IRAFEM
International Review of Applied Finance, Economics and Management

International Review of Applied Finance, Economics, and Management *IRAFEM VOL. 1 Issue 2 (2025)*

Non-linear Dynamics of Household Natural Gas Use: Environmental and Demographic Challenges in the Energy Transition

Mohamed Jaouad Malzi[1*], Angeliki N. Menegaki[2], Said Hanchane[3], Aziz Ettahir[4]

[1] *Cadi Ayyad University, Marrakesh, Morocco*
[2] *Agricultural University of Athens, Greece*
[3] *Mohammed VI Polytechnic University (UM6P), Benguerir, Morocco*
[4] *Mohammed V University, Rabat, Morocco*

*Corresponding author: j.malzi@uca.ac.ma

Abstract

Nonlinearities cause fuzziness and vagueness, characteristics that energy policy makers wish to downsize if not eliminate. This study investigates the existence of nonlinear relationship between gas demand and its main determinants. To this end, the study applies nonparametric and semi-parametric panel data model on a large panel of OECD countries over the period 1980-2024. The analysis demonstrates the existence of a non-linear relationship between gas demand and its main determinants. The inverted-U hypothesis is totally confirmed in the case of population density and urbanization. This study also finds that the negative impact of the environmental policy stringency on natural gas use is very strong in the long term. Finally, a positive relationship exists between elderly population and the use of gas as the curve is revealing a positive trend. Based on the empirical evidence this analysis emphasizes on the proper implementation of the undertaken environmental and demographic policies towards the sustainable development agendas.

Keywords: Residential Natural Gas Use; Environmental Policy; Nonparametric model, Kuznets curve, energy transition.

1. Introduction

The demand for natural gas was often analyzed using classical parametric models considering a linear linkage between gas demand and its main determinants (Verhulst.1950, Balestra and Nerlove. 1966, Green. 1987, Asche et al. 2008, Alberini et al. 2011, Wadud et al. 2011 and Zeng et al. 2018). The most commonly used natural gas demand model in the literature is the log-linear specification, where the demand for natural gas is considered as a linear function of real income and gas prices. Other explanatory variables may also be involved such as price of substitutes, temperature, household and demographic characteristics, (Andersen et al. 2011, Wadud et al. 2011, Yu et al. 2014, Ota et al. 2018, Malzi et al. 2019; 2020). Research studies so far have typically used the logged form of variables, while the coefficients on gas price and income are straightforward interpretations of price and income elasticities. The main issue of the log-linear model is the assumption of a strict linear relationship between natural gas demand and all exogenous variables. Hence, the demand elasticities are considered to be homogeneous across all groups.

However, this assumption is far from being realistic, on the contrary is rather questionable since price, income (and all other explanatory variables) elasticities of natural gas demand, tend to be quite heterogeneous across different countries and regions considering this heterogeneity departing from weather, natural resources, energy regulation and environmental policies. Another popular specification of natural gas demand analysis in the literature is the trans-log specification which incorporates quadratic terms of natural gas price and income along with an interaction term of the two explanatory variables. Within this model, income and price elasticities may be derived as linear functions of real income and gas prices. As a result, instead of being constant as supposed in the log-linear analysis, the demand elasticities analyzed from the trans-log model may vary with real income and gas prices.

In fact, not only income and natural gas prices are the main origin of heterogeneity in explaining demand elasticities. Other factors apart from income and price are leaning to distort the elasticities of natural gas demand in different contexts. For example, to the same increase in gas prices, households may behave differently within a group of countries from another group, depending on the electric appliances available in the country, the possibilities and capacity of fuel switching and the standard of living. It depends also on the demographic structure and population characteristics of the country, the urbanization rate and other socio-economic factors (Sardianou, 2007). Besides, in the same country, urban and rural residents react differently to the same increase/decrease in gas prices. This heterogeneity can be captured in the trans-log model by adding an interaction term between natural gas prices and household location. However, by appending additional interaction terms to the model in order to control for other factors, more degrees of freedom would have to be given up and that will undoubtedly bias the estimation. Accordingly, finding a more flexible specification form to estimate natural gas demand is mandatory.

Moreover, economic phenomena are volatile and complex. Thus, the relationships among different economic variables are complex. However, the majority of previous studies have been based on the linear assumption, resulting in the employment of linear specifications to question the relationships among economic variables. Literally, there exist many nonlinear associations among economic variables (Granger and Terasvirta, 1993). Overlooking the existence of non-linear associations and handling linear models to investigate the relationships among economic variables will disturb the robustness of the specification and cause biased parameter estimation. In this paper, in order to capture heterogeneity and estimate the natural gas demand efficiently, we specify a nonparametric and semi-parametric panel data model.

Furthermore, classical consumption demand cannot explain the behavior of countries in transition. In these countries, for example, we notice that the profile of the demand is gradually disassociates from the level of prices and becomes an issue of several other parameters. In other words, the classical microeconomic relationship between the demanded quantity for goods and the respective prices (the two magnitudes being of an inverse relationship) is not verified in the case of natural gas. At a certain point in the demand curve, natural gas becomes an inferior good. Before this transition, the necessity of the good is strongly proven. However, the relationship is inverted in the transition period. As a result, the demand function for natural gas assumes the form of a bell-shaped curve with respect to the price. Historically, political, social and economic conditions can be the main factors of this inverse shape of the curve.

Although some research studies (Nguyen van 2010) have demonstrated the nonlinear relationship between energy consumption and other determinants such as price and income, no previous study has generated empirical evidence of the nonlinearity relationship between natural gas demand and its major determinants by means of nonparametric models. Hence, in this paper, we suggest the use of nonparametric and semi-parametric models with panel data to test for the nonlinearity of the relationship between gas consumption and its main determinants. The advantage of the non-parametric models is to test the relationship between the dependent variable and all exogenous variables without requiring assumptions of a predetermined form of the relationship. In fact, the parametric modeling allows estimation of a finite number of parameters (number of explanatory variables +1), whereas nonparametric modeling allows having the best adjustment of the scatter plot for the relationship.

Indeed, this specification has the particular advantage of using estimation techniques without requiring any restrictions on the functional form of the relationship and as aforementioned, it requires no prior knowledge of the nature of the relationship, which is derived directly under minor assumptions. Moreover, the nonparametric models are data-driven models. Hence, the relationship between economic variables is completely determined by the data themselves (Xu and Lin, 2015). Thus, with nonparametric modeling, we are required to estimate a function *g(z)* called "*link function*" which replaces the estimation of parameters of parametric modeling.
The rest of the current paper is organized as follows: After this introduction (Part 1), is part 2 that comprises a brief literature review, part 3 involves methods and data, part 4 offers the results and part 5 concludes the paper.

2. A Brief literature review

The supply of natural gas has increased due to the hydraulic fracturing possibilities in the USA and Asian countries and the building of additional gas pipelines. Natural gas has been regarded as a backup technology that contributed to electricity generation in peak hours, while the baseload came from other resources such as coal, oil etc. Nowadays natural gas is also regarded as baseload due to its lower prices and the belief that it will continue to be available in high capacity and low prices. The required infrastructure for its transportation and storage has been established and this has enabled prices to drop. It is also a reliable energy source for cooking and heating.

Natural gas has also competing uses in vehicles and industry, but this is not studied in the current paper since we are interested only in the household demand function. The sample of countries we have used represents countries belonging to the OECD, so participating countries come from all three major regional markets for natural gas: North America, Europe and Asia. Countries that do not have domestic supplies or do not make use of hydraulic fracturing have higher gas prices. Given the utility offered to individuals by the usage of appliances, the consumer may be indifferent about the type of energy supplying those appliances.

Gas consumption as all energy types is subject to a three stage decision process (Hartman, 1979 ; Stevens 2000 ; Bhattacharyya, 2006). First the household decides whether to opt for this type of energy, second the household decides what type of appliances to use, third it decides how much and how long to operate each appliance. In the second stage, there are various factors that can affect the decision. It is income size and its regularity, alternative uses of money, the reference point towards which consumers make decisions, past experience with the energy type, what other fuels are available etc. Natural gas has various thermal qualities and therefore its efficiency varies from country to country and appliances. Many works have studied the demand for natural gas in different areas by emphasizing the determinants of gas consumption in the world, regional, national and city level, in all sectors including residential, commercial and industrial level.

Verhulst (1950), examined the demand for natural gas in French gas industry taking into consideration price and income responses for a sample of 46 firms. However, the first experimental study of demand analysis, using panel data estimators, was conducted by Balestra and Nerlove (1966) in the residential and commercial sectors. After a period of time, Tinic et al. (1973) determined a demand function for natural gas in rural Alberta and then measured the price eleasticity to evaluate the rural gasification plan economically. In 1977, Berndt and Watkins (1977) studied natural gas demand in the residential and commercial sectors in Ontario and British Columbia by generalizing the econometric model developed earlier by Balestra and Nerlove. In 1981 Beierlein et al. (1981) analyzed electricity and natural gas demand for the residential, commercial and industrial sectors of the northeastern United States. Based on Consumer Expenditure Survey Barnes et al. (1982) studied the short-run demand for natural gas. Baker et al. (1989) modelled household demand for gas and electricity in Great Britain. In

2002 Krichene (2002) analyzed natural gas and oil markets by examining demand and supply elasticities in the world. Asche et al. (2008) studied the dynamic of household natural gas demand in European countries, by estimating short and long-run elasticities. Payne et al. (2011) studied residential natural gas demand in the state of Illinois by estimating the long-run and short-run elasticities. In the same year. Bernstein and Madlener (2011) examined the demand for natural gas in the residential sector in selected OECD countries. Bilgili (2013) analyzed natural gas consumption function by evaluating the sensitivity of per capita natural gas consumption regarding natural gas prices and per capita income using a panel data in some OECD countries. Orlov (2015) estimated the elasticity of natural gas across various sectors of the Russian Economy.

Liu et al. (2018) analyzed the demand for natural gas by studying the factors impacting natural gas consumption in the household sector of 30 provinces in China by applying a Generalized Least Squares (GLS) Method between 2006 and 2015, Malzi et al. (2019) investigated residential natural gas demand in OECD countries by analyzing the main factors influencing natural gas consumption in these countries, including price, income, urbanization and demographic factors.

The topic of nonlinearities in natural gas consumption has been very scarcely studied in up to date literature. Ozmen (2023) have used multivariate adaptive regression splines in a nonparametric framework. They find that the nonparametric approach outperforms the parametric with reliable one day and one year ahead forecasts. Ding et al. (2022) employ probability density forecasts using mixed-frequency dynamic factors to solve imbalances between the supply and the demand. They employ a hydrid model for China which involves mode decomposition, mixed data sampling and support vector machine regression. They find that the most predictive dynamic factors: gas industry index, electricity industry index, Daqing crude oil prices, Qinhuangdao steam coal prices, West Texas Intermediate crude oil prices, and Australia steam coal prices. They find that nonlinearity contributes to 21%in the forecasting of a benchmark model.

In another study by Yonh-Hong and Hui (2018), they use data with small sample size, nonlinearity, randomness and fuzziness and they amploy a least squares support vector machine model based on grey related analysis for China. Their results also the higher accuracy of the proposed new method but they are not very verbose on the implications of their findings. In a study on Brazil, Costa et al. (2018) a robust least square method combined with a log-linear Cobb–Douglas model. Overall, the scant studies in the field are mostly comparisons of various methods that attempt to make predictions more robust, they do not however, contain the richness of the data in the current paper and they do not reach policy conclusions.

3. Methods and data

3.1. Traditional nonparametric estimation techniques

This section provides a brief overview of the first nonparametric estimation techniques. First, we introduce the Kernel regression estimator that was first proposed by Nadaraya (1964) and Watson (1964) and the moving regression called the Locally Estimated Scatterplot Smoothing (Loess) or the Locally Weighted Scatterplot Smothing (Lowess) developed Cleveland (1979) and Cleveland and Devlin (1988). Second, we introduce our selected nonparametric and semi-parametric estimation techniques.

Kernel regression developed by Nadaraya (1964) and Waston (1964) is inspired by the Smoothed Moving Average. However, the only difference is that in Kernel regression, the link function *g(z)* is not equally weighted. In fact, in the case of the Smoothed Moving Average the weighting of parameters (z_i) is the same while in Kernel estimator the closer points (z_i) is to z, the higher is their weight. The link function *g(z)* of the kernel estimator can be defined as a weighted average of observations yi where the weight is $w_j(z) = \frac{K\left(\frac{[z_i - z]}{\lambda}\right)}{\sum_{j=1}^{N} K\left(\frac{[z_j - z]}{\lambda}\right)}$

λ is the smoothing parameter. Thus, the Kernel estimator of the link function is the following:

$$\hat{g}(z) = \sum_{i=1}^{N} w_i(z) y_i$$

Kernel estimator is equipped with more advantages than the moving average estimator. In fact, the moving average estimator converges in probability under three conditions: the window size k should be large enough, its rate on the sample size should be small enough and the sample size should be large enough. Whereas, the Kernel estimator converges in probability as long as the sample size is large enough without any additional condition on the smoothing parameter. A second estimation technique is the aforementioned Loess or Lowess.

This local polynomial smoothing technique consists of fitting a linear regression within a chosen neighborhood V(z) of the point z. Parameters generated from this regression vary according to the reference point z. This non-parametric method was initially proposed by Cleveland (1979). Thus, the Loess method can be written as follows:

$$\begin{cases} \hat{g}(z) = \hat{\beta}_1(z) + z\hat{\beta}_2(z) \\ \{\hat{\beta}_1(z), \hat{\beta}_2(z)\} = \underset{\{\beta_1(z),\beta_2(z)\}}{MIN} \sum_{z_i \in V(Z)} [y_i - \beta_1(z) - z_i\beta_2(z)]^2 \end{cases}$$

The method consists of identifying the parameters β1 and β2 by minimizing the least squares of each neighborhood point (*z*). Hence, the parameters vary according to the reference point *z*. In this technique, all observation points have the same weight.

The Lowess proposed by Cleveland and Develin (1988) is the general form of Loess. This technique consists of weighting each point z_i by a weigh function taking the form of Kernel. Parameters $\beta_1(z)$ and $\beta_2(z)$ of the local regression are determined by the general equation:

$$\begin{cases} \hat{g}(z) = \hat{\beta}_1(z) + z\hat{\beta}_2(z) \\ \{\hat{\beta}_1(z), \hat{\beta}_2(z)\} = \underset{\{\beta_1(z),\beta_2(z)\}}{MIN} \sum_{z_i \in V(Z)} K\left(\frac{[z_i - z]}{\lambda}\right)[y_i - \beta_1(z) - z_i\beta_2(z)]^2 \end{cases}$$

3.2. Nonparametric and semi-parametric panel data models

With the convenience of nonparametric and semi-parametric panel data models, research studies with nonparametric models have expanded (Rodríguez-Poo and Soberon, 2017; Lv et al., 2017) concerning univariate and multivariate regressions. Indeed, in the univariate regression the parameter *z* of the link function *g(z)* refers to only one exogenous variable, whereas in the multivariate regression, it refers to a vector of dimension «*p*» indicating the number of the selected exogenous variables such as $p>1$. The multivariate regression concerns an adjustment of multidimensional relationships considering the nonlinearity caused by the interaction among exogenous variables.

Although nonparametric modeling has been introduced since the 1950s (Rosenblatt, 1956), the use of panel data under such estimation techniques frameworks, is relatively recent. The first studies in this perspective are by Li and Stengos (1996), Li and Ullah (1998). Additionally, Henderson et al. (2008) has suggested a new nonparametric and semi-parametric estimation technique on panel data based on an iterative procedure. Zhou and Li (2011) generalized this method for unbalanced panels where not all the statistical units are observed in the same periods.

In the current paper we adopt panel data nonparametric and semi-parametric specifications suggested by Zhou and Li (2011) considered as suitable for the nature of our data, since the estimation is adapted to cylinder and non-cylinder data. In essence, our analysis employs the multivariate regression as the general framework where *z* refers to a dimensional vector «*p*», then we will deduce the univariate regression assuming «$p = 1$». This flexible methodological framework allows testing the nonlinearity of the relationship taking advantages of panel data and nonparametric modeling.

$$y_{it} = g(z_{it}) + u_i + \varepsilon_{it} \qquad t = 1,2,\dots,m_i; \quad i = 1,2,\dots n. \qquad (1)$$

$$y_{it} = g(z_{it}) + x'_{it}\gamma + u_i + \varepsilon_{it} \quad t = 1,2,\dots,m_i; \quad i = 1,2,\dots n. \qquad (2)$$

Equations (1) and (2) represent the nonparametric and semiparametric panel data models with fixed effects, respectively where (*y*) refers to the endogenous regressor and (*z*) to the vector of exogenous variables «p». The link function *g*(.) linking the vector *z* to the dependent variable *y* is a non-specified function to estimate and defined as $\Re^p \rightarrow \Re$. In the case of the semi-parametric model other control variables "*x*" are considered, where *y* is a vector of «*q*» parameters to estimate.

In our study, we consider the case of a cylinder sample where each group « I » has m_i observations. Individual effects u_i are fixed effects and correlated with *z* while the form of this correlation in undefined. Error terms ε_{it} are assumed to be independent identically distributed (*i.i.d*), with zero mean and a variance σ_ε^2 where:

$$E(\varepsilon_{it}|z_{it}) = 0$$

We consider Ik an identity matrix of k dimension and e_ka unit vector k×1. By defining $\tilde{\varepsilon}_i = (\tilde{\varepsilon}_{i2}, \dots, \tilde{\varepsilon}_{im_i})'$ where $\tilde{\varepsilon}_{it} = \varepsilon_{it} - \varepsilon_{i1}$, we can define the covariance matrix of $\tilde{\varepsilon}_i \langle \sum_i \rangle$ and its inverse $\langle \sum_i^{-1} \rangle$ depending on σ_ε^2 as follows:

$$\begin{cases} \Sigma_i = \sigma_\varepsilon^2 \left(I_{m_i-1} + e_{m_i-1} e'_{m_i-1} \right) \\ \Sigma_i^{-1} = \sigma_\varepsilon^{-2} \left(I_{m_i-1} - e_{m_i-1} e'_{m_i-1} / m_i \right) \end{cases}$$

By denoting $g_{it} = g(z_{it})$, model (1) becomes then $y_{it} = g_{it} + u_i + \varepsilon_{it}$. For the case where « t =1 » we get $y_{i1} = g_{i1} + u_i + \varepsilon_{i1}$.

Also by denoting $\tilde{y}_i = (\tilde{y}_{i2}, \dots, \tilde{y}_{im_i})'$and $g_i = (g_{i2}, \dots, g_{im_i})'$where$\tilde{y}_{it} = y_{it} - y_{i1}$, we can express $\tilde{\varepsilon}_i$ as a function of $\tilde{y}_i$ $g_i$and g_{i1} such as:

$$y_{it} = g(z_{it}) + u_i + \varepsilon_{it} = g_{it} + u_i + \varepsilon_{it}$$
$$\Rightarrow \tilde{y}_{it} + y_{i1} = g_{it} + u_i + \tilde{\varepsilon}_{it} + \varepsilon_{i1}$$
$$\Rightarrow \tilde{y}_{it} = g_{it} - g_{i1} + \tilde{\varepsilon}_{it}$$
$$\Rightarrow \tilde{y}_i = g_i - g_{i1} e_{m_i-1} + \tilde{\varepsilon}_i$$

And $\tilde{\varepsilon}_i = \tilde{y}_i - g_i + g_{i1} e_{m_i-1}$

Similarly to Wang (2003), Lin and Carroll (2006) and Henderson (2008), we use the iterative method based on the profiled likelihood called *Profile Likelihood*. We firstly develop the first derivative of the individual likelihood «$L_i(\cdot)$» regarding g_{it} denoted $L_{it}^g = \partial L_i(\cdot)/\partial g_{it}$. Thus, we can develop the individual likelihood by formulating $\tilde{\varepsilon}_i$as follows:

$$L_i(\cdot) = -\frac{1}{2} \tilde{\varepsilon}'_i \Sigma_i^{-1} \tilde{\varepsilon}_i \quad , \quad i = 1,2, \dots, n.$$

$$L_i(\cdot) = -\frac{1}{2} \left(\tilde{y}_i - g_i + g_{i1} e_{m_i-1} \right)' \Sigma_i^{-1} \left(\tilde{y}_i - g_i + g_{i1} e_{m_i-1} \right), \quad i = 1,2, \dots, n.$$

$$L_{it}^{g} = \frac{\partial L_i(\cdot)}{\partial g_{it}} = \begin{cases} -e'_{m_i-1}\Sigma_i^{-1}\left(\tilde{y}_i - g_i + g_{i1}e_{m_i-1}\right), \ t = 1 \\ c'_{i,t-1}\Sigma_i^{-1}\left(\tilde{y}_i - g_i + g_{i1}e_{m_i-1}\right) \quad , \ t \geq 2 \end{cases}$$

Where $c_{i,t-1}$is a dimensional vector (mi – 1) × 1 whose elements are all equal to 0 except for (t-1)th element which is equal to 1.

By defining:

$$\binom{\alpha_0}{\alpha_1} = \binom{g(z)}{\partial g(z)/\partial z} = \binom{g(z)}{g^{(1)}(z)}$$

and

$$G_{it} = \binom{1}{(z_{it} - z)/h}$$

We can estimate $\binom{\alpha_0}{\alpha_1}$ by solving the first-order condition of the profiled likelihood applying an iterative technique as follows:

$$\sum_{i=1}^{n}\frac{1}{m_i}\sum_{t=1}^{m_i} K_h(z_{it} - z) \ G_{it}L_{it}^{g}\left(\hat{g}_{[l-1]}(z_{i1}), \dots, G_{it}(\alpha_0, \alpha_1)', \dots, \hat{g}_{[l-1]}\left(z_{im_i}\right)\right) = 0$$

Where $\hat{g}_{[l-1]}(z_{is})$ is the estimation of g(zis) for the (l-1)th iteration and $k_h(\upsilon) = h^{-1}k(\upsilon/h)$and $k(\cdot)$ is a Kernel function.

Accordingly, we can define the estimation for the Ith iteration depending on the (I-1th iteration:

$\binom{\alpha_0}{\alpha_1} = \begin{pmatrix}\hat{g}_{[l]}(z) \\ \hat{g}_{[l]}^{(1)}(z)\end{pmatrix} = \frac{(A_1+A_2)}{A_3}$ such as:

$$\begin{cases} A_1 = \sum_{i=1}^{n}\frac{1}{m_i}\left(e'_{m_i-1}\Sigma_i^{-1}e_{m_i-1}K_h(z_{i1} - z)G_{i1}\hat{g}_{[l-1]}(z_{i1}) + \sum_{t=2}^{m_i} c'_{i,t-1}\Sigma_i^{-1}c_{i,t-1}K_h(z_{it} - z)G_{it}\hat{g}_{[l-1]}(z_{it})\right) \\ A_2 = \sum_{i=1}^{n}\frac{1}{m_i}\left(-K_h(z_{i1} - z)G_{i1}e'_{m_i-1}\Sigma_i^{-1}H_{i,[l-1]} + \sum_{t=2}^{m_i} K_h(z_{i1} - z)G_{it}c'_{i,t-1}\Sigma_i^{-1}H_{i,[l-1]}\right) \\ A_3 = \sum_{i=1}^{n}\frac{1}{m_i}\left(e'_{m_i-1}\Sigma_i^{-1}e_{m_i-1}K_h(z_{i1} - z)G_{i1}G'_{i1} + \sum_{t=2}^{m_i} c'_{i,t-1}\Sigma_i^{-1}c_{i,t-1}K_h(z_{it} - z)G_{it}G'_{it}\right) \end{cases}$$

Where $H_{i,[l-1]}$is a vector with dimension $(m_i - 1) \times 1$ whose elements denoted «$h_{is,[l-1]}$» are such as $h_{is,[l-1]} = \left(\tilde{y}_{it} - \left(\hat{g}_{[l-1]}(z_{it}) - \hat{g}_{[l-1]}(z_{i1})\right)\right), \ t = 1,2, \dots, m_i$

The initial estimator of $g(\cdot)$ is obtained based on time series, whereas the last iteration is reached when the convergence criterion is verified:

$$\frac{\sum_{i=1}^{n}\frac{1}{m_i}\sum_{t=2}^{m_i}\left(\hat{g}_{[l]}(z_{it})-\hat{g}_{[l-1]}(z_{it})\right)^2}{\sum_{i=1}^{n}\frac{1}{m_i}\sum_{t=2}^{m_i}\hat{g}^2{}_{[l-1]}(z_{it})}\leq 0.01$$

Besides, the variance σ_ε^2 is estimated by:

$$\hat{\sigma}_\varepsilon^2=\frac{1}{2n}\sum_{i=1}^{n}\frac{1}{m_i-1}\sum_{t=2}^{m_i}\left(y_{it}-y_{i1}-\left(\hat{g}_{[l]}(z_{it})-\hat{g}_{[l-1]}(z_{it})\right)\right)^2$$

The variance of the estimator $\hat{g}(z)$ is calculated by: $\kappa\left(nh\hat{\Omega}(z)\right)^{-1}$

Where $\kappa=\int k^2(v)\,dv$n and $\hat{\Omega}(z)=\sum_{i=1}^{n}\frac{m_i-1}{m_i}\sum_{t=2}^{m_i}K_h(z_{i1}-z)/\,\hat{\sigma}_\varepsilon^2$

For estimating the semi-parametric model, we define the nonparametric estimator of « q » control variables $\hat{g}_x(\cdot)=\left(\hat{g}_{x,1}(\cdot),\dots,\hat{g}_{x,q}(\cdot)\right)'$ and the endogenous variable $\hat{g}_y(\cdot)$ defines the estimator. Thus, the estimation of γ with $q\times 1$ dimension is the following:

$$\hat{\gamma}=\left(\sum_{i=1}^{n}\frac{\tilde{x}_{i^*}'\Sigma_i^{-1}\tilde{x}_{i^*}}{m_i}\right)^{-1}\left(\sum_{i=1}^{n}\frac{\tilde{x}_{i^*}'\Sigma_i^{-1}\tilde{y}_{i^*}}{m_i}\right)$$

Where $\tilde{x}_{i^*}$and $\tilde{y}_{i^*}$ are two matrices of dimension$(m_i-1)\times q$and $(m_i-1)\times 1$ respectively such as the sth is defined as: $\tilde{x}_{is^*}=\tilde{x}_{is}-\left(\hat{g}_x(z_{is})-\hat{g}_x(z_{i1})\right)$ and $\tilde{y}_{is^*}=\tilde{y}_{is}-\left(\hat{g}_y(z_{is})-\hat{g}_y(z_{i1})\right)$.

The nonparametric component of the semi-parametric model is deduced by replacing $\tilde{y}_{it}$ by $\tilde{y}_{it}-x_{it}'\hat{\gamma}$

3.3. Household natural gas demand: empirical evidence of non-linearity

To our knowledge, previous studies have not considered the non-linear nature of the relationship between household gas demand and its determinants; even less when it comes to demographic and environmental variables. No empirical examination has been performed in that sense. This confirms the scarcity of studies raising these issues.

Our objective is to test the non-linearity of the relationship while using a nonparametric and semi-parametric modeling unexplored so far. Our approach provides both advantages; a flexible smoothing of non-parametric shapes and the robustness of panel data estimates. Accordingly, we begin by a parametric polynomial technique that will contribute to the identification of the overall non-linear nature of the relationship.

3.3.1. Data and sources

Table 1. Variable definition

Variable	Definition	Source
Measures		
Natural gas use (NGU)	Per capita household natural gas consumption (Terajoules)	International Energy Statistics (EIA, 2024).
Policy Variable		
Environmental Policy *(EP)*	The environmental policy stringency index is used as the proxy of the environmental policy. This index measures the stringency of environmental policy or the degree to which environmental policies put an explicit or implicit price on polluting or environmentally harmful behavior by combining quantitative and qualitative information related to environmental policy. The index ranges from level 0 (non-demanding) to 6 (very demanding).	OECD environmental statistic database
Economic and Climate Variables		
Prices (P)	Natural gas end-user price (US$ per MWh)	EIA, 2024
Income per capita (INC)	Per capita income (current US $)	WDI, 2024
Temperature (TEM)	The average temperature of the coldest month of each year (*Celsius °C*)	WDI, 2024
Demographic Variable		
Elderly population (ELD)	Population over age 65 among the total population (%)	WDI, 2024
Population Density (DEN)	Population density (inhabitants/km2)	WDI, 2024
Urbanization (URB)	Population living in urban areas among the total population (%)	WDI, 2024

3.3.2. A Nonlinear polynomial model

To empirically check whether a nonlinear relationship exists between gas demand and its main determinants, we firstly test a parametric model including higher order powers. More specifically, we insert a quadratic and a cubic polynomial function of our explanatory variables in order to capture the overall behavior of the relationship in different levels.

The nonlinear polynomial modeling allows identifying possible concavity changes by simply studying the signs of linear, quadratic and cubic terms. This method allows capturing the adequate pace of the relationship over a wide range of well-known shapes (U-shaped curve, inverted U-shaped curve, N-shaped curve, inverted N curve, etc.).
Our polynomial model of degree k is the following:

$$\begin{cases} y_{it} = \sum_{k=0}^{K+1} \beta_k \quad (MV_{it-5})^k + \gamma X_{it-5} + u_i + \varepsilon_{it} \\ where \quad t = 1980,1985,\ldots,2024; \quad i = 1,2,\ldots 29 \end{cases} \quad (M1)$$

Where *MV* refers to the main explanatory variable in our model. It can be population ages 65 or above (Elderly), Population Density, Environmental Policy or Urbanization.

$$\begin{cases} y_{it} = \sum_{k=0}^{K+1} \beta_k \quad (Eld_{it-5})^k + \gamma X_{it-5} + u_i + \varepsilon_{it} \\ where \quad t = 1980,1985,\ldots,2024; \quad i = 1,2,\ldots 29 \end{cases} \quad (The\ Elderly\ Model)$$

$$\begin{cases} y_{it} = \sum_{k=0}^{K+1} \beta_k \quad (den_{it-5})^k + \gamma X_{it-5} + u_i + \varepsilon_{it} \\ where \quad t = 1980,1985,\ldots,2024; \quad i = 1,2,\ldots 29 \end{cases} \quad (The\ Population\ Density\ model)$$

$$\begin{cases} y_{it} = \sum_{k=0}^{K+1} \beta_k \quad (EP_{it-5})^k + \gamma X_{it-5} + u_i + \varepsilon_{it} \\ where \quad t = 1980,1985,\ldots,2024; \quad i = 1,2,\ldots 29 \end{cases} \quad (The\ Environmental\ Policy\ model)$$

$$\begin{cases} y_{it} = \sum_{k=0}^{K+1} \beta_k \quad (URB_{it-5})^k + \gamma X_{it-5} + u_i + \varepsilon_{it} \\ where \quad t = 1980,1985,\ldots,2024; \quad i = 1,2,\ldots 29 \end{cases} \quad (The\ Urbanization\ model)$$

From each polynomial model, we can derive three parametric sub-models. The linear model, quadratic and cubic polynomial model acknowledging the focus on different variables:

a. <u>The elderly Model</u>

$$y_{it} = \beta_0 + \beta_1 eld_{it-5} + \gamma X_{it-5} + u_i + \varepsilon_{it} \quad (Linear)$$
$$y_{it} = \beta_0 + \beta_1 eld_{it-5} + \beta_2 eld_{it-5}{}^2 + \gamma X_{it-5} + u_i + \varepsilon_{it} \quad (Quadratic)$$
$$y_{it} = \beta_0 + \beta_1 eld_{it-5} + \beta_2 eld_{it-5}{}^2 + \beta_3 eld_{it-5}{}^3 + \gamma X_{it-5} + u_i + \varepsilon_{it} \quad (Cubic)$$
$$where \quad X_{it-5} = (P_{it-5}, INC_{it-5}, T_{it-5}, Den_{it-5}, URB_{it-5}, EP_{it-5})'$$
$$t = 1980,1985,\ \ldots,\ 2024; \quad i = 1,2,\ldots 29$$

b. The population density Mode

$$y_{it} = \beta_0 + \beta_1 den_{it-5} + \gamma X_{it-5} + u_i + \varepsilon_{it} \quad (Linear)$$
$$y_{it} = \beta_0 + \beta_1 den_{it-5} + \beta_2 {den_{it-5}}^2 + \gamma X_{it-5} + u_i + \varepsilon_{it} \quad (Quadratic)$$
$$y_{it} = \beta_0 + \beta_1 den_{it-5} + \beta_2 {den_{it-5}}^2 + \beta_3 {den_{it-5}}^3 + \gamma X_{it-5} + u_i + \varepsilon_{it} \ (Cubic)$$
$$where \quad X_{it-5} = (P_{it-5}, INC_{it-5}, T_{it-5}, Eld_{it-5}, URB_{it-5}, EP_{it-5})'$$
$$t = 1980,1985,\ \ldots,\ 2024; \quad i = 1,2,\ldots 29$$

c. The Urbanization model

$$y_{it} = \beta_0 + \beta_1 URB_{it-5} + \gamma X_{it-5} + u_i + \varepsilon_{it} \quad (Linear)$$
$$y_{it} = \beta_0 + \beta_1 URB_{it-5} + \beta_2 {URB_{it-5}}^2 + \gamma X_{it-5} + u_i + \varepsilon_{it} \quad (Quadratic)$$
$$y_{it} = \beta_0 + \beta_1 URB_{it-5} + \beta_2 {URB_{it-5}}^2 + \beta_3 {URB_{it-5}}^3 + \gamma X_{it-5} + u_i + \varepsilon_{it} (Cubic)$$
$$where \quad X_{it-5} = (P_{it-5}, INC_{it-5}, T_{it-5}, Eld_{it-5}, Den_{it-5}, EP_{it-5})'$$
$$t = 1980,1985,\ \ldots,\ 2024; \quad i = 1,2,\ldots 29$$

d. The Environmental Policy model

$$y_{it} = \beta_0 + \beta_1 EP_{it-5} + \gamma X_{it-5} + u_i + \varepsilon_{it} \quad (Linear)$$
$$y_{it} = \beta_0 + \beta_1 EP_{it-5} + \beta_2 {EP_{it-5}}^2 + \gamma X_{it-5} + u_i + \varepsilon_{it} \quad (Quadratic)$$
$$y_{it} = \beta_0 + \beta_1 EP_{it-5} + \beta_2 {EP_{it-5}}^2 + \beta_3 {EP_{it-5}}^3 + \gamma X_{it-5} + u_i + \varepsilon_{it} \quad (Cubic)$$
$$where \quad X_{it-5} = (P_{it-5}, INC_{it-5}, T_{it-5}, Eld_{it-5}, Den_{it-5} URB_{it-5},)'$$
$$t = 1980,1985,\ \ldots,\ 2024; \quad i = 1,2,\ldots 29$$

Table 2 represents the estimation results of the three parametric specifications, for each sub-model, applied on a panel of 29 countries. The 1st column contains the result from the linear model as our reference model. The parameter of elderly, density and urbanization are positive and significant at 5%, 5% and 10%, whereas the parameter of environmental policy is negative and significant at 5%. The demographic factors augment the use of natural gas while EP reduces the use of gas.

The parameters $\gamma 1$ $\gamma 2$ $\gamma 3$ $\gamma 4$ $\gamma 5$ $\gamma 6$ corresponding to the control variables URB, ELD, EP (we set the main variable from these four variables: Density, Elderly, Urbanization and Environmental policy, and let the other three remaining variables as control variables to test the effect of each variable on gas consumption), P, INC, TEM, respectively, are all statistically significant. Thus, these results are in line with previous research indicating a negative impact of prices and a positive impact of the income on the quantity demanded for natural gas. The negative relationship between gas demanded quantity and its own prices brings the substitution effect into the foreground especially in the long term. That is, when gas prices are higher, consumers shift to the use of another source of energy especially electricity that can perfectly substitute the use of gas in the household sector.

The parameter $\gamma 6$ which is significantly negative, reveals that temperature slows down the consumption of gas in OECD countries. In fact, when the temperature rises, the use of the heating system, as the mean source of gas consumption in the household sector, decreases. However, this assumption is no more robust and consistent since the gas is recently used for heating as well as for cooling purposes. The columns 2, 5, 6 and 11 in Table 2 show the results of the quadratic polynomial models considering the main four variables. For the density (column 2), and environmental policy (column 11) models, the significant parameters reveal the existence of a U-shaped curve since the quadratic term is negative. As for the model regarding the impact of elderly people (column 5) and Urbanization (column 8), the quadratic term shows a positive trend.

Our results confirm the trend of the relationship to be highly inversed and the coefficients of the polynomial models are significant. The columns 3, 6, 9 and 12 in Table 2 contain the results of the cubic polynomial models for the four variables. The results remain significant for the cubic estimation. These results support the assumption that the relationship leans to a non-linear adjustment. The signs of the parameters and the significance of the linear relationship and the cubic terms for density, urbanization and environmental policy variables confirm the non-linear inversed trend towards gas demand. The different stages of development explain changes in the curve shapes.

3.3.3. Nonparametric and semi-parametric models

It was very insightful to apply nonlinear polynomial modeling before using nonparametric and semi-parametric models. The use of a polynomial model allows examining whether a nonlinear relationship exists. However, some forms of nonlinearity might be ignored while introducing only quadratic and cubic terms. In fact, the use of nonparametric and semi-parametric modeling is less restrictive since it does not require a prior functional form. This approach allows having the best adjustment of the scatter plot for the relationship while benefiting from flexible and smoothed nonparametric functions instead of a predetermined polynomial function.

Indeed, we are analyzing the relationship between gas demand and its main determinants using nonparametric and semi-parametric models in a panel data framework of 29 OECD countries observed over five-year intervals over the period 1980-2024. Thus, by replacing the polynomial function $\sum_{k=0}^{K+1} \beta_k (eld_{it-5})^k$ in the polynomial model (*M1*) with the link function $g(MV_{it-5})$ to estimate, we can move to the semi-parametric model. Moreover, by removing the parametric part γX_{it-5} containing the control variable, our model becomes non-parametric. Thus, we denote our nonparametric and semiparametric model as follows:

$$\begin{cases} y_{it} = g(MV_{it-5}) + u_i + \varepsilon_{it} \\ where \quad t = 1980,\ 1985,\ \dots,\ 2024; \quad i = 1,\ 2,\ \dots\ 29 \end{cases} \qquad (M2)$$

$$\begin{cases} y_{it} = g(MV_{it-5}) + \gamma X_{it-5} + u_i + \varepsilon_{it} \\ where \quad t = 1980,\ 1985,\ \dots,\ 2024; \quad i = 1,\ 2,\ \dots\ 29 \end{cases} \qquad (M3)$$

Figures 1-8 represent the results from the nonparametric estimation (*M2*).

Table 2. Parametric estimation results

	Dependent variable : NGU											
	Linear	**Quadratic**	**Cubic**	**Linear**	**Quadratic**	**Cubic**	**Linear**	**Quadratic**	**Cubic**	**Linear**	**Quadratic**	**Cubic**
	(1)[a]	(2)[a]	(3)[a]	(1)[b]	(2)[b]	(3)[b]	(1)[c]	(2)[c]	(3)[c]	(1)[d]	(2)[d]	(3)[d]
Intercept	8.1***	8.6***	9.89**	7.37**	10.13**	11.01**	6.83***	9.089***	9.931**	8.004**	9.401**	9.894**
DEN	0.18**	0.20***	0.35**	0.41**	1.02**	0.95***	2.356**	2.745	3.641***	1.916**	1.692***	3.01*
DEN2		-2.15***	-0.60**									
DEN3			-0.8**									
ELD	-0.44**	-8.04***	-4.4*	-1.3***	-5.1***	-7.3*	-3.19**	-2.109*	-3.38***	-0.476	-3.33**	-4.64**
ELD2					1.8**	2.02**						
ELD3						2.91**						
URB	9.77*	0.003***	0.12**	2.04**	1.31**	0.5***	1.7**	1.54***	1.94*	1.21**	1.66***	0.98**
URB2								0.7***	0.82**			
URB3									-0.012*			
EP	0.41***	-0.15***	-017**	0.61***	-0.32***	-0.9***	1.54**	-0.912***	-0.7***	0.31**	-0.899**	-0.57**
EP2											-1.45**	-1.8**
EP3												-2.342**
P	-4.08**	-5.14***	-4.7***	-4.69	-6.52***	-4.211	-4.35**	-4.328**	-5.410***	-6.01**	-4.401**	-2.77***
INC	0.66**	0.98***	1.52**	1.6**	0.126***	1.001***	1.02***	1.91**	1.914**	0.8***	2.601**	1.56**
TEMP	-0.010	-0.019**	-0.61*	-0.14*	-1.64*	0.09**	-0.53**	-0.62***	-0.87***	-0.09*	-0.353***	-1.32**

*** statistically significant at 1% ,** statistically significant at 5%, * statistically significant at 10%.

[a] Density is the main variable, [b] Elderly is the main variable, [c] Urbanization is the main variable, [d] Environmental policy is the main variable

Table 3. Model specification tests

Model	Hypotheses	I_n statistic (*p*-value)	Selected model
Model 1	H_0: Quadratic H_1: Nonparametric	9.203 (0.003)	Nonparametric
	H_0: Cubic H_1: Nonparametric	5.900 (0.001)	Nonparametric
Model 2	H_0: Quadratic H_1: semiparametric	11.09 (0.003)	Semiparametric
	H_0: Cubic H_1: semiparametric	8.434 (0.001)	Semiparametric
Model 1 and 2	H_0: Nonparametric (1) H_1: Semiparametric (2)	9.009 (0.002)	Nonparametric

4. Results

Table 4: Nonparametric estimation of g(.) at different points of DEN

Quantile of Z= DEN		Non-parametric model		Semi-parametric model	
%	z	g(z)	Std. err.	g(z)	Std. err.
2.5	2.7195	0.0053	2.3481	0.0034	2.2482
25	53.1262	0.0061	1.5904	0.0051	1.3904
50	109.2422	0.0082	1.5308	0.0077	1.4307
75	179.5977	0.0113	2.2279	0.0103	2.1278
95	425.3294	0.0128	1.1930	0.0116	1.1432
97.5	487.5404	0.0087	1.9318	0.0071	1.7920

In Table 4, 5, 6 and 7 the nonparametric function g(.) is estimated at some quantile points of Density, Elderly, Urbanization and Environmental policy by using a nonparametric model (Eq. 2) and a semi-parametric model (Eq.3). In all cases, the semi-parametric estimates are slightly smaller than their nonparametric counterparts, implying that the overall impact of the control variables on per capita natural gas use is not important. In fact, these economics and policy characteristic variables can undeniably impact natural gas use.

Table 5: Nonparametric estimation of g(.) at different points of ELD

Quantile of Z= ELD		Non-parametric model		Semi-parametric model	
%	z	g(z)	Std. err.	g(z)	Std. err.
2.5	6.0131	- 0.0034	1.9417	-0.0031	1.9019
25	10.950	0.0082	2.8206	0.0080	2.7005
50	14.145	0.0085	1.1906	0.0081	1.1505
75	16.187	0.0103	1.6806	0.0106	1.4303
95	18.824	0.0113	1.3211	0.0113	1.2622
97.5	21.178	0.0123	1.2314	0.0113	1.0803

Table 6: Nonparametric estimation of g(.) at different points of URB

Quantile of Z= URB		Non-parametric model		Semi-parametric model	
%	z	g(z)	Std. err.	g(z)	Std. err.
2.5	55.0491	0.0079	1.7545	0.0074	1.7546
25	67.1264	0.0099	2.5637	0.0089	2.5637
50	74.9630	0.0083	2.3214	0.0080	2.3210
75	81.0452	0.0076	2.1977	0.0073	2.1975
95	87.8694	0.0073	1.9005	0.0069	1.9002
97.5	90.0824	0.0070	2.3126	0.0067	2.3125

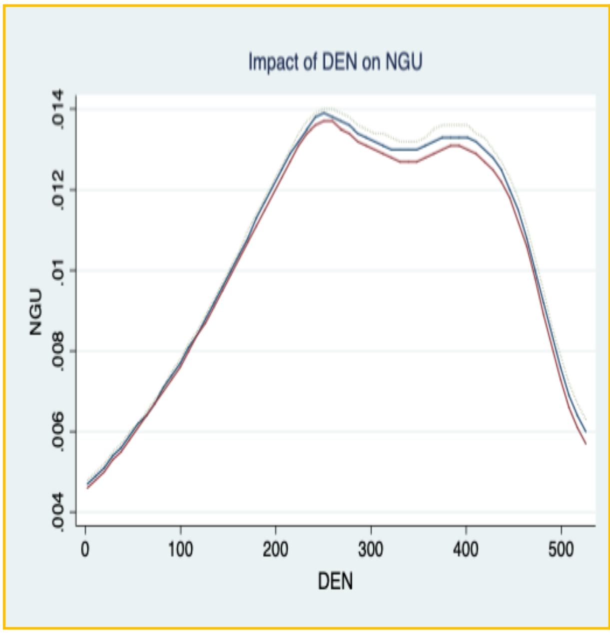
Table 7: Nonparametric estimation of g(.) at different points of EP

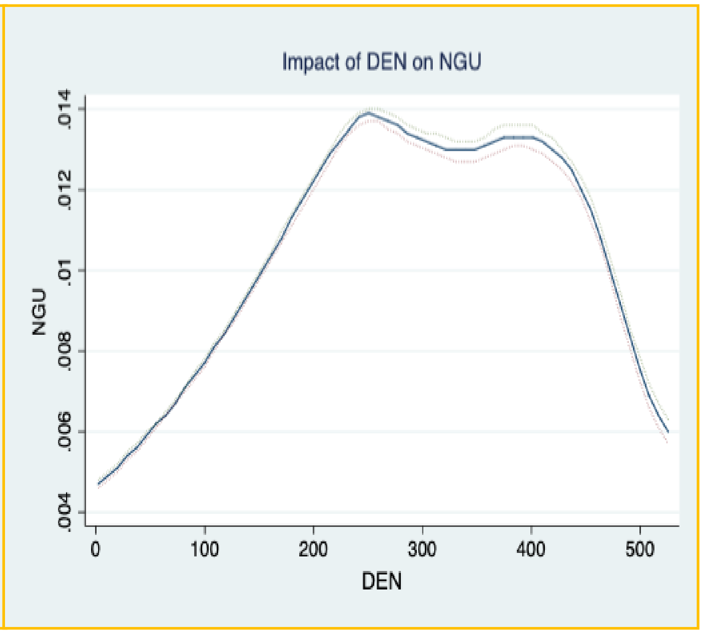
Quantile of Z= EP		Non-parametric model		Semi-parametric model	
%	z	g(z)	Std. err.	g(z)	Std. err.
2.5	0.5612	0.0059	1.6570	0.0055	1.5266
25	0.7837	0.0091	1.1022	0.0100	1.0902
50	1.6059	0.0097	1.9421	0.0104	1.8470
75	2.3557	0.0095	1.6920	0.0098	1.6860
95	3.4071	0.0084	1.5013	0.0087	1.6000
97.5	3.7200	0.0073	1.2017	0.0091	1.1820

Figures 1 and 2 show the nonparametric estimation of g(.) in the nonparametric and semi-parametric models (1) and (2), respectively (density model). In these figures we have also outlined lower and upper bounds of 95% confidence intervals. In both figures, the estimates are acceptable, because the estimation has boundary effects. Furthermore, in both figures, the curves of g(.) are almost similar, implying that despite the fact that control variables can have an overall impact on gas use, they play a small role in the estimation under a nonlinear shape of g(.). Huang (2004) has already supported these findings. The estimation is robust to the control variables and the inverted-U hypothesis is totally confirmed. For all OECD countries, per-capita gas use increases rapidly with the evolution of density, remains stable for a certain density level (between 75% and 95% quantile, see table 4) and then begins to decrease when density is higher, thought very significantly, with a very narrow confidence interval. This implies that the inverted-U hypothesis is held from the lower stage of density evolution. In the early stage of development when countries get more and more denser, the use of gas increases drastically. However, the development of energy efficiency technologies for the household sector, especially the central heating, drives per-capita consumption of gas to slow down.

The main implication of our results is the existence of returns to scale driven by energy efficiency solutions and innovative and smart development plans requiring investments in collective energy supplies that facilitate the sustainable energy transition in OECD countries.

Figure. 1. g(·) from nonparametric model (1) **Figure. 2. g(·) from semiparametric model (2)**

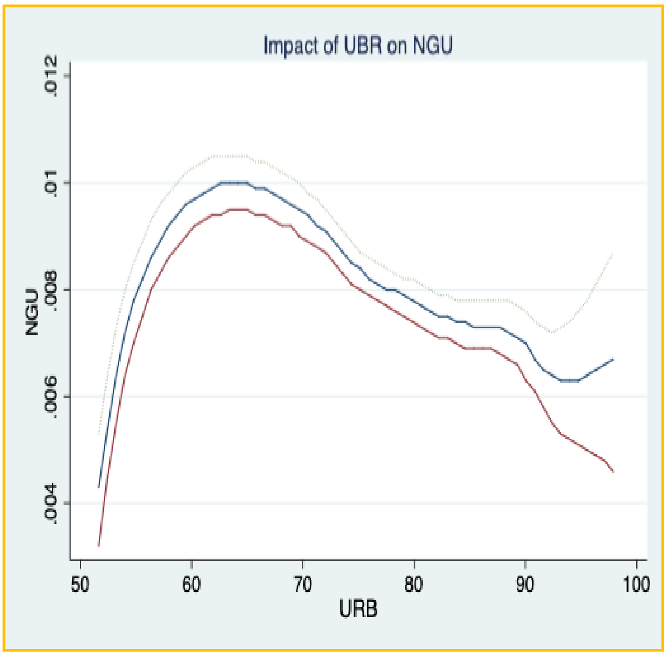


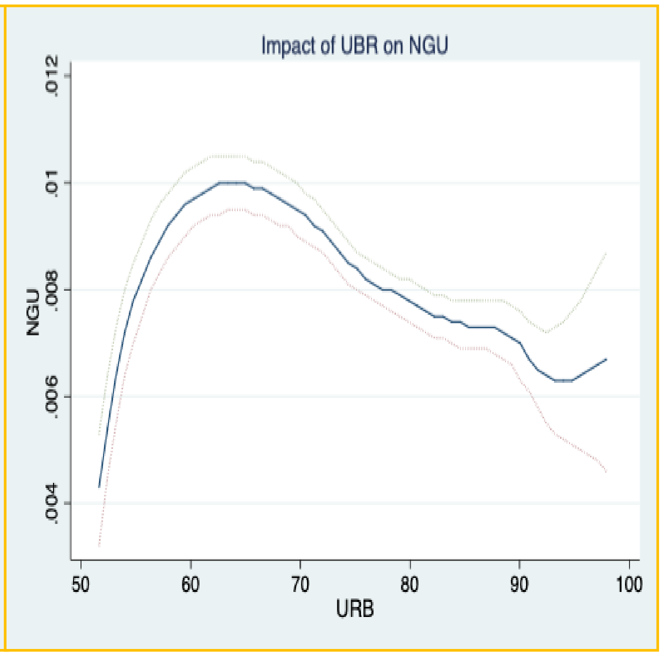


Figures 3 and 4 present the nonparametric estimation of g(.) in the nonparametric and semi-parametric models (1) and (2), respectively, in the urbanization model. The curve tends to be positive at the beginning of the urbanization process (urbanization ≤ 60%) and then the relationship becomes negative as the urbanization starts increasing. This result reflects the effect of the shift from rural to urban areas characterized by high gas consumption at the beginning of the transition period before investing in solutions that reduce gas consumption and promote its efficiency.

This result can be added to the previous one (density curve), since the use of gas is very important and necessary in the first stage of urbanization as people use more gas in urban areas compared to rural areas where the use of coal, wood and other traditional fuels is prevailing. However, the urbanization drives density to increase, and subsequently the use of natural gas declines gradually.

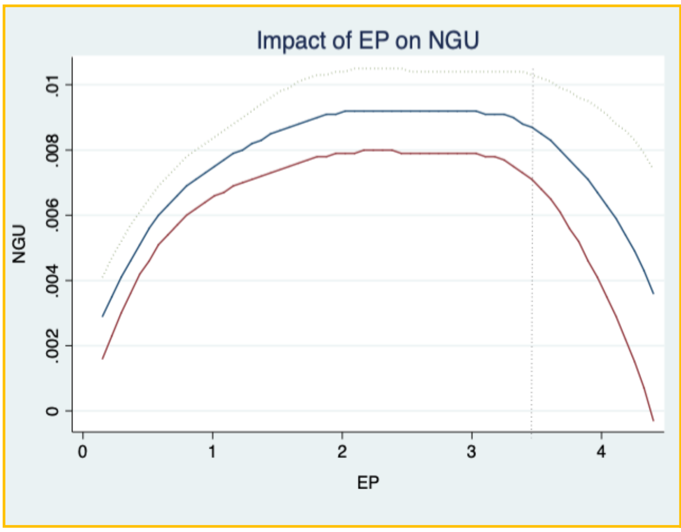
Figure. 3. g(·) from nonparametric model (1) **Figure.4. g(·) from semiparametric model (2)**

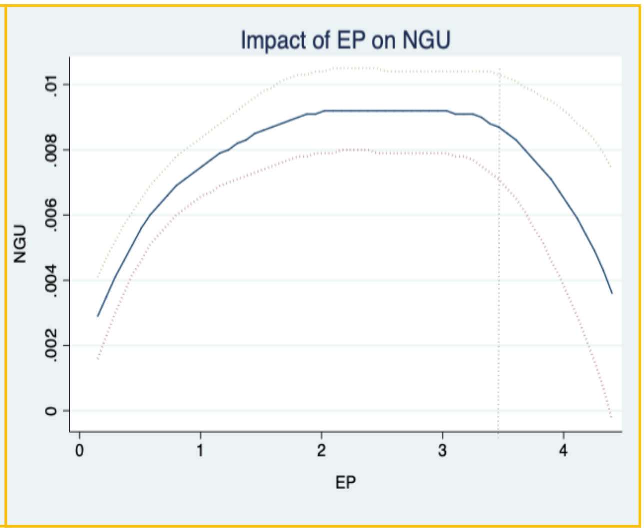




Figures 5 and 6 reveal the nonparametric estimation of g(.) in the nonparametric and semi-parametric models (1) and (2), respectively, of the environmental policy model. The estimated nonparametric curve is confirming an inverted U-shaped which reflects the strong effect of environmental policies on the consumption of natural gas.

Figure.5. g(·) from nonparametric model(1) **Figure.6. g(·) from semiparametric model (2)**





This implies that the impact of the environmental policy on natural gas use is not very high in the short run. Besides, the negative effect starts especially when the EP index is high (higher than 3), meaning that the implemented environmental policy is demanding and imposes strict measures. This is explained by the fact that although natural gas is more environmentally friendly and emits relatively less Carbon Dioxide (CO2) compared to other fuels, notably coal and oil, it remains a fossil fuel and is beyond comparison to renewable energy sources with respect to greenhouse gas emissions.

Figure.7. g(·) from nonparametric model (1)

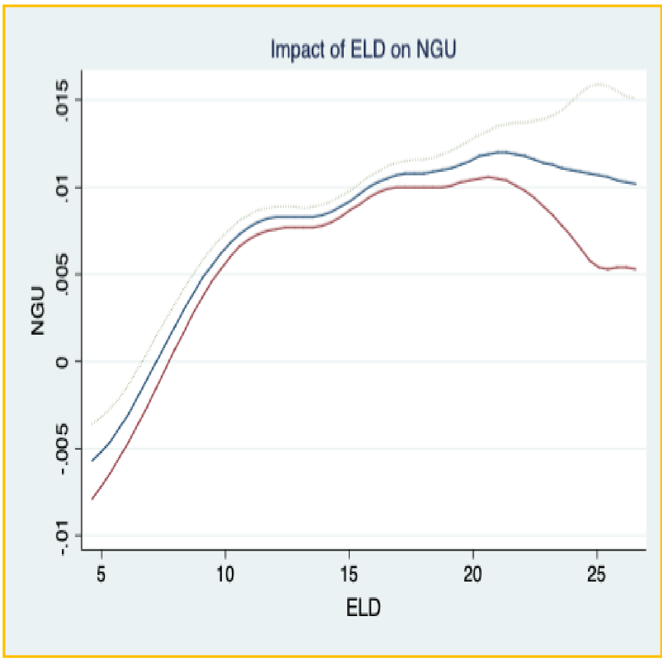
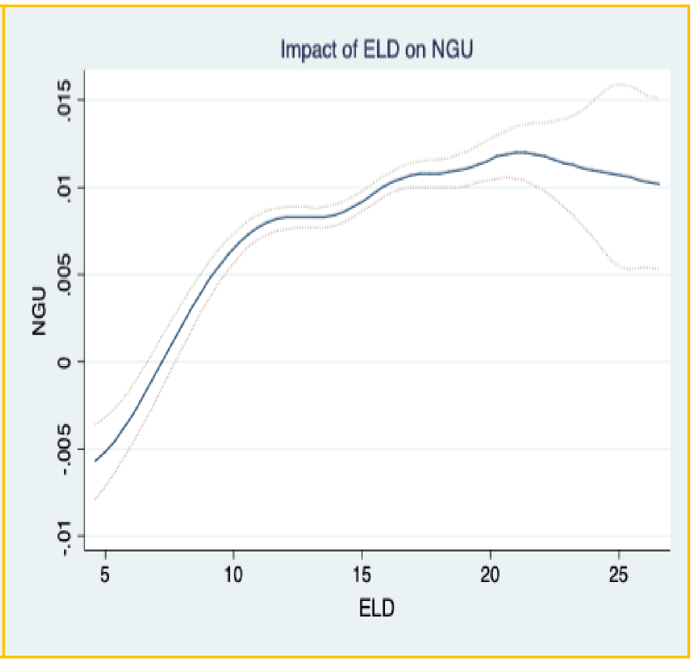


Figure. 8. g(·) from semiparametric model (2)



The main lesson to take from this important finding is that natural gas cannot be regarded as long-term energy source, but rather a transitional energy, a backstop technology or a bridge fuel towards renewable energies. Figures 7 and 8 present the nonparametric estimation of g (.) in the nonparametric and semi-parametric models (1) and (2), respectively, of the elderly population model. We note a positive relationship between elderly population and the use of gas, as the curve is revealing a positive trend. Our nonparametric specification confirms the results found in the literature (Ota et al. 2018, Malzi et al. 2020). Hence, as the population gets older, the demand for natural gas grows, since older people tend to consume more gas compared to younger people. The sensitivity to temperature variations, more indoor activities and the preference to use more gas appliances compared to the young population, are the main factors driving old people to consume more natural gas especially for interior heating and water heating.

Another implication of this result is that old population has more rigid behavior concerning changes in its consumption patterns as it prefers to continue using gas heating rather than new electrical equipment that may require installation changes and therefore costly investments. Table 3 reveals all the tests of model 1 and model 2. The tests reject the null hypotheses with a significance level of 1% implying that the semi-parametric model is not adequate, and the nonparametric specification is more suitable for our analysis.

5. Conclusion and policy implications

In this paper, we have investigated the existence of a nonlinear relationship between gas demand and its main determinants from a large panel of OECD countries over the period 1980-2024. For this purpose, we specify nonparametric and semi-parametric panel data model which do not require any predetermined form of the relation. We have first estimated polynomial models integrating powers of order greater than 1. These models confirmed that the relationship can be nonlinear with an inverse trend between per capita residential natural gas demand and its determinants changing of concavity according to the level of the dependent variables. The employed tests have confirmed that the parametric model is not adequate, and the nonparametric specification is more suitable for our analysis. These nonlinear models allow capturing heterogeneity and thus, estimate our natural gas demand relationship in an efficient way.

Subsequently, estimating a non-parametric model for OECD members allowed us to detect the nature of this concavity. We reveal that the inverted-U hypothesis is totally confirmed in the case of population density and pertains from the lower stage of density evolution. In the early stage of development when countries get more and more denser, the use of gas increases drastically. In the case of urbanization, the curve tends to positivity in the beginning of the urbanization process and then becomes negative as the urbanization starts increasing. This result can be added to the previous result since the use of gas is important in the first stage of urbanization, as people use more gas in urban areas compared to rural areas where the use of coal, wood and other traditional fuels is prevailing.

As far as environmental policy stringency is concerned, the estimated nonparametric curve is confirming an inverted U-shaped implying that the impact of the environmental policy stringency on natural gas use is very strong in the long term. Besides, the negative effect starts especially when the EPS index is high meaning that the implemented environmental policy is demanding and requires taking strict measures. In this study, we also note a positive relationship between elderly population and the use of gas as the curve is revealing a positive trend. Hence, as the population gets older, the demand for natural gas grows since old people tend to consume more gas compared to younger people. The sensitivity to temperature, more indoor activities and the preference to use gas appliances are the main factors driving old people to consumer more natural gas especially for space and water heating.

References

Alberini, A., Gans, W., & Velez-Lopez, D. (2011). Residential consumption of gas and electricity in the US: The role of prices and income. *Energy Economics*, *33*(5), 870–881. https://doi.org/10.1016/j.eneco.2011.01.015

Andersen, T. B., Nilsen, O. B., & Tveterås, R. (2011). How is demand for natural gas determined across European industrial sectors? *Energy Policy*, *39*(9), 5499–5508. https://doi.org/10.1016/j.enpol.2011.05.042

Asche, F., Nilsen, O. B., & Tveterås, R. (2008). Natural gas demand in the European household sector. *The Energy Journal, 29*(3), 27–46. https://doi.org/10.5547/ISSN0195-6574-EJ-Vol29-No3-2

Baker, P., Blundell, R., & Micklewright, J. (1989). Modelling household energy expenditures using micro-data. *The Economic Journal, 99*(397), 720–738. https://doi.org/10.2307/2233760

Balestra, P., & Nerlove, M. (1966). Pooling cross section and time series data in the estimation of a dynamic model: The demand for natural gas. *Econometrica, 34*(3), 585–612. https://doi.org/10.2307/1909771

Barnes, R., Gillingham, R., & Hagemann, R. (1982). The short-run residential demand for natural gas. *The Energy Journal, 3*(1), 59–72. https://doi.org/10.5547/ISSN0195-6574-EJ-Vol3-No1-5

Beierlein, J. G., Dunn, J. W., & McConnon, J. C. (1981). The demand for electricity and natural gas in the northeastern United States. *The Review of Economics and Statistics, 63*(3), 403–408. https://doi.org/10.2307/1924259

Belaïd, F. (2016). Understanding the spectrum of domestic energy consumption: Empirical evidence from France. *Energy Policy, 92*, 220–233. https://doi.org/10.1016/j.enpol.2016.02.004

Berkhout, P. H., Ferrer-i-Carbonell, A., & Muskens, J. C. (2004). The ex post impact of an energy tax on household energy demand. *Energy Economics, 26*(3), 297–317. https://doi.org/10.1016/j.eneco.2004.04.001

Berndt, E. R., & Watkins, G. C. (1977). Demand for natural gas: Residential and commercial markets in Ontario and British Columbia. *Canadian Journal of Economics, 10*(1), 97–111. https://doi.org/10.2307/134626

Bernstein, R., & Madlener, R. (2011). Residential natural gas demand elasticities in OECD countries: An ARDL bounds testing approach. *FCN Working Paper No. 8/2011*, Institute for Future Energy Consumer Needs and Behavior.

Bhattacharyya, S. C. (2006). Renewable energies and the poor: Niche or nexus? *Energy Policy, 34*(6), 659–663. https://doi.org/10.1016/j.enpol.2005.11.030

Bilgili, F. (2014). Long run elasticities of demand for natural gas: OECD panel data evidence. *Energy Sources, Part B: Economics, Planning, and Policy, 9*(4), 334–341. https://doi.org/10.1080/15567240903118287

Burke, P. J., & Yang, H. (2016). The price and income elasticities of natural gas demand: International evidence. *Energy Economics, 59*, 466–474. https://doi.org/10.1016/j.eneco.2016.08.025

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association, 74*(368), 829–836. https://doi.org/10.1080/01621459.1979.10481038

Cleveland, W. S., & Devlin, S. J. (1988). Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association, 83*(403), 596–610. https://doi.org/10.1080/01621459.1988.10478639

Costa, O. L., de Oliveira Ribeiro, C., Ho, L. L., Rego, E. E., Parente, V., & Toro, J. (2020). A robust least square approach for forecasting models: An application to Brazil's natural gas demand. *Energy Systems, 11*, 1111–1135. https://doi.org/10.1007/s12667-019-00356-w

Dagher, L. (2012). Natural gas demand at the utility level: An application of dynamic elasticities. *Energy Economics, 34*(4), 961–969. https://doi.org/10.1016/j.eneco.2011.09.007

Ding, L., Zhao, Z., & Wang, L. (2022). Probability density forecasts for natural gas demand in China: Do mixed-frequency dynamic factors matter? *Applied Energy, 312*, 118756. https://doi.org/10.1016/j.apenergy.2022.118756

Flaig, G. (1990). Household production and the short-and long-run demand for electricity. *Energy Economics, 12*(2), 116–121. https://doi.org/10.1016/0140-9883(90)90037-W

Granger, C. W., & Terasvirta, T. (1993). *Modelling non-linear economic relationships*. Oxford University Press.

Green, R. D. (1987). Regional variations in US consumer response to price changes in home heating fuels: The Northeast and the South. *Applied Economics, 19*(9), 1261–1268.

Haas, R., & Schipper, L. (1998). Residential energy demand in OECD countries and the role of irreversible efficiency improvements. *Energy Economics, 20*(4), 421–442.

Haas, R., & Biermayr, P. (2000). The rebound effect for space heating: Empirical evidence from Austria. *Energy Policy, 28*, 403–410.

Harold, J., Lyons, S., & Cullinan, J. (2015). The determinants of residential gas demand in Ireland. *Energy Economics, 51*, 475–483.

Hartman, R. S. (1979). Frontiers in energy demand modeling. *Annual Review of Energy, 4*(1), 433–466.

Henderson, D. J., Carroll, R. J., & Li, Q. (2008). Nonparametric estimation and testing of fixed effects panel data models. *Journal of Econometrics, 144*, 257–275.

Krichene, N. (2002). World crude oil and natural gas: A demand and supply model. *Energy Economics, 24*(6), 557–576.

Liao, H. C., & Chang, T. F. (2002). Space-heating and water-heating energy demands of the aged in the US. *Energy Economics, 24*(3), 267–284.

Liu, B. C. (1983). Natural gas price elasticities: Variations by region and by sector in the USA. *Energy Economics, 5*(3), 195–201.

Li, Q., & Stengos, T. (1996). Semiparametric estimation of partially linear regression models. *Journal of Econometrics, 71*, 389–397.

Li, Q., & Ullah, A. (1998). Estimating partially linear panel data models with one-way error components. *Econometric Reviews, 17*(2), 145–166.

Lin, X., & Carroll, R. J. (2006). Semiparametric estimation in general repeated measures problems. *Journal of the Royal Statistical Society: Series B, 68*, 68–88.

Liu, G., Dong, X., Jiang, Q., Dong, C., & Li, J. (2018). Natural gas consumption of urban households in China and corresponding influencing factors. *Energy Policy, 122*, 17–26.

Malzi, M. J., Ettahir, A., & Hanchane, S. (2019). Responsiveness of residential natural gas demand to elderly, urban population and density: Evidence from OECD countries. *International Journal of Energy Economics and Policy, 9*(4), 388–395.

Malzi, M. J., Sohag, K., Vasbieva, D. G., & Ettahir, A. (2020). Environmental policy effectiveness on residential natural gas use in OECD countries. *Resources Policy, 66*, 101651.

Meier, H., & Rehdanz, K. (2010). Determinants of residential space heating expenditures in Great Britain. *Energy Economics, 32*(5), 949–959.

Nguyen-Van, P. (2010). Energy consumption and income: A semiparametric panel data analysis. *Energy Economics, 32*(3), 557–563.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and Its Applications, 9*, 141–142.

Orlov, A. (2015). An assessment of optimal gas pricing in Russia: A CGE approach. *Energy Economics, 49*, 492–506.

Ota, T., Kakinaka, M., & Kotani, K. (2018). Demographic effects on residential electricity and city gas consumption in the aging society of Japan. *Energy Policy, 115*, 503–513.

Özmen, A. (2023). Sparse regression modeling for short-and long-term natural gas demand prediction. *Annals of Operations Research, 322*(2), 921–946.

Payne, J. E., Loomis, D., & Wilson, R. (2011). Residential natural gas demand in Illinois: Evidence from the ARDL bounds testing approach. *Journal of Regional Analysis & Policy, 41*(2), 138–147.

Rodríguez-Poo, J. M., & Soberon, A. (2017). Nonparametric and semiparametric panel data models: Recent developments. *Journal of Economic Surveys, 31*(4), 923–960.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics, 27*(3), 832–837.

Santos, G. F. (2013). Fuel demand in Brazil in a dynamic panel data approach. *Energy Economics, 36*, 229–240.

Sardianou, E. (2007). Estimating energy conservation patterns of Greek households. *Energy Policy, 35*, 3778–3791.

Stevens, P. (2000). An introduction to energy economics. In *The Economics of Energy* (pp. 1–25).

Thomas, V., Wang, Y., & Fan, X. (2003). Measuring education inequality: Gini coefficients of education for 140 countries (1960–2000). *Journal of Education Planning and Administration, 17*(1), 5–33.

Tinic, S. M., Harnden, B. M., & Janssen, C. T. L. (1973). Estimation of rural demand for natural gas. *Management Science, 20*(4-part-ii), 604–616.

Verhulst, M. J. (1950). The theory of demand applied to the French gas industry. *Econometrica, 18*(1), 45–55.

Wadud, Z., Dey, H. S., Kabir, M. A., & Khan, S. I. (2011). Modeling and forecasting natural gas demand in Bangladesh. *Energy Policy, 39*, 7372–7380.

Watson, G. S. (1964). Smooth regression analysis. *Sankhya: The Indian Journal of Statistics, Series A, 26*, 359–372.

Wu, Y. H., & Shen, H. (2018). Grey-related least squares support vector machine optimization model and its application in predicting natural gas consumption demand. *Journal of Computational and Applied Mathematics, 338*, 212–220. https://doi.org/10.1016/j.cam.2017.07.038

Xu, B., & Lin, B. (2015). How industrialization and urbanization process impacts on CO_2 emissions in China: Evidence from nonparametric additive regression models. *Energy Economics, 48*, 188–202. https://doi.org/10.1016/j.eneco.2014.12.022

Yu, Y., Zheng, X., & Han, Y. (2014). On the demand for natural gas in urban China. *Energy Policy, 70*, 57–63. https://doi.org/10.1016/j.enpol.2014.03.008

Zeng, S., Chen, Z. M., Alsaedi, A., & Hayat, T. (2018). Price elasticity, block tariffs, and equity of natural gas demand in China: Investigation based on household-level survey data. *Journal of Cleaner Production, 179*, 441–449. https://doi.org/10.1016/j.jclepro.2018.01.102

Zhou, X., & Li, K. W. (2011). Inequality and development: Evidence from semiparametric estimation with panel data. *Economics Letters, 113*(2), 203–207. https://doi.org/10.1016/j.econlet.2011.07.016