (to) news sentiment to (from) stock returns and trading volumes for both clean energy and traditional energy companies. We also examine the contemporaneous effects using a regression approach (Azar and Lo 2016; Siganos et al. 2017). We find that for Clean Energy companies, news sentiment has a relatively small impact on stock returns and trading volumes, and conversely, stock returns and trading volumes also have a limited impact on news sentiment. Notably, the size of spillover effects are asymmetric, whereby the impact of News sentiment on stock returns and trading volumes is larger in magnitude. These findings are consistent with the findings of Reboredo and Ugolini (2018). More importantly, we find that Traditional Energy companies exhibit a similar pattern in the relationship between News sentiment, stock returns and trading volumes. In particular, spillovers from News sentiment to stock returns and trading volumes are greater than the spillovers in the opposite directions, and spillovers from News sentiment to stock returns are relatively greater than to trading volumes. This is confirmed through correlation analysis, and our regression results also confirm the positive impact of News

[1] Institutional investors constituted 82.63% of the aggregate shareholding on the final day of our sample.

sentiment on stock returns and trading volumes.

These findings suggest that investor sentiment plays a significant role in the price formation of both Clean Energy and Traditional Energy sectors, with investor sentiment playing a marginally more important role in the Clean Energy sector. Whilst Investor sentiment is significant in the price formation of Clean Energy and Traditional Energy companies, the magnitude of impact is relatively limited, with 2.81% and 2.71% of returns attributed to the impact of investor sentiment for Clean Energy companies and Traditional Energy companies, respectively.

Through the use of an expanded sample, our findings affirm the earlier work of Nofer and Hinz (2015) and Reboredo and Ugolini (2018), providing additional empirical evidence of the role that investor sentiment plays within asset pricing. Given the difference from Nofer and Hinz (2015) and Reboredo and Ugolini (2018) in investor sentiment sources (Twitter vs Bloomberg News articles), yet similarity in findings, our findings support Mao et al. (2011)'s conclusion that, if appropriately broad, different information sources should yield a comparable impact of sentiment on stock returns. Moreover, through the study of European companies, our findings make an important contribution to the literature, as we are able to confirm the role News sentiment plays in price formation and trading of European Clean Energy stocks as comparable with U.S. Clean Energy stocks. Additionally, our findings indicate that no discernible difference in the impact of News sentiment exists between Clean Energy and Traditional Energy stocks in Europe. As such, this paper provides valuable evidence to the ongoing debate concerning the impact of News sentiment on stock price formation within different Industry sectors (Uygur and Taş 2014; Khan et al. 2020; Niu et al. 2021).

The reminder of the paper is structured as follows. Section 4.2 outlines the econometric approach used to characterize the spillover effects. Section 4.3 provides detailed information on the sample and data collection. Section 4.4 presents our main results, and Section 4.5 concludes.

4.2. Econometric Approach

In this section, we introduce the empirical model used to examine the impact of news sentiment on both Traditional and Clean Energy stock prices. Specifically, we adopt an 3-variable vector autoregression (VAR) framework with exogenous variables developed by Diebold and Yilmaz (2009), Diebold and Yilmaz (2012), and Diebold and Yılmaz (2014) to measure the spillover effects for each company in both Clean Energy and Traditional Energy groups.

4.2.1. Generalized VAR model

Diebold and Yilmaz (2009) introduce a dynamic spillover measure based on the notion of forecast error variance decompositions from vector autoregressions (VARs). Since the spillover measure relies on variable ordering and only addresses the total spillovers rather than directional spillovers, Diebold and Yilmaz (2012) and Diebold and Yılmaz (2014) extend their spillover index and develop a generalized VAR framework that eliminates the possible dependence of the results on ordering. This VAR-based network methodology has been applied in a range of economics studies, such as stock market inter-dependencies, business cycle synchronization, volatility spillover, and sentiment spillover (Baruník et al. 2016; Corbet et al. 2018; Demirer et al. 2018; Wiesen et al. 2018; Antonakakis et al. 2018; Reboredo and Ugolini 2018; Zhang and Broadstock 2020).

Following Diebold and Yılmaz (2014) and Reboredo and Ugolini (2018), we consider an N-dimensional covariance-stationary VAR(p) with orthogonal shocks. Let y_t as a column vector of time series variables which contain daily information on asset k at time t, including stock returns ($r_{k,t}$), trading volumes ($v_{k,t}$), and news sentiment ($sent_{k,t}$). As such, $y_t = (r_{k,t}, v_{k,t}, sent_{k,t})$'. Given that these variables may be endogenously determined, the VAR model allows feedback to occur between these variables and captures their interrelations. Notably, the VAR model also accounts for the effect of contemporaneous exogenous variables in determining the value of endogenous variables. The reduced-form VAR model with p lags is specified as:

$$y_t = \Phi_0 + \sum_{i=1}^{p} \Phi_i y_{t-i} + \Psi x_t + \varepsilon_t \tag{4.1}$$

where Φ_i are (3×3) coefficient matrices for $i = 0, 1, \ldots, p$; $\varepsilon_t \sim (0, \Sigma)$ is a vector of independently and identically distributed disturbances; Ψ is a $N \times N$ parameter matrix; x_t represents an $(N \times 1)$ matrix of exogenous variables. Following previous studies (Reboredo and Ugolini 2018; Geng et al. 2021), we include a range of market factors as our exogenous variables. More specifically, market returns, as characterised by the SX5E Index, market volatility, as characterised by the V2X Index, and the European natural gas price, as characterised by the Dutch Title Transfer Facility (TTF) day ahead price. Since the sample includes securities denominated in Polish Zloty, Swiss Franc and Euro, we deviate from Reboredo and Ugolini 2018 by not specifying a currency index as a market factor.

The moving average representation of Eq.(4.1) is written as:

$$y_t = \gamma + \sum_{i=0}^{\infty} A_i \varepsilon_{t-i} + \sum_{i=0}^{\infty} A_i \Psi x_{t-i} \tag{4.2}$$

where the 3×3 coefficient matrices A_i follow the recursion $A_i = \Phi_1 A_{i-1} + \Phi_2 A_{i-2} + \ldots + \Phi_p A_{i-p}$, with A_0 being an 3×3 identity matrix and with $A_i = 0$ for $i < 0$; $\gamma = \sum_{i=0}^{\infty} A_i \Phi_0$.

4.2.2. Spillover Measure

The moving average coefficients (or variance decomposition) in Eq.(4.2) are the key to understanding the dynamics of the system. Specifically, contemporaneous aspects of connectedness are summarized in $\{A_0\}$, and the dynamics in $\{A_1, A_2, \ldots\}$. Variance decomposition allows us to achieve a greater understanding of the spillover effect by parsing the forecast error variances of each variable into components which are attributable to the various systems. In particular, the variance decomposition can be used to assess the fraction of the H-step-ahead error variance in forecasting variable i that is due to shocks in variable j, $\forall i \neq j$, for each i. Following the generalized variance decomposition (GVD) framework of Koop et al. (1996), Pesaran and Shin (1998), and Diebold and Yılmaz (2014), we define the H-step-ahead generalized error variance decomposition for

the i variable as:

$$\theta_{i\leftarrow j}^{H} = \frac{\sigma_{jj}^{-1} \sum_{h=0}^{H-1} (e_i{}' A_h \sum e_j)^2}{\sum_{h=0}^{H-1} (e_i{}' A_h \sum A_h{}' e_i)^2} \tag{4.3}$$

where i, j denote the endogenous variables ($r_{k,t}$, $v_{k,t}$, $sent_{k,t}$); e_j is a selection vector with one as j^th element and zeros elsewhere; A_h is the coefficient matrix multiplying the h-lagged shock vector in the infinite moving-average representation of the non-orthogonalized VAR (Eq.(4.2)); $\sum$ is the covariance matrix of the shock vector in the non-orthogonalized VAR; σ_{jj} is the j^{th} diagonal element of $\sum$.

The parameter $\theta_{i\leftarrow j}^{H}$ provides information on the fraction of the H-step-ahead forecast error for variable i that is attributable to shocks in variable j. Since the information is invariant to the variable ordering and shocks are not necessarily orthogonal in the GVD framework, the sum of forecast error variance for different variables is not necessarily equal to one. Diebold and Yilmaz 2012; Diebold and Yılmaz 2014 further suggest that in order to use the information provided in $\theta_{i\leftarrow j}^{H}$ to identify the spillover effects between endogenous variables, Eq.(4.3) can be normalized as:

$$\widetilde{\theta}_{i\leftarrow j}^{H} = \frac{\theta_{i\leftarrow j}^{H}}{\sum_{j=1}^{D} \theta_{i\leftarrow j}^{H}} \tag{4.4}$$

where D denotes the number of endogenous variables. In this study, the endogenous variables include stock returns ($r_{k,t}$), trading volumes ($v_{k,t}$), and news sentiment ($sent_{k,t}$), therefore, D equals 3. By construction, $\sum_{j=1}^{D} \widetilde{\theta}_{i\leftarrow j}^{H} = 1$, which means the total directional connectedness (spillover effects) from variable j to all others counts for 100% . This generalized VAR approach enables us to measure the pairwise directional contribution of the j^{th} variable to i^{th} variable at horizon h, and vice versa. For example, $\widetilde{\theta}_{i\leftarrow j}^{1}0$ with a value of 22, means the directional connectedness is from j to i is 22% out of 100% within the predictive horizon of 10 days. Therefore, in our study, we employ this approach to measure the contribution of news sentiment to firms' stock returns and trading volumes, and the contribution of stock returns or trading volumes to news sentiment for both Traditional and Clean Energy firms.

4.3. Data

4.3.1. Sample

To obtain a representative sample of the European equities market, we used daily information from the STOXX Europe Total Market Index (BKXP). The BKXP Index is chosen as it has a variable number of components and it covers approximately 95% of the free float market capitalisation of stocks across 18 European countries [2]. As constituents of the BKXP Index are actively traded, with Average Daily Traded Value (ADTV) assessed as a requirement for Index membership [3], and a mean analyst coverage of 6.24 is maintained throughout the sample period, we have rich daily news information regarding constituents.

In order to identify appropriate groups of Traditional Energy and Clean Energy companies, the BKXP constituents are assessed by Global Industry Classification Standard (GICS), including GICS sector, and GICS subsector. The GICS methodology is selected as the GICS methodology remains consistent throughout the sample period, whereas the Industry Classification Benchmark (ICB) methodology grew from 114 subsectors in 2015 to 173 subsectors in 2019 [4], introducing bias into the classification. Furthermore, GICS considers both 'earnings and market perception' within the classification methodology [5]. This is particularly important when considering companies which are undergoing a transition from the Traditional Energy classification to the Clean Energy classification, as the GICS Sub Industry classification would re-classify this security earlier than the ICB Classification.

Using GICS sector and GICS subsector, the Clean Energy classification is comprised of securities within the "Utilities" (551010) GICS Sector, with subsector of "Electric Utilities" (55101010), "Independent Power Producers & Energy Traders" (55105010), "Gas Utilities" (55102010) and "Renewable Electricity" (55105020) considered. Conversely, the

[2] Stoxx Europe Total Market Index Factsheet
[3] Stoxx Index Methodology
[4] FTSE Russell Industry Classification Benchmark History
[5] The Global Industry Classification Standard (GICS) Handbook

Traditional Energy classification considers securities from the "Energy" (101020) GICS Sector, with securities from the subsectors of "Integrated Oil & Gas" (10102010), "Oil & Gas Drilling" (10101010), "Oil & Gas Equipment & Services" (10101020), "Oil & Gas Exploration & Production" (10102020), "Oil & Gas Refining & Marketing" (10102030), "Oil & Gas Storage & Transportation" (10102040), and "Coal & Consumable Fuels" (10102050) present in the subsample of Traditional Energy companies.

The sample period spans 1st January 2015 to 1st April 2022, with the starting date determined by data availability. Of the 58 BKXP constituents classified as Traditional Energy companies by GICS subsector, we excluded the 8 firms which have incomplete daily news sentiment records. Similarly, of the 46 BXKP constituents classified as Clean Energy companies, we excluded 11 firms which had incomplete news sentiment data, leaving 50 Traditional Energy companies and 35 Clean Energy companies in our sample.[6] The sample companies accounted for 86% and 76% of the Traditional Energy and Clean Energy classifications as of the sample end date, respectively. Table 4.1 lists the number of company from each subsector after excluding the firms with incomplete sentiment data.[7]

[6] Anderloni and Tanda (2017) study the Initial Public Offers (IPOs) of 144 energy firms listed on the stock exchanges across 13 European countries during 2000-2014. Their sample is made of 86 (59.7%) traditional and 58 (40.3%) green energy companies. The sample of this paper is the firms listed in the STOXX Europe Total Market Index during 2015 –2022 with non-missing sentiment data and trading data. Using the GICS, we classify 50 (58.8%) traditional and 35 (41.2%) clean energy companies. The ratio between traditional and clean energy is comparable to the study of Anderloni and Tanda (2017).

[7] To further mitigate potential sample bias driven by classification, we re-structured the sample for Clean Energy by classifying the GICS subsector - "Independent Power Producers & Energy Traders" (55105010) as Traditional Energy, which leads to 32 Clean Energy and 52 Traditional Energy companies. Then, we re-estimated the empirical analysis, and found that the results remain robust to the alternative sample classification. Tables for robustness tests are provided in Appendix B.

Table 4.1. "Traditional Energy" and "Clean Energy" Classification

Clean Energy Classification - GICS sector "Utilities 551010"		
GICS Subsector name	**GICS Subsector code**	**No.**
Electric Utilities	55101010	20
Independent Power Producers & Energy Traders	55105010	3
Gas Utilities	55102010	5
Renewable Electricity	55105020	7
Total number of Clean Energy companies		35
Traditional Energy Classification - GICS sector "Energy 101020"		
GICS Subsector name	**GICS Subsector code**	**No.**
Integrated Oil & Gas	10102010	9
Oil & Gas Drilling	10101010	3
Oil & Gas Equipment & Services	10101020	15
Oil & Gas Exploration & Production	10102020	11
Oil & Gas Refining & Marketing	10102030	5
Oil & Gas Storage & Transportation	10102040	7
Coal & Consumable Fuels	10102050	0
Total number of Traditional Energy companies		50

4.3.2. Sentiment Data

We then obtained information on news articles for each company from Bloomberg. Bloomberg uses supervised machine learning techniques to emulate human cognition in processing textual information. First, each news article is assigned a categorical polarity score, e.g., 1, 0, -1, which indicates positive, neutral or negative sentiment, respectively. The labelling is based on the question "If an investor having a long position in the security mentioned were to read this news or tweet, is he/she bullish, bearish or neutral on his/her holdings?" If an investor is bullish, then the categorical value assigned is 1. If an investor is bearish, then the categorical value is -1, otherwise 0 is assigned for a neutral attitude. Then, the annotated data is fed into machine-learning models, such as a support vector machine (SVM). A confidence level, which is considered as the probability associated with the article being negative, neutral or positive, is assigned to each news article. Once the model is trained, when new information comes, the model automatically assigns a probability of being positive, negative or neutral to each news story.

We then use a company-level News Sentiment Index score (NSI), which is the confidence-weighted average of news sentiment recorded on that day. NSI delivers a numerical value ranging from -1 to 1, with -1 representing extremely negative sentiment, 1 representing extremely positive sentiment, and 0 being neutral sentiment. Specifically, NSI for each company is computed by multiplying the categorical polarity score of each news article with it's associated probability score (Gupta and Banerjee 2019):

$$NSI = \sum_{i=0}^{Z} 1 \times Prob.(Positive)_i + \sum_{i=0}^{Z} (-1) \times Prob.(Negative)_i \in [-1; 1] \tag{4.5}$$

where Z refers to the total number of daily news stories published by Bloomberg corresponding to company k, $Prob.$ refers to the corresponding confidence probability assigned by the SVM, and $\sum Prob.$ equals 100%. $NSI_{k,t}$ represents daily News sentiment for asset k at time t ($sent_{k,t}$) in Eq.(4.1).

4.3.3. Security Data

Data on daily closing price and daily trading volumes is obtained from Bloomberg for each security, whereby the daily closing price and daily trading volumes were adjusted to reflect Spin-Offs, Stock Splits, Stock Consolidations, Stock Dividends and Rights Offerings. The use of corporate action adjusted price and volume information is considered theoretically appropriate, as this study aims to isolate the relationship between News sentiment and stock price innovation. As such, inclusion of price innovation driven by exogenous events, such as Corporate Actions, may distort the empirical results of the study.

Daily return ($LogReturn$) for each security is computed as the first difference of the natural logarithm of the corporate action adjusted closing prices, and daily trading volumes ($LogVolume$) are defined as the natural logarithm of the number of shares traded on the primary exchange for each security. In our empirical analysis, $LogReturn_{k,t}$ refers to stock return for asset k at time t ($r_{k,t}$) and $LogVolume_{k,t}$ represents daily trading volumes

for asset k at time t ($\upsilon_{k,t}$) in Eq.(4.1).

4.3.4. Market Factor Data

Finally, we collect data from Bloomberg on several factors included in the vector x_t in Eq.(4.1), which may be relevant in determining the values of the endogenous variables. We consider a range of market factors; stock market returns, as proxied by the first log difference of the EURO STOXX 50 Index (SX5E), stock market volatility, as proxied by the EURO STOXX 50 Volatility Index (V2X Index) and a representative price of European natural gas, Dutch Title Transfer Facility Day Ahead Price (TTFGDAHD Index). Specifically, the EURO STOXX 50 Index (*SX5E*) is selected, as the Index is comprised of the Eurozone's largest and most traded companies, representing a diversified exposure to the largest companies of each ICB Supersector [8]. Additionally, the EURO STOXX 50 Volatility Index (*V2X*) is considered as a representative measure of volatility within Eurozone equities, as the V2X Index considers the implied volatility on EURO STOXX 50 Index options within a rolling 30 day expiry [9]. We also consider changes in TTF Day Ahead prices (*TTFGDAHD*), as natural gas price is known to have a long-term relationship with renewable energy demand (Berry 2005; Fell and Kaffine 2018), and by extension, the stock price of renewable energy companies.

Table B.4 provides summary statistics for the exogenous market variables. All three variables have a mean value close to zero, yet exhibit a large dispersion. The EURO STOXX 50 Index (*SX5E*) has a standard deviation of 1.28% and its distribution is moderately left skewed (-0.95). The European natural gas prices Index (*TTFGDAHD*) exhibits higher variability than the (*SX5E*) with a standard deviation of 5.7% and the EuroSTOXX 50 Volatility Index *V2X* has the highest standard deviation of 7.6%. The distributions of both variables are characterised by fat-tails.

[8] EuroSTOXX 50 Index Factsheet
[9] EuroSTOXX 50 Volatility Index Factsheet

Table 4.2. Exogenous Market Factor Variables

This table presents the descriptive statistics for exogenous market factor data: the first difference of the log EURO STOXX 50 Index (*SX5E*), the EURO STOXX 50 Volatility Index (*V2X*), and the natural gas price (*TTFGDAHD*) for the sample period of 1st January 2015 to 1st April 2022. S.D., Min and Max denote standard deviation, minimum and maximum values, respectively.

Variables	Obs.	Mean	S.D.	Skewness	Kurtosis	Min	Max
SX5E	146,964	0.0001	0.0128	-0.9531	15.3541	-0.1324	0.0883
V2X	146,964	0.0001	0.0763	0.7599	7.0803	-0.4347	0.4859
TTFGDAHD	146,964	0.0014	0.0567	-2.3415	69.8578	-1.0165	0.5219

4.4. Empirical Results

4.4.1. Descriptive Statistics

Table 4.3 presents the descriptive statistics of the key variables in our study (*LogReturn*, *LogVolume*, and *NSI*) for each company over the sample period 1st January 2015 - 1st April 2022. Panel A summarizes the statistics of Clean Energy companies, and Panel B displays statistics of Traditional Energy companies. Whilst the average daily returns over the sample period are negative for the Traditional Energy segment (-0.00005), the Clean Energy segment exhibits a small, positive return on average (0.00025). Additionally, 27 of the 35 Clean Energy companies exhibited a positive mean daily return, compared with 24 of the 50 Traditional Energy companies. This divergence is likely due to the concomitant factors of additional strategic investment in the Clean Energy sector, and increased financing costs of Traditional Energy projects (Pickl 2019). Notably, on average, the dispersion of daily returns within the Traditional Energy segment (3.17%) is substantially larger than that of the Clean Energy segment (2.41%), with the Traditional Energy segment containing both the highest (VBKGY) and lowest (CGGFP) average return securities within the sample. The standard deviations of daily returns indicate that stock price volatility is more pronounced for Traditional Energy companies than Clean Energy companies.

Regarding liquidity, characterised by the average *LogVolume*, Clean Energy and Tra-

ditional Energy companies display similar dynamics. Traditional Energy companies are marginally more liquid than Clean Energy companies, recording mean values of 2.61 and 2.54 respectively. Conversely, the standard deviation of *LogVolume* follows a markedly different profile, as Clean Energy companies (23.42%) exhibited a much larger standard deviation in *LogVolume* than Traditional Energy companies (14.73%). Taken in unison, this implies that Clean Energy companies are, on average, less liquid and more susceptible to large deviations in traded volume. This could be due to the comparative free float of Traditional and Clean Energy companies, whereby Clean Energy companies held a lower mean free float throughout the sample period, thus less shares were available to be traded.

The mean values of *NSI* for both the Clean Energy segment and Traditional Energy segment are positive, indicating that throughout our sample period, an aggregate positive sentiment was extracted from news articles. Notably, the average *NSI* of Clean Energy (0.0235) companies is moderately larger than that exhibited by Traditional Energy (0.0197) companies. Given the increased social awareness of climate change and European energy policy objectives, it is expected that news articles relating to Clean Energy firms are likely to be more positive than news articles relating to Traditional Energy companies. The highest positive News sentiment (0.15) is held by Equinor (EQNRNO), which is classified as a Clean Energy company. This is likely due to Equinor's relatively large market capitalisation, and by extension, prominent media coverage, combined with the high Environmental, Air Quality and Climate Change disclosure scores throughout the sample period. Additionally, on average, the variance of *NSI* is larger for Traditional Energy (16.83%) companies than Clean Energy (15.80%) companies, which is likely due to the comparatively larger analyst coverage of Traditional Energy companies throughout the sample, and relative maturity of the two industries (Newbery 2016).

Table 4.3. Descriptive Statistics

This table reports means and standard deviations in parentheses for returns, trading volumes and News sentiment index for 35 Clean Energy firms in Panel A and 50 Traditional Energy firms in Panel B for the sample period of 1st January 2015 to 1st April 2022.

Ticker	Company	*LogReturn*	*LogVolume*	*NSI*
Panel A. Clean Energy Companies.				
ABIOFP	ABIO FP Equity	0.0005 (0.0183)	2.3403 (0.0764)	0.0073 (0.0686)
ANASQ	ANA SQ Equity	0.0006 (0.0181)	2.4682 (0.0471)	0.0292 (0.1428)
BKWSE	BKW SE Equity	0.0008 (0.0135)	2.3325 (0.0760)	0.0029 (0.0835)
DRXLN	DRX LN Equity	0.0003 (0.0251)	2.6319 (0.0406)	-0.0014 (0.1510)
EBKGY	EBK GY Equity	0.0007 (0.0199)	1.6775 (0.2954)	0.0075 (0.1038)
ECVGY	ECV GY Equity	0.0008 (0.0216)	2.4852 (0.0810)	0.0239 (0.1329)
EDFFP	EDF FP Equity	-0.0005 (0.0232)	2.6925 (0.0330)	0.0028 (0.1018)
EDPPL	EDP PL Equity	0.0002 (0.0160)	2.7521 (0.0291)	0.0720 (0.2253)
EDPRPL	EDPR PL Equity	0.0008 (0.0170)	2.5339 (0.0729)	0.1019 (0.2378)
ELESQ	ELE SQ Equity	0.0001 (0.0141)	2.6601 (0.0444)	0.0286 (0.1695)
ELIBB	ELI BB Equity	0.0007 (0.0145)	2.3530 (0.0570)	0.0047 (0.0677)
ENAPW	ENA PW Equity	-0.0003 (0.0242)	2.5692 (0.0508)	0.0062 (0.1723)
ENELIM	ENEL IM Equity	0.0003 (0.0163)	2.8459 (0.0241)	0.0415 (0.1568)
ENGPW	ENG PW Equity	-0.0006 (0.0195)	2.5387 (0.1258)	0.0122 (0.2031)
ENGSQ	ENG SQ Equity	-0.0001 (0.0140)	2.6363 (0.0425)	0.0370 (0.2007)
ERGIM	ERG IM Equity	0.0006 (0.0175)	2.5087 (0.0453)	0.0068 (0.0729)
EVNAV	EVN AV Equity	0.0005 (0.0147)	2.3496 (0.0673)	0.0060 (0.0863)
FORTUMFH	FORTUM FH Equity	-0.0001 (0.0175)	2.6682 (0.0314)	0.0380 (0.1780)
IBESQ	IBE SQ Equity	0.0003 (0.0136)	2.8138 (0.0328)	0.0418 (0.1877)
IGIM	IG IM Equity	0.0003 (0.0154)	2.6759 (0.0329)	0.0202 (0.1657)
NEOENFP	NEOEN FP Equity	0.0011 (0.0243)	2.4156 (0.0999)	0.0158 (0.1280)
NTGYSQ	NTGY SQ Equity	0.0001 (0.0156)	2.6491 (0.0493)	0.0624 (0.2021)
ORSTEDDC	ORSTED DC Equity	-0.0015 (0.0972)	2.5619 (0.0419)	0.0290 (0.1421)
PEPPW	PEP PW Equity	0.0005 (0.0275)	2.1091 (0.2034)	0.0019 (0.1009)
PGEPW	PGE PW Equity	-0.0004 (0.0268)	2.6789 (0.0394)	0.0184 (0.1720)
REESQ	REE SQ Equity	0.0000 (0.0126)	2.6680 (0.0445)	0.0123 (0.1946)
RUIFP	RUI FP Equity	0.0001 (0.0177)	2.4919 (0.0399)	0.0423 (0.1611)
SCATCNO	SCATC NO Equity	0.0008 (0.0266)	2.5229 (0.0671)	0.0087 (0.0904)
SRGIM	SRG IM Equity	0.0002 (0.0152)	2.7782 (0.0263)	0.0498 (0.1833)
SSELN	SSE LN Equity	0.0000 (0.0160)	2.6954 (0.0293)	0.0083 (0.2003)
TPEPW	TPE PW Equity	-0.0003 (0.0258)	2.7157 (0.0408)	-0.0003 (0.2000)
TRNIM	TRN IM Equity	0.0004 (0.0141)	2.7461 (0.0258)	0.0358 (0.1512)
UN01GY	UN01 GY Equity	0.0006 (0.0178)	2.5729 (0.0509)	0.0137 (0.1631)
VERAV	VER AV Equity	0.0010 (0.0211)	2.4783 (0.0459)	0.0138 (0.1175)
VLTSAFP	VLTSA FP Equity	0.0005 (0.0204)	2.2118 (0.1628)	0.0100 (0.0883)
Summary: Clean Energy		0.0002 (0.0241)	2.5423 (0.2342)	0.0235 (0.1580)
Panel B. Traditional Energy Companies.				
AKASTNO	AKAST NO Equity	-0.0005 (0.0300)	2.4634 (0.1036)	0.0083 (0.0910)
AKRBPNO	AKRBP NO Equity	0.0012 (0.0283)	2.5949 (0.0421)	0.0131 (0.1239)

Continued on next page

Table 4.3 - continued from previous page

Ticker	Company	*LogReturn*	*LogVolume*	*NSI*
AKSONO	AKSO NO Equity	-0.0001 (0.0344)	2.6372 (0.0562)	0.0354 (0.1701)
BORRNO	BORR NO Equity	-0.0017 (0.0617)	2.4566 (0.2189)	0.0011 (0.1204)
BPLN	BP/ LN Equity	0.0000 (0.0204)	2.8533 (0.0230)	-0.0061 (0.2078)
BWLPGNO	BWLPG NO Equity	0.0001 (0.0311)	2.5873 (0.0432)	0.0210 (0.1267)
BWONO	BWO NO Equity	-0.0005 (0.0417)	2.5670 (0.0770)	0.0215 (0.1322)
CGGFP	CGG FP Equity	-0.0019 (0.0447)	2.6831 (0.1050)	0.0216 (0.1894)
CNELN	CNE LN Equity	0.0000 (0.0294)	2.6368 (0.0374)	-0.0102 (0.1802)
DECLN	DEC LN Equity	0.0005 (0.0230)	2.5929 (0.1180)	0.0075 (0.0847)
DNONO	DNO NO Equity	-0.0001 (0.0343)	2.7451 (0.0350)	0.0303 (0.1843)
DRLCODC	DRLCO DC Equity	-0.0005 (0.0327)	2.4453 (0.0576)	0.0102 (0.1157)
ENIIM	ENI IM Equity	0.0000 (0.0186)	2.8049 (0.0241)	0.0067 (0.1912)
ENOGLN	ENOG LN Equity	0.0010 (0.0330)	2.4474 (0.1011)	0.0059 (0.0824)
ENQSS	ENQ SS Equity	0.0000 (0.0399)	2.7169 (0.0526)	0.0078 (0.1510)
EQNRNO	EQNR NO Equity	0.0005 (0.0195)	2.7141 (0.0286)	0.1518 (0.3025)
EURNBB	EURN BB Equity	0.0000 (0.0225)	2.5770 (0.0451)	0.0210 (0.1812)
FRONO	FRO NO Equity	-0.0001 (0.0350)	2.5851 (0.0556)	0.0175 (0.2126)
GALPPL	GALP PL Equity	0.0002 (0.0199)	2.6580 (0.0317)	0.0264 (0.1847)
GTTFP	GTT FP Equity	0.0004 (0.0220)	2.3891 (0.0498)	0.0066 (0.0636)
HAFNINO	HAFNI NO Equity	-0.0002 (0.0296)	2.4823 (0.0724)	0.0248 (0.1469)
HBRLN	HBR LN Equity	-0.0011 (0.0539)	2.5661 (0.0611)	0.0249 (0.1659)
IPCOSS	IPCO SS Equity	0.0008 (0.0323)	2.5584 (0.0644)	0.0184 (0.1436)
LTSPW	LTS PW Equity	0.0005 (0.0226)	2.5354 (0.0480)	0.0169 (0.1817)
LUNESS	LUNE SS Equity	0.0007 (0.0242)	2.6180 (0.0359)	0.0606 (0.1753)
MGNNO	MGN NO Equity	0.0000 (0.0391)	2.2463 (0.2890)	0.0064 (0.0616)
NESTEFH	NESTE FH Equity	0.0010 (0.0214)	2.6528 (0.0369)	0.0235 (0.1520)
ODLNO	ODL NO Equity	0.0004 (0.0334)	2.5076 (0.1003)	0.0049 (0.0803)
OMVAV	OMV AV Equity	0.0004 (0.0223)	2.5604 (0.0349)	0.0171 (0.1647)
PENNO	PEN NO Equity	0.0004 (0.0421)	2.5127 (0.0946)	0.0055 (0.0621)
PGNPW	PGN PW Equity	0.0002 (0.0213)	2.7211 (0.0333)	0.0236 (0.1773)
PGSNO	PGS NO Equity	-0.0014 (0.0527)	2.6982 (0.0441)	0.0029 (0.1605)
PKNPW	PKN PW Equity	0.0003 (0.0217)	2.6206 (0.0374)	0.0122 (0.1595)
REPSQ	REP SQ Equity	-0.0001 (0.0212)	2.7655 (0.0342)	0.0331 (0.2027)
SBMONA	SBMO NA Equity	0.0002 (0.0223)	2.6190 (0.0418)	0.0005 (0.1591)
SBOAV	SBO AV Equity	-0.0001 (0.0267)	2.3425 (0.0555)	0.0086 (0.1247)
SHELLNA	SHELL NA Equity	-0.0001 (0.0200)	2.7767 (0.0287)	0.0020 (0.1402)
SPMIM	SPM IM Equity	-0.0012 (0.0308)	2.7701 (0.0362)	0.0116 (0.2232)
SRSIM	SRS IM Equity	-0.0001 (0.0269)	2.7376 (0.0449)	0.0111 (0.1102)
SUBCNO	SUBC NO Equity	0.0000 (0.0271)	2.6450 (0.0397)	0.0747 (0.2471)
TEFP	TE FP Equity	-0.0004 (0.0282)	2.5845 (0.0489)	0.0008 (0.1205)
TENIM	TEN IM Equity	0.0001 (0.0235)	2.7156 (0.0264)	0.0193 (0.2646)
TGSNO	TGS NO Equity	-0.0001 (0.0274)	2.5525 (0.0389)	0.0191 (0.1580)
TLWLN	TLW LN Equity	-0.0010 (0.0545)	2.7737 (0.0373)	0.0059 (0.2146)
TRMDADC	TRMDA DC Equity	-0.0003 (0.0223)	2.4215 (0.0890)	0.0063 (0.0952)
TTEFP	TTE FP Equity	0.0001 (0.0180)	2.7484 (0.0261)	0.0183 (0.1656)
VBKGY	VBK GY Equity	0.0023 (0.0339)	2.4243 (0.0788)	0.0087 (0.0919)

Continued on next page

Table 4.3 - continued from previous page

Ticker	Company	*LogReturn*	*LogVolume*	*NSI*
VKFP	VK FP Equity	-0.0015 (0.0406)	2.5608 (0.0550)	0.0516 (0.1888)
VPKNA	VPK NA Equity	-0.0002 (0.0162)	2.5448 (0.0362)	0.0141 (0.1337)
WGLN	WG/ LN Equity	-0.0007 (0.0295)	2.6642 (0.0436)	0.0234 (0.1898)
Summary:	**Traditional Energy**	-0.0001 (0.0317)	2.6097 (0.1472)	0.0197 (0.1683)

4.4.2. Correlation

Table 4.4 displays the Pearson correlation values for *NSI* with returns and trading volumes. Panel A shows the results for the 35 Clean Energy firms. Using the time series data of each Clean Energy firm, we find that 85% (29 out of 35) of the firms' stock returns are positively related to *NSI*, and the rest exhibit insignificant correlation between stock returns and *NSI*. We also find that 54% (19 out of 35) of the firms' trading volumes are positively associated with *NSI*, a negative correlation for SCATCNO only and insignificant correlation for the rest of the firms. A reason for Scatec ASA's (SCATCNO) negative correlation between stock returns and *NSI* is the failure to meet earnings projections in 2021 and 2022. Although Scatec ASA (SCATCNO) missed earnings projections, with a subsequent fall in stock price, the *NSI* value remained positive within the period, with news articles focusing on Scatec ASA's growth potential. Across the panel data of all 35 Clean Energy firms, on average, we find a significant positive correlation between stock returns and *NSI* with a value of 0.053, and a significant positive correlation between trading volumes and *NSI* with a value of 0.061.

Table 4.4. Correlation: Clean Energy

This table reports the Pearson correlation coefficient for News sentiments index with returns and trading volumes for 35 Clean Energy companies for the sample period of 1st January 2015 to 1st April 2022. $***$, $**$, and $*$ denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	NSI	
Ticker	***LogReturn***	***LogVolume***
ABIOFP	0.049**	0.079***
ANASQ	0.116***	0.186***
BKWSE	0.148***	0.074***
DRXLN	0.034	0.049*
EBKGY	0.098***	0.009
ECVGY	0.063***	0.034
EDFFP	0.061***	-0.019
EDPPL	0.038*	-0.01
EDPRPL	0.087***	0.049**
ELESQ	0.122***	0.006
ELIBB	0.131***	0.072***
ENAPW	0.021	0.173***
ENELIM	0.097***	0.048**
ENGPW	0.069***	0.021
ENGSQ	0.096***	0.01
ERGIM	0.048**	0.028
EVNAV	0.066***	0.078***
FORTUMFH	0.122***	0.065***
IBESQ	-0.019	0.092***
IGIM	-0.026	0.060**
NEOENFP	0.144***	0.062***
NTGYSQ	0.099***	0.017
ORSTEDDC	0.067***	0.060**
PEPPW	0.088***	-0.003
PGEPW	0.050**	0.032
REESQ	0.084***	-0.019
RUIFP	0.060**	0.015
SCATCNO	0.01	-0.050**
SRGIM	0.070***	0.040*
SSELN	0.063***	0.151***
TPEPW	0.049*	0.075***
TRNIM	0.022	-0.029
UN01GY	0.105***	0.012
VERAV	0.068***	0.066***
VLTSAFP	0.0775***	0.0985***
Summary: Clean Energy	0.053***	0.061***

Table 4.5 presents the correlation results for Traditional Energy firms. We find that 27 of the 50 companies display a significant positive correlation between stock returns and *NSI*. Notable exceptions to this include GazTransport et Technigaz (GTTFP) and Lundin Energy (LUNESS), which exhibit a negative correlation between *NSI* and stock returns throughout the sample period. This is likely due to the nature of GazTransport et Technigaz's primary business, offering cargo containment systems for Liquified Natural Gas (LNG) carriers, and the pending merger between Aker (AKRBPNO) and Lundin Energy (LUNESS), to form Orron Energy AB (ORRONSS) [10]. We also find that 21 firms' stock trading volumes are positively correlated with *NSI*, and 7 firms show a negative correlation. These 7 firms are Aker Solutions (AKSONO), British Petroleum (BPLN), CGG SA (CGGFP), Hafnia Ltd (HAFNINO), Shell PLC (SHELLNA), TGS ASA (TGSNO) and John Wood Group (WGLN). Based on a panel of all 50 Traditional Energy firms, the result shows that stock returns or trading volumes are significantly positive correlated to *NSI* with respective correlation values of 0.049 and 0.025. Taken in unison, it can be concluded that *NSI* has a more consistent relationship with returns than daily trading volume throughout the sample, which corroborates Li et al. 2014's finding of a positive relationship between News sentiment and asset returns.

When comparing the results of Table 4.4 (Clean Energy) and Table 4.5 (Traditional Energy), it is apparent that, on average, *NSI* is positively correlated with stock price returns within the sample period, which is consistent with investor sentiment theory (Qiu and Welch 2004). Moreover, the relationship between *NSI* and daily trading volumes, whilst less consistent at the firm level, is significantly positive, supporting the earlier findings of Joseph et al. 2011; Ryu et al. 2017.

[10] Merger between Aker BP and Lundin Energy's E&P business completed

Table 4.5. Correlation: Traditional Energy

This table reports the Pearson correlation coefficient for News sentiments index with returns and trading volumes for 50 Traditional Energy firms for the sample period of 1st January 2015 to 1st April 2022. $***$, $**$, and $*$ denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	NSI	
Ticker	***LogReturn***	***LogVolume***
AKASTNO	0.050**	0.040*
AKRBPNO	0.021	0.015
AKSONO	0.056**	-0.078***
BORRNO	0.03	-0.016
BPLN	0.002	-0.044*
BWLPGNO	0.074***	0.113***
BWONO	0.095***	0.129***
CGGFP	0.028	-0.103***
CNELN	0.004	-0.002
DECLN	0.080***	0.022
DNONO	-0.037	0.129***
DRLCODC	0.027	-0.036
ENIIM	-0.006	0.02
ENOGLN	0.018	0.012
ENQSS	0.032	0.062***
EQNRNO	0.045*	0.111***
EURNBB	0.025	-0.025
FRONO	0.037	-0.004
GALPPL	0.068***	0.154***
GTTFP	-0.085**	0.122***
HAFNINO	0.014	-0.059**
HBRLN	0.025	0.133***
IPCOSS	0.092***	0.041*
LTSPW	0.073***	0.139***
LUNESS	-0.104***	0.039*
MGNNO	-0.004	-0.023
NESTEFH	0.069***	0.120***
ODLNO	0.034	0.021
OMVAV	0.023	0.101***
PENNO	0.098***	0.023
PGNPW	0.125***	0.026
PGSNO	0.055**	0.028
PKNPW	0.061***	0.012
REPSQ	0.134***	0.091***
SBMONA	0.102***	0.032
SBOAV	0.090***	0.031
SHELLNA	0.146***	-0.096***
SPMIM	0.057**	0.093***
SRSIM	0.029	-0.002
SUBCNO	0.167***	0.039

Continued on next page

Table 4.5 - continued from previous page

	NSI	
Ticker	***LogReturn***	***LogVolume***
TEFP	0.080***	-0.014
TENIM	0.026	0.005
TGSNO	0.016	-0.086***
TLWLN	0.153***	0.044*
TRMDADC	-0.003	0.051**
TTEFP	0.147***	0.079***
VBKGY	0.055**	0.026
VKFP	0.078***	0.099***
VPKNA	0.112***	0.061***
WGLN	0.038*	-0.052**
Summary: Traditional Energy	0.049***	0.025***

4.4.3. Spillover Effects

To compute the value of the spillover parameter for each firm, we first estimate the generalised VAR model (Eq.(4.1)) using the endogenous variables, *LogReturn*, *LogVolume*, and *NSI*. The lag order in the VAR model is determined by the Bayesian information criterion (BIC) (Reboredo and Ugolini 2018). Then, we use the information provided by the estimation and calculate the directional connectedness for a horizon of 10 days ($\widetilde{\theta}^{10}_{i\leftarrow j}$) following Eq.(4.4).

Table 4.6 reports the spillover parameters for the 35 Clean Energy firms. Columns (1)–(3) show the spillover effect between stock returns and *NSI*. Specifically, Column (1) shows the spillover effect from *NSI* to stock returns ($\widetilde{\theta}_{LogReturn\leftarrow NSI}$) and Column (2) is the spillover effect to *NSI* from stock returns ($\widetilde{\theta}_{NSI\leftarrow LogReturn}$). Column (3) is the net effect of spillover, which is measured as the difference between $\widetilde{\theta}_{LogReturn\leftarrow NSI}$ and $\widetilde{\theta}_{NSI\leftarrow LogReturn}$. The firm with highest spillover from *NSI* to stock returns is DRXLN ($\widetilde{\theta}_{LogReturn\leftarrow NSI} = 7.55\%$), while its spillover from stock returns to *NSI* ($\widetilde{\theta}_{NSI\leftarrow LogReturn}$) is only 1.53%. This evidence implies that DRXLN's News sentiment has a large impact on stock returns, while the information embedded within stock returns is not fully reflected in News sentiment, precipitating an asymmetric influence of News sentiment on DRXLN's stock price. Notably, ELIBB ($\widetilde{\theta}_{LogReturn\leftarrow NSI} = 5.77\%$) and EDFFP ($\widetilde{\theta}_{LogReturn\leftarrow NSI} = 5.37\%$) have the second and third largest spillover from *NSI* to stock returns, and the

reverse spillovers ($\widetilde{\theta}_{NSI \leftarrow LogReturn}$) are comparable with respective values of 5.2% and 5.33%. This finding indicates that News sentiment impacts stock returns and in return, stock returns feed back to News sentiment. In addition, we notice that ENELIM has the lowest spillovers from and to stock returns ($\widetilde{\theta}_{LogReturn \leftarrow NSI} = 0.68\%$, $\widetilde{\theta}_{NSI \leftarrow LogReturn} = 0.85\%$). In general, for stock returns of Clean Energy firms, the empirical evidence indicates that on average, the spillover from *NSI* to stock returns is 2.81%. In return, the spillover from stock returns to *NSI* is 2.31%, leading to a net spillover value of 0.5% from *NSI* to stock returns.

Columns (4)–(6) present the spillover effect between daily trading volumes and *NSI*. Similarly, Column (4) reports the spillover effect from *NSI* to trading volumes ($\widetilde{\theta}_{LogVolume \leftarrow NSI}$), Column (5) shows the reverse spillover from trading volumes to *NSI* ($\widetilde{\theta}_{NSI \leftarrow LogVolume}$), and Column (6) is the difference between $\widetilde{\theta}_{LogVolume \leftarrow NSI}$ and $\widetilde{\theta}_{NSI \leftarrow LogVolume}$. The firms with the top 3 highest spillovers from *NSI* to trading volumes are RUIFP (4.06%), UN01GY (3.39%), and EDFFP (3.37%). Their pairwise reverse spillovers from trading volumes to *NSI* are RUIFP (3.37%), UN01GY (1%), and EDFFP (3.53%). Firm ELESQ has the lowest spillovers from and to trading volumes ($\widetilde{\theta}_{LogVolume \leftarrow NSI} = 0.7\%$, $\widetilde{\theta}_{NSI \leftarrow LogVolume} = 0.64\%$). On average, for daily trading volumes of Clean Energy firms, the results show that spillovers from and to *NSI* are 1.86% and 1.73% respectively, leading to a net spillover from *NSI* to trading volumes of 0.13%.

So far, our empirical evidence on European Clean Energy firms is consistent with Reboredo and Ugolini 2018's finding from a sample of 17 U.S. Clean Energy firms that: 1) News sentiment has a relatively small impact on stock returns and trading volumes, 2) stock returns and trading volumes also have a limited impact on News sentiment, and 3) spillovers are asymmetric. However, in our sample of 35 European Clean Energy firms, we highlight that sentiment has stronger impact on stock returns rather than trading volumes.

Table 4.6. Spillover Effects: Clean Energy

This table reports spillover effects calculated from Eq. (4.4) between the News sentiment index (*NSI*), returns (*LogReturn*) and trading volumes (*LogVolume*) for 35 Clean Energy firms. The number of lags for the VAR model is selected using the Bayesian information criterion. $\widetilde{\theta}_{LogReturn \leftarrow NSI}$ refers to the spillover effect from NSI to stock returns and $\widetilde{\theta}_{LogVolume \leftarrow NSI}$ is to the spillover effect from NSI to trading volumes. Conversely, $\widetilde{\theta}_{NSI \leftarrow LogReturn}$ refers to the spillover effect from stock returns to NSI, and $\widetilde{\theta}_{NSI \leftarrow LogVolume}$ is the spillover effect from trading volumes to NSI. The values of spillover effects, ranging from 0% to 100% are presented in Columns (1)–(6).

		Returns			**Trading volumes**		
		$\widetilde{\theta}_{LogReturn \leftarrow NSI}$	$\widetilde{\theta}_{NSI \leftarrow LogReturn}$	**Net**	$\widetilde{\theta}_{LogVolume \leftarrow NSI}$	$\widetilde{\theta}_{NSI \leftarrow LogVolume}$	**Net**
Ticker	**Lag**	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
ABIOFP	4	4.06	3.18	0.88	2.77	2.45	0.32
ANASQ	5	4.44	4.10	0.34	3.19	1.63	1.56
BKWSE	6	4.49	4.17	0.32	2.84	3.20	-0.36
DRXLN	2	7.55	1.53	6.02	1.18	0.83	0.35
EBKGY	4	2.11	1.69	0.42	1.13	1.20	-0.07
ECVGY	6	2.36	1.87	0.49	2.02	1.84	0.18
EDFFP	3	5.37	5.33	0.04	3.37	3.53	-0.16
EDPPL	5	1.69	1.71	-0.02	0.93	0.77	0.16
EDPRPL	2	1.98	1.70	0.28	2.21	1.89	0.32
ELESQ	2	2.90	1.42	1.48	0.70	0.64	0.06
ELIBB	5	5.77	5.20	0.57	1.33	2.02	-0.69
ENAPW	3	3.04	2.57	0.47	2.20	2.82	-0.62
ENELIM	4	0.68	0.85	-0.17	2.10	2.49	-0.39
ENGPW	5	3.24	2.51	0.73	1.60	1.95	-0.35
ENGSQ	6	2.03	2.07	-0.04	1.75	0.88	0.87
ERGIM	5	2.80	2.53	0.27	2.94	2.97	-0.03
EVNAV	4	1.11	0.99	0.12	1.22	1.09	0.13
FORTUMFH	5	1.82	0.91	0.91	1.93	1.39	0.54
IBESQ	6	2.10	1.56	0.54	1.02	0.85	0.17
IGIM	3	2.04	2.94	-0.90	1.82	2.39	-0.57
NEOENFP	6	1.02	1.03	-0.01	0.80	0.99	-0.19
NTGYSQ	6	2.85	2.40	0.45	0.86	0.79	0.07
ORSTEDDC	3	2.23	1.78	0.45	1.74	1.39	0.35
PEPPW	2	3.98	2.17	1.81	1.57	1.39	0.18
PGEPW	3	2.85	2.94	-0.09	1.41	1.31	0.10
REESQ	6	1.40	1.80	-0.40	0.82	0.62	0.20
RUIFP	4	3.16	3.02	0.14	4.06	3.37	0.69
SCATCNO	6	3.00	3.01	-0.01	2.75	3.67	-0.92
SRGIM	4	1.76	1.14	0.62	1.46	0.97	0.49
SSELN	4	2.94	0.62	2.32	1.35	1.00	0.35
TPEPW	4	3.02	2.58	0.44	1.70	1.68	0.02
TRNIM	5	1.09	1.20	-0.11	1.24	0.98	0.26
UN01GY	4	2.16	1.66	0.50	3.39	1.00	2.39
VERAV	5	1.70	1.91	-0.21	1.23	1.23	0.00
VLTSAFP	4	3.50	4.63	-1.13	2.50	3.23	-0.73

Continued on next page

Table 4.6 - continued from previous page

		Returns			Trading volumes		
		$\widetilde{\theta}_{LogReturn \leftarrow NSI}$	$\widetilde{\theta}_{NSI \leftarrow LogReturn}$	**Net**	$\widetilde{\theta}_{LogVolume \leftarrow NSI}$	$\widetilde{\theta}_{NSI \leftarrow LogVolume}$	**Net**
Ticker	**Lag**	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Clean Energy		2.81	2.31	0.50	1.86	1.73	0.13

In light of Reboredo and Ugolini 2018's findings, we extend their study to provide a comparative analysis between Clean Energy companies, and Traditional Energy companies in Europe. Table 4.7 displays the results. When considering stock returns, the firms with the top 3 largest spillovers from and to *NSI* are BPLN (5.75% and 0.75%), VBKGY (5.06% and 2.97%), and VKFP (4.57% and 1.98%). On average, the result shows that spillovers from and to *NSI* are 2.71% and 1.65% respectively, leading to a net spillover from *NSI* to stock returns of 1.06% for Traditional Energy firms. In terms of trading volumes, the firms with the top 3 largest spillovers from and to *NSI* are GTTFP (4.84% and 4.94%), LUNESS (3.9% and 1.82%), and VKFP (3.63% and 3.8%). Interestingly, owing to Vallourec SA's (VKFP) well documented sales contraction in 2018 Q3 [11], corresponding 24% fall in share price, and subsequent debt restructuring, VKFP's stock returns and trading volumes are strongly influenced by News sentiment within our sample period. Moreover, ENOGLN has the lowest spillovers for both stock returns ($\widetilde{\theta}_{LogReturn \leftarrow NSI} = 0.5\%$ and $\widetilde{\theta}_{NSI \leftarrow LogReturn} = 0.62\%$) and trading volumes ($\widetilde{\theta}_{LogVolume \leftarrow NSI} = 0.6\%$ and $\widetilde{\theta}_{NSI \leftarrow LogVolume} = 0.4\%$). Overall, for trading volumes of Traditional Energy firms, the evidence shows that spillovers from and to *NSI* are 1.95% and 1.62% respectively, leading to a net spillover from *NSI* to trading volumes of 0.33%.

[11] Vallourec SA Financial Results

Table 4.7. Spillover Effects: Traditional Energy

This table reports spillover effects calculated from Eq. (4.4) between the News sentiment index (*NSI*), returns (*LogReturn*) and trading volumes (*LogVolume*) for 50 Traditional Energy firms. The number of lags for the VAR model is selected using the Bayesian information criterion. $\widetilde{\theta}_{LogReturn \leftarrow NSI}$ refers to the spillover effect from NSI to stock returns and $\widetilde{\theta}_{LogVolume \leftarrow NSI}$ is to the spillover effect from NSI to trading volumes. Conversely, $\widetilde{\theta}_{NSI \leftarrow LogReturn}$ refers to the spillover effect from stock returns to NSI, and $\widetilde{\theta}_{NSI \leftarrow LogVolume}$ is the spillover effect from trading volumes to NSI. The values of spillover effects, ranging from 0% to 100% are presented in Conlumns (1)–(6).

		Returns			**Trading volumes**		
		$\widetilde{\theta}_{LogReturn \leftarrow NSI}$	$\widetilde{\theta}_{NSI \leftarrow LogReturn}$	**Net**	$\widetilde{\theta}_{LogVolume \leftarrow NSI}$	$\widetilde{\theta}_{NSI \leftarrow LogVolume}$	**Net**
Ticker	**Lag**	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
AKASTNO	6	2.19	1.82	0.37	1.25	1.27	-0.02
AKRBPNO	5	2.23	1.64	0.59	1.30	1.12	0.18
AKSONO	2	1.25	0.99	0.26	1.13	1.96	-0.83
BORRNO	2	1.06	1.10	-0.04	2.09	1.28	0.81
BPLN	2	5.75	0.75	5.00	1.95	1.66	0.29
BWLPGNO	4	3.41	2.11	1.30	3.44	3.14	0.30
BWONO	6	3.25	2.64	0.61	3.24	2.52	0.72
CGGFP	4	2.63	1.22	1.41	2.20	1.82	0.38
CNELN	2	1.75	0.80	0.95	1.36	1.14	0.22
DECLN	4	1.15	0.70	0.45	1.48	0.45	1.03
DNONO	3	2.10	1.82	0.28	1.95	1.69	0.26
DRLCODC	6	2.11	1.15	0.96	1.96	2.11	-0.15
ENIIM	3	2.06	1.05	1.01	1.53	1.21	0.32
ENOGLN	4	0.50	0.62	-0.12	0.60	0.40	0.20
ENQSS	3	2.44	1.18	1.26	1.45	0.86	0.59
EQNRNO	6	1.04	0.71	0.33	2.39	1.37	1.02
EURNBB	4	1.54	1.46	0.08	1.46	1.32	0.14
FRONO	3	4.86	2.56	2.30	2.39	1.32	1.07
GALPPL	5	1.03	0.63	0.40	0.80	1.33	-0.53
GTTFP	2	4.59	4.04	0.55	4.84	4.94	-0.10
HAFNINO	6	0.81	1.32	-0.51	1.00	0.87	0.13
HBRLN	2	3.36	1.00	2.36	2.21	1.21	1.00
IPCOSS	5	2.55	0.60	1.95	1.12	0.76	0.36
LTSPW	6	3.65	1.65	2.00	1.87	1.03	0.84
LUNESS	2	2.92	1.54	1.38	3.90	1.82	2.08
MGNNO	5	4.53	4.52	0.01	2.39	2.02	0.37
NESTEFH	3	4.23	3.32	0.91	1.54	0.91	0.63
ODLNO	4	3.66	2.45	1.21	1.85	2.44	-0.59
OMVAV	6	1.03	0.83	0.20	0.90	0.89	0.01
PENNO	4	2.25	1.53	0.72	2.17	2.16	0.01
PGNPW	5	1.44	1.68	-0.24	0.99	0.77	0.22
PGSNO	6	3.56	2.94	0.62	1.58	1.28	0.30
PKNPW	2	1.91	0.73	1.18	1.56	1.41	0.15
REPSQ	5	1.57	1.15	0.42	1.12	0.94	0.18
SBMONA	3	3.06	2.43	0.63	2.36	1.57	0.79

Continued on next page

Table 4.7 - continued from previous page

		Returns			Trading volumes		
		$\widetilde{\theta}_{LogReturn \leftarrow NSI}$	$\widetilde{\theta}_{NSI \leftarrow LogReturn}$	**Net**	$\widetilde{\theta}_{LogVolume \leftarrow NSI}$	$\widetilde{\theta}_{NSI \leftarrow LogVolume}$	**Net**
Ticker	**Lag**	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
SBOAV	3	2.90	2.40	0.50	1.07	1.24	-0.17
SHELLNA	6	3.77	0.58	3.19	2.49	1.35	1.14
SPMIM	2	3.46	2.21	1.25	1.55	1.26	0.29
SRSIM	5	3.17	2.47	0.70	1.79	2.26	-0.47
SUBCNO	5	2.55	1.47	1.08	3.58	1.98	1.60
TEFP	6	1.60	1.12	0.48	0.77	2.21	-1.44
TENIM	3	3.61	1.38	2.23	1.18	0.83	0.35
TGSNO	6	1.00	0.73	0.27	3.42	2.88	0.54
TLWLN	4	4.17	0.60	3.57	2.00	0.67	1.33
TRMDADC	3	3.05	3.84	-0.79	3.02	3.39	-0.37
TTEFP	6	1.30	0.41	0.89	1.85	1.12	0.73
VBKGY	3	5.06	2.97	2.09	2.22	2.28	-0.06
VKFP	6	4.57	1.98	2.59	3.63	3.80	-0.17
VPKNA	5	4.19	2.74	1.45	2.49	1.82	0.67
WGLN	4	3.45	0.97	2.48	1.12	0.75	0.37
Traditional Energy		2.71	1.65	1.06	1.95	1.62	0.33

Taken together, comparing the spillovers in Table 4.6 and Table 4.7, our findings indicate that for both Clean Energy and Traditional Energy firms: 1) News sentiment affects stock returns and trading volumes, 2) spillovers from News sentiment to stock returns and trading volumes are greater than the spillovers in the opposite direction, 3) spillovers from News sentiment to stock returns are relatively greater than to trading volumes, and 4) spillovers are asymmetric and moderate. These findings are consistent with investor sentiment literature (Tetlock 2007; Baker and Wurgler 2007), indicating that News sentiment, as characterised by Bloomberg News articles, asymmetrically impacts stock returns. Notably, our findings are consistent with Reboredo and Ugolini 2018, indicating that News sentiment plays a similar role in price formation of European Clean Energy and U.S. Clean Energy stocks. More importantly, considering the body of literature detailing the sector specific impact of News sentiment on stock price (Uygur and Taş 2014; Khan et al. 2020; Niu et al. 2021), we find that News sentiment has a comparable impact on the price formation and trading of Clean Energy and Traditional Energy stocks.

Table 4.8. Comparative Analysis: Clean v.s. Traditional Energy

This table presents the t-statistics of a comparative analysis: whether the spillover effects between Clean Energy companies and Traditional Energy companies are significantly different. The sample for this test includes 35 Clean Energy companies and 50 Traditional Energy companies. $***$, $**$, and $*$ denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Variables	Clean Energy Mean of 35 Obs.	Traditional Energy Mean of 50 Obs.	Diff.	T-test (p-value)
$\widetilde{\theta}_{LogReturn \leftarrow NSI}$	2.81	2.71	0.10	0.738
$\widetilde{\theta}_{NSI \leftarrow LogReturn}$	2.31	1.65	0.66	0.996
$\widetilde{\theta}_{LogVolume \leftarrow NSI}$	1.86	1.95	-0.09	0.648
$\widetilde{\theta}_{NSI \leftarrow LogVolume}$	1.73	1.62	0.11	0.580

Table B.4 presents the results of t-tests on the spillovers between *NSI*, *LogReturn* and *LogVolume*. Our findings indicate that the t-tests are statistically insignificant when comparing the impact of *NSI* on *LogReturn* and *LogVolume*, suggesting that News Sentiment has a similar impact on the price formation of Clean Energy companies and Traditional Energy companies. Similarly, the impact of *LogReturn* and *LogVolume* on *NSI* is statistically insignificant. Taken in unison, the findings presented in Table B.4 shows that Clean Energy companies and Traditional Energy companies exhibit similar dynamics in their respective relationships with *NSI*. Notably, this finding augments the body of literature outlining the sector specific impact of sentiment on price formation (Uygur and Taş 2014; Khan et al. 2020; Niu et al. 2021), as we are the first to show that Traditional Energy companies and Renewable Energy companies share similar dynamics with respect to News Sentiment.

4.4.4. Contemporaneous Effects

The results presented in the prior section display that News sentiment holds a significant information component, which can impact stock returns and trading volumes. To further examine whether the small magnitude of spillover effects can be explained the Efficient Market Hypothesis (Fama 1960), whereby an efficient market rapidly incorporates all public information, including the impact News sentiment, into the stock prices, we

follow Reboredo and Ugolini 2018 and assume that the News sentiment index (*NSI*) is an exogenous variable. For each firm k, we estimate the following specification:

$$y_{k,t} = \alpha + \lambda y_{k,t-i} + \beta NSI_{k,t} + \Psi x_{k,t} + \varepsilon_{k,t} \tag{4.6}$$

where $y_{k,t}$ is the dependent variable (*LogReturn* or *LogVolume*), $NSI_{k,t}$ is the daily News Sentiment Index, and $x_{k,t}$ are the exogenous variables including *SX5E*, *V2X*, and *TTFGDAHD*.

Table B.5 presents the results of the linear regression of stock returns on *NSI* and control variables for 35 Clean Energy companies. The empirical evidence shows that News sentiment has significant positive contemporaneous effects on stock returns for 21 of 35 Clean Energy firms. Using all 35 Clean Energy firms' data, the coefficient of *NSI* is statistically significant at the 1% level with a positive value of 0.007.

Table 4.10 also suggests that *NSI* has a significant positive contemporaneous effect on stock returns for 25 of 50 traditional energy firms with exceptions of firms BPLN and IPCOSS. The results of the panel regression show that the size of the impact of *NSI* on Traditional Energy firms' stock return are the same as that of Clean Energy firms. These results affirm the positive correlations shown in Table 4.5 and are also consistent with the finding in Table 4.7 that spillovers from *NSI* to stock returns of Clean Energy firms (2.81%) is comparable to that of Traditional Energy firms (2.71%).

Table 4.9. Linear Regression of Stock Returns on NSI: Clean Energy

This table reports the results of linear regression of stock returns on *NSI* for each of the 35 Clean Energy firms over the sample period of 1st January 2015 to 1st April 2022. The dependent variable is $LogReturn_t$, the variable of interest is NSI_t, and the control variables are $LogReturn_{t-1}$, $SX5E_t$, $V2X_t$, and $TTFGDAHD_t$. t-statistics are presented in the parentheses. $***$, $**$, and $*$ denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	$LogReturn_t$					
Ticker	$\boldsymbol{NSI_t}$ **(1)**	$\boldsymbol{LogReturn_{t-1}}$ **(2)**	$\boldsymbol{SX5E_t}$ **(3)**	$\boldsymbol{V2X_t}$ **(4)**	$\boldsymbol{TTFGDAHD_t}$ **(5)**	**Constant (6)**
ABIOFP	0.013	-0.019	-0.045	0.002	0.002	0.000
	[0.91]	[-0.49]	[-0.96]	[0.22]	[0.28]	[0.74]
ANASQ	0.015***	-0.039	0.022	0.010	0.015**	0.000
	[2.65]	[-0.58]	[0.40]	[1.16]	[2.22]	[0.40]
BKWSE	0.028**	-0.064*	0.054	0.005	0.003	0.000
	[2.30]	[-1.75]	[1.40]	[0.84]	[0.51]	[1.47]
DRXLN	0.003	0.017	0.023	0.005	0.005	0.000
	[0.60]	[0.36]	[0.38]	[0.53]	[0.32]	[0.38]
EBKGY	0.003	-0.070	-0.041	-0.010	-0.014	0.001
	[0.54]	[-1.47]	[-0.80]	[-1.07]	[-1.54]	[1.00]
ECVGY	0.018***	-0.078**	0.028	0.003	0.017*	0.000
	[2.83]	[-2.18]	[0.44]	[0.28]	[1.85]	[0.58]
EDFFP	0.014*	0.015	-0.080	-0.009	-0.007	0.000
	[1.79]	[0.44]	[-1.15]	[-0.71]	[-0.71]	[0.21]
EDPPL	0.002	0.017	-0.069	-0.007	-0.009	0.000
	[1.08]	[0.44]	[-1.24]	[-0.79]	[-1.14]	[0.51]
EDPRPL	0.003	0.030	0.005	-0.007	-0.003	0.001
	[1.50]	[0.89]	[0.05]	[-0.62]	[-0.47]	[1.39]
ELESQ	0.007***	0.047	0.027	0.005	0.005	0.000
	[2.97]	[1.34]	[0.65]	[0.75]	[0.99]	[0.13]
ELIBB	0.021***	-0.012	0.025	0.004	0.006	0.001**
	[3.26]	[-0.24]	[0.59]	[0.62]	[0.61]	[2.14]
ENAPW	0.015***	0.036	0.063	0.001	0.003	-0.001
	[3.69]	[1.17]	[0.90]	[0.09]	[0.26]	[-1.45]
ENELIM	0.001	-0.081*	-0.012	0.006	-0.003	0.001
	[0.55]	[-1.70]	[-0.22]	[0.69]	[-0.42]	[1.42]
ENGPW	0.010***	0.005	-0.040	-0.002	-0.008	-0.001
	[2.73]	[0.17]	[-0.71]	[-0.21]	[-0.77]	[-1.45]
ENGSQ	0.004**	-0.011	0.014	0.007	-0.007	-0.000
	[1.99]	[-0.19]	[0.34]	[0.94]	[-1.01]	[-0.72]
ERGIM	0.012	-0.061	0.134***	0.024***	-0.005	0.001**
	[1.03]	[-0.83]	[2.70]	[2.84]	[-0.65]	[2.16]
EVNAV	0.004	-0.047	0.292***	0.018	-0.005	0.001*
	[0.72]	[-1.15]	[2.87]	[1.61]	[-0.58]	[1.75]
FORTUMFH	0.006**	0.042	-0.086	-0.006	0.004	-0.000
	[2.44]	[1.13]	[-1.61]	[-0.59]	[0.51]	[-0.57]
IBESQ	0.001	-0.061	0.005	-0.004	0.004	0.000

Continued on next page

Table 4.9 - continued from previous page

	$LogReturn_t$					
Ticker	NSI_t **(1)**	$LogReturn_{t-1}$ **(2)**	$SX5E_t$ **(3)**	$V2X_t$ **(4)**	$TTFGDAHD_t$ **(5)**	**Constant (6)**
	[0.64]	[-1.42]	[0.14]	[-0.66]	[0.75]	[0.92]
IGIM	-0.005	-0.051	0.110**	0.013*	-0.011	0.000
	[-1.00]	[-0.93]	[2.32]	[1.67]	[-1.48]	[1.10]
NEOENFP	-0.003	0.044	0.030	0.022	0.009	0.001
	[-0.35]	[0.83]	[0.38]	[1.23]	[0.76]	[1.17]
NTGYSQ	0.005**	-0.024	0.009	-0.008	0.027	-0.000
	[2.15]	[-0.46]	[0.14]	[-0.67]	[0.81]	[-0.58]
ORSTEDDC	0.017**	-0.008**	-0.285	-0.005	0.006	-0.003
	[2.42]	[-2.57]	[-0.87]	[-0.34]	[0.31]	[-0.79]
PEPPW	0.023**	0.013	-0.156*	-0.018	0.005	0.000
	[2.53]	[0.22]	[-1.86]	[-1.34]	[0.63]	[0.41]
PGEPW	0.010*	0.046	-0.027	0.004	-0.008	-0.001
	[1.74]	[1.24]	[-0.33]	[0.35]	[-1.03]	[-1.53]
REESQ	0.003**	-0.012	-0.028	-0.007	0.000	-0.000
	[2.04]	[-0.24]	[-0.71]	[-1.02]	[0.02]	[-0.32]
RUIFP	0.016***	-0.025	0.027	0.001	0.003	-0.001
	[3.06]	[-0.58]	[0.48]	[0.06]	[0.50]	[-1.16]
SCATCNO	0.029**	-0.009	0.023	0.014	-0.004	0.000
	[2.25]	[-0.28]	[0.29]	[1.20]	[-0.35]	[0.60]
SRGIM	0.005**	-0.139**	-0.017	-0.000	0.015**	0.000
	[2.11]	[-2.44]	[-0.44]	[-0.03]	[2.29]	[0.46]
SSELN	0.003	-0.051	0.068	-0.001	0.000	0.000
	[1.55]	[-0.86]	[1.33]	[-0.08]	[0.04]	[0.39]
TPEPW	0.018***	0.072*	0.130	0.007	-0.003	-0.001
	[4.37]	[1.96]	[1.50]	[0.54]	[-0.23]	[-0.89]
TRNIM	-0.001	-0.129***	-0.047	-0.017**	0.003	0.001**
	[-0.32]	[-3.16]	[-1.29]	[-2.34]	[0.64]	[2.07]
UN01GY	0.016***	0.153**	-0.037	0.001	-0.015**	0.000
	[3.81]	[2.04]	[-0.37]	[0.09]	[-2.46]	[0.56]
VERAV	0.006	0.008	0.041	0.006	0.002	0.001**
	[1.40]	[0.19]	[0.63]	[0.57]	[0.27]	[1.96]
VLTSAFP	0.020**	0.042	0.053	0.019*	-0.012	0.000
	[2.11]	[0.97]	[0.92]	[1.92]	[-1.17]	[0.03]
Clean Energy	0.007***	-0.004	0.005	0.002	0.001	0.000
	[11.18]	[-0.70]	[0.41]	[1.07]	[0.53]	[0.65]

Table 4.10. Linear Regression of Stock Returns on NSI: Traditional Energy

This table reports the results of linear regression of stock returns on *NSI* for each of the 50 Traditional Energy firms over the sample period of 1st January 2015 to 1st April 2022. The dependent variable is $LogReturn_t$, the variable of interest is NSI_t, and the control variables are $LogReturn_{t-1}$, $SX5E_t$, $V2X_t$, and $TTFGDAHD_t$. t-statistics are presented in the parentheses. $***$, $**$, and $*$ denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	$LogReturn_t$					
Ticker	NSI_t **(1)**	$LogReturn_{t-1}$ **(2)**	$SX5E_t$ **(3)**	$V2X_t$ **(4)**	$TTFGDAHD_t$ **(5)**	**Constant (6)**
AKASTNO	0.011	-0.022	0.692***	0.004	0.032***	-0.000
	[1.21]	[-0.71]	[5.06]	[0.20]	[2.65]	[-0.16]
AKRBPNO	0.005	0.027	0.535***	-0.061***	0.056**	0.001
	[0.85]	[0.84]	[3.85]	[-3.52]	[2.46]	[1.45]
AKSONO	0.008*	0.001	0.721***	-0.034	0.052***	0.000
	[1.88]	[0.02]	[4.83]	[-1.62]	[2.82]	[0.15]
BORRNO	0.012	0.063	1.325***	0.023	0.069**	-0.002
	[0.95]	[1.02]	[2.96]	[0.45]	[1.96]	[-0.97]
BPLN	-0.003*	0.057*	0.861***	-0.008	0.023**	-0.000
	[-1.91]	[1.67]	[6.45]	[-0.51]	[2.23]	[-0.99]
BWLPGNO	0.016**	0.020	0.513***	-0.027	0.025	-0.000
	[2.14]	[0.65]	[4.73]	[-1.62]	[1.64]	[-0.23]
BWONO	0.018*	0.087**	0.742***	-0.056**	0.050**	-0.001
	[1.77]	[2.02]	[4.30]	[-2.40]	[2.32]	[-0.59]
CGGFP	0.011*	0.045	0.910***	-0.050*	0.024	-0.003***
	[1.78]	[1.20]	[5.39]	[-1.86]	[1.50]	[-2.60]
CNELN	0.001	0.065	0.819***	-0.019	0.022*	0.000
	[0.12]	[1.28]	[4.41]	[-0.92]	[1.68]	[0.34]
DECLN	0.008	-0.018	0.488***	-0.001	0.000	-0.000
	[1.19]	[-0.47]	[3.40]	[-0.08]	[0.05]	[-0.60]
DNONO	0.009*	0.025	0.614***	-0.066***	0.068***	-0.000
	[1.91]	[0.63]	[3.69]	[-2.95]	[2.99]	[-0.54]
DRLCODC	-0.004	0.067*	0.606***	-0.049*	0.053**	-0.001
	[-0.32]	[1.77]	[2.82]	[-1.69]	[2.13]	[-0.85]
ENIIM	0.003*	-0.024	1.063***	0.006	0.022**	0.000
	[1.85]	[-0.96]	[11.56]	[0.59]	[2.05]	[0.08]
ENOGLN	-0.009	0.024	1.357***	0.061*	0.061***	0.002*
	[-1.16]	[0.34]	[4.84]	[1.81]	[2.96]	[1.68]
ENQSS	0.005	0.057*	0.937***	-0.034	0.079***	-0.001
	[0.92]	[1.91]	[4.44]	[-1.32]	[2.63]	[-0.76]
EQNRNO	0.001	-0.017	0.498***	-0.027**	0.059***	0.000
	[0.41]	[-0.52]	[5.61]	[-2.25]	[3.47]	[0.54]
EURNBB	0.006*	0.008	0.237**	-0.073***	0.007	-0.001
	[1.73]	[0.30]	[2.33]	[-5.93]	[0.48]	[-1.18]
FRONO	0.018***	-0.045	0.287*	-0.065***	0.026	-0.001
	[3.63]	[-1.16]	[1.86]	[-2.97]	[1.55]	[-1.46]
GALPPL	0.003	0.072**	0.673***	-0.027**	0.020	-0.000

Continued on next page

Table 4.10 - continued from previous page

	$LogReturn_t$					
Ticker	NSI_t **(1)**	$LogReturn_{t-1}$ **(2)**	$SX5E_t$ **(3)**	$V2X_t$ **(4)**	$TTFGDAHD_t$ **(5)**	**Constant (6)**
	[1.40]	[2.56]	[7.62]	[-2.49]	[1.30]	[-0.19]
GTTFP	0.008	0.006	0.508***	-0.024**	0.011	-0.000
	[0.36]	[0.17]	[6.41]	[-2.12]	[1.34]	[-0.32]
HAFNINO	-0.011	-0.033	0.356**	-0.041	0.018	0.000
	[-0.88]	[-0.57]	[2.11]	[-1.42]	[0.96]	[0.31]
HBRLN	-0.003	0.030	1.541***	-0.013	0.066**	-0.001
	[-0.57]	[0.71]	[3.58]	[-0.29]	[2.47]	[-1.18]
IPCOSS	-0.016**	0.019	1.148***	-0.011	0.061**	0.001
	[-2.18]	[0.61]	[4.15]	[-0.40]	[2.16]	[0.69]
LTSPW	0.011***	0.023	0.384***	-0.044***	0.008	-0.000
	[3.35]	[0.67]	[4.07]	[-3.04]	[0.59]	[-0.39]
LUNESS	0.010**	-0.046*	0.717***	-0.024	0.064***	-0.000
	[2.46]	[-1.65]	[5.30]	[-1.61]	[3.18]	[-0.42]
MGNNO	-0.070	-0.074	0.047	-0.064**	0.007	0.000
	[-0.93]	[-1.39]	[0.31]	[-2.54]	[0.46]	[0.35]
NESTEFH	0.019***	-0.016	0.636***	-0.024*	0.010	0.000
	[4.16]	[-0.36]	[7.16]	[-1.80]	[0.88]	[0.73]
ODLNO	0.011	-0.011	0.525***	-0.052***	0.037***	0.001
	[0.96]	[-0.33]	[3.73]	[-2.88]	[2.64]	[0.72]
OMVAV	0.002	0.042	0.982***	-0.006	0.014	0.000
	[0.66]	[0.80]	[7.71]	[-0.38]	[1.61]	[0.80]
PENNO	0.055*	0.013	0.742***	-0.055**	0.061***	0.000
	[1.90]	[0.44]	[4.34]	[-2.13]	[3.55]	[0.26]
PGNPW	0.009**	-0.021	0.228**	-0.045***	0.027*	-0.000
	[2.16]	[-0.67]	[2.27]	[-3.33]	[1.74]	[-0.67]
PGSNO	0.041***	-0.018	0.801***	-0.057**	0.076***	-0.002
	[4.13]	[-0.41]	[4.71]	[-2.15]	[2.86]	[-1.43]
PKNPW	0.008**	-0.003	0.378***	-0.044***	0.007	-0.001
	[2.15]	[-0.09]	[4.69]	[-3.45]	[0.53]	[-1.55]
REPSQ	0.006**	0.041	0.974***	-0.009	0.032***	-0.000
	[2.36]	[1.15]	[10.77]	[-0.84]	[2.86]	[-0.91]
SBMONA	0.008*	-0.014	0.693***	-0.022**	0.017***	-0.000
	[1.84]	[-0.56]	[8.27]	[-2.12]	[2.68]	[-0.23]
SBOAV	0.024***	0.082**	0.545***	-0.055***	0.028**	-0.001
	[3.02]	[2.33]	[3.59]	[-3.01]	[2.46]	[-0.98]
SHELLNA	-0.000	0.034	0.986***	-0.001	0.032***	-0.000
	[-0.14]	[1.01]	[8.15]	[-0.09]	[2.82]	[-1.04]
SPMIM	0.012***	-0.002	0.763***	-0.040***	0.035***	-0.002***
	[3.39]	[-0.06]	[7.79]	[-3.08]	[2.74]	[-3.52]
SRSIM	0.014	0.072***	0.586***	-0.026*	0.006	-0.001
	[1.44]	[2.60]	[6.76]	[-1.95]	[0.55]	[-1.51]
SUBCNO	0.005*	-0.059**	0.691***	-0.037**	0.035***	-0.001
	[1.71]	[-1.99]	[4.87]	[-2.17]	[2.62]	[-0.81]
TEFP	0.029**	0.067	0.692*	-0.008	-0.035	-0.001
	[2.04]	[0.66]	[1.92]	[-0.19]	[-1.14]	[-0.41]

Continued on next page

Table 4.10 - continued from previous page

	$LogReturn_t$					
Ticker	NSI_t **(1)**	$LogReturn_{t-1}$ **(2)**	$SX5E_t$ **(3)**	$V2X_t$ **(4)**	$TTFGDAHD_t$ **(5)**	**Constant (6)**
TENIM	0.005***	-0.016	0.736***	-0.047***	0.038***	-0.001
	[2.82]	[-0.68]	[7.07]	[-3.64]	[2.58]	[-1.05]
TGSNO	0.010**	-0.046	0.474***	-0.045***	0.042***	-0.001
	[2.17]	[-1.53]	[4.03]	[-2.83]	[3.15]	[-1.45]
TLWLN	0.002	0.007	1.180***	-0.036	0.075***	-0.000
	[0.38]	[0.14]	[3.86]	[-1.00]	[2.67]	[-0.20]
TRMDADC	0.007	-0.003	0.159	-0.066***	0.015	-0.001
	[0.59]	[-0.09]	[1.35]	[-4.08]	[1.21]	[-1.06]
TTEFP	-0.001	0.041	1.045***	0.009	0.019**	0.000
	[-0.60]	[1.39]	[13.04]	[0.92]	[2.17]	[0.23]
VBKGY	0.042**	-0.022	0.636***	-0.066***	0.011	0.001
	[2.16]	[-0.66]	[4.92]	[-3.63]	[0.71]	[1.06]
VKFP	0.007	0.070**	0.765***	-0.066***	0.018	-0.002**
	[0.84]	[2.17]	[5.08]	[-2.84]	[0.77]	[-2.22]
VPKNA	0.010*	0.023	0.329***	-0.036***	-0.009	-0.000
	[1.88]	[0.64]	[4.14]	[-3.08]	[-1.35]	[-0.77]
WGLN	0.000	0.054	0.914***	-0.006	0.027**	-0.001
	[0.12]	[1.46]	[5.93]	[-0.33]	[2.17]	[-1.63]
Traditional Energy	0.007***	0.018**	0.698***	-0.033***	0.032***	-0.000***
	[9.03]	[2.29]	[28.02]	[-10.66]	[13.82]	[-3.97]

Table 4.11 and Table 4.12 show the results of regression of trading volumes on *NSI* for 35 Clean Energy companies and 50 Traditional Energy companies, respectively. The result displays a positive effect of *NSI* on trading volumes for 11 Clean Energy firms, and for 13 Traditional Energy firms. Notably, there exists a negative association between *NSI* and trading volumes among firms BORRMO, ENIIM, EQNRNO, and SUBCNO. Furthermore, the coefficients of the panel regression for both Clean Energy firms (0.01) and Traditional Energy firms (0.003) suggest a significant positive contemporaneous impact of *NSI* on trading volumes. In contrast, the evidence indicates the impact of *NSI* on trading volumes is stronger for Clean Energy firms, whereas Section 4.4.3 shows that the size of the spillover effect of *NSI* to trading volumes is similar for Clean Energy firms (1.86%) and Traditional Energy firms (1.95%). Nevertheless, the positive contemporaneous effects are consistent with the positive correlations.

Table 4.11. Linear Regression of Trading Volumes on NSI: Clean Energy

This table reports the results of linear regression of trading volumes on *NSI* for each of the 35 Clean Energy firms over the sample period of 1st January 2015 to 1st April 2022. The dependent variable is $LogVolume_t$, the variable of interest is NSI_t, and the control variables are $LogVolume_{t-1}$, $SX5E_t$, $V2X_t$, and $TTFGDAHD_t$. t-statistics are presented in the parentheses. $***$, $**$, and $*$ denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	$\boldsymbol{LogVolume_t}$					
Ticker	$\boldsymbol{NSI_t}$ **(1)**	$\boldsymbol{LogVolume_{t-1}}$ **(2)**	$\boldsymbol{SX5E_t}$ **(3)**	$\boldsymbol{V2X_t}$ **(4)**	$\boldsymbol{TTFGDAHD_t}$ **(5)**	**Constant (6)**
ABIOFP	0.003	0.634***	-0.035	0.036**	0.011	0.938***
	[0.86]	[27.46]	[-0.40]	[2.20]	[0.86]	[15.85]
ANASQ	0.005	0.520***	-0.057	-0.048**	0.007	1.234***
	[0.56]	[14.90]	[-0.61]	[-2.50]	[0.41]	[13.84]
BKWSE	0.096***	0.688***	0.018	0.052	-0.035	0.784***
	[3.38]	[29.75]	[0.09]	[1.40]	[-1.55]	[13.34]
DRXLN	0.032	0.385***	0.029	0.002	0.099	1.296***
	[0.78]	[4.77]	[0.06]	[0.02]	[1.52]	[7.49]
EBKGY	-0.001	0.514***	-0.083	-0.007	-0.012	1.306***
	[-0.10]	[20.81]	[-0.85]	[-0.42]	[-0.94]	[19.74]
ECVGY	0.003	0.511***	-0.030	0.012	0.008	1.334***
	[0.61]	[20.14]	[-0.35]	[0.78]	[0.68]	[19.33]
EDFFP	0.015***	0.708***	0.101	0.035**	0.011	0.791***
	[2.92]	[35.31]	[1.11]	[2.15]	[0.81]	[14.63]
EDPPL	0.005	0.493***	0.060	0.036**	0.030**	1.331***
	[0.83]	[19.99]	[0.65]	[2.18]	[1.97]	[20.54]
EDPRPL	0.000	0.475***	0.225**	0.022	0.045***	1.404***
	[0.10]	[18.36]	[2.16]	[1.22]	[2.94]	[20.41]
ELESQ	0.001	0.535***	-0.062	0.029*	-0.023	1.289***
	[0.31]	[22.61]	[-0.61]	[1.90]	[-1.40]	[19.72]
ELIBB	0.020***	0.616***	0.087	0.013	0.017	0.957***
	[2.72]	[24.57]	[0.90]	[0.75]	[1.10]	[15.32]
ENAPW	0.009*	0.756***	0.116	0.056***	0.009	0.641***
	[1.71]	[38.31]	[1.39]	[3.68]	[0.80]	[12.38]
ENELIM	0.015	0.609***	0.124	0.086***	-0.010	0.918***
	[1.41]	[27.08]	[0.84]	[3.54]	[-0.43]	[17.39]
ENGPW	0.028*	0.756***	-0.018	0.024	-0.035	0.618***
	[1.73]	[32.94]	[-0.13]	[1.01]	[-1.61]	[10.62]
ENGSQ	-0.006	0.702***	-0.042	0.048***	0.017	0.830***
	[-1.41]	[33.69]	[-0.52]	[3.75]	[1.45]	[14.33]
ERGIM	-0.005	0.682***	0.169*	0.053***	-0.007	0.881***
	[-1.39]	[34.06]	[1.95]	[3.84]	[-0.62]	[15.85]
EVNAV	0.004	0.639***	0.040	0.008	0.005	1.006***
	[1.55]	[28.07]	[0.56]	[0.64]	[0.43]	[15.94]
FORTUMFH	0.002	0.672***	0.195**	0.056***	0.011	0.898***
	[0.18]	[35.86]	[2.18]	[3.23]	[0.81]	[17.43]
IBESQ	0.001	0.573***	-0.017	0.004	-0.009	1.152***

Continued on next page

Table 4.11 - continued from previous page

	$LogVolume_t$					
Ticker	NSI_t **(1)**	$LogVolume_{t-1}$ **(2)**	$SX5E_t$ **(3)**	$V2X_t$ **(4)**	$TTFGDAHD_t$ **(5)**	**Constant (6)**
	[0.19]	[22.86]	[-0.21]	[0.29]	[-0.78]	[17.02]
IGIM	0.007**	0.687***	-0.162**	0.002	0.017	0.830***
	[2.56]	[32.23]	[-2.05]	[0.16]	[1.21]	[14.71]
NEOENFP	0.026**	0.662***	0.473	0.143***	0.051	0.875***
	[2.09]	[12.19]	[1.40]	[3.03]	[1.28]	[6.22]
NTGYSQ	0.001	0.573***	0.058	0.042***	0.001	1.162***
	[0.56]	[26.16]	[0.82]	[3.60]	[0.09]	[19.50]
ORSTEDDC	-0.008	0.599***	0.052	0.030*	0.002	1.027***
	[-1.25]	[23.45]	[0.50]	[1.75]	[0.13]	[15.78]
PEPPW	-0.003	0.724***	0.093	0.034**	0.013	0.767***
	[-1.19]	[33.75]	[0.94]	[2.21]	[0.96]	[12.88]
PGEPW	0.005	0.592***	0.011	0.012	-0.002	1.110***
	[1.05]	[24.47]	[0.10]	[0.70]	[-0.12]	[16.88]
REESQ	0.029*	0.744***	-0.146	-0.012	0.006	0.620***
	[1.74]	[36.39]	[-0.84]	[-0.39]	[0.25]	[12.39]
RUIFP	-0.001	0.619***	-0.010	0.003	0.012	1.049***
	[-0.40]	[29.05]	[-0.16]	[0.29]	[1.49]	[17.94]
SCATCNO	0.001	0.636***	0.010	0.057***	0.013	1.003***
	[0.21]	[24.80]	[0.10]	[4.28]	[1.05]	[14.21]
SRGIM	0.004	0.786***	0.128	0.017	0.005	0.554***
	[0.72]	[32.24]	[1.41]	[1.00]	[0.40]	[8.84]
SSELN	0.013	0.691***	0.267	0.085***	0.006	0.749***
	[0.57]	[25.42]	[1.59]	[3.32]	[0.28]	[11.29]
TPEPW	0.006	0.588***	0.019	0.012	-0.040	1.023***
	[0.59]	[24.40]	[0.20]	[0.73]	[-1.44]	[17.14]
TRNIM	0.013**	0.761***	0.007	0.051***	-0.002	0.613***
	[1.98]	[41.13]	[0.07]	[2.68]	[-0.11]	[12.91]
UN01GY	0.127***	0.731***	0.164	0.030	0.032	0.593***
	[4.75]	[28.32]	[0.41]	[0.51]	[0.53]	[10.20]
VERAV	0.022***	0.590***	0.014	0.039**	0.015	1.047***
	[2.80]	[20.84]	[0.16]	[2.46]	[1.21]	[14.55]
VLTSAFP	-0.003	0.744***	0.080	0.045***	0.007	0.683***
	[-0.67]	[29.82]	[0.93]	[3.01]	[0.52]	[10.28]
Clean Energy	0.010***	0.948***	0.051	0.009	0.001	0.134***
	[5.94]	[124.33]	[1.04]	[1.16]	[0.09]	[6.82]

Table 4.12. Linear Regression of Trading Volumes on NSI: Traditional Energy

This table reports the results of linear regression of trading volumes on *NSI* for each of the 50 Traditional Energy firms over the sample period of 1st January 2015 to 1st April 2022. The dependent variable is $LogVolume_t$, the variable of interest is NSI_t, and the control variables are $LogVolume_{t-1}$, $SX5E_t$, $V2X_t$, and $TTFGDAHD_t$. t-statistics are presented in the parentheses. $***$, $**$, and $*$ denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	$\boldsymbol{LogVolume_t}$					
Ticker	$\boldsymbol{NSI_t}$ **(1)**	$\boldsymbol{LogVolume_{t-1}}$ **(2)**	$\boldsymbol{SX5E_t}$ **(3)**	$\boldsymbol{V2X_t}$ **(4)**	$\boldsymbol{TTFGDAHD_t}$ **(5)**	**Constant (6)**
AKASTNO	0.043	0.617***	0.042	-0.011	0.027	0.895***
	[1.21]	[26.63]	[0.23]	[-0.33]	[1.03]	[16.43]
AKRBPNO	0.029	0.368***	-0.133	-0.032	0.010	1.558***
	[1.64]	[2.77]	[-0.41]	[-0.50]	[0.26]	[4.74]
AKSONO	0.003	0.714***	0.023	0.033**	0.009	0.742***
	[0.61]	[29.81]	[0.22]	[1.98]	[0.64]	[11.94]
BORRNO	-0.015***	0.766***	0.006	0.005	0.013	0.621***
	[-2.63]	[46.73]	[0.06]	[0.24]	[1.07]	[14.35]
BPLN	0.036***	0.643***	-0.143	-0.028	-0.009	0.881***
	[4.97]	[28.37]	[-1.16]	[-1.41]	[-0.67]	[15.77]
BWLPGNO	0.062***	0.707***	-0.108	-0.012	0.044*	0.685***
	[3.04]	[29.17]	[-0.67]	[-0.42]	[1.86]	[12.01]
BWONO	0.007	0.783***	0.924**	0.205***	0.038	0.535***
	[0.49]	[16.26]	[2.51]	[2.76]	[1.11]	[4.39]
CGGFP	-0.002	0.633***	-0.062	0.032***	0.016	1.049***
	[-0.76]	[27.44]	[-0.87]	[2.82]	[1.57]	[15.94]
CNELN	0.022***	0.677***	0.040	0.029	0.015	0.836***
	[2.62]	[33.56]	[0.36]	[1.63]	[1.17]	[16.00]
DECLN	0.015*	0.834***	0.220*	0.060***	0.018	0.427***
	[1.92]	[50.79]	[1.75]	[2.68]	[1.24]	[10.07]
DNONO	-0.004	0.943***	0.032	0.033	0.036**	0.153***
	[-0.73]	[97.91]	[0.24]	[1.41]	[2.57]	[5.84]
DRLCODC	-0.006	0.523***	0.196*	0.039**	0.008	1.258***
	[-1.07]	[20.03]	[1.72]	[2.23]	[0.44]	[18.26]
ENIIM	-0.031*	0.639***	-0.027	0.043	0.010	0.937***
	[-1.92]	[19.03]	[-0.10]	[0.74]	[0.40]	[10.62]
ENOGLN	-0.002	0.683***	0.008	0.029*	0.010	0.870***
	[-0.56]	[34.52]	[0.09]	[1.84]	[0.93]	[15.99]
ENQSS	0.013	0.716***	0.300**	0.096***	0.001	0.695***
	[1.10]	[24.54]	[2.14]	[3.14]	[0.05]	[9.70]
EQNRNO	-0.011**	0.659***	-0.102	-0.008	0.004	0.899***
	[-2.34]	[27.98]	[-1.05]	[-0.51]	[0.31]	[14.48]
EURNBB	0.134**	0.177***	1.334	0.241	-0.122	1.374***
	[2.03]	[5.58]	[1.05]	[1.28]	[-0.98]	[24.59]
FRONO	0.008	0.791***	0.223*	0.044*	0.021	0.520***
	[0.74]	[44.31]	[1.77]	[1.85]	[1.23]	[11.63]
GALPPL	0.004	0.588***	-0.087	-0.007	0.003	1.113***

Continued on next page

Table 4.12 - continued from previous page

	$LogVolume_t$					
Ticker	NSI_t **(1)**	$LogVolume_{t-1}$ **(2)**	$SX5E_t$ **(3)**	$V2X_t$ **(4)**	$TTFGDAHD_t$ **(5)**	**Constant (6)**
	[0.43]	[24.50]	[-1.15]	[-0.47]	[0.21]	[17.24]
GTTFP	-0.001	0.615***	0.104	-0.001	-0.007	1.063***
	[-0.61]	[23.59]	[1.51]	[-0.07]	[-0.79]	[14.83]
HAFNINO	-0.003	0.781***	-0.007	0.002	-0.004	0.558***
	[-0.57]	[42.02]	[-0.06]	[0.08]	[-0.24]	[11.79]
HBRLN	0.010*	0.527***	-0.191	-0.023	-0.002	1.260***
	[1.77]	[22.26]	[-1.32]	[-1.09]	[-0.16]	[20.08]
IPCOSS	0.004	0.667***	0.175	0.001	-0.015	0.786***
	[0.18]	[26.40]	[1.47]	[0.07]	[-0.76]	[13.21]
LTSPW	0.019***	0.527***	0.029	0.022	-0.016	1.218***
	[2.83]	[21.08]	[0.24]	[1.13]	[-0.81]	[18.94]
LUNESS	0.008***	0.672***	-0.001	0.009	0.008	0.936***
	[2.58]	[33.59]	[-0.01]	[1.01]	[1.06]	[16.45]
MGNNO	0.009	0.909***	-0.036	0.002	0.010	0.233***
	[1.31]	[49.27]	[-0.20]	[0.07]	[0.40]	[4.90]
NESTEFH	0.003	0.513***	0.088	-0.002	0.012	1.286***
	[0.66]	[19.05]	[0.77]	[-0.09]	[0.73]	[18.18]
ODLNO	0.000	0.694***	0.053	0.059***	0.004	0.859***
	[0.06]	[38.54]	[0.65]	[4.97]	[0.47]	[17.00]
OMVAV	0.004	0.562***	0.279	0.079	0.077*	1.071***
	[0.13]	[13.45]	[0.86]	[1.47]	[1.94]	[10.40]
PENNO	-0.000	0.764***	0.039	0.053***	0.041**	0.639***
	[-0.09]	[41.43]	[0.32]	[2.71]	[2.45]	[12.69]
PGNPW	0.002	0.679***	0.018	0.042***	-0.002	0.874***
	[0.89]	[31.61]	[0.26]	[2.79]	[-0.27]	[15.01]
PGSNO	0.045**	0.547***	-0.054	-0.020	0.014	1.138***
	[2.34]	[25.54]	[-0.54]	[-1.11]	[0.99]	[21.18]
PKNPW	0.002	0.712***	-0.008	0.035**	0.023	0.743***
	[0.40]	[38.10]	[-0.07]	[1.98]	[1.44]	[15.39]
REPSQ	0.005	0.584***	-0.170	0.014	-0.010	0.980***
	[0.32]	[26.58]	[-1.09]	[0.53]	[-0.40]	[18.91]
SBMONA	0.002	0.667***	-0.200***	-0.032***	0.011	0.890***
	[0.73]	[32.24]	[-2.80]	[-2.67]	[1.08]	[16.11]
SBOAV	-0.006	0.764***	0.005	0.024	-0.006	0.611***
	[-1.25]	[42.13]	[0.04]	[1.12]	[-0.43]	[12.99]
SHELLNA	-0.002	0.631***	0.009	0.036***	0.011	0.983***
	[-0.69]	[24.15]	[0.12]	[2.90]	[0.90]	[14.12]
SPMIM	0.111***	0.535***	0.014	0.031	0.020	1.112***
	[4.81]	[22.77]	[0.09]	[1.32]	[0.79]	[19.80]
SRSIM	0.029	0.518***	0.065	0.100**	-0.011	1.194***
	[1.43]	[10.35]	[0.26]	[2.02]	[-0.27]	[9.58]
SUBCNO	-0.018***	0.808***	0.084	0.045**	-0.012	0.495***
	[-3.42]	[46.86]	[0.67]	[2.12]	[-0.60]	[11.21]
TEFP	0.013***	0.529***	0.016	-0.005	-0.011	1.328***
	[3.84]	[21.89]	[0.18]	[-0.36]	[-0.90]	[19.58]

Continued on next page

Table 4.12 - continued from previous page

	$Log Volume_t$					
Ticker	NSI_t **(1)**	$Log Volume_{t-1}$ **(2)**	$SX5E_t$ **(3)**	$V2X_t$ **(4)**	$TTFGDAHD_t$ **(5)**	**Constant (6)**
TENIM	0.010	0.650***	-0.047	-0.002	0.000	0.938***
	[1.65]	[24.01]	[-0.51]	[-0.15]	[0.04]	[12.97]
TGSNO	0.007	0.725***	0.129	0.080***	0.022	0.704***
	[0.93]	[31.44]	[0.84]	[2.94]	[1.34]	[11.84]
TLWLN	0.004	0.578***	0.149	0.043**	0.045***	1.073***
	[0.80]	[25.93]	[1.30]	[2.16]	[2.84]	[18.97]
TRMDADC	0.008	0.686***	-0.031	0.035**	0.012	0.822***
	[1.61]	[28.00]	[-0.35]	[2.51]	[0.91]	[12.81]
TTEFP	0.150*	0.470***	0.655	0.107	0.141	1.195***
	[1.88]	[12.12]	[0.96]	[0.77]	[1.49]	[13.35]
VBKGY	0.006	0.813***	-0.375	0.002	-0.033	0.450***
	[0.38]	[27.07]	[-1.30]	[0.05]	[-1.30]	[6.12]
VKFP	0.002	0.711***	0.093	0.043***	0.018*	0.768***
	[0.44]	[35.87]	[1.07]	[3.03]	[1.69]	[14.60]
VPKNA	0.011**	0.592***	0.002	0.015	0.002	1.082***
	[2.27]	[22.20]	[0.02]	[0.78]	[0.11]	[15.37]
WGLN	0.030	0.819***	0.120	0.044	0.010	0.456***
	[1.54]	[31.87]	[0.60]	[1.62]	[0.41]	[6.99]
Traditional Energy	0.003***	0.910***	0.075***	0.047***	0.015***	0.236***
	[3.63]	[114.08]	[3.06]	[9.58]	[4.57]	[11.25]

4.5. Conclusion

This study aims to test and compare the impact of News sentiment on pricing and trading for Clean Energy companies and Traditional Energy companies, based on a sample of 35 European Clean Energy firms and 50 European Traditional Energy firms. Using daily news extracted from Bloomberg, we examine the dynamic spillover effects between News sentiment, stock returns and trading volumes. Specifically, we employ the dynamic connectedness framework developed by Diebold and Yılmaz 2014, and estimate a VAR model with exogenous market factors for each firm. We find that Clean Energy firms and Traditional Energy firms possess similar patterns; that News sentiment positively affects both stock returns and trading volumes, conversely, stock returns and trading volumes have a limited impact on News sentiment. Nevertheless, the spillovers from News sentiment to stock return (2.81%) and trading volumes (1.86%) are relatively moderate, and the spillovers in the opposite directions are smaller. Moreover, our regression results

further affirm the findings of our correlation analysis; that News sentiment is positively related to stock returns and trading volumes, indicating that News sentiment conveys a certain level of information in forecasting prices and trading behaviour.

Overall, our study provides additional empirical evidence that investor sentiment plays a role in asset pricing (Baker and Wurgler 2007; Kaplanski et al. 2015; Siganos et al. 2017). By disentangling the relationship between News sentiment and European Energy stocks, this study addresses three gaps in the literature. Firstly, through extraction of sentiment embedded in Bloomberg News articles, this study provides an alternative sentiment measure, which is more appropriate for professional or institutional investors. Secondly, through use of a sample of European companies, our study is the first to address the relationship between News Sentiment, trading volumes and returns in Europe. Finally, whilst prior studies Nofer and Hinz 2015; Reboredo and Ugolini 2018 consider the impact of News sentiment on the trading volumes and returns of Clean Energy companies, little context is provided for these findings. Our study provides additional context for the findings, through analysis of the impact of News sentiment on Clean Energy and Traditional Energy companies in Europe.

Our evidence corroborates Nofer and Hinz 2015; Reboredo and Ugolini 2018's finding that the impact of News sentiment on pricing and trading is limited for the Clean Energy sector. Given the different sources used by Nofer and Hinz 2015; Reboredo and Ugolini 2018 to isolate investor sentiment (Twitter vs Bloomberg News articles), our findings lend weight to Mao et al. 2011's conclusion that, if appropriately broad, different information sources should yield a comparable impact of sentiment on stock returns.

Taken in unison with the earlier work of Reboredo and Ugolini 2018, our findings provide evidence that News sentiment plays a comparable role in the price formation and trading of European and U.S. Clean Energy stocks. The comparable impact of News sentiment on returns between European and U.S. Clean Energy stocks holds important implications for portfolio construction of Clean Energy portfolios. Owing to the similar impact of News sentiment on returns, professional investors may incorporate European securities into their portfolios, with limited additional sentiment risk.

Most importantly, our findings indicate that there is no discernible difference in the impact of News sentiment on Clean Energy and Traditional Energy stocks in Europe. Although News sentiment has a marginally larger impact on the returns of Clean Energy stocks (2.81%) than Traditional Energy stocks (2.71%), no significant difference exists between the impact of News sentiment on Clean Energy and Traditional Energy stocks. As such, this paper provides valuable evidence to the ongoing debate concerning the impact of News sentiment on stock price formation within different Industry sectors (Uygur and Taş 2014; Khan et al. 2020; Niu et al. 2021).

Chapter 5

Conclusions

This thesis seeks to contribute to the literature on European Natural Gas market integration and more broadly, Energy market efficiency. Owing to the current European energy crisis, the reliability and timeliness of market monitoring mechanisms is critical in maintaining normal market function, and enforcing competition. As such, this thesis contributes to the literature by providing short-term market monitoring mechanisms; monitoring the degree to which price integration is being practically achieved, and prediction of congestion within the physical network. The first framework for dynamic assessment of the level of market integration is capable of isolating both physical and non-physical barriers to market integration. However, the second framework enables prediction of the underlying reasons of low market integration, such as capacity constraints. In addition this thesis expands the empirical evidence on the impact of News Sentiment on energy securities. Chapter 2– 4 in this thesis, present the three individual papers in which we aim to answer the following important yet understudied research questions:

1. To what extent is the European Natural Gas market integrated? Which national markets play the most (least) important roles in price formation?

2. Do physical barriers to European Natural Gas market integration exist? Can short term physical barriers to European Natural Gas market integration be reliably predicted?

3. To what extent does News Sentiment influence the pricing of “Clean” and “Traditional” Energy securities?

In Chapter 2, we examine the degree to which a single European market for Natural Gas has been achieved. Through the construction of dynamic networks, where the nodes correspond to the daily gas prices at 12 EU trading hubs, we find a number of spikes in network density, suggesting short periods of improved connectivity of European gas markets, which appear to be driven by exogenous factors. Through application of a Markov regime-switching model to the network density term, we find that abnormally positive or negative changes in network density are equally distributed in magnitude and volume. The validity of the technique is established through identification of historical events, such as unseasonal weather patterns, seismic activity and capacity reductions through pronounced changes in 'in-strengths', 'out-strengths' and 'network density' terms.

The findings of this paper highlight the time varying nature of European gas market dynamics, thus the importance of continual monitoring of market evolution. Moreover, the conclusions drawn in relation to physical and non-physical barriers to market integration are pertinent when considering Gas Infrastructure Europe's Ten-Year Development Plans, as our findings of physical-barriers to market integration are consistent with the locations of planned infrastructure projects. Additionally, our finding of non-physical disconnections infers that additional alignment of network codes is required for some markets. As our findings are consistent with Gas Infrastructure Europe, we propose that the model's ability to isolate physical and non-physical barriers to market integration provides valuable policy advice for European regulatory bodies.

In Chapter 3, we provide a framework for the identification and prediction of physical barriers to European Natural Gas market integration. Owing to the time-varying nature of demand, physical constraints and complicated spatial interdependencies inherent within the underlying physical infrastructure, forecasting Natural Gas pipeline flows is a challenging application of spatiotemporal forecasting. To address this, the infrastructure network is learnt as a graph, and a deep-learning framework, based on the topology of the infrastructure network, is applied to learn the interactions between different pipelines. The empirical results show that the model outperforms baseline methods in predicting gas pipeline flows, and provides a granularity unseen in extant literature.

Taken in unison with Chapter 2, the resulting outcomes provide deep insight into the reason behind short periods of high (low) market integration. The framework presented in Chapter 3 is able to predict periods of high infrastructure utilisation, thus, periods of limited market inter-connectivity, owing to capacity constraints, or physical barriers. If these physical disconnections are consistently predicted, the framework presented in Chapter 2 can be used to evaluate their impact on market integration, which provides valuable policy advice for European regulators. As such, areas of the network which exhibit a limited degree of market integration, owing to capacity constraints, can be isolated, with the direct impact on market integration measured. This information can provide valuable policy advice when considering the location of future gas infrastructure projects.

In Chapter 4, we study the relationship between News Sentiment, price formation and trading volumes of Clean Energy and Traditional Energy companies. Within a sample from the STOXX Europe Total Market Index, we find that European Clean Energy firms and Traditional Energy firms share the same patterns; that News sentiment positively affects both stock returns and trading volumes, conversely, stock returns and trading volumes have a limited impact on News sentiment. Nevertheless, the spillovers are relatively moderate and asymmetric. Our study closes the research gap by providing empirical evidence to two strands of literature; Energy market literature and News sentiment literature, and is the first to draw a comparative analysis on the impact of News Sentiment on Clean Energy and Traditional Energy stocks.

Our research provides valuable advice for investors when considering the role of news sentiment on stock prices. Although some investors may assume that green energy companies hold a different relationship with news sentiment than traditional energy companies, we find that the relationship with news sentiment is similar for green and traditional energy companies. This study contributes to the corporate finance literature, as the comparative relationships between green and traditional energy companies, and news sentiment, remains unstudied. We close this research gap, and provide additional information for investors engaged in portfolio construction.

For future work, there are many adaptations, methods and tests which can be conducted. Firstly, due to data availability, Chapter 2 – 3 focus on the European Natural Gas market. It could be interesting to apply the methods detailed in Chapter 2 – 3 to measure the relative levels of regional integration of European, North American and Asian Natural Gas market integration. Secondly, given LNG's increasing prominence as a swing supplier, it could be interesting to integrate vessel data into the model presented in Chapter 3. Finally, to determine if the impact of News Sentiment on pricing and trading volumes is homogeneous globally, it could be interesting to extend the study conducted in Chapter 4 to include European, North American and Asian domiciled Clean Energy and Traditional Energy companies.

Appendix A

Appendices to Chapter 3

Appendix A

Table A.1. Performance Metrics

A list of Mean Forecast Error (MFE), Mean Bias Error (MBE), Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) values relating to each pipeline within the dataset.

Node Name	MFE	MBE	RMSE	MAE
Arnoldstein/Tarvisio	-0.14%	0.14%	1.84%	1.28%
Murfeld/Cersak	0.25%	-0.25%	1.49%	1.14%
Gorizia/Sempeter	0.39%	-0.39%	3.52%	1.15%
Rogatec	11.58%	-11.58%	13.40%	11.58%
Baumgarten	-0.03%	0.03%	1.57%	1.15%
Bizzarone	1.58%	-1.58%	1.91%	1.66%
Dravaszerdahely	0.29%	-0.29%	1.71%	1.37%
Mosonmagyarovar	1.30%	-1.30%	2.12%	1.75%
Balassagyarmat/Velke Zlievce	13.44%	-13.44%	16.10%	13.47%
Lanzhot	-11.40%	11.40%	14.73%	11.78%
Oberkappel	-3.64%	3.64%	4.89%	4.04%
Passo Gries/Griespass	-0.66%	0.66%	2.10%	1.25%
Beregdaroc	2.43%	-2.43%	2.92%	2.50%
Cesky Tesin/Cieszyn	-0.67%	0.67%	3.45%	2.40%
Waidhaus	-6.27%	6.27%	8.15%	6.74%
Hora Svate Kateriny/Deutschneudorf	1.80%	-1.80%	6.83%	1.81%
Hora Svate Kateriny/Olbernhau	-2.55%	2.55%	4.37%	3.17%
Brandov Stegal/Stegal	2.49%	-2.49%	4.74%	2.49%

Continued on next page

Table A.1 - continued from previous page

Node Name	MFE	MBE	RMSE	MAE
Uberackern ABG/Uberackern 1	2.14%	-2.14%	5.54%	3.75%
Uberackern Sudal/ Uberackern 2	0.65%	-0.65%	2.01%	1.20%
Jura	4.81%	-4.81%	5.66%	4.96%
Oltingue/Rodersdorf	-0.53%	0.53%	3.48%	1.84%
Wallbach	2.96%	-2.96%	3.83%	3.11%
RC Basel	1.04%	-1.04%	2.52%	1.40%
Gela	1.70%	-1.70%	2.19%	1.82%
Maraza del Vallo	-1.58%	1.58%	2.19%	1.79%
Drozdowicze	-2.92%	2.92%	4.13%	3.53%
Wysokoje	-3.33%	3.33%	3.98%	3.40%
Lampertheim IV	4.06%	-4.06%	4.87%	4.11%
Keinbaum	-3.94%	3.94%	4.70%	4.05%
GCP Gaz System/Ontras	-1.98%	1.98%	3.16%	2.51%
RC Lindau	1.81%	-1.81%	2.30%	1.91%
Pirineos	-0.46%	0.46%	1.49%	1.08%
Obergailbach/Medelsheim	-5.32%	5.32%	5.71%	5.32%
Gernsheim	-2.73%	2.73%	12.58%	6.09%
Uzhgorod	-0.55%	0.55%	2.75%	1.06%
Mallnow	0.27%	-0.27%	2.20%	1.61%
Steinitz	-0.74%	0.74%	1.44%	0.98%
Griefswald	1.95%	-1.95%	2.50%	2.26%
Iberico	-1.65%	1.65%	2.97%	2.17%
Almeria	0.55%	-0.55%	2.07%	1.51%
Tarifa	-1.22%	1.22%	1.86%	1.45%
Alveringhem	-0.56%	0.56%	2.06%	1.19%
Eynatten 1/Lichtenbusch	-0.80%	0.80%	2.09%	1.20%
Eynatten 2/Raeren	0.65%	-0.65%	2.64%	1.72%
Virtualys	-0.18%	0.18%	2.02%	1.40%
Broichweiden Sud	-0.27%	0.27%	1.98%	1.17%
Kondratki	-0.01%	0.01%	1.47%	1.00%
Tieterowka	0.31%	-0.31%	1.12%	0.86%
Steinbrink	1.85%	-1.85%	4.13%	2.12%
Ahlten	0.13%	-0.13%	1.97%	1.04%
Zeebrugge IZT	-0.29%	0.29%	1.60%	1.10%
Zelzate Pijpleiding	2.65%	-2.65%	3.38%	2.90%

Continued on next page

Table A.1 - continued from previous page

Node Name	MFE	MBE	RMSE	MAE
Zelzate	0.92%	-0.92%	1.69%	1.22%
Bocholtz	-0.32%	0.32%	1.15%	0.80%
Bocholtz-Vetshau	0.20%	-0.20%	0.92%	0.22%
Hilvarenbeek	2.42%	-2.42%	2.91%	2.55%
Tegelen	0.44%	-0.44%	1.19%	0.94%
Achim II	-1.77%	1.77%	2.97%	2.24%
Drohne Nowal	6.01%	-6.01%	8.20%	6.01%
Moffat	0.14%	-0.14%	2.95%	2.12%
Zeebrugge ZTP	-1.66%	1.66%	4.80%	2.98%
Zandvleit	1.13%	-1.13%	2.25%	1.42%
Haanrade	0.03%	-0.03%	0.94%	0.68%
Gravenvoeren Dilsen/Obbicht	-0.27%	0.27%	1.15%	0.83%
Zevenaar	0.14%	-0.14%	2.14%	1.66%
Ellund	0.31%	-0.31%	1.25%	0.77%
Emsburen Berge	0.42%	-0.42%	1.31%	0.76%
Nordlohne	0.11%	-0.11%	0.19%	0.12%
Winterswijk	0.98%	-0.98%	1.86%	1.44%
Dinxperlo	0.28%	-0.28%	0.56%	0.41%
Bunder Tief	2.03%	-2.03%	3.18%	2.04%
Etzel	-0.06%	0.06%	0.60%	0.38%
Oude Standjil/Bunde	2.35%	-2.35%	3.09%	2.55%
Vleighuis	0.70%	-0.70%	2.43%	1.41%
Emden	0.85%	-0.85%	2.27%	1.72%
Dornum	3.54%	-3.54%	4.27%	3.78%
Julianadorp	-1.49%	1.49%	3.25%	2.14%
Bacton	1.19%	-1.19%	8.09%	2.87%

Table A.2. Maximal Technical Capacity

A list of representative maximal technical capacities (MTC) associated with each border point specified within this study, with the Mean Absolute Error (MAE) expressed in MWh/h values, providing a standardised error magnitude.

Node Name	**MTC (MWh/h)**	**MAE (MWh/h)**
Arnoldstein/Tarvisio	47938.54	614.50
Murfeld/Cersak	4688.61	53.65
Gorizia/Sempeter	895.03	10.30
Rogatec	2208.33	255.62
Baumgarten	67166.67	769.62
Bizzarone	537.50	8.94
Dravaszerdahely	3166.67	43.46
Mosonmagyarovar	6379.17	111.38
Balassagyarmat/Velke Zlievce	5291.67	712.74
Lanzhot	16683.33	1965.07
Oberkappel	8308.33	335.61
Passo Gries/Griespass	26445.83	329.58
Beregdaroc	25216.67	629.82
Cesky Tesin/Cieszyn	1166.67	28.03
Waidhaus	37787.50	2545.40
Hora Svate Kateriny/Deutschneudorf	5637.50	101.77
Hora Svate Kateriny/Olbernhau	13320.83	422.87
Brandov Stegal/Stegal	4166.67	103.70
Uberackern ABG/Uberackern 1	3500.00	131.42
Uberackern Sudal/ Uberackern 2	7554.17	90.54
Jura	1558.33	77.26
Oltingue/Rodersdorf	9708.33	178.18
Wallbach	22295.83	693.91
RC Basel	366.67	5.14
Gela	21387.50	388.36
Maraza del Vallo	50137.50	896.82
Drozdowicze	5650.00	199.44
Wysokoje	7045.83	239.29
Lampertheim IV	3016.67	124.06
Keinbaum	5550.00	224.77
GCP Gaz System/Ontras	2029.17	50.89
RC Lindau	1008.33	19.31

Continued on next page

Table A.2 - continued from previous page

Node Name	MTC (MWh/h)	MAE (MWh/h)
Pirineos	9375.00	101.44
Obergailbach/Medelsheim	23825.00	1267.40
Gernsheim	4441.67	270.50
Uzhgorod	86666.67	918.77
Mallnow	38812.50	624.01
Steinitz	11100.00	109.06
Griefswald	65429.17	1479.42
Iberico	6000.00	129.97
Almeria	11083.33	167.46
Tarifa	18500.00	268.02
Alveringhem	11250.00	134.05
Eynatten 1/Lichtenbusch	5908.33	70.90

Appendix B

Appendices to Chapter 4

Appendix B

Table B.1. Robustness: "Traditional" and "Clean" Energy Classification

Clean Energy Classification - GICS sector "Utilities 551010"		
GICS Subsector name	**GICS Subsector code**	**No.**
Electric Utilities	55101010	20
Gas Utilities	55102010	5
Renewable Electricity	55105020	7
Total number of Clean Energy companies		32
Traditional Energy Classification - GICS sector "Energy 101020"		
GICS Subsector name	**GICS Subsector code**	**No.**
Independent Power Producers & Energy Traders	55105010	3
Integrated Oil & Gas	10102010	9
Oil & Gas Drilling	10101010	3
Oil & Gas Equipment & Services	10101020	15
Oil & Gas Exploration & Production	10102020	11
Oil & Gas Refining & Marketing	10102030	5
Oil & Gas Storage & Transportation	10102040	7
Coal & Consumable Fuels	10102050	0
Total number of Traditional Energy companies		53

Table B.2. Robustness: Pearson Correlation

This table reports the Pearson correlation coefficient for News sentiments index with returns and trading volumes for 32 Clean Energy and 53 Traditional Energy firms for the sample period of 1st January 2015 to 1st April 2022. $***$, $**$, and $*$ denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	NSI	
	LogReturn	***LogVolume***
Summary: Clean Energy	0.053***	0.064***
Summary: Traditional Energy	0.045***	0.033***

Table B.3. Robustness: Spillover Effects

This table reports spillover effects calculated from Eq.(4.4) between the News sentiment index (*NSI*), returns (*LogReturn*) and trading volumes (*LogVolume*) for 32 Clean and 53 Traditional Energy firms. The number of lags for the VAR model is selected using the Bayesian information criterion. $\widetilde{\theta}_{LogReturn \leftarrow NSI}$ refers to the spillover effect from NSI to stock returns and $\widetilde{\theta}_{LogVolume \leftarrow NSI}$ is to the spillover effect from NSI to trading volumes. Conversely, $\widetilde{\theta}_{NSI \leftarrow LogReturn}$ refers to the spillover effect from stock returns to NSI, and $\widetilde{\theta}_{NSI \leftarrow LogVolume}$ is the spillover effect from trading volumes to NSI. The values of spillover effects, ranging from 0% to 100% are presented in Columns (1)–(6).

	Returns			**Trading volumes**		
	$\widetilde{\theta}_{LogReturn \leftarrow NSI}$	$\widetilde{\theta}_{NSI \leftarrow LogReturn}$	**Net**	$\widetilde{\theta}_{LogVolume \leftarrow NSI}$	$\widetilde{\theta}_{NSI \leftarrow LogVolume}$	**Net**
Ticker Lag	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Clean Energy	2.68	2.34	0.34	1.80	1.74	0.06
Traditional Energy	2.79	1.67	1.12	1.98	1.62	0.37

Table B.4. Robustness: Comparative Analysis

This table presents the t-statistics of a comparative analysis: whether the spillover effects between Clean Energy companies and Traditional Energy companies are significantly different. The sample for this test includes 32 Clean Energy companies and 53 Traditional Energy companies. $***$, $**$, and $*$ denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Variables	**Clean Energy Mean of 32 Obs.**	**Traditional Energy Mean of 53 Obs.**	**Diff.**	**T-test (p-value)**
$\widetilde{\theta}_{LogReturn \leftarrow NSI}$	2.68	2.79	-0.11	0.686
$\widetilde{\theta}_{NSI \leftarrow LogReturn}$	2.34	1.67	0.67	0.003
$\widetilde{\theta}_{LogVolume \leftarrow NSI}$	1.80	1.98	-0.18	0.412
$\widetilde{\theta}_{NSI \leftarrow LogVolume}$	1.74	1.62	0.12	0.251

Table B.5. Linear Regression of Stock Returns and Trading Volumes on NSI

This table reports the results of linear regression of stock returns and trading volumes on *NSI* for each of the 32 Clean Energy and 53 Traditional Energy firms over the sample period of 1st January 2015 to 1st April 2022. The dependent variable is $LogReturn_t$ in Panel A and $LogVolume_t$ in Panel B, the variable of interest is NSI_t, and the control variables are $LogReturn_{t-1}$, $SX5E_t$, $V2X_t$, and $TTFGDAHD_t$. t-statistics are presented in the parentheses. $***$, $**$, and $*$ denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A. Linear Regression of Stock Returns on NSI.

	$LogReturn_t$					
	NSI_t (1)	$LogReturn_{t-1}$ (2)	$SX5E_t$ (3)	$V2X_t$ (4)	$TTFGDAHD_t$ (5)	Constant (6)
Clean Energy	0.007***	-0.005	0.003	0.001	0.001	0.000
	[10.80]	[-1.05]	[0.18]	[0.58]	[0.73]	[0.26]
Traditional Energy	0.007***	0.020**	0.673***	-0.026***	0.033***	-0.000***
	[8.43]	[2.37]	[24.83]	[-7.91]	[13.20]	[-3.19]

Panel B. Linear Regression of Trading Volumes on NSI.

	$LogVolume_t$					
	NSI_t (1)	$LogVolume_{t-1}$ (2)	$SX5E_t$ (3)	$V2X_t$ (4)	$TTFGDAHD_t$ (5)	Constant (6)
Clean Energy	0.011***	0.948***	0.056	0.009	0.000	0.132***
	[6.12]	[123.40]	[1.06]	[1.13]	[0.01]	[6.68]
Traditional Energy	0.004***	0.895***	0.066**	0.043***	0.014***	0.277***
	[3.80]	[89.77]	[2.42]	[7.77]	[3.88]	[10.55]

Bibliography

ACER, CEER (2018). *Annual Report on the Results of Monitoring the Internal Electricity and Natural Gas Markets in 2017-Electricity Wholesale Markets Volume.*

ACER, CEER and CEER (2017). *Annual Report on the Results of Monitoring the Internal Electricity and Gas Markets in 2016. Electricity Wholesale Market Volume.*

Anderloni, Luisa and Alessandra Tanda (2017). "Green energy companies: Stock performance and IPO returns". *Research in International Business and Finance* 39, pages 546–552.

Antonakakis, Nikolaos et al. (2018). "Dynamic connectedness of uncertainty across developed economies: A time-varying approach". *Economics Letters* 166, pages 63–75.

Arouri, Mohamed El Hedi et al. (2011). "Volatility spillovers between oil prices and stock sector returns: Implications for portfolio management". *Journal of International money and finance* 30.7, pages 1387–1405.

Asche, Frank et al. (2000). "Market integration for natural gas in Europe".

— (2002). "European market integration for gas? Volume flexibility and political risk". *Energy Economics* 24.3, pages 249–265.

Azar, Pablo D and Andrew W Lo (2016). "The wisdom of Twitter crowds: Predicting stock market reactions to FOMC meetings via Twitter feeds". *The Journal of Portfolio Management* 42.5, pages 123–134.

Baker, Malcolm and Jeffrey Wurgler (2006). "Investor sentiment and the cross-section of stock returns". *The journal of Finance* 61.4, pages 1645–1680.

— (2007). "Investor sentiment in the stock market". *Journal of economic perspectives* 21.2, pages 129–152.

Baruník, Jozef et al. (2016). "Asymmetric connectedness on the US stock market: Bad and good volatility spillovers". *Journal of Financial Markets* 27, pages 55–78.

Bastianin, Andrea et al. (2019). "Convergence of European natural gas prices". *Energy Economics* 81, pages 793–811.

Beenstock, Michael et al. (2019). "Spatial vector autoregressions". *The Econometric Analysis of Non-Stationary Spatial Panel Data*, pages 129–161.

Berry, David (2005). "Renewable energy as a natural gas price hedge: the case of wind". *Energy Policy* 33.6, pages 799–807.

Bollen, Johan et al. (2011). "Twitter mood predicts the stock market". *Journal of computational science* 2.1, pages 1–8.

Bourbonnais, Régis and Patrice Geoffron (2007). "Delineation of energy markets with cointegration techniques". *The Econometrics of Energy Systems*. Springer, pages 168–185.

Bros, T (2018). *Reflection on the Baumgarten Gas Explosion: Markets are Working.*

Brown, Stephen PA and Mine K Yucel (2008). "What drives natural gas prices?": *The Energy Journal* 29.2.

Burg, John Parker (1967). *Maximum entropy spectral analysis.* Stanford University.

Chen, Hailiang et al. (2011). "Sentiment revealed in social media and its effect on the stock market". *2011 IEEE Statistical Signal Processing Workshop (SSP)*. IEEE, pages 25–28.

Contreras, Javier et al. (2003). "ARIMA models to predict next-day electricity prices". *IEEE transactions on power systems* 18.3, pages 1014–1020.

Corbet, Shaen et al. (2018). "Exploring the dynamic relationships between cryptocurrencies and other financial assets". *Economics Letters* 165, pages 28–34.

Creti, Anna and Duc Khuong Nguyen (2015). *Energy markets financialization, risk spillovers, and pricing models*.

Cuddington, John T and Zhongmin Wang (2006). "Assessing the degree of spot market integration for US natural gas: evidence from daily price data". *Journal of Regulatory Economics* 29.2, pages 195–210.

Da, Zhi et al. (2011). "In search of attention". *The journal of finance* 66.5, pages 1461–1499.

De Vany, Arthur and W David Walls (1993). "Pipeline access and market integration in the natural gas industry: Evidence from cointegration tests". *The Energy Journal* 14.4.

Defferrard, Michaël et al. (2016). "Convolutional neural networks on graphs with fast localized spectral filtering". *Advances in neural information processing systems* 29.

Demirer, Mert et al. (2018). "Estimating global bank network connectedness". *Journal of Applied Econometrics* 33, pages 1–15.

Diebold, Francis X and Kamil Yilmaz (2009). "Measuring financial asset return and volatility spillovers, with application to global equity markets". *The Economic Journal* 119, pages 158–171.

— (2012). "Better to give than to receive: Predictive directional measurement of volatility spillovers". *International Journal of Forecasting* 28, pages 57–66.

Diebold, Francis X and Kamil Yılmaz (2014). "On the network topology of variance decompositions: Measuring the connectedness of financial firms". *Journal of Econometrics* 182, pages 119–134.

Dieckhöner, Caroline et al. (2013). "European natural gas infrastructure: the impact of market developments on gas flows and physical market integration". *Applied energy* 102, pages 994–1003.

Fabra, Natalia et al. (2020). *Preparing for the next crisis: How to secure the supply of essential goods and services*. Centre for Economic Policy Research.

Fama, Eugene F (1960). "Efficient market hypothesis". *Diss. PhD Thesis, Ph. D. dissertation*.

Fell, Harrison and Daniel T Kaffine (2018). "The fall of coal: Joint impacts of fuel prices and renewables on generation and emissions". *American Economic Journal: Economic Policy* 10.2, pages 90–116.

Geng, Jiang-Bo et al. (2014). "A dynamic analysis on global natural gas trade network". *Applied Energy* 132, pages 23–33.

Geng, Jiang-Bo et al. (2021). "Network connectedness between natural gas markets, uncertainty and stock markets". *Energy Economics* 95, page 105001.

Gers, Felix A et al. (2002). "Learning precise timing with LSTM recurrent networks". *Journal of machine learning research* 3.Aug, pages 115–143.

Geweke, John (1982). "Measurement of linear dependence and feedback between multiple time series". *Journal of the American statistical association* 77.378, pages 304–313.

Ghosh, Bikramaditya and Elie Bouri (2022). "Long memory and fractality in the universe of volatility indices". *Complexity* 2022.

Granger, Clive WJ (1969). "Investigating causal relations by econometric models and cross-spectral methods". *Econometrica: journal of the Econometric Society*, pages 424–438.

Granger, Clive WJ and Ramu Ramanathan (1984). "Improved methods of combining forecasts". *Journal of forecasting* 3.2, pages 197–204.

Growitsch, Christian et al. (2015). "Price convergence and information efficiency in German natural gas markets". *German Economic Review* 16.1, pages 87–103.

Grundmann, Tina et al. (2016). "Forecasting the natural gas price trend-evaluation of a sentiment analysis". *2016 IEEE 4th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW)*. IEEE, pages 160–164.

Guo, Jian-Feng and Qiang Ji (2013). "How does market concern derived from the Internet affect oil prices?": *Applied energy* 112, pages 1536–1543.

Gupta, Kartick and Rajabrata Banerjee (2019). "Does OPEC news sentiment influence stock returns of energy firms in the United States?": *Energy Economics* 77, pages 34–45.

Hamilton, James D (1983). "Oil and the macroeconomy since World War II". *Journal of political economy* 91.2, pages 228–248.

Han, Jiawei et al. (2011). "Data mining: concepts and". *Techniques (3rd ed), Morgan Kauffman*.

Han, Liyan et al. (2017). "Can investor attention predict oil prices?": *Energy Economics* 66, pages 547–558.

Heather, P and B Petrovich (2017). *European traded gas hubs: an updated analysis on liquidity, maturity and barriers to market integration*.

Heather, Patrick (2015). *The evolution of European traded gas hubs*. Oxford Institute for Energy Studies Oxford.

— (2017). "European traded gas hubs: an updated analysis on liquidity, maturity and barriers to market integration".

Hochreiter, Sepp and Jürgen Schmidhuber (1997). "Long short-term memory". *Neural computation* 9.8, pages 1735–1780.

Huang, Roger D et al. (1996). "Energy shocks and financial markets". *Journal of Futures markets* 16.1, pages 1–27.

Hulshof, Daan et al. (2016). "Market fundamentals, competition and natural-gas prices". *Energy policy* 94, pages 480–491.

Hunt, Chelsie and Olaf Weber (2019). "Fossil fuel divestment strategies: Financial and carbon-related consequences". *Organization & Environment* 32.1, pages 41–61.

Ji, Qiang and Ying Fan (2016). "Evolution of the world crude oil market integration: A graph theory analysis". *Energy Economics* 53, pages 90–100.

Ji, Qiang et al. (2019). "Measuring the interdependence between investor sentiment and crude oil returns: New evidence from the CFTC's disaggregated reports". *Finance Research Letters* 30, pages 420–425.

Jones, Charles M and Gautam Kaul (1996). "Oil and the stock markets". *The journal of Finance* 51.2, pages 463–491.

Joseph, Kissan et al. (2011). "Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search". *International Journal of Forecasting* 27, pages 1116–1127.

Kaminker, Christopher and Fiona Stewart (2012). "The role of institutional investors in financing clean energy".

Kaplanski, Guy et al. (2015). "Do happy people make optimistic investors?": *Journal of Financial and Quantitative Analysis* 50.1-2, pages 145–168.

Karabulut, Yigitcan (2013). "Can facebook predict stock market activity?": *AFA 2013 San Diego meetings paper*.

Khan, Muhammad Asif et al. (2020). "Time and frequency relationship between household investors' sentiment index and US industry stock returns". *Finance Research Letters* 36, page 101318.

King, Marlin and Milan Cuc (1996). "Price convergence in North American natural gas spot markets". *The Energy Journal* 17.2.

Kipf, Thomas N and Max Welling (2016). "Semi-supervised classification with graph convolutional networks". *arXiv preprint arXiv:1609.02907*.

Koop, Gary et al. (1996). "Impulse response analysis in nonlinear multivariate models". *Journal of Econometrics* 74, pages 119–147.

Kuang, Wei (2021). "Which clean energy sectors are attractive? A portfolio diversification perspective". *Energy Economics* 104, page 105644.

Lemmon, Michael and Evgenia Portniaguina (2006). "Consumer confidence and asset prices: Some empirical evidence". *Review of Financial Studies* 19.4, pages 1499–1529.

Li, Xiaodong et al. (2014). "News impact on stock price return via sentiment analysis". *Knowledge-Based Systems* 69, pages 14–23.

Lochner, Stefan and David Bothe (2007). *From Russia with gas: an analysis of the Nord Stream pipeline's impact on the European Gas Transmission System with the TIGER-Model*. Tech. rep. EWI Working Paper.

Mankiw, N Gregory (2020). *Principles of economics*. Cengage Learning.

Mao, Huina et al. (2011). "Predicting financial markets: Comparing survey, news, twitter and search engine data". *arXiv preprint arXiv:1112.1051*.

Monforti, F and A Szikszai (2010). "A MonteCarlo approach for assessing the adequacy of the European gas transmission system under supply crisis conditions". *Energy Policy* 38.5, pages 2486–2498.

Nahar, Syfun and Brett Inder (2002). "Testing convergence in economic growth for OECD countries". *Applied Economics* 34.16, pages 2011–2022.

Neumann, A and A Cullmann (2012). "What's the story with natural gas markets in Europe? Empirical evidence from spot trade data". *2012 9th International Conference on the European Energy Market*. IEEE, pages 1–6.

Neumann, Anne et al. (2006). "Convergence of European spot market prices for natural gas? A real-time analysis of market integration using the Kalman Filter". *Applied Economics Letters* 13.11, pages 727–732.

Newbery, David M (2016). "Towards a green energy economy? The EU Energy Union's transition to a low-carbon zero subsidy electricity system–Lessons from the UK's Electricity Market Reform". *Applied Energy* 179, pages 1321–1330.

Niu, Hongli et al. (2021). "Does investor sentiment differently affect stocks in different sectors? Evidence from China". *International Journal of Emerging Markets*.

Nofer, Michael and Oliver Hinz (2015). "Using twitter to predict the stock market". *Business & Information Systems Engineering* 57.4, pages 229–242.

Pesaran, H Hashem and Yongcheol Shin (1998). "Generalized impulse response analysis in linear multivariate models". *Economics Letters* 58, pages 17–29.

Pickl, Matthias J (2019). "The renewable energy strategies of oil majors–From oil to energy?": *Energy Strategy Reviews* 26, page 100370.

Qiu, Lily and Ivo Welch (2004). *Investor sentiment measures*.

Reboredo, Juan C and Andrea Ugolini (2018). "The impact of twitter sentiment on renewable energy stocks". *Energy Economics* 76, pages 153–169.

Renou-Maissant, Patricia (2012). "Toward the integration of European natural gas markets: A time-varying approach". *Energy Policy* 51, pages 779–790.

Ritchie, Justin, Hadi Dowlatabadi, et al. (2015). "Divest from the carbon bubble? Reviewing the implications and limitations of fossil fuel divestment for institutional investors". *Review of Economics & Finance* 5.2, pages 59–80.

Ritz, Robert A (2019). "A strategic perspective on competition between pipeline gas and LNG". *The Energy Journal* 40.5.

Robinson, Terry (2007). "Have European gas prices converged?": *Energy Policy* 35.4, pages 2347–2351.

Ryu, Doojin et al. (2017). "Investor sentiment, trading behavior and stock returns". *Applied Economics Letters* 24, pages 826–830.

Schlögl, Alois and Gernot Supp (2006). "Analyzing event-related EEG data with multivariate autoregressive parameters". *Progress in brain research* 159, pages 135–147.

Serletis, Apostolos (1997). "Is there an east-west split in North American natural gas markets?": *The Energy Journal* 18.1.

Seth, Anil K (2010). "A MATLAB toolbox for Granger causal connectivity analysis". *Journal of neuroscience methods* 186.2, pages 262–273.

Siganos, Antonios et al. (2014). "Facebook's daily sentiment and international stock markets". *Journal of Economic Behavior & Organization* 107, pages 730–743.

— (2017). "Divergence of sentiment and stock market trading". *Journal of Banking & Finance* 78, pages 130–141.

Song, Yingjie et al. (2019). "The dynamic dependence of fossil energy, investor sentiment and renewable energy stock markets". *Energy Economics* 84, page 104564.

Tetlock, Paul C (2007). "Giving content to investor sentiment: The role of media in the stock market". *The Journal of finance* 62.3, pages 1139–1168.

Tieleman, Tijmen and Geoffrey Hinton (2012). "Rmsprop: Divide the gradient by a running average of its recent magnitude. coursera: Neural networks for machine learning". *COURSERA Neural Networks Mach. Learn.*

Uygur, Utku and Oktay Taş (2014). "The impacts of investor sentiment on different economic sectors: Evidence from Istanbul Stock Exchange". *Borsa Istanbul Review* 14.4, pages 236–241.

Wallach, Daniel and Bruno Goffinet (1989). "Mean squared error of prediction as a criterion for evaluating and comparing system models". *Ecological modelling* 44.3-4, pages 299–306.

Walls, W David (1994). "Price convergence across natural gas fields and city markets". *The Energy Journal* 15.4.

Wiesen, Thomas FP et al. (2018). "Are generalized spillover indices overstating connectedness?": *Economics Letters* 173, pages 131–134.

Woroniuk, David et al. (2019). "European Gas Markets, Trading Hubs, and Price Formation: A Network Perspective".

Zhang, Dayong and David C Broadstock (2020). "Global financial crisis and rising connectedness in the international commodity markets". *International Review of Financial Analysis* 68, page 101239.