# How are day-ahead prices informative for predicting the next day's consumption of natural gas? Evidence from France

*Arthur Thomas*[*], *Olivier Massol*[**], *Benoît Sévi*[***]

**Abstract**

The purpose of this paper is to investigate, for the first time, whether the next day's consumption of natural gas can be accurately forecast using a simple model that solely incorporates the information contained in day-ahead market data. Hence, unlike standard models that use a number of meteorological variables, we only consider two predictors: the price of natural gas and the spark ratio measuring the relative price of electricity to gas. We develop a suitable modeling approach that captures the essential features of daily gas consumption and, in particular, the nonlinearities resulting from power dispatching and apply it to the case of France. Our results document the existence of a long-run relation between demand and spot prices and provide estimates of the marginal impacts that these price variables have on observed demand levels. We also provide evidence of the pivotal role of the spark ratio in the short run which is found to have an asymmetric and highly nonlinear impact on demand variations. Lastly, we show that our simple model is sufficient to generate predictions that are considerably more accurate than the forecasts published by infrastructure operators.

**Keywords:** Natural gas markets, day-ahead prices, load forecasting
**JEL:** L95, Q41, Q47, C22, C53

## 1. INTRODUCTION

The accuracy of the gas demand forecasts issued by Transmission System Operator (TSO) is now becoming an important matter in regulatory debates and has motivated the adoption of dedicated incentive schemes in some countries.[1] In response, TSOs have implemented advanced forecasting tools combining several methodologies (e.g., time series, neural network, adaptative logic networks) along with a plethora of variables (e.g., temperatures, wind speeds, rain, snow, cloud cover, forecasted

This paper has greatly benefited from the judicious comments of the editor and four anonymous referees. We are also indebted to Kevin Forbes and Yannick Le Pen for insightful discussions and suggestions. We have also benefited from helpful discussions with conference participants at the IAEE international conference (Montreal, 2019) and the Trans-Atlantic Infraday (Washington DC, 2019). Of course, any remaining errors are ours. This research has been supported by the Chair "The Economics of Natural Gas." The views expressed herein are strictly those of the authors and should not be construed to reflect the opinions of their respective institutions.

[*]Corresponding author. Department of Economics, IFP Énergies Nouvelles, 1-4 av. de Bois Préau, F-92852 Rueil-Malmaison, & LEMNA, Université de Nantes, BP. 52231 Chemin de la Censive du Tertre, 44322 Nantes Cedex, IFP Énergies Nouvelles & Université de Nantes, LEMNA. E-mail:arthur.thomas@univ-nantes.fr.

[**]IFP Énergies Nouvelles, 1-4 av. de Bois Préau, F-92852 Rueil-Malmaison, & Center for Energy Economics and Management, IFP School, 228-232 av. Napoléon Bonaparte, F-92852 Rueil-Malmaison, France, & Department of Economics, City, University of London, Northampton Square, London EC1V 0HB, UK. IFP Énergies Nouvelles, IFP School & City, University of London

[***]Université de Nantes, LEMNA, BP 52231 Chemin de la Censive du Tertre, 44322 Nantes Cedex, France.

[1]For example, in the UK, a dedicated annual incentive scheme has been implemented so that, depending on the observed average annual forecast error, the TSO can earn up to a maximum of £10 million (in case of 100% accuracy of the published day-ahead demand forecasts) or lose up to £1.5 million (Grid (2018)). In Italy, the regulatory authority monitors the forecasting error of the next day's load and uses it as a performance indicator to assess the quality of the information transmitted to the market (ENTSOG (2017) p.36).

power demand). Yet, despite these efforts, forecasting the next day's consumption of natural gas remains a challenging task.[2] By testing an alternative forecasting approach using the information in day-ahead prices, the present paper usefully contributes to the ongoing discussion on the performance of short-term consumption forecasts used in the gas industry.

Over the last two decades, a series of European regulatory reforms have prompted the emergence of a collection of day-ahead wholesale markets for natural gas, the so-called "gas hubs," that turned out to become an important source of gas procurement as the previously monopolized industry structure gradually became more fragmented (Miriello and Polo, 2015). By construction, these markets have been developed to cope with local network balancing needs and allow an optimal scheduling of resources. Their functioning is thus closely affected by the detailed balancing rules used by the TSO. An important milestone in the design of these balancing procedures occurred in 2014 when the European Commission imposed a unified network code on TSOs.[3] Yet, despite that harmonization, market analysts recurrently point to significant differences in the perceived degree of trading liquidity observed at the European gas hubs (Heather and Petrovich, 2017). Thus, a fundamental public policy issue is whether the current market design generates transparent spot prices that reflect the market participation of all concerned economic agents (suppliers, trading firms, and consumers).

In the electricity sector, Forbes and Zampelli (2014) proposed an original approach to examining the informational content of day-ahead electricity prices. They hypothesized that if day-ahead markets for electricity were efficient then these prices should reflect the processed information of all market participants regarding the next day's load. That consideration led them to test whether it was possible to improve the predictions of the next day's electricity load using only the information contained in the day-ahead price series. The authors examined California's PG&E aggregation area and applied traditional time-series techniques (namely a linear ARMAX specification) to model the next day's load as a function of a single explanatory variable: the day-ahead spark ratio defined as the electricity to gas price ratio. Their results reveal that this approach is sufficient for computing very accurate forecasts which outperform those published by the system operators. Remarkably, their results document the major informational content of day-ahead prices in the case of electricity.

Our paper is the first econometric study of the daily interactions between day-ahead prices and the natural gas demand observed at a given hub. In some respects, it extends Forbes and Zampelli (2014)'s approach in highlighting what necessary specific dimensions must be considered to produce accurate natural gas demand forecasts. However, we acknowledge that dealing with gas prices requires modeling specificities that differ from the ones used for electricity. Essentially, three main characteristics are typical of the gas market: (i) the fact that the aggregate gas demand emanates from both end-users and thermoelectric generation, (ii) the expected nonlinearity in the relationship between price and demand incidental to the level of the relative price of electricity to natural gas (spark ratio), and (iii) the time series (unit-root) properties of the data. We further elaborate on these three distinctive features in section 2 as they provide the essential justifications for our modeling choice and are thus key in our analysis. In light of these features, we consider two nonlinear specifications that are extensions of the well-known Autoregressive Distributed Lag (ARDL) model: the Nonlinear ARDL (NARDL) and the Threshold ARDL (TARDL), that we propose here, and that is a straightforward extension of the genuine ARDL model. By doing so, we investigate the presence of a long-run relationship between consumption, day-ahead price, and spark ratio and explore the potential asymmetric influence of the spark ratio on observed consumption levels.[4]

[2]For example, in the French southern gas balancing zone, the root mean squared error of the day-ahead consumption forecasts issued at 5pm by GRTGaz – the largest TSO – was approximately 23 GWh over the year 2015 corresponding to about 6.6% of the average daily load. That year, one working day out of four (the exact proportion is 25.2%) experienced an absolute forecasting error larger than 5% in relative terms (source: smart.grtgaz.com).

[3]See Commission Regulation (EU) No 312/2014 of March 26, 2014, OJ L 91, 27.3.2014.

[4]Notably, our modeling approach does not include any meteorological variables as it posits that all available information about the weather should be reflected in the trading decisions taken by market participants and thus in the day-ahead prices. It

Our application deals with two French wholesale markets – namely, the Point d'Échange de Gaz Nord (PEG Nord) and the Trading Region South (TRS) – over the period 2015-2018. This allows us to present a series of original findings. First, we provide evidence of a symmetric and significant long-run relationship between the daily demand level, the spot price of natural gas, and the relative price of electricity to gas. Second, we document the magnitude of the reaction of daily gas consumption to the price of natural gas in the short run. Third, we show that, in the short run, the spark ratio has an asymmetric and nonlinear impact on observed demand levels. In each market, the reported relationship obtained with the TARDL model is sufficiently robust to producing day-ahead forecasts that are considerably more accurate than those published by network operators.

Our main contribution to the literature is to show that publicly available information, such as day-ahead prices, can be used to produce efficient forecasts of tomorrow's consumption by relying on a quite simple econometric model. The fact that our demand forecasts are much better than those provided by TSOs appears quite puzzling as TSOs are expected to hold superior information. Why such superior information does not translate into better demand predictions remains an open question.

Our framework can provide useful guidance to a large audience interested in the dynamics of natural gas demand in the short run and in the reaction of that demand to market prices. While a large literature in applied econometrics literature has approached the question using medium to low frequency data (e.g., monthly, quarterly or annual), that reaction has never been examined using daily data. In principle, the use of daily data is much more relevant for eliciting the short run effects from lagged changes in prices on the observed demand. Geweke (1978) stresses that estimation over broader data intervals can result in significant bias. His analysis indicates that aggregation over time can create some kind of omitted variables bias problem because the intertemporal lag distribution is not properly specified. In our case, the use of daily data may provide more reliable estimates of the marginal impacts that gas and electricity prices have on natural gas demand in both the long- and the short-run, and thus on the dynamics of the reactions of natural gas demand to energy price changes. As these marginal impacts play an important role in the models developed to examine the effects of a possible sudden temporary disruption in gas supplies on optimal import policies,[5] our modeling approach usefully contributes to the policy discussions related to the security of foreign-controlled gas supplies in importing nations.

Though our discussion is confined to the French case, we believe that the results are pertinent for other countries engaged in a transition toward less carbon-intensive energy systems. In France, the gas consumption emanating from the power sector exhibits large and sudden variations because Combined Cycle Gas Turbines (CCGT) plants are primarily dispatched as peaking units, which leads to large flow variations in the gas network as these plants ramp up and down. That situation is likely to prefigure the new role assigned to gas-fired power plants when a previously thermoelectric dominated power system experiences a massive penetration of renewable generation. Because of their almost zero marginal costs of production, solar and wind generators are placed at the beginning of the electricity 'merit order,' which greatly reduces the need to dispatch gas-fueled generators as baseload or mid-merit units (Green and Vasilakos, 2010) and thus leads to large variations in the gas demand emanating from these plants (Qadrdan et al., 2010).

The present analysis has important implications for the cost-efficient operation of a natural gas pipeline system and its economic regulation in the broader context of decarbonized energy systems. As gas-fired generation is increasingly used as a backup technology in power systems, the short-term variability observed in the power sector is increasingly transferred to gas infrastructures. Because of both the increased variability of the next days' loads and the possibility of demand forecasting errors, gas TSOs' are forced to adopt precautionary network management strategies based on the buildup and discharge of a pipeline inventory named linepack (Gopalakrishnan and Biegler, 2013; Tran et al.,

should be emphasized that, following Forbes and Zampelli (2014), our aim is to explore the informational content of day-ahead markets rather than investigating the potential contribution of additional variables to the prediction of gas consumption.

[5] See: (Manne et al., 1986; Hoel and Strom, 1987; Markandya and Pemberton, 2010; Abada and Massol, 2011)

2018). The linepack storage is an important source of short-term flexibility that could be extremely valuable in decarbonized energy systems as shown in Sun et al. (2012) and Arvesen et al. (2013). Yet, from a market design perspective, that storage service is seldom sold to market participants in Europe and a substantial share of its cost is socialized by means of transport tariffs (Keyaerts et al., 2011; Hallack and Vazquez, 2013). Because of these factors, there is a heightened interest in the accuracy of the gas demand forecasts published by the TSOs. That topic is now considered to be an important regulatory policy issue that has recently motivated the implementation of specific incentive schemes in some countries (e.g., the UK and Italy, see footnote 1 for more details).

The remaining sections of this paper are organized as follows. In the next section, we present some attributes that are typical for gas markets and which inspire our econometric approach which is exposed in section 3. In section 4, we present the data along with some preliminary analysis of the series. The empirical findings are provided in section 5. Finally, section 6 concludes.

## 2. DISTINCTIVE FEATURES IN MODELING AND FORECASTING SHORT-TERM NATURAL GAS DEMAND

At least three distinctive features are worth considering when attempting to model and forecast the daily consumption of natural gas. The first one is related to the specific nature of natural gas as a source of energy. The original work of Forbes and Zampelli (2014) focuses on electricity, which is a type of energy predominantly consumed by end-users. In contrast, natural gas is either directly consumed by end-users (e.g., by households or industries to produce heat) or is converted into electricity. While one can assume that the consumption emanating from the former users is directly influenced by the price of natural gas, those from the power sector are likely to be influenced by the relative price of natural gas and electricity. Given the importance of the power sector's demand for natural gas, one can hardly overlook the role of electricity prices in the observed aggregate demand for natural gas. Hence, the present analysis suggests considering two variables to predict the next day's load: the day-ahead price of natural gas along with the day-ahead spark ratio. In line with Forbes and Zampelli (2014), our results provide evidence of the key role of the spark ratio in predicting tomorrow's consumption.

The second feature is related to the use of natural gas in power generation. On a given power system, gas-fueled generation is seldom the unique technology available to generate electricity. Thus, depending on the observed level of the relative price of electricity and gas, it is likely that the consumption of natural gas in the power sector differs. Arguably, for low levels of the spark ratio, the revenue derived from gas-fueled generation is not sufficient to compensate both the thermal losses and other operating costs. One can thus conjecture that the consumption of natural gas in the power sector remains circumscribed to a few cogeneration plants (i.e., Combined Heat and Power plants) that must run to supply heat at industrial sites or district heating systems and is thus not very responsive to the electricity/gas ratio. In contrast, whenever the spark ratio is large enough, large gas-fueled thermal power stations consume natural gas, which suggests a stronger positive relationship between the gas load and the spark price ratio in this case. Overall, that discussion suggests opting for a modeling approach that can incorporate nonlinearities and is suitable for detecting the possible presence of asymmetries between the observed consumption levels of natural gas and the value of the spark ratio. In our application, our estimates strongly support the nonlinear assumption. Moreover, nonlinearity is shown to be critical to our analysis and estimates related to nonlinearity are consistently highly significant.

The third specific feature has a rather methodological nature. In recent years, a large empirical literature has examined the time series properties of day-ahead electricity prices.[6] In these studies, spot power prices are commonly found to be mean reverting and unlikely to have a unit root. In contrast, the day-ahead price series of natural gas (Vany and Walls, 1993; Serletis and Herbert,

[6] See, e.g., (Serletis and Herbert, 1999; Lucia and Schwartz, 2002; Knittel and Roberts, 2005; Worthington et al., 2005; Bunn and Gianfreda, 2010; de Menezes and Houllier, 2016; Gianfreda and Bunn, 2018).

1999; Renou-Maissant, 2012; Thoenes, 2014) and the daily consumption of natural gas, as examined in Giulietti et al. (2012), are often found to be non-stationary and an integrated process of order one, $I(1)$. As a result, one cannot directly estimate a regression equation involving the variables in levels without conveying the risk of generating spurious results. To overcome that problem, we consider the ARDL model suggested in Pesaran and Shin (1999) and Pesaran et al. (2001) which is a single cointegration and error correction approach that yields valid results regardless of whether the underlying variables are integrated in different orders (one, zero, or a combination of both). This method has two main advantages. First, it allows us to test for the presence of a long-run relationship among the variables without any prior knowledge of their order of integration, which partly avoids problems associated with unit root testing.[7] Second, it offers a parsimonious modeling approach that can easily be extended to incorporate nonlinearities using partial sum decompositions as in the Nonlinear ARDL (NARDL) model proposed by Shin et al. (2014) or the Threshold ARDL (TARDL) model newly formulated in this paper. We provide evidence that our data series either have a unit root or are stationary, thereby duly justifying the use of an ARDL specification.

Overall, as will be shown below, our empirical results confirm that accounting for these three features is highly consequential for modeling and accurately predicting the next day's consumption of natural gas. In this respect, our empirical modeling approach thus noticeably departs from the one in Forbes and Zampelli (2014) although our aim remains quite similar in spirit.

## 3. ECONOMETRIC APPROACH

In this section, we first provide a condensed review of the standard ARDL model before presenting two extensions: the NARDL and the TARDL. We let $q_t$ denote the quantity demanded on day $t$ in a given wholesale market.[8] We aim to model the quantity demanded as a function of the day-ahead price of natural gas that is delivered that day $p_t$, the spark ratio $s_t := p_t^E/p_t$ where $p_t^E$ is the day-ahead price of electricity delivered during the peak-load block of day $t$, and $\mathrm{M}_i$, with $i = 1 \cdots 11$, 11 monthly dummy variables. Hence, our model has the following form: $q_t = f\left(p_t, s_t, \mathrm{M}_{i,t}\right)$.

### 3.1 ARDL model

The linear ARDL model of Pesaran et al. (2001) enables interpretation based on the short- and long-run effects of the explanatory variables on the dependent variable. This approach has an important advantage over other cointegration techniques such as those of Engle and Granger (1987) or Johansen and Juselius (1990), as it can be applied regardless of whether all the variables share the same order of integration, which is not possible under alternative cointegration models.[9]

In our framework, the specification of a linear ARDL model is as follows:

$$\begin{aligned}\Delta q_t &= \alpha + \rho q_{t-1} + \theta_1 s_{t-1} + \theta_2 p_{t-1} + \sum_{i=1}^{p} \phi_i \Delta q_{t-i} + \sum_{i=0}^{q_1-1} \gamma_i \Delta s_{t-i} \\ &+ \sum_{i=0}^{q_2-1} \delta_i \Delta p_{t-i} + \sum_{i=1}^{11} \kappa_i \mathrm{M}_{i,t} + \sum_{i=1}^{11} \zeta_i \Delta \mathrm{M}_{i,t} + \mu_t\end{aligned} \tag{1}$$

where: $\Delta$ is the first difference operator; $\alpha$ denotes an intercept; $\rho$ is the feedback coefficient (expected

[7]As we discuss below, unit-root testing is nevertheless useful for checking that the series are not $I(2)$, i.e., integrated or order 2.

[8]For simplicity, we abstract from indexing variables with respect to the market under scrutiny. Our notation system works in the same way for the two markets examined in the application presented below.

[9]In an ARDL model, the variables can be either stationary (i.e., integrated of order zero $I(0)$) or integrated of order one $I(1)$. However, this model is not valid when there are $I(2)$ variables.

to be negative); $\theta_1$ and $\theta_2$ represent the long-run coefficients; $\phi_i$, $\gamma_i$ and $\delta_i$ are the short-run coefficients; $p$, $q_1$ and $q_2$ are the respective lag orders for the dependent and explanatory variables; $\kappa_i$ and $\zeta_i$ represent the long- and short-run effect of the monthly dummy variables; and $\mu_t$ is the error term. These coefficients can be combined to obtain the long-run multipliers[10] is the long-run multiplier associated with the i$^{th}$ monthly dummies. $\beta_1 := -\theta_1/\rho$ and $\beta_2 := -\theta_2/\rho$. The coefficients $\gamma_i$ (respectively $\delta_i$) capture the short-run adjustments of natural gas demand to spark ratio (respectively gas price) shocks. In particular, $\gamma_0$ and $\delta_0$ measure the contemporaneous impacts of these changes on natural gas demand variations.

As we suspect the possible presence of endogeneity between the price and consumption series, the results obtained from an OLS estimation of eq. (1) can yield inaccurate standard deviations of the estimated parameters (Pesaran and Shin, 1999). To correct for that, we follow the discussion in Section 3 of Pesaran and Shin (1999) and estimate this ARDL model using the delta-method ($\Delta$-method hereafter).

The ARDL model can be used to examine the existence of a long-run (i.e., cointegration) relationship among the underlying variables by employing the bound testing approach described in Pesaran et al. (2001). This amounts to testing the null hypothesis of no-cointegration among the underlying variables, that is, $H_0 : (\rho = \theta_1 = \theta_2 = 0)$. To test this hypothesis, the authors propose a non-standard $F$-test that takes into account the stationarity properties of the variables and evaluate bounds for the critical values at any significance level. The lower bound assumes that all variables are $I(0)$, whereas the upper bound assumes that all variables are $I(1)$. If the test statistic is larger than the upper bound critical value, the null hypothesis of no-cointegration is rejected, which means that a cointegrating relationship among the underlying variables can be ascertained. Conversely, if the test statistic is lower than the lower bound then this null hypothesis is not rejected, which indicates that the underlying variables are not cointegrated. Lastly, if the test statistic falls between these two bounds, the inference remains inconclusive.

By construction, the long-run marginal impact of the electricity price to natural gas demand is $\beta_1$, which is expected to be positive, as one can presume both energies will be substitutable. Also, the long-run marginal impact of natural gas price on natural gas demand is $\beta_2 - \beta_1$ , which is expected to be negative.[11]

### 3.2 NARDL model

Although the ARDL model enables the investigation of the short- and long-run relationships between variables, it presumes that all exogenous variables have symmetric effects on the dependent variable and thus becomes unsuitable when these linkages are nonlinear and/or asymmetric. To overcome that limitation, Shin et al. (2014) developed the nonlinear autoregressive distributed lag (NARDL) model that offers an asymmetric expansion of the original ARDL model.

In the NARDL model, short-run and long-run nonlinearities are introduced via positive and negative partial sum decompositions of the explanatory variables. In the present paper, we focus solely on the asymmetric impact that the spark ratio $s_t$ can have on the consumption of natural gas.[12]

[10] By gathering the terms in levels, one can use that definition of the multipliers to express the specification in the usual ECM form: $\Delta q_t = \alpha + \rho \left( q_{t-1} - \beta_1 s_{t-1} - \beta_2 p_{t-1} - \sum_{i=1}^{11} \beta_{3,i} \mathrm{M}_{i,t} \right) + \sum_{i=1}^{p} \phi_i \Delta q_{t-i} + \sum_{i=0}^{q_1-1} \gamma_i \Delta s_{t-i} + \sum_{i=0}^{q_2-1} \delta_i \Delta p_{t-i} + \sum_{i=1}^{11} \zeta_i \Delta \mathrm{M}_{i,t} + \mu_t$, where $\beta_{3,i} := -\kappa_i/\rho$

[11] Here, the long-run multipliers represent the percentage change in the quantity demanded in the long-run resulting from a *ceteris paribus* 1 percent change in the change of the other explanatory variable (e.g., the spark ratio). Following a remark raised by two of the referees, we agree that to derive elasticity estimates, it is necessary to rely on a full-fledged structural econometric model. Though the interpretation of ARDL estimates as elasticities is quite common in the applied econometric literature in the context of modeling energy demand (Adeyemi and Hunt, 2014; Dergiades and Tsoulfidis, 2008; Hunt and Ninomiya, 2003) here we abstract from such an interpretation.

[12] Our focus on the asymmetric response of gas consumption to the spark-ratio is derived from a series of empirical investigations conducted on the two French wholesale markets. In both cases, we consistently observed that the null hypothesis of a symmetric response of gas consumption to variations in the day-ahead price of natural gas was always rejected in the short

We thus decompose the spark ratio $s_t$ as follows:

$$s_t = s_0 + s_t^+ + s_t^- \tag{2}$$

where $s_0$ is an arbitrary initial value and $s_t^+$ and $s_t^-$ denote partial sum processes which accumulate positive and negative changes in $s_t$ respectively. These partial sum processes are derived from the first differences $\Delta s_j := s_{j+1} - s_j$ using the following definitions:

$$s_t^+ = \sum_{j=1}^{t} max\left(\Delta s_j, 0\right)$$

$$s_t^- = \sum_{j=1}^{t} min\left(\Delta s_j, 0\right)$$

This decomposition is then introduced in (1) to obtain the NARDL model, which is an immediate generalization of the genuine ARDL but allows for the presence of short- and long-run asymmetries which are seen to be very useful in our case. The NARDL model can be written as follows:

$$\begin{aligned}\Delta q_t &= \alpha + \rho q_{t-1} + \left(\theta_1^+ s_{t-1}^+ + \theta_1^- s_{t-1}^-\right) + \theta_2 p_{t-1}^G + \sum_{i=1}^{p} \phi_i \Delta q_{t-i} \\ &+ \sum_{i=0}^{q_1-1} \left(\gamma_i^+ \Delta s_{t-i}^+ + \gamma_i^- \Delta s_{t-i}^-\right) + \sum_{i=0}^{q_2-1} \delta_i \Delta p_{t-i}^G + \sum_{i=1}^{11} \kappa_i \mathrm{M}_{i,t} + \sum_{i=1}^{11} \zeta_i \Delta \mathrm{M}_{i,t} + \mu_t\end{aligned} \tag{3}$$

where $\theta_1^+$ and $\theta_1^-$ are the long-run asymmetric coefficients and $\gamma_i^+$ and $\gamma_i^-$ are the short-run asymmetric coefficients representing the contemporaneous impacts of positive and negative changes in the spark ratio on natural gas demand variations. The asymmetric long-run multipliers are $\beta_1^+ := -\theta_1^+/\rho$ and $\beta_1^- := -\theta_1^-/\rho$.

In this framework, the non-standard bounds-based $F$-test of Pesaran et al. (2001) are still valid for examining the presence of an asymmetric long-run relationship among the variables in levels. More specifically, the null hypothesis becomes $H_0 : \left(\rho = \theta_1^+ = \theta_1^- = \theta_2 = 0\right)$.

By construction, the NARDL specification nests three special cases: (i) a symmetric long-run relationship, that is, the null hypothesis $H_0 : \left(\theta_1^+ = \theta_1^-\right)$; (ii) a symmetric short-run relationship, that is: $H_0 : \left(\gamma_i^+ = \gamma_i^-, \forall i \in \{1, ..., (q_1 - 1)\}\right)$; and (iii) the joint presence of long- and short-run symmetry as in the original ARDL model. These three restrictions can be tested using standard Wald tests.

### 3.3 Threshold ARDL model

The NARDL model implicitly uses a zero threshold value to define the partial sum processes as an observed variation is thought to be either positive or negative. While the use of such a zero threshold value may be appealing in macroeconomics or in finance,[13] one can question its relevance for the present application. Arguably, the technological considerations in section 2 suggest that the observed level of the spark ratio is likely to have a nonlinear influence on the power sector's demand for natural gas. Indeed, for low levels of that ratio, the impact is likely to be of minor importance. The other way round, when the value of the spark ratio is large enough to compensate the costs for generating electricity at large gas-fired power stations, we expect it to have a larger influence on gas demand.

run, but not in the long run. These results, which motivate our econometric specifications, are available from the authors upon request.

[13]For instance, to capture the potential asymmetries occurring during expansionary or contractionary periods of the business cycle or to model the effect of positive and negative financial news

These considerations lead us to consider a different decomposition of the exogenous variable $s_t$ that explicitly refers to a possibly non-zero threshold. Several recent contributions in economics have suggested decomposing the explanatory variable to allow for likely different regimes.[14] By construction, these earlier approaches compare the value of the first-differenced variable (here $\Delta s_t$) with a threshold and directly use that comparison to define the partial sum processes. However, the discussion above suggests that in the present case, it would be preferable to examine whether an observed variation in the spark ratio does or does not have the same impact on the observed demand variation if the level attained by that explanatory variable exceeds, or not, a given threshold $Th$. Formally, this leads us to consider the following new decomposition of the spark ratio $s_t$:

$$s_t = s_0' + {s_t}^{>Th} + {s_t}^{\leq Th} \tag{4}$$

$${s_t}^{>Th} = \sum_{j=1}^{t} \Delta {s_j}^{>Th} = \sum_{j=1}^{t} \Delta s_j I_{s_j > Th} \tag{5}$$

$${s_t}^{\leq Th} = \sum_{j=1}^{t} \Delta {s_j}^{\leq Th} = \sum_{j=1}^{t} \Delta s_j \left(1 - I_{s_j > Th}\right) \tag{6}$$

there $s_0'$ is an arbitrary initial value (different from the one obtained in the NARDL) and $I_{s_j > Th}$ is the indicator function that takes the value 1 if the condition $s_j > Th$ is satisfied and 0 otherwise.

The specification of the Threshold ARDL model is then obtained by introducing the decomposition in (1), that is:

$$\begin{aligned} \Delta q_t = {} & \alpha + \rho q_{t-1} + \left( {\theta_1}^{>Th} s_{t-1}^{>Th} + {\theta_1}^{\leq Th} s_{t-1}^{\leq Th} \right) + \theta_2 p_{t-1}^{G} + \sum_{i=1}^{p} \phi_i \Delta q_{t-i} \\ & + \sum_{i=0}^{q_1 - 1} \left( {\gamma_i}^{>Th} \Delta s_{t-i}^{>Th} + {\gamma_i}^{\leq Th} \Delta s_{t-i}^{\leq Th} \right) + \sum_{i=0}^{q_2 - 1} \delta_i \Delta p_{t-i}^{G} + \sum_{i=1}^{11} \kappa_i \mathrm{M}_{i,t} + \sum_{i=1}^{11} \zeta_i \Delta \mathrm{M}_{i,t} + \mu_t \end{aligned} \tag{7}$$

where ${\theta_1}^{>Th}$ and ${\theta_1}^{\leq Th}$ are the long-run asymmetric coefficients and ${\gamma_i}^{>Th}$ and ${\gamma_i}^{\leq Th}$ are the short-run asymmetric coefficients. The associated asymmetric long-run multipliers are ${\beta_1}^{>Th} := -{\theta_1}^{>Th}/\rho$ and ${\beta_1}^{\leq Th} := -{\theta_1}^{\leq Th}/\rho$.

For a given value of the threshold parameter $Th$, that specification can be estimated using the $\Delta$-method. The value of that threshold parameter can be determined using a grid search algorithm over the spark-ratio variable. In this paper, we consider the grid set defined by the percentiles of $s_t$ – after trimming for the $10^{th}$ and $90^{th}$ percentiles so as to maintain a sufficient number of observations in each regime – and select the threshold value that minimizes the sum of squared residuals. That said, this procedure provides only one value of $Th$ which is an uncertain parameter. So, we use a block bootstrapping approach to obtain both the estimator of $Th$ and the associated standard deviation.Because of the nuisance parameter associated with the threshold value, it is important to keep in mind that the validity of the Wald-test procedure can be jeopardized (Hansen, 1996). In the present paper, the long-run and short-run symmetry restrictions are thus tested using the expF statistic described in (Andrews and Ploberger, 1994).

[14] See, e.g., (Greenwood-Nimmo, 2011; Pal and Mitra, 2015; Bagnai and Ospina, 2018).

## 4. PRELIMINARY ANALYSIS OF THE DATA

### 4.1 Data

We focus on the two gas-balancing zones in France, namely PEG Nord and TRS, that are respectively associated with the country's northern and southern wholesale markets for natural gas. In France, transit to and from other countries represents a modest share of the flows transported on the gas pipeline systems. One can thus expect that the price formed at such a hub reflects the interaction of supply and demand prevailing in that specific balancing zone. From a regulatory perspective, these two gas hubs share a common market design and are monitored by the same regulatory authority: the Commission de Régulation de l'Énergie (CRE). Despite that institutional similarity, PEG Nord attracts a greater number of active market participants and is reputed to provide a more liquid trading environment than TRS (Heather and Petrovich, 2017).

We consider the period covering April 1, 2015, to February 2, 2018. The starting date is the day on which trading operations commenced at the TRS balancing zone (CRE, 2012). During that period, the PEG Nord and the TRS experienced a steady institutional environment comprising unchanged infrastructure access rules and balancing conditions. In both markets, the main provisions stipulated in the EU's network code on the gas balancing of transmission networks were already implemented at that time (ACER-ENTSOG, 2014).

The daily data used in this paper are collected from the sources indicated in Table 1. GRTGaz is the unique TSO operating in the PEG Nord balancing zone and provides the consumption data observed in that zone. In contrast, two TSOs operate in the TRS region – GRTGaz and Teréga (formerly named TIGF) – and our daily consumption series is obtained by adding the figures reported by these two TSOs. In France, there is a unique wholesale market for electricity that covers the entire metropolitan territory. We use the day-ahead price of electricity for the peak-load block that covers the hours from 9am to 8pm on the next working day.[15] In each gas-balancing zone, the spark ratio is computed by dividing the day-ahead electricity price by the corresponding day-ahead price of natural gas.

Following the approach retained in numerous empirical analyses of energy markets (Ramanathan et al., 1997; Karakatsani and Bunn, 2008; Bordignon et al., 2013), we concentrate our attention on the working days and remove the weekends and the bank holidays from the data. Hence, the day-ahead gas prices formed on a Friday value the gas to be delivered on the following Monday.

The whole dataset has a total of 625 observations. In the sequel, that data set is further divided into two parts. The first part, covering the period April 1, 2015, to December 31, 2016 (i.e., 396 observations), is only used for model estimation. The remaining part comprises 229 observations and is used for evaluating out-of-sample forecasts.

**Table 1: Data sources.**

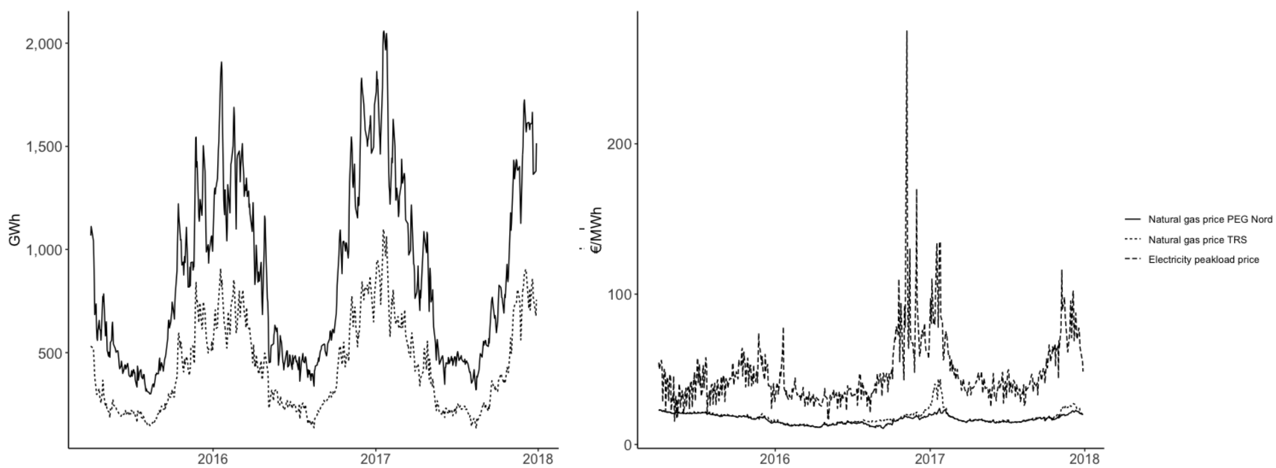
| Market | Data | Source | Specification | Unit |
|---|---|---|---|---|
| Natural gas PEG | Daily consumption | smart.grtgaz.com (GRTGaz) | Consumption North | GWh |
| | Day-ahead price | Bloomberg | PEG Nord, End of Day | €/MWh |
| Natural gas TRS | Daily consumption | smart.grtgaz.com (GRTGaz) opendata.reseaux-energies.fr (Teréga) | Consumption South | GWh |
| | Day-ahead price | Bloomberg | TRS, End of Day | €/MWh |
| Electricity | Day-ahead price | Bloomberg | Powernext Peak-load, End of Day | €/MWh |

[15]Hence, the prices are formed during the day preceding the delivery (and thus the consumption) of natural gas.

### 4.2 Descriptive statistics

Figure 1 provides plots of the price and consumption series in levels for the entire sample period. A visual inspection of the consumption plot indicates that the PEG Nord and TRS demand series present a smooth yearly seasonality linked to the gradual variation in weather conditions and the use of natural gas for space heating.

**Figure 1: Daily consumption of natural gas (left) and day-ahead energy prices in France (right).**




The price plots show that the day-ahead price of electricity measured during the peak-load block is both larger and more variable than that of natural gas. One can also note that the prices formed at the southern gas hub (TRS) are very similar to the ones at PEG Nord (except during winter 2016–2017).

Table 9 (in Appendix A) summarizes the descriptive statistics for the series in levels during the estimation period. The average consumption figures indicate that PEG Nord is roughly twice larger than TRS (i.e., 814.8 GWh compared with 401.8 GWh) which is not surprising as it is home to a larger share of the population, gathers the biggest industrial sites, and is generally affected by cooler weather. The coefficient of variation of these two series are quite large and attain 50.6% and 52.1% in PEG Nord and TRS, respectively. The slightly larger coefficient observed in the southern gas balancing zone could be related to that region's smaller industrial base, as the consumption observed in large industrial sites usually exhibits limited seasonal variations.

The distributional properties of these series show some signs of non-normality (see the highly significant Jarque-Bera statistics in Table 9 (Appendix A). Thus, in what follows, all the variables are transformed into natural logarithms to facilitate the economic interpretation of the estimated coefficients and to partly deal with the fact that the empirical distributions are not Gaussian.

### 4.3 Unit roots

The ARDL modeling framework relaxes the usual assumption in the cointegration analysis that all variables must be integrated of the same order. However, it is necessary to check the unit root properties of the data series as that method is not valid in the presence of $I(2)$ variables.[16]

[16]Recall that the natural gas consumption series exhibit a smooth yearly seasonality and that the presence of that seasonal pattern may bias the testing procedure used to detect the presence of a unit root. To correct for the effects of that seasonal pattern, the test results reported in Table 2 have been obtained with the deseasonalized consumption series resulting from a regression with monthly dummy variables.

Here, the integration properties of the data are investigated using two standard unit root tests – namely, the Augmented Dickey-Fuller (ADF) and the Phillips and Perron (PP) tests – and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test for the presence of stationarity.

From the results presented in Table 2, it is found that none of the variables are integrated in order two which allows us to use an ARDL approach to investigate the presence of cointegration. For the electricity and gas price series, the findings are consistent with the characteristics highlighted in earlier studies evoked in the introduction: the spot price of electricity is found to be mean reverting and the natural gas prices are non-stationary. Unsurprisingly, the properties of the spark ratio series, which is defined as the quotient of these two price series, is not straightforward to interpret. For PEG Nord, the KPSS test suggests that we are dealing with an $I(1)$ variable whereas the ADF and PP unit roots tests firmly reject the presence of a unit root. A similar issue occurs when examining the results concerning the daily consumption of natural gas. The KPSS tests strongly reject the stationarity hypotheses, whereas the ADF and PP test results suggest that the series are $I(0)$. These results may leave us wondering whether these series are stationary or not. Facing such a dilemma, we follow Cochrane (1991) who argues that the important question is not whether the time series should be unequivocally classified in the unit root or the stationary categories but, rather, how to develop the appropriate inferential procedure for the data being analyzed. In the present case, there is strong evidence that the first-differenced consumption series are stationary. Therefore, even if the support in favor of a unit root is not strong, modeling the consumption data using first-differenced series reduces the risk of spurious regression. As such, we follow Hendry and Juselius (2000) who note that *"even though a variable is stationary, but with root close to unity [...] it is often a good idea to act as if there are unit roots to obtain robust statistical inference."*(p. 21).

**Table 2: Unit root tests.**

| | Levels | | | First differences | | |
|---|---|---|---|---|---|---|
| | ADF | PP | KPSS | ADF | PP | KPSS |
| **PEG Nord** | | | | | | |
| $q_t$ | -8.41*** | -7.77*** | $0.74_C$*** | -19.77*** | -38.33*** | $0.06_C$ |
| $s_t$ | $-3.72_C$*** | $-12.20_{C\&T}$*** | $0.27_{C\&T}$ *** | -13.81*** | -69.38*** | $0.06_C$ |
| $p_t^G$ | $-2.09_C$ | $-2.18_C$ | $0.45_{C\&T}$*** | -29.42*** | -29.56*** | $0.23_C$ |
| **TRS** | | | | | | |
| $q_t$ | -3.62*** | -2.94*** | $2.38_C$*** | -16.27*** | -33.52*** | $0.09_C$ |
| $s_t$ | $-4.69_{C\&T}$*** | $-12.66_{C\&T}$*** | $0.19_{C\&T}$ * | -13.51*** | -66.33*** | $0.07_C$ |
| $p_t^G$ | $-2.19_C$ | $-2.32_C$ | $0.36_{C\&T}$*** | -30.88*** | -31.43*** | $0.19_C$ |
| **Electricity** | | | | | | |
| $p_t^E$ | $-3.47_C$*** | $-9.34_{C\&T}$*** | $0.11_{C\&T}$ | -13.40*** | -59.46*** | $0.08_C$ |

Notes: For the ADF test, we apply the lag structure suggested by the Schwartz information criterion. For the PP test, the truncation lags are decided by Newey-West default. For the KPSS test, the bandwidth is selected by the Newey-West automatic selection procedure using the Bartlett kernel. The $p$-values are provided in parentheses. Here, $C$ (respectively $T$) indicates that a constant (respectively a trend) is included in the test equation. Asterisks indicate rejection of the null hypothesis at 0.05** and 0.01*** significance levels, respectively.

## 5. EMPIRICAL FINDINGS

### 5.1 Estimation results

The lag structure is determined using the Bayesian Information Criterion (BIC). For each specification, we allow a maximum lag-length of 20 for each variable and then compare the BIC scores obtained for each possible combination of lagged values. For each model, the selected specification is the one that provides the lowest BIC score. With that selection procedure, the obtained optimal lag structure is identical for the ARDL, NARDL and TARDL models, that is: $p = 1$, $q_1 = 1$ and $q_2 = 1$, which indicates that the short-run dynamics solely consider the instantaneous impacts that the gas price and the spark ratio have on the contemporary demand changes.

For the TARDL model, the block bootstrap procedure finds that, on average, the sum of squared residuals is minimized for $Th = 0.60$ (s.d.= 0.08) in PEG Nord and $Th = 0.59$ (s.d.= 0.01) in TRS.[17] Of particular interest is the relationship between these threshold values and the ones obtained from power engineering considerations. After exponentiation, these values correspond to a price of electricity in levels that amount to 1.82 (respectively 1.80) times those of natural gas in the PEG Nord (respectively the TRS) zone. Clearly, these values are commensurate with the heat rates of the large gas-fired power plants installed in France as the nominal thermal efficiencies (i.e., the inverse of the heat rate) of these CCGT plants are in the $[50\%, 62\%]$ range.

We now examine whether the presence of short- and long-run asymmetries is supported by the data or not. From the Wald statistics $W_{SR,LR}$ and the expF statistics $expF_{SR,LR}$ reported in Table 3, we can notice that, whatever the alternative nonlinear model under scrutiny (NARDL or TARDL), the null hypothesis of a fully symmetric ARDL model (i.e., $H_0 : \left(\theta_1^+ = \theta_1^-\right)$ and $\left(\gamma_0^+ = \gamma_0^-\right)$ and $H_0 : \left({\theta_1}^{>Th} = {\theta_1}^{\leq Th}\right)$ and $\left({\gamma_0}^{>Th} = {\gamma_0}^{\leq Th}\right)$ respectively) is mildly rejected (particularly for PEG Nord). Therefore, we further explore whether partial restrictions in the form of either short- or long-run symmetry are supported by the data or not.

The test results in Table 3 convey similar findings for the two nonlinear models and the two markets. We observe low values for both the Wald statistics $W_{LR}$ and the expF statistics $expF_{SR}$ so that we cannot reject symmetry for the coefficients of the spark ratio in the long-run equation. In contrast, the highly significant statistics $W_{SR}$ and $expF_{SR}$ for the short-run dynamics unambiguously confirm the presence of asymmetry in the short run.

Altogether, these findings reveal: (i) that one should not overlook the presence of asymmetry in the short-run dynamics, and (ii) that it is preferable to opt for symmetric – and thus more parsimonious – specifications for the long-run coefficient associated with the spark-ratio variable. Accordingly, in what follows, our preferred NARDL and TARDL specifications include a symmetric long-run coefficient for that variable and possibly asymmetric ones in the short run.

**Table 3: Asymmetric tests**

| | **PEG Nord** | | **TRS** | |
|---|---|---|---|---|
| | stat. | *p*-value | stat. | *p*-value |
| NARDL | | | | |
| $W_{SR,LR}$ | 12.26 | (0.06)* | 13.90 | (<0.01)*** |
| $W_{LR}$ | 0.02 | (0.88) | 1.13 | (0.28) |
| $W_{SR}$ | 14.84 | (<0.01)*** | 22.28 | (<0.01)*** |
| TARDL | | | | |
| $expF_{SR,LR}$ | 5.32 | (0.09)* | 7.39 | (0.04)** |
| $expF_{LR}$ | 44.77 | (1) | 38.61 | (0.37) |
| $expF_{SR}$ | 6.94 | (<0.01)*** | 5.60 | (0.02)** |

Notes: The statistics $W_{SR,LR}$ and $expF_{SR,LR}$ test the null hypothesis of a fully symmetric ARDL model. The statistics $W_{LR}$ and $expF_{LR}$ are for the null hypothesis of long-run symmetry restrictions (respectively $\theta_1^+ = \theta_1^-$ for the NARDL model and ${\theta_1}^{>Th} = {\theta_1}^{\leq Th}$ for the TARDL model). The test statistics $W_{SR}$ and $expF_{SR}$ are for the null hypothesis of short-run symmetry restrictions (respectively $\gamma_i^+ = \gamma_i^-, \forall i$ for the NARDL model and ${\gamma_i}^{>Th} = {\gamma_i}^{\leq Th}, \forall i$ for the TARDL model). Asterisks indicate rejection of the null hypothesis at 0.10*, 0.05** and 0.01*** significance levels, respectively.

Our preferred specifications were subjected to several time series diagnostic tests (see Appendix B). The test results indicate that the models are properly specified. In particular, we find no indication of serial correlation in the residuals. This finding is important for the validity of our

[17] These values have been obtained using a time series block bootstrap procedure that comprises a total of 1,000 replications randomly drawn from the original data set using a 20-day time window.

estimates. In case of autocorrelated residuals, one could consider the approach in Lewbel and Ng (2005) that deals with the issue of nonstationarity (and nonlinearities).[18]

**Table 4: Estimation and test results for the PEG Nord market.**

| **Estimation results** | **NARDL** | | | **TARDL** | | |
|---|---|---|---|---|---|---|
| | Estimate | Std. Error | t-stat | Estimate | Std. Error | t-stat |
| *Short-run coefficients* | | | | | | |
| Constant | 5.42 | (0.54) | 10.03*** | 5.34 | (0.52) | 10.29*** |
| $Coint_{t-1}$ | -0.28 | (0.03) | -10.03*** | -0.28 | (0.02) | -10.29*** |
| $\Delta q_{t-1}$ | 0.08 | (0.04) | 2.01* | 0.09 | (0.04) | 2.24* |
| $\Delta s^{+}$ | 0.16 | (0.02) | 9.18 *** | | | |
| $\Delta s^{-}$ | 0.06 | (0.05) | 1.23*** | | | |
| $\Delta {s_t}^{>Th}$ | | | | 0.17 | (0.02) | 8.45*** |
| $\Delta {s_t}^{\leq Th}$ | | | | 0.11 | (0.03) | 4.20*** |
| $\Delta p^{G}$ | 0.15 | (0.13) | 1.14 | 0.16 | (0.13) | 1.22 |
| $\Delta M_1$:Jan | 0.23 | (0.07) | 3.35*** | 0.25 | (0.07) | 3.62*** |
| $\Delta M_2$:Feb | 0.06 | (0.09) | 0.67 | 0.08 | (0.09) | 0.87 |
| $\Delta M_3$:Mar | 0.09 | (0.10) | 0.82 | 0.10 | (0.10) | 1.00 |
| $\Delta M_4$:Apr | 0.07 | (0.10) | 0.61 | 0.08 | (0.11) | 0.79 |
| $\Delta M_5$:May | -0.18 | (0.10) | -1.69* | -0.15 | (0.10) | -1.47 |
| $\Delta M_6$:Jun | -0.17 | (0.10) | -1.63* | -0.15 | (0.10) | -1.48 |
| $\Delta M_7$:Jul | -0.17 | (0.09) | -1.68* | -0.16 | (0.09) | -1.65* |
| $\Delta M_8$:Au | -0.33 | (0.09) | -3.63*** | -0.32 | (0.09) | -3.60*** |
| $\Delta M_9$:Sep | -0.25 | (0.08) | -3.00*** | -0.24 | (0.08) | -3.00*** |
| $\Delta M_{10}$:Oct | -0.12 | (0.07) | -1.70 | -0.12 | (0.07) | -1.67* |
| $\Delta M_{11}$:Nov | -0.04 | (0.05) | -0.75 | -0.04 | (0.05) | -0.79 |
| *Long-run multipliers* | | | | | | |
| $s_t$ | 0.43 | (0.02) | 29.30*** | 0.54 | (0.01) | 39.9*** |
| $p_t^G$ | 0.34 | (0.07) | 4.74*** | 0.45 | (0.07) | 8.02*** |
| $M_1$:Jan | 0.15 | (0.07) | 2.09* | 0.24 | (0.06) | 3.78*** |
| $M_2$:Feb | 0.26 | (0.07) | 3.02*** | 0.36 | (0.06) | 5.66*** |
| $M_3$:Mar | 0.15 | (0.06) | 2.40* | 0.27 | (0.06) | 4.62*** |
| $M_4$:Apr | -0.18 | (0.05) | -3.63*** | -0.17 | (0.04) | -3.76*** |
| $M_5$:May | -0.61 | (0.05) | -12.63*** | -0.49 | (0.04) | -11.48*** |
| $M_6$:Jun | -0.87 | (0.05) | -18.6*** | -0.79 | (0.04) | -18.91*** |
| $M_7$:Jul | -0.99 | (0.05) | -20.68*** | -0.94 | (0.04) | -21.83*** |
| $M_8$:Aug | -1.05 | (0.05) | -22.74*** | -0.99 | (0.04) | -23.92*** |
| $M_9$:Sep | -0.79 | (0.05) | -16.13*** | -0.75 | (0.04) | -17.27*** |
| $M_{10}$:Oct | -0.35 | (0.05) | -6.95 *** | -0.34 | (0.06) | -7.52*** |
| $M_{11}$:Nov | -0.17 | (0.05) | -3.40*** | -0.16 | (0.04) | -3.58*** |
| **Bounds Test** | $F$-statistic | $I(0)$ | $I(1)$ | $F$-statistic | $I(0)$ | $I(1)$ |
| $W_{coint}$ | 10.12*** | 4.29 | 5.61 | 12.13*** | 4.29 | 5.61 |
| *Adjust* $R^2$ | 0.95 | | | 0.98 | | |
| Observations | 395 | | | 395 | | |

Notes: The table reports the estimation results obtained with the Δ-method for PEG Nord for the period April 1, 2015, to December 31, 2016. Asterisks indicate significance at 0.10*, 0.05** and 0.01*** levels, respectively. The table also reports the non-standard $F$-test of Pesaran et al. (2001) and the two critical bounds corresponding to the 1% critical level. If that test statistic is lower than the lower bound critical value, the test fails to reject the null of no cointegration. If the test statistic is higher than the upper critical value, the null of no cointegration among the variables is rejected. Asterisks indicate the rejection of the null hypothesis at the 0.01*** level.

The estimation results obtained with our preferred specifications are reported in Table 4 for PEG Nord and in Table 5 for the TRS market. One can note that all models exhibit comparable – and high – explanatory powers.

[18] In addition, following a referee's comment, we also tested the usefulness of including a proxy of economic activity, as in Agnolucci et al. (2017), but the latter was unsuccessful in fitting the data better.

**Table 5: Estimation and test results for the TRS market.**

| **Estimation results** | | **NARDL** | | | **TARDL** | |
|---|---|---|---|---|---|---|
| | Estimate | Std. Error | t-stat | Estimate | Std. Error | t-stat |
| *Short-run coefficients* | | | | | | |
| (Constant) | 6.44 | (0.51) | 12.53*** | 6.73 | (0.55) | 12.06*** |
| $Coint_{t-1}$ | -0.34 | (0.02) | -12.53*** | -0.34 | (0.03) | -12.06*** |
| $\Delta q_{t-1}$ | 0.09 | (0.04) | 2.20** | 0.12 | (0.04) | 2.81** |
| $\Delta s^{+}$ | 0.15 | (0.02) | 8.42*** | | | |
| $\Delta s^{-}$ | 0.07 | (0.05) | 1.54 | | | |
| $\Delta s_t^{>Th}$ | | | | 0.15 | (0.02) | 6.96*** |
| $\Delta s_t^{\leq Th}$ | | | | 0.11 | (0.02) | 4.56*** |
| $\Delta p_t^G$ | 0.32 | (0.13) | 2.40** | 0.30 | (0.13) | 2.18** |
| $\Delta M_1$:Jan | 0.18 | (0.07) | 2.57*** | 0.17 | (0.07) | 2.34** |
| $\Delta M_2$:Feb | 0.04 | (0.09) | 0.41 | 0.014 | (0.09) | 0.15 |
| $\Delta M_3$:Mar | 0.01 | (0.10) | 0.04 | -0.02 | (0.11) | -0.22 |
| $\Delta M_4$:Apr | 0.10 | (0.11) | 0.92 | 0.06 | (0.11) | 0.59 |
| $\Delta M_5$:May | -0.03 | (0.10) | -0.30 | -0.06 | (0.11) | -0.59 |
| $\Delta M_6$:Jun | -0.04 | (0.10) | -0.35 | -0.06 | (0.11) | -0.59 |
| $\Delta M_7$:Jul | -0.02 | (0.10) | -0.20* | -0.05 | (0.10) | -0.48 |
| $\Delta M_8$:Au | -0.23 | (0.09) | -2.42* | -0.24 | (0.09) | -2.56** |
| $\Delta M_9$:Sep | -0.17 | (0.08) | -1.98* | -0.18 | (0.09) | -2.13** |
| $\Delta M_{10}$:Oct | -0.08 | (0.07) | -1.09 | -0.09 | (0.07) | -1.28 |
| $\Delta M_{11}$:Nov | -0.08 | (0.05) | -1.68* | -0.10 | (0.05) | -1.98* |
| *Long-run multipliers* | | | | | | |
| $s_t$ | 0.51 | (0.02) | 34.24*** | 0.51 | (0.01) | 40.90*** |
| $p^G$ | 0.33 | (0.07) | 4.39*** | 0.13 | (0.07) | 1.83* |
| $M_1$:Jan | 0.09 | (0.07) | 1.37 | 0.06 | (1.00) | 0.31 |
| $M_2$:Feb | 0.23 | (0.07) | 3.50*** | 0.17 | (0.06) | 2.68*** |
| $M_3$:Mar | 0.09 | (0.06) | 1.46 | 0.04 | (0.06) | 0.70 |
| $M_4$:Apr | -0.20 | (0.05) | -4.31*** | -0.39 | (0.04) | -8.43*** |
| $M_5$:May | -0.58 | (0.05) | -12.87*** | -0.65 | (0.04) | -14.66*** |
| $M_6$:Jun | -0.84 | (0.04) | -19.08*** | -0.87 | (0.04) | -20.51*** |
| $M_7$:Jul | -0.99 | (0.05) | -21.81*** | -1.01 | (0.04) | -23.03*** |
| $M_8$:Aug | -1.12 | (0.04) | -25.74*** | -1.14 | (0.04) | -27.11*** |
| $M_9$:Sep | -0.81 | (0.05) | -17.73*** | -0.83 | (0.04) | -18.82*** |
| $M_{10}$:Oct | -0.45 | (0.05) | -9.43*** | -0.47 | (0.04) | -9.96*** |
| $M_{11}$:Nov | -0.23 | (0.05) | -4.73*** | -0.23 | (0.04) | -4.83*** |
| **Bounds Test** | $F$-statistic | $I(0)$ | $I(1)$ | $F$-statistic | $I(0)$ | $I(1)$ |
| $W_{coint}$ | 16.58*** | 4.29 | 5.61 | 18.46*** | 4.29 | 5.61 |
| *Adjust* $R^2$ | 0.97 | | | 0.97 | | |
| Observations | 395 | | | 395 | | |

Notes: The table reports the estimation results obtained with the Δ-method for TRS for the period April 1, 2015, to December 31, 2016. Asterisks indicate significance at 0.10*, 0.05** and 0.01*** levels, respectively. The table also reports the non-standard $F$-test of Pesaran et al. (2001) and the two critical bounds corresponding to the 1% critical level. If that test statistic is lower than the lower bound critical value, the test fails to reject the null of no cointegration. If the test statistic is higher than the upper critical value, the null of no cointegration among the variables is rejected. Asterisks indicate the rejection of the null hypothesis at the 0.01*** level.

From the estimates, three relevant series of findings can be derived. The first series of findings concern the presence of cointegration. To check the presence of a long-run relationship, we follow the bounds test procedure proposed by Pesaran et al. (2001). In all cases, the $F$-test statistics reported in the tables systematically exceed the upper bounds critical values at the 1% level, which indicates the presence of cointegration among the daily consumption of natural gas, the price of natural gas, and the relative price of electricity and natural gas. Furthermore, the estimated feedback coefficient in the short-run dynamics is negative (as expected) for all models.

Our second series of findings focuses on the estimated long-run multipliers. In both markets, the multipliers of the spark ratio variable are, as expected, positive, and their estimated values are statistically significant at the 1% level. Regarding the gas price coefficient (that is, $\beta_2$), it is important to keep in mind that this multiplier does not fully capture the long-run influence of the price of natural gas on daily gas demand because that variable is also present in the spark ratio. Hence, the long-run multiplier that captures the long-run reaction of daily gas demand to the price of natural gas is given by the difference $\beta_2 - \beta_1$. We report these differences in Table 6. The obtained values are negative, as expected, but have very low magnitudes (particularly for PEG Nord). This finding indicates that, in the long run, the daily gas consumption is not very sensitive to the gas price. As one may wonder whether that consumption could be completely unaffected by that price in the long run, we also report the test results for the null hypotheses $\beta_2 - \beta_1 = 0$ in Table 6. For each market, we found that this hypothesis is firmly rejected by the data for the NARDL and the TARDL models. Thus, in the long run, the gas price has a small but significant impact on observed consumption levels.

**Table 6: Long-run reaction of natural gas demand to the price of natural gas.**

| | PEG Nord | | | TRS | | |
|---|---|---|---|---|---|---|
| | value | stat. | p-value | value | stat. | p-value |
| **NARDL** | | | | | | |
| $\epsilon_{Gas}^{LR} := (\beta_2 - \beta_1)$ | -0.09 | | | -0.18 | | |
| $H_0 : (\beta_1 - \beta_2 = 0)$ | | 15.44 | <0.01*** | | 151.70 | <0.01*** |
| **TARDL** | | | | | | |
| $\epsilon_{Gas}^{LR} := (\beta_2 - \beta_1)$ | -0.09 | | | -0.38 | | |
| $H_0 : (\beta_1 - \beta_2 = 0)$ | | 166.68 | <0.01*** | | 174.17 | <0.01*** |

Notes: The table reports the difference $(\beta_2 - \beta_1)$ computed from the estimation results. It also reports the result of a Wald test for NARDL model and the expF test statistic of Andrews and Ploberger (1994) for the null hypothesis of zero long-run reaction of demand to gas price. Asterisks indicate rejection of the null at 0.01*** level.

Lastly, with regards to short-run dynamics, we observe evidence of asymmetry in the estimated coefficients of the TARDL models. In the two markets, we observe that the instantaneous impact of a change in the spark ratio on gas demand variations is positive and highly significant and that the magnitude of that impact is substantially larger when the spark ratio is larger than the threshold value. These results are fully consistent with the expectations derived from the technological considerations discussed in section 2. For the NARDL model, the asymmetric coefficients obtained in the two markets are positive. In both markets, a positive change of the spark ratio has a more pronounced instantaneous impact on natural gas demand variations than a negative change.

Overall, our results provide evidence of a positive (respectively negative) long-run relationship between natural gas consumption and electricity (respectively natural gas) price and of short-run dynamics that are clearly asymmetric, as was expected from the technology considerations above.

### 5.2 Out-of-sample analysis

As the NARDL and TARDL models are not nested, it is easy to decide which model should be used using standard tests. To compare the predictive performance of these two models, we now use the estimates from our preferred NARDL and TARDL specifications to generate out-of-sample forecasts for the evaluation period.[19]

To gain insight on their predictive accuracy, we use three benchmarks. The first is given by a simple AR(1) model, which is a standard reference commonly used in the forecasting literature. The second benchmark is the linear ARDL model presented above. That benchmark should provide useful insights on the performance gains resulting from supplementing an ARDL specification with nonlinear components. Lastly, our third benchmark is formed by the next day's demand forecast

[19] Recall that this period has 229 observations, which is large enough to compare the predictions obtained with different models.

published each day at 5pm – i.e., directly after the closure of the day-ahead market when closing prices are already known – by infrastructure operators. In the PEG Nord market, GRTGaz is the unique TSO and its forecast can be readily used as a benchmark. In the TRS zone, there are two pipeline operators (GRTGaz and Terėga) that serve different territories. The forecast of the next day's consumption in the southern zone is thus obtained by summing up the two individual forecasts issued by these two TSOs on their respective websites. To ease comparisons (and obtain error figures measured in energy units), the results presented in this subsection are based on the exponentiated (detransformed) predicted values of the next day's consumption.

Table 7 reports a common measure of accuracy of the predicted values of the next day's consumption: the Root Mean Square Error (RMSE) which is measured in GWh over the entire evaluation period. As the consumption of natural gas is substantially larger during the winter season than during the summer, when gas-fueled heating systems are turned off, this table also reports the prediction errors obtained for these two specific seasons in our out-of-sample validation period.

**Table 7: Prediction error statistics values**

| | | **NARDL** | **TARDL** | **AR(1)** | **ARDL** | **TSO** |
|---|---|---|---|---|---|---|
| **Validation period** | PEG Nord | 26.18 | **11.77** | 55.06 | 39.15 | 47.14 |
| | TRS | 23.59 | **3.77** | 59.70 | 17.01 | 23.76 |
| **Summer subperiod** | PEG Nord | 31.04 | **14.75** | 69.43 | 37.76 | 31.43 |
| | TRS | 24.77 | **2.78** | 75.14 | 17.08 | 19.94 |
| **Winter subperiod** | PEG Nord | 20.35 | **8.20** | 70.12 | 40.70 | 65.57 |
| | TRS | 10.51 | **5.32** | 33.07 | 16.93 | 27.59 |

Notes: The RMSE, measured in GWh, is successively evaluated: for the entire evaluation period (January 1, 2017, to February 2, 2018), for the "gas winter" subperiod gathering all the observations between Nov. 1 and March 31 and for the "gas summer" subperiod (i.e., all the observations between April 1 and Oct. 31). The figures in bold indicate that this model has the lowest error value among the three models.

Overall, we note that the results obtained for PEG Nord and TRS during the out-of-sample forecast validation period and during the two subperiods are qualitatively similar. In most cases, the RMSE statistics obtained for the PEG Nord market are larger than the ones obtained for the TRS market, which is consistent with the relative sizes of these two markets.

An examination of the performances of our three benchmarks conveys the following findings. Observe that, as can be expected, whatever the market and the period under scrutiny, the forecasts issued by the TSOs and the ones obtained with a linear ARDL model are substantially more accurate than the ones emanating from a simple AR(1) specification. This finding confirms that a simple autoregressive structure is not adapted to predict the next day's load. Interestingly, one can also note that, in all cases, the RMSE statistics obtained with a linear ARDL model are markedly lower than the ones computed from the forecasts published by the TSOs. Given the excellent performance of the linear ARDL model, it is not surprising to observe that the two nonlinear extensions of that model also compete very favorably with the TSO benchmark. Overall, these findings suggest that the forecasting performance of the tools currently used by infrastructure operators remains poor, which justifies the current heightened regulatory attention paid to that issue.

We now examine the performance of our preferred nonlinear models. For both markets, we observe that our TARDL model systematically provides the lowest RMSE figures and outperforms both the three benchmarks and the NARDL model. In contrast, one can note that a nonlinear specification based on a positive and negative partial sum decomposition of the spark ratio hardly surpasses the ARDL and the TSO models. Indeed, the prediction errors obtained with the NARDL model are either commensurate or larger than these two benchmarks (cf. Table 7: the RMSE figures obtained for PEG Nord during the summer and the ones for TRS during the entire validation period). These findings suggest that one should prefer the TARDL model to the NARDL model and thus use that model to assess the impacts that gas and electricity prices have on daily gas demand.

In order to statistically compare the predictive accuracy of our models, we also perform the test procedure proposed by Diebold and Mariano (1995) and extended in Harvey et al. (1997). The null hypothesis is that the two forecasting models have the same forecast accuracy and the alternative hypothesis is that a baseline model (labeled B) is more accurate than the reference model (labeled A).[20] A negative value for the Diebold-Mariano statistic indicates that the baseline model (model B) delivers significantly better forecasts. The Diebold-Mariano statistics reported in Table 8 make use of the TARDL specification as a reference model. These test results confirm the preceding findings as the TARDL model: (i) performs significantly better than the AR(1), the ARDL and the NARDL model in all cases, and (ii) is also significantly more accurate than the forecasts issued by infrastructure operators during the winter subperiod.

**Table 8: Diebold-Mariano test statistics**

| | | NARDL | TARDL | AR(1) | ARDL | TSO |
|---|---|---|---|---|---|---|
| Validation period | PEG NORD | 38.91*** | - | 12.95*** | 15.79*** | 1.31* |
| | TRS | 28.44*** | - | 13.6*** | 21.20*** | 1.43* |
| Summer subperiod | PEG NORD | 31.47*** | - | 9.81*** | 15.19*** | 0.73 |
| | TRS | 22.70*** | - | 10.37*** | 18.54*** | 1.15* |
| Winter subperiod | PEG NORD | 27.72*** | - | 8.99*** | 7.84*** | 6.49*** |
| | TRS | 5.58*** | - | 3.47*** | 5.20*** | 7.33*** |

Notes: This table reports the Diebold and Mariano (1995) (hereafter DM) test of the null hypothesis of no difference in the accuracy of the compared forecasts. The TARDL forecasts are used as a baseline model. The test statistics are evaluated: for the entire evaluation period (January 1, 2017, to February 2, 2018), for the "gas winter" subperiod gathering all the observations between Nov. 1 and March 31 and for the "gas summer" subperiod (i.e., all the observations between April 1 and Oct. 31). Our implementation of that test uses a loss differential defined as the difference between squared forecast errors. Numbers are DM tests statistics. Asterisks indicate rejection of the null hypothesis at 0.10*, 0.05** and 0.01*** significance levels, respectively.

This excellent performance of the TARDL model emphasizes the pivotal role of the spark ratio for the purpose of forecasting natural gas demand. Our findings confirm the intuition that allowing for a possibly non-zero threshold does help in forecasting the natural gas demand as natural gas consumption unequivocally depends on the relative price of electricity to natural gas. Moreover, the good performance of the TARDL relative to the NARDL shows that while the presence of nonlinearity is clearly validated by the data, accounting for this feature alone is not sufficient to outperform the forecasting accuracy of TSOs. We thus conclude that the spark ratio contains much information pertaining to natural gas demand.

## 6. CONCLUDING REMARKS

Ideally, the prices formed at spot markets for natural gas should reflect the processed information of a large number of market participants. However, anxiety over the liquidity and maturity of some of the European gas hubs has emerged in recent years. The question examined in this paper is therefore whether the information in these day-ahead prices is substantial enough to provide reasonably accurate predictions of the next day's consumption of natural gas.

To answer this question, we examine for the first time the daily interactions between day-ahead prices and daily consumption for two French hubs over the period 2015–2018. Importantly, given the unit-root property of the time series and technological considerations related to the dispatching of gas-fired power plants in the electricity sector, we propose a new nonlinear extension to the ARDL model that permits the response of the natural gas demand to vary with respect to the spark ratio depending on whether that ratio attains or not a certain level: the TARDL model. Our results have shown that its forecasting performance outperforms both those of the usual nonlinear ARDL model – a model that considers the possibly different impacts of positive and negative variations of the spark

[20]The Diebold-Mariano test is known to have poor performance in the case of nested models but this limitation is not a concern in the present application because the NARDL and TARDL models are not nested.

ratio on gas demand – or those of the tools routinely used by infrastructure operators.

On the whole, our findings have important policy implications, particularly with respect to the quality of the demand forecasts produced by infrastructure operators. The accuracy of these predictions is now emerging as a source of regulatory concern and has recently motivated the adoption of dedicated incentive schemes in some countries (the UK, Italy). Indeed, it has very important implications for: (i) the cost-efficient operation of the gas transportation network, (ii) the quality of the information given to infrastructure users for within-day flow balancing purposes, and (iii) the possibility to use existing gas pipeline infrastructures to supply short-term, linepack-based, flexibility services to a renewable-dominated power sector. Our results suggest that accounting for the information contained in day-ahead prices represents a promising avenue to improve the performance of these demand forecasts. That said, the fact that a relatively simple econometric model, based solely on publicly available spot prices, provides a more accurate forecasting procedure certainly points to some deficiencies in the operators' forecasting activities.

Beyond predictive soundness, our research also gives rise to important empirical findings on the economic determinants of the daily demand for natural gas in France. Our results confirm the existence of a long-run relation between the observed demand levels and the spot prices and indicate that this long-run relation is consistent with the conjectures derived from standard microeconomics. Our results indicate that, in the long run, the marginal impact of natural gas price on daily gas consumption is negative and that we are dealing with a very small price-responsive demand. In the long run, the daily gas demand is also positively affected by the price of electricity. Regarding the short run, our empirical analysis documents the nonlinear nature of the short-run interactions between the observed demand variations and the relative price of electricity to gas. As expected, a positive variation of the spark ratio instantaneously leads to a demand increase but it should be emphasized that the results gained with the TARDL model reveal that its impact is larger when the spark ratio is above a given threshold. Critically, we observe that the value of that threshold is commensurate with the heat rate of the gas-fired power plants.

Notwithstanding the value of our findings, it should be borne in mind that our empirical analysis can be extended in several directions. First of all, it could be interesting to check whether similar results are obtained when applying that methodology to examine the situation prevailing in other European markets. Another strand of research could be the application of that methodology to sectorally disaggregated datasets. Indeed, all consumers do not make optimal demand decisions under the same constraints and do not necessarily demand the same services from natural gas. Hence, the determinants of natural gas demand might differ among different economic sectors. Should disaggregated daily consumption data become publicly available for the various sectors, an analysis of the determinant of the daily demand for natural gas observed in each sector could offer an interesting avenue for future research. Lastly, as noted in footnote 11, future research could also consider a structural approach *à la* Lewbel and Ng (2005) to properly estimate demand elasticities and verify whether their magnitudes are commensurate with the values of the marginal impacts discussed in the present paper.

## APPENDIX A: DESCRIPTIVE STATISTICS

**Table 9: Descriptive statistics for the price of natural gas $p_t^G$, the electricity price $p_t^E$ and the consumption of natural gas $q_t$.**

| | **PEG Nord** | | **TRS** | | |
|---|---|---|---|---|---|
| | $p_t^G$ | $q_t$ | $p_t^G$ | $q_t$ | $p_t^E$ |
| Mean | 16.61 | 814.8 | 17.55 | 401.8 | 44.50 |
| Min. | 10.65 | 299.4 | 11.38 | 135.5 | 15.50 |
| Max. | 23.16 | 190.9 | 25.40 | 907.5 | 275.00 |
| Std. dev. | 3.32 | 412.947 | 3.19 | 209.52 | 21.44 |
| Skewness | 0.01 | 0.67 | 0.09 | 0.65 | 4.49 |
| Kurtosis | 1.58 | 2.24 | 1.76 | 2.08 | 39.49 |
| JB | 33.28*** | 31.02*** | 25.88*** | 62.31*** | 24 147*** |

Notes: The table presents: the mean, min-max, the standard deviation, the skewness, the kurtosis, and the Jarque-Bera test statistics for the series in levels during the model estimation period. Asterisks indicate rejection of the null hypothesis of normality at 0.01*** significance level. For readability, consumption data are measured in GWh.

## APPENDIX B: DIAGNOSTICS TESTS

The models were subjected to several diagnostic tests to detect: the presence serial correlation, the presence of heteroscedasticity, and a possible functional misspecification. The test results are detailed in Table 10.

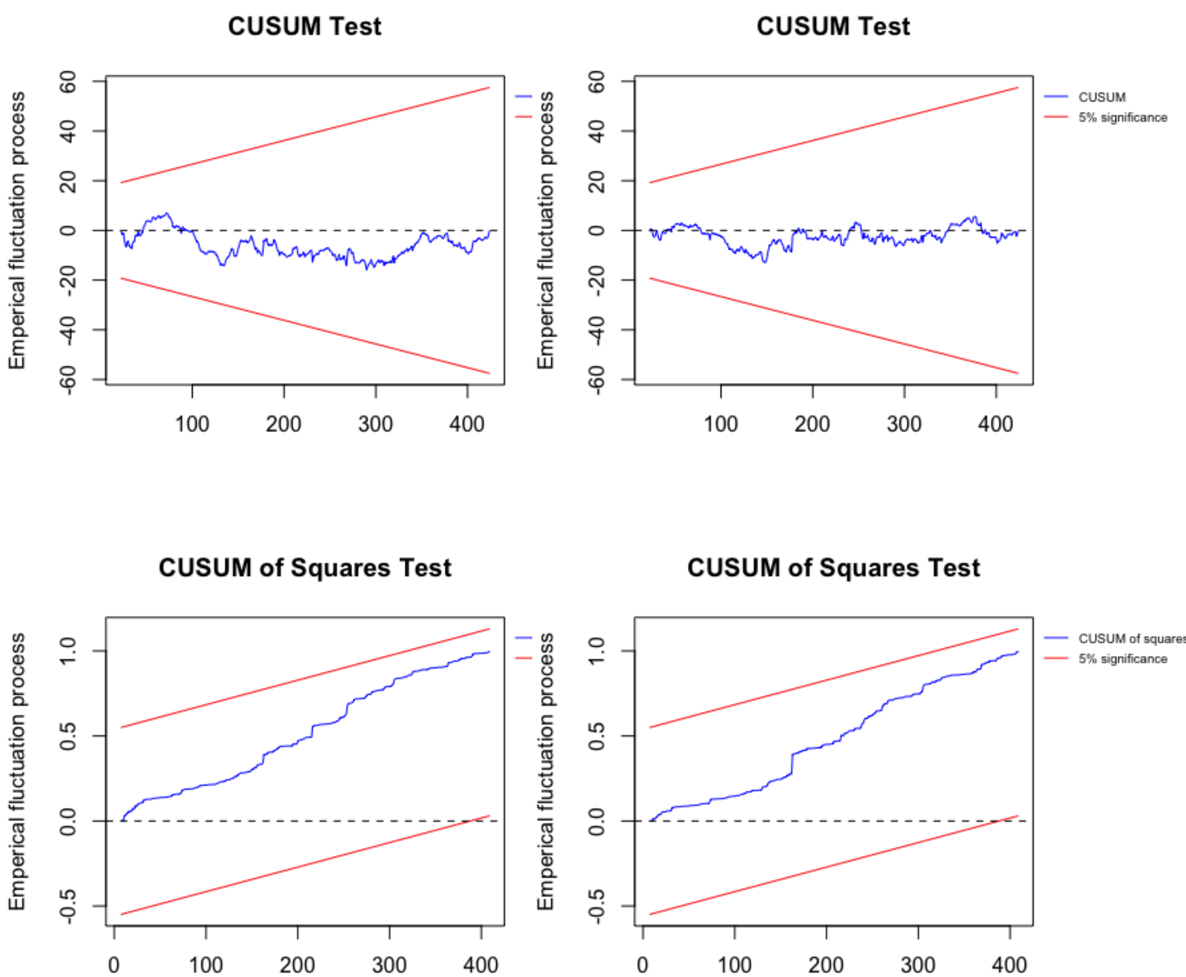
**Table 10: Diagnostics test.**

| | **NARDL model** | | **TARDL model** | |
|---|---|---|---|---|
| | Statistic | $p$-value | Statistic | $p$-value |
| **PEG Nord** | | | | |
| A: Serial correlation L.-B.(5) | 1.70 | 0.19 | 1.47 | 0.22 |
| B: Heteroscedasticity ARCH(5) | 0.46 | 0.99 | 0.59 | 0.98 |
| C: Functional form RESET(2,3) | 1.26 | 0.26 | 0.87 | 0.35 |
| D: KS-GMM. | - | 0.22 | - | 0.57 |
| **TRS** | | | | |
| A: Serial correlation L.-B.(5) | 0.53 | 0.46 | 0.24 | 0.62 |
| B: Heteroscedasticity ARCH(5) | 0.05 | 1.00 | 0.04 | 1.00 |
| C: Functional form RESET(2,3) | 1.85 | 0.15 | 1.00 | 0.36 |
| D: KS-GMM | - | 0.67 | - | 0.78 |

Notes: L.-B.(5) is the Ljung-Box Q-statistic for the null hypothesis of no serial correlation up to the 5th order. ARCH(5) is the LM-test for the absence of autoregressive conditional heteroscedasticity with 5 lags. RESET(2,3) is the Ramsey Regression Equation Specification Error Test for model misspecification based on the square and cube of fitted values. KS-GMM is the Kolmogorov-Smirnov test for the null hypothesis of a Gaussian Mixture Model based on Smirnov (1948). As the KS-GMM estimation is based on a bootstrap procedure, we solely report the p-value. Asterisks indicate rejection of the null hypothesis at 0.10*, 0.05** and 0.01*** significance levels, respectively.

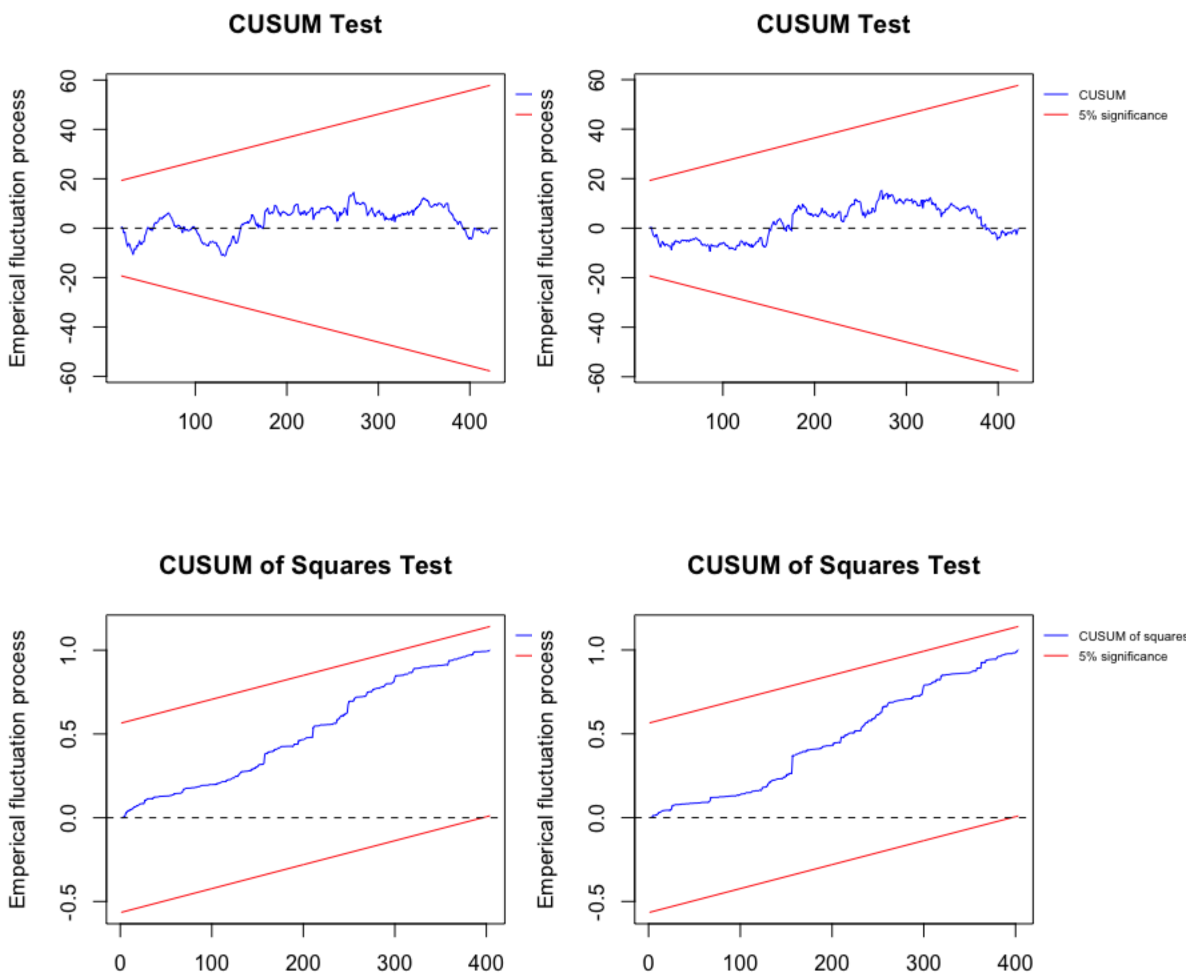
The autoregressive structures of the estimated models are statistically adequate since there is no evidence of residual autocorrelation (see the Ljung-Box statistic for up to the fifth order). The ARCH test confirms the absence of conditional heteroscedasticity. The Ramsey RESET test for model misspecification based on the powers of the fitted value of consumption shows no sign of functional misspecification for both models in each market. To examine the temporal stability of the estimated coefficients, we have also evaluated the cumulative sum of recursive residuals (CUSUM) and CUSUM of square test statistics. The test results are presented in Appendix A (see Fig. 2 and Fig. 3). In all cases, the test statistics are well within the 5% critical bounds which indicates that there is no evidence of parameter instability in any of the models over the estimation period. The results gained from Kolmogorov-Smirnov tests indicate that, in all cases, a Gaussian Mixture Model adequately describes the distributional properties of the residual series.[21]

**Figure 2: CUSUM & CUSUMQ test with the NARDL models for PEG Nord (on the left) and TRS (on the right)**




[21]From a technical perspective, these distributional properties are compatible with the theroretical conditions in Pesaran and Shin (1999) for the validity of the model estimates as Pesaran and Shin (1999) simply stipulate that the error terms should be *"serially uncorrelated disturbances with zero means and constant variance-covariances."*

**Figure 3: CUSUM & CUSUMQ tests with the TARDL models for PEG Nord (on the left) and TRS (on the right)**



## References

Abada, I., Massol, O., 2011. Security of supply and retail competition in the european gas market: some model-based insights. Energy Policy 39, 4077–4088.

ACER-ENTSOG, 2014. ACER-ENTSOG Report on the early implementation of the Balancing Network Code. Technical Report. European network of transmission system operators for gas: Brussels.

Adeyemi, O.I., Hunt, L.C., 2014. Accounting for asymmetric price responses and underlying energy demand trends in oecd industrial energy demand. Energy Economics 45, 435 – 444.

Agnolucci, P., De Lipsis, V., Arvanitopoulos, T., 2017. "modelling uk sub-sector industrial energy demand". Energy Economics 67, 366–374.

Andrews, D., Ploberger, W., 1994. Optimal tests when a nuisance parameter is present only under the alternative. Econometrica 62, 1383–1414.

Arvesen, o., Medbø, V., Fleten, S.E., Tomasgard, A., Westgaard, S., 2013. Linepack storage valuation under price uncertainty. Energy 52, 155–164.

Bagnai, A., Ospina, C.A.M., 2018. Asymmetries, outliers and structural stability in the us gasoline market. Energy Economics 69, 250 – 260.

Bordignon, S., Bunn, D.W., Lisi, F., Nan, F., 2013. Combining day-ahead forecasts for british electricity prices. Energy Economics 35, 88 – 103. Quantitative Analysis of Energy Markets.

Bunn, D.W., Gianfreda, A., 2010. Integration and shock transmissions across european electricity forward markets. Energy Economics 32, 278 – 291.

Cochrane, J.H., 1991. Production-based asset pricing and the link between stock returns and economic fluctuations. The Journal of Finance 46, 209–237.

CRE, 2012. Délibération de la Commission de Régulation de l'Énergie du 13 décembre 2012 portant décision sur le tarif d'utilisation des réseaux de transport de gaz naturel. Technical Report. Commission de Régulation de l'Énergie: Paris.

Dergiades, T., Tsoulfidis, L., 2008. Estimating residential demand for electricity in the united states, 1965–2006. Energy Economics 30, 2722 – 2730.

Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. Journal of Business & Economic Statistics 13, 253–263.

Engle, R., Granger, C., 1987. Co-integration and error correction: Representation, estimation, and testing. Econometrica 55, 251–76.

ENTSOG, 2017. Balancing network code. Implementation and effect monitoring report. Technical Report. European Network of Transmission System Operators for Gas: Brussels.

Forbes, K.F., Zampelli, E.M., 2014. Do day-ahead electricity prices reflect economic fundamentals? evidence from the california iso. The Energy Journal 35, 129–144.

Geweke, J., 1978. Temporal aggregation in the multiple regression model. Econometrica 46, 643–661.

Gianfreda, A., Bunn, D., 2018. A stochastic latent moment model for electricity price formation. Operations Research 66, 1189–1203.

Giulietti, M., Grossi, L., Waterson, M., 2012. A rough analysis: valuing gas storage. The Energy Journal 33, 119–141.

Gopalakrishnan, A., Biegler, L.T., 2013. Economic nonlinear model predictive control for periodic optimal operation of gas pipeline networks. Computers  Chemical Engineering 52, 90 – 99.

Green, R., Vasilakos, N., 2010. Market behaviour with large amounts of intermittent generation. Energy Policy 38, 3211 – 3220. Large-scale wind power in electricity markets with Regular Papers.

Greenwood-Nimmo, M., 2011. The asymmetric ardl model with multiple unknown threshold decompositions : an application to the phillips curve in canada.

Grid, N., 2018. National Grid Gas (NTS) System Operator Incentives. 2018/19 Supporting Information. Technical Report. National Grid UK.

Hallack, M., Vazquez, M., 2013. European union regulation of gas transmission services: Challenges in the allocation of network resources through entry/exit schemes. Utilities Policy 25, 23 – 32.

Hansen, B.E., 1996. Inference when a nuisance parameter is not identified under the null hypothesis. Econometrica 64, 413–430.

Harvey, D., Leybourne, S., Newbold, P., 1997. Testing the equality of prediction mean squared errors. International Journal of Forecasting 13, 281 – 291.

Heather, P., Petrovich, B., 2017. European traded gas hubs: an updated analysis on liquidity, maturity and barriers to market integration. Technical Report. Oxford Institute for Energy Studies.

Hendry, D.F., Juselius, K., 2000. Explaining cointegration analysis: part 1. The Energy Journal 21, 1–42.

Hoel, M., Strom, S., 1987. Supply security and import diversification of natural gas. Technical Report. R. Golombek, J. Vislie, M. Hoel (Eds.), Natural Gas Markets and Contracts", North-Holland.

Hunt, L.C., Ninomiya, Y., 2003. Unravelling trends and seasonality: a structural time series analysis of transport oil demand in the uk and japan. The Energy Journal 24, 63–96.

Johansen, S., Juselius, K., 1990. Maximum likelihood estimation and inference on cointegration - with applications to demand for money. Oxford Bulletin of Economics and Statistics 52, 169–210.

Karakatsani, N.V., Bunn, D.W., 2008. Forecasting electricity prices: the impact of fundamentals and time-varying coefficients. International Journal of Forecasting 24, 764 – 785. Energy Forecasting.

Keyaerts, N., Hallack, M., Glachant, J.M., D'haeseleer, W., 2011. Gas market distorting effects of imbalanced gas balancing rules: Inefficient regulation of pipeline flexibility. Energy Policy 39, 865 – 876.

Knittel, C.R., Roberts, M.R., 2005. An empirical examination of restructured electricity prices. Energy Economics 27, 791 – 817.

Lewbel, A., Ng, S., 2005. Demand systems with nonstationary prices. The Review of Economics and Statistics 87, 479–494.

Lucia, J.J., Schwartz, E.S., 2002. Electricity prices and power derivatives: evidence from the nordic power exchange. Review of Derivatives Research 5, 5–50.

Manne, A.S., Roland, K., Stephan, G., 1986. Security of supply in the western european market for natural gas. Energy Policy 14, 52 – 64.

Markandya, A., Pemberton, M., 2010. Energy security, energy modelling and uncertainty. Energy Policy 38, 1609 – 1613.

de Menezes, L.M., Houllier, M.A., 2016. Reassessing the integration of european electricity markets: A fractional cointegration analysis. Energy Economics 53, 132 – 150.

Miriello, C., Polo, M., 2015. The development of gas hubs in europe. Energy Policy 84, 177 – 190.

Pal, D., Mitra, S.K., 2015. Asymmetric impact of crude price on oil product pricing in the united states: An application of multiple threshold nonlinear autoregressive distributed lag model. Economic Modelling 51, 436 – 443.

Pesaran, M.H., Shin, Y., 1999. An autoregressive distributed-lag modelling approach to cointegration analysis. Cambridge University Press. Econometric Society Monographs, p. 371–413.

Pesaran, M.H., Shin, Y., Smith, R.J., 2001. Bounds testing approaches to the analysis of level relationships. Journal of Applied Econometrics 16, 289–326.

Qadrdan, M., Chaudry, M., Wu, J., Jenkins, N., Ekanayake, J., 2010. Impact of a large penetration of wind generation on the gb gas network. Energy Policy 38, 5684 – 5695. The socio-economic transition towards a hydrogen economy - findings from European research, with regular papers.

Ramanathan, R., Engle, R., Granger, C.W., Vahid-Araghi, F., Brace, C., 1997. Short-run forecasts of electricity loads and peaks. International Journal of Forecasting 13, 161 – 174.

Renou-Maissant, P., 2012. Toward the integration of european natural gas markets:a time-varying approach. Energy Policy 51, 779 – 790. Renewable Energy in China.

Serletis, A., Herbert, J., 1999. The message in north american energy prices. Energy Economics 21, 471 – 483.

Shin, Y., Yu, B., Greenwood-Nimmo, M., 2014. Modelling asymmetric cointegration and dynamic multipliers in a nonlinear ARDL framework. Springer. pp. 281–314.

Smirnov, N., 1948. Table for estimating the goodness of fit of empirical distributions. The Annals of Mathematical Statistics 19, 279–281.

Sun, X., Huang, D., Wu, G., 2012. The current state of offshore wind energy technology development. Energy 41, 298 – 312. 23rd International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems, ECOS 2010.

Thoenes, S., 2014. Understanding the determinants of electricity prices and the impact of the German nuclear moratorium in 2011. The Energy Journal 0.

Tran, T.H., French, S., Ashman, R., Kent, E., 2018. Linepack planning models for gas transmission network under uncertainty. European Journal of Operational Research 268, 688 – 702.

Vany, A.D., Walls, W.D., 1993. Pipeline access and market integration in the natural gas industry: evidence from cointegration tests. The Energy Journal 14, 1–20.

Worthington, A., Kay-Spratley, A., Higgs, H., 2005. Transmission of prices and price volatility in australian electricity spot markets: a multivariate garch analysis. Energy Economics 27, 337 – 350. Special Issue on Electricity Markets.