A Comparison of Stochastic Models of Natural Gas Consumption

Serli Kiremitçiyan
e-mail: serli.kiremitciyan@boun.edu.tr
Ahmet Göncü
e-mail: Ahmet.Goncu@xjtlu.edu.cn
Tolga Umut Kuzubaş
e-mail: umut.kuzubas@boun.edu.tr

January 7, 2015

Abstract

We model dynamic behaviour of natural gas consumption using continuous-time stochastic models which offer a wide set of choices for the drift and volatility terms and can be used for pricing of contingent claims on natural gas consumption as they yield analytical solutions for any forecast horizon. We apply One-factor mean-reverting and stochastic Gompertz diffusion models for the empirical analysis of daily natural gas consumption in Istanbul, Turkey. Both models perform well in reflecting the empirical properties of consumption data including stationarity, strong seasonality, mean reversion, and serial correlation. Based on the comparisons of forecast performances,we show that One-factor mean-reverting process improves upon the Gompertz diffusion process due to different specifications of the drift term and estimation procedure.

Keywords: Gompertz diffusion process, One-factor mean reverting process, Natural gas consumption, forecasting

JEL Classification Numbers: Q47, Q54, C15

1 Introduction

Natural gas is an energy source that exhibited an increasing share in the global energy consumption over the last decades. Recent developments in the gas production technologies pave the way to a higher share of the natural gas in total energy mix. As an emerging country with growing energy needs, Turkey's natural gas consumption has increased rapidly over the last two decades.[1]

The Turkish natural gas industry has undergone a process of reconstruction to create a competitive natural gas market. For this purpose, the Natural Gas Market Law, was enacted in 2001 which aimed to change the monopolistic structure of the market and form a more competitive market by establishing an independent regulatory authority, the Energy Market Regulatory Authority.[2] Liberalization of the energy markets necessitates an accurate short term forecasting and demand management of natural gas. Furthermore, it is plausible to assume that new contingent claims on the consumption amounts or prices of natural gas will probably become available.

Accurate modelling and forecasting of natural gas consumption are crucial for efficient management of resources and have been extensively studied in the literature. A comprehensive review on modelling and forecasting of natural gas consumption is given by Soldo (2012. The dominant approach is to use time series models with autoregressive structure of natural gas consumption with/without other explanatory variables such as heating degree days or temperature. Ediger and Akar (2007) use autoregressive integrated moving average and seasonal moving average models to forecast energy demand in Turkey. Aras and Aras (2004) estimate aggregate natural gas demand in residential areas of Eskişehir, Turkey using monthly data. They estimate separate autoregressive time series models for heating and non-heating months. Gümrah et al. (2001) and Sarak and Satman (2003) utilize degree days to explain the relation between natural gas demand and temperature levels. Erdoğdu (2010) employs an ARIMA model to forecast natural gas demand using quarterly data over the period of 1988 to 2005. Liu and Lin (1991) employ multiple-input transfer function models to study the relationship between natural gas consumption, temperature and price using monthly and quarterly data for Taiwan. Sanchez-Ubeda and Berzosa (2007) develops a flexible prediction method where the forecast is obtained by estimating the trend, seasonality, and transitory components. Crompton and Wu (2005) utilize a Bayesian vector autoregressive methodology to forecast energy demand for China, including demand for natural gas.

To the best of our knowledge, only the studies by Göncü et al. (2013) and Gutierrez et al. (2005) rely on stochastic processes driven by Brownian motion to model natural gas consumption. Continuous time stochastic models provide important advantages over statistical or econometric models. First, analytical formulas for the conditional expectation and variance can be derived for any forecast horizon. Second, the models can be solved analytically and contingent claims dependent on the path of the natural gas consumption can be priced relatively easily. Third, empirical characteristics of natural gas consumption data can be described using a large set of choices for drift and volatility terms. Therefore, continuous time models provide an important alternative for energy modelling. Additionally, forecasts obtained from time series models with high-frequency data will not be reliable since explanatory variables such as the macroeconomic variables or temperatures are difficult to

[1] See [EPDK (2012)]

[2] EMRA(EPDK in Turkish). See [Erdogdu (2010b)]

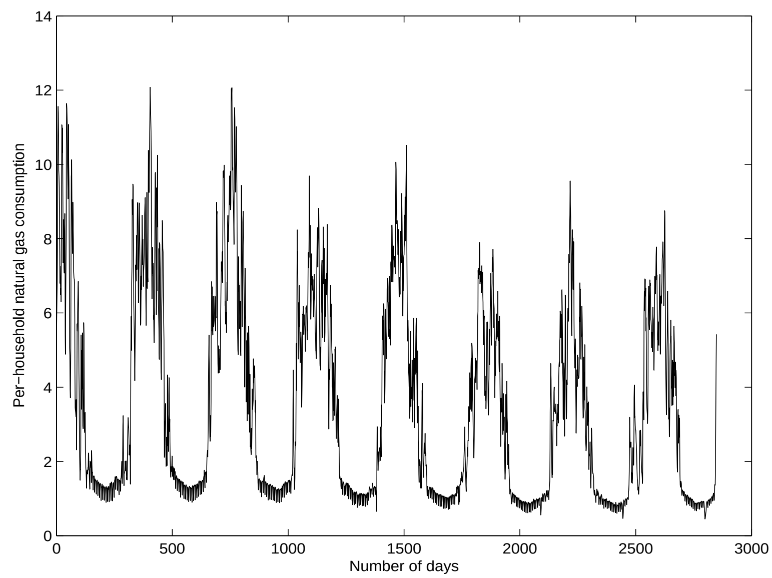
predict.

In this paper, we consider two continuous-time models used in the literature for modelling natural gas consumption; Göncü et al. (2013), which adopts the model in Lucia and Schwartz (2002), and Gutierrez et al. (2005). Our results provide important insights for future work aiming at extending these models to fit to empirical characteristics of natural gas consumption. In Section 2, we present empirical properties of data used in estimation. In Section 3 and 4, we discuss properties of the theoretical models considered in our analysis. In Section 5, we compare them in terms of their forecasting powers in capturing the empirical properties of consumption. Section 6 concludes.

2 Data

The data for natural gas consumption is obtained from IGDAS, the only natural gas distributor in Istanbul, Turkey, which contain 2848 daily observations of residential and commercial natural gas consumption in urban areas[3] and the number of consumers for the time period between January 1, 2004 to October 18, 2011.

Figure 1: Per-Household Daily Natural Gas Consumption



Per-consumer natural gas consumption is plotted in Figure 1, which exhibits a seasonal pattern and mean reversion to the seasonal mean. In particular, the seasonal pattern and mean reversion are stronger during summer months of the year, where the natural gas consumption is low. During winter months, deviations are larger with high volatility around the seasonal pattern. Therefore, a clear seasonality with slow mean reversion is expected.

[3]Industrial use of natural gas consumption is not included in the dataset.

We also observe that holidays, including weekends, have important effects on natural gas consumption and thus, a holiday dummy variable is included. In addition to stationarity, strong serial correlation exists where the first order lag is statistically significant. We use the logarithm of the per-consumer natural gas consumption as our dependent variable.

3 One-Factor Mean-Reverting Process

Following the one-factor model in [Göncü et al. (2013)], we decompose per-consumer natural gas consumption C_t as:

$$C_t = \exp(f(t) + Y_t) \tag{1}$$

where $f(t)$ is a bounded deterministic function of time and Y_t is a mean reverting stochastic process driven by standard Brownian motion $\{W_t\}_0^\infty$, which is defined on the probability space $(\Omega, \mathcal{F}, P)$ with filtration $\{\mathcal{F}_t\}_0^\infty$, and the initial value of the process is $Y_s = y_s$. The process Y_t follows:

$$dY_t = -\kappa Y_t dt + \sigma dW_t \tag{2}$$

where $\kappa > 0$ is the speed of mean reversion and $\sigma > 0$ is the volatility of the process. The solution of Equation (2) is given by:

$$Y_t = y_s e^{-\kappa(t-s)} + \int_s^t e^{-\kappa(t-u)} \sigma_u dW_u \tag{3}$$

and thus $Y_t \sim N\left(y_s e^{-\kappa(t-s)}, \int_s^t e^{-2\kappa(t-u)} \sigma_u^2 du\right)$.

For the deterministic function $f(t)$, we assume the following form to capture seasonality in natural gas consumption:

$$f(t) = \beta_0 + \beta_1 H_t + \sum_{i=1}^{p} \alpha_i \sin(iwt) + \gamma_i \cos(iwt), \tag{4}$$

where the holiday dummy variable $H(t) = 1$, if date t is weekend or holiday, and $H(t) = 0$ otherwise, $w = 2\pi/365$, and p is the number of sine and cosine terms, which is taken as 2.

In many developing countries such as Turkey, natural gas prices are centrally determined, thus fluctuations in these prices are not frequent, at least in the short run. It is feasible to add the logarithm of prices into Equation (4); however, we choose to leave the effect of prices in the stochastic part.

The per-consumer natural gas consumption C_t is now obtained as:

$$C_t = \exp(f(t) + y_s e^{-\kappa(t-s)} + \int_s^t e^{-\kappa(t-u)} \sigma dW_u) \tag{5}$$

where $f(t)$ is expressed as in Equation (4).

The conditional expectation and variance of natural gas consumption with respect to the filtration $\mathcal{F}_s$ at time $t > s > 0$ are given by:

$$E[C_t|\mathcal{F}_s] = \exp\left(f(t) + (\ln c_s - f(s))e^{-\kappa(t-s)} + \frac{1}{2}\int_s^t e^{-2\kappa(t-u)}\sigma_u^2 du\right) \tag{6}$$

and

$$var(C_t|\mathcal{F}_s) = \left[\exp(\int_s^t e^{-2\kappa(t-u)}\sigma_u^2 du) - 1\right] \times \tag{7}$$

$$\exp\left[2f(t) + 2(\ln c_s - f(s))e^{-\kappa(t-s)} + \int_s^t e^{-2\kappa(t-u)}\sigma_u^2 du\right]$$

For the case of constant volatility, conditional mean and variance simplify to:

$$E[C_t|\mathcal{F}_s] = \exp\left(f(t) + (\ln c_s - f(s))e^{-\kappa(t-s)} + \frac{\sigma^2}{4\kappa}(1 - e^{-2\kappa(t-s)})\right) \tag{8}$$

and

$$var(C_t|\mathcal{F}_s) = \exp\left(2f(t) + 2(\ln c_s - f(s))e^{-\kappa(t-s)} + \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa(t-s)})\right) \times \tag{9}$$

$$\left[\exp(\frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa(t-s)})) - 1\right],$$

respectively.

In our empirical tests, we assume a constant volatility and use Equations (8) and (9)to obtain confidence intervals for our forecasts.

3.1 Estimation of Model Parameters

We specify natural gas consumption as an Autoregressive Distributed Lag (ADL) model with lags of deterministic components and the first order lag of consumption. By substituting $f(t)$ in Equation (1) and discretizing Equation (2),we obtain:

$$\ln(C_t) = z_t = \beta_0 + \beta_1 H_t + \sum_{i=1}^{2} \alpha_i \sin(iwt) + \gamma_i \cos(iwt) + Y_t \tag{10}$$

$$Y_t = \phi Y_{t-1} + u_t, \qquad u_t \sim N(0, \sigma^2) \tag{11}$$

$$z_t = \phi z_{t-1} + H(\Phi, x_t) - \phi H(\Phi, x_{t-1}) + u_t - u_{t-1} \tag{12}$$

where H is a function of the vector of explanatory variables x_t and vector of parameters Φ, and z_t is the dependent variable. The parameters are estimated simultaneously by using non-linear least squares procedure and the mean reversion parameter is given by $\hat{\kappa} = 1 - \hat{\phi}$.

Table 1: Estimated parameters for the One-Factor Mean Reverting process (all parameters are significant at 95% confidence level)

β_0	β_1	α_1	α_2	γ_1	γ_2	κ	σ
0.960	-0.115	0.508	0.028	0.934	0.046	0.096	0.134

4 Gompertz diffusion process

The stochastic Gompertz diffusion process has two versions as the Homogenous, which implies the use of only the consumption data, and the Non-Homogenous case, which implies the use of other exogenous factors. Following the study by [Gutierrez et al. (2005)], the stochastic Homogenous Gompertz diffusion model is expressed as[4]:

$$dC_t = (\alpha - \beta \ln C_t)C_t dt + \sigma C_t dW_t, \quad C_s = c_s \tag{13}$$

where C_t is the natural gas consumption at time t, and W_t is standard Brownian motion process defined on the probability space $(\Omega, \mathcal{F}, P)$ with filtration $\{\mathcal{F}_t\}_0^\infty$. By applying Itô's formula to the transformation $e^{\beta t} \ln C_t$ and denoting $\gamma = \alpha - \sigma^2/2$, the solution of Equation (13) is obtained as:

$$C_t = \exp\left(\ln(c_s)e^{-\beta(t-s)} + \frac{\gamma}{\beta}(1 - e^{-\beta(t-s)}) + \sigma \int_s^t e^{-\beta(t-\tau)} dW_\tau\right). \tag{14}$$

Then, the conditional expectation under this process is given by:

$$E[C_t|\mathcal{F}_s] = \exp\left(\ln c_s e^{-\beta(t-s)} + \frac{\gamma}{\beta}(1 - e^{-\beta(t-s)}) + \frac{\sigma^2}{4\beta}(1 - e^{-2\beta(t-s)})\right), \tag{15}$$

which is used to forecast the natural gas consumption.

4.1 Estimation Of Model Parameters

In the study by [Gutierrez et al. (2005)], the likelihood estimators of the drift parameters, a and b in Equation (13) are given as:

$$\hat{\alpha} = \frac{\left(\int_0^T \log(C_t)^2 dt\right)\left(\int_0^T \frac{dC_t}{C_t}\right) - \left(\int_0^T \log(C_t)dt\right)\left(\int_0^T \frac{\log(C_t)}{C_t} dC_t\right)}{T\int_0^T \log^2(C_t)dt - \left(\int_0^T \log(C_t)dt\right)^2}, \tag{16}$$

[4]To simplify notation, we denote volatility by σ for both models even though two models does not necessarily have the same volatility.

and

$$\hat{\beta} = \frac{\left(\int_0^T \log(C_t)dt\right)\left(\int_0^T \frac{dC_t}{C_t}\right) - T\left(\int_0^T \frac{\log(C_t)}{C_t}dC_t\right)}{T\int_0^T \log^2(C_t)dt - \left(\int_0^T \log(C_t)dt\right)^2}, \tag{17}$$

respectively. As stated in [Gutierrez et al. (2005)], the integrals can be written as Riemann sums by applying Itô's formula and evaluated numerically with the trapezoidal rule. We also follow this approach to obtain the estimators of α and β.

The volatility σ is estimated by:

$$\hat{\sigma} = \frac{1}{T-1}\sum_{t=2}^{T} \frac{|C_t - C_{t-1}|}{\sqrt{tC_tC_{t-1}}}. \tag{18}$$

As an alternative estimation method, we discretize the stochastic differential equation given in (13) as:

$$\ln C_{t+1} = \omega_0 + \omega_1 \ln C_t + \eta_t, \qquad \eta_t \sim N(0, \sigma^2\Delta t), \tag{19}$$

where $\eta \sim N(0, \sigma^2\Delta t)$ denotes the noise component. This equation can be considered as a least-squares fitting problem where $\omega_0 = a\Delta t$ and $\omega_1 = 1 - b\Delta t$. Equation (19) can be rewritten as:

$$C_{t+1} = e^{\omega_0+\eta_t} C_t^{\omega_1} \tag{20}$$

and

$$E[C_{t+1}|\mathcal{F}_t] = C_t^{\omega_1}\exp(\omega_0 + \frac{\sigma^2\Delta t}{2}) \tag{21}$$

Similarly, for any forecast horizon h:

$$E[C_{t+h}|\mathcal{F}_t] = C_t^{\omega_1^h}\exp(\omega_0\sum_{i=1}^{h}\omega_1^{i-1} + \frac{\sigma^2\Delta t}{2}\sum_{i=1}^{h}\omega_1^{i-1}) \tag{22}$$

Estimated parameters a, b and σ are given in table below.

Table 2: Estimated Parameters for the Homogenous Gompertz Diffusion Model (all parameters are significant at 95% confidence level)

Likelihood Estimators		
α	β	σ
2.0765×10^{-5}	3.3486×10^{-5}	0.0040
Least Squares Estimates		
α	β	σ
0.0151	0.0166	0.0040

4.2 Gompertz Diffusion Process: Non-Homogenous Case

[Gutierrez et al. (2006)] generalize the model in [Gutierrez et al. (2005)] by including exogenous factors to the model in Equation (13). They specify [Gutierrez et al. (2006)], the Non-Homogenous Gompertz diffusion model as:

$$dC_t = (h(t) - \beta \ln C_t)C_t dt + \sigma C_t dW_t, \quad C_s = c_s. \tag{23}$$

Possible exogenous factors that affect growth of natural gas consumption are included as a time dependent function in $h(t) = \alpha_0 + \sum_{i=1}^{q} \alpha_i g_i(t)$ where $g_i(t)$ (exogenous variables) are continuous functions (w.r.t. to time) in $[t_0, T]$. $W(t)$ is a standard Brownian motion process.

The parameters β, σ, and α_i for $i = 0, 1, 2, ..., q$ are time-independent and should be estimated from the data using Maximum Likelihood Estimation (MLE). By applying Itô's formula to the transformation $y_t = e^{\beta t} \log(C_t)$, the solution of Equation (23) is obtained as:

$$C_t = \exp\left(e^{-\beta(t-s)} \log(C_s) + \int_s^t (h(\tau) - \sigma^2/2)e^{-\beta(t-\tau)} d\tau + \sigma \int_s^t e^{-\beta(t-\tau)} dW_\tau\right) \tag{24}$$

The conditional expectation of C_t becomes:

$$\begin{aligned} E[C_t|\mathcal{F}_s] = \exp\left(\log(c_s)e^{-\beta(t-s)} + \frac{\alpha_0 - \frac{\sigma^2}{2}}{\beta}(1 - e^{-\beta(t-s)}) + \frac{\sigma^2}{4\beta}(1 - e^{-2\beta(t-s)})\right) \cdot \\ \exp\left(\sum_{i=1}^{q} \alpha_i \int_s^t g_i(\tau)e^{-\beta(t-\tau)} d\tau\right) \end{aligned} \tag{25}$$

Again, we can discretize the stochastic differential equation given in (23) as:

$$\ln C_{t+1} = a_0 + \sum_{i=1}^{q} a_i g_i(t) + \beta \ln C_t + \eta_t, \qquad \eta_t \sim N(0, \sigma^2 \Delta t) \tag{26}$$

where $a_0 = \alpha_0 \Delta t$, $a_i = \alpha_i \Delta t$, $b = (1 - \beta \Delta t)$. Equation (26) can be written as

$$C_{t+1} = e^{h(t)\Delta t + \eta_t} C_t^b \tag{27}$$

and

$$E[C_{t+1}|\mathcal{F}_t] = C_t^{\alpha_1} \exp(\alpha_0 + \frac{\sigma^2 \Delta t}{2}) \tag{28}$$

Similarly, for any forecast horizon h:

$$E[C_{t+h}|\mathcal{F}_t] = C_t^{(1-\beta\Delta t)^h} \exp(h(t)\Delta t \sum_{i=1}^{h} (1 - \beta\Delta t)^{i-1} + \frac{\sigma^2 \Delta t}{2} \sum_{i=1}^{h} (1 - \beta\Delta t)^{i-1}) \tag{29}$$

Using the above estimators, the estimated parameters are given in table below.

Table 3: Estimated Parameters for the Non-Homogenous Gompertz Diffusion Model (all parameters are significant at 95% confidence level)

Likelihood Estimators						
α_0	α_1	α_2	α_3	α_4	β	σ
0.1094	6.9415	-2.5431	0.1965	-0.0614	0.1185	0.0234
Least Squares Estimates						
α_0	α_1	α_2	α_3	α_4	β	σ
0.1028	6.5412	-2.3281	0.1843	-0.0541	0.1114	0.1445

5 Empirical Results

We evaluate forecasting performances of the two stochastic models using the formulas for the conditional expectations and variances which are used to obtain confidence intervals for our point forecasts. In this section, we implement the backtesting method for three different forecast horizons: daily, weekly and monthly. Observations from initial two-years period is used for the estimation and the method is applied for the remaining data points by iteratively expanding the estimation window by one sample at a time. Table 4 presents the Relative Mean Square Errors (RMSE) obtained from the backtesting method.

Table 4: Comparison of RMSE's for the considered stochastic models: One-factor mean reverting stochastic process versus the Gompertz diffusion process

One-Factor Mean-Reverting Process		
Daily	Weekly	Monthly
0.0176	0.1230	0.2194
Homogenous Gompertz Diffusion Model, Likelihood Estimators		
Daily	Weekly	Monthly
0.0201	0.1186	0.4595
Least Squares Estimates		
Daily	Weekly	Monthly
0.0202	0.1114	0.4316

Table 4 shows that both non-homogenous Gompertz diffusion and One-factor mean reverting processes improve upon the fit of homogenous Gompertz diffusion process, especially for longer forecast horizons, with lower Relative Mean Square Errors (RMSE).Furthermore, including sinusoidal variables as exogenous factors increases prediction powers of the models(at the monthly horizon, 0.2194 in the One-factor mean-reverting process and 0.2003 in

Table 4: Continued.
Non-Homogenous Gompertz Diffusion Model,

Likelihood Estimators		
Daily	Weekly	Monthly
0.0208	0.1210	0.2003
Least Squares Estimates		
Daily	Weekly	Monthly
0.0221	0.1931	0.3782

the Non-Homogenous Gompertz Diffusion model), and the role of deterministic trends become more pronounced at longer horizons. Another significant difference in RMSE's stems from the estimation method, as we observe that likelihood estimators perform better in the backtesting procedure, hence provide more reliable forecasts.

6 Conclusion

In this paper, we employ and compare two stochastic diffusion models used in the literature aiming to explain the dynamic behaviour of natural gas consumption, the One-factor mean-reverting and stochastic Gompertz diffusion models. We apply our methodology to model and forecast daily natural gas consumption in Istanbul, Turkey. We compare forecast performance of both models using backtesting method. We show that the One-factor mean reverting process improves upon the fit of homogenous Gompertz diffusion model, especially for longer horizons. Model selection mainly depends on the specific area of application. For example, in the context of pricing, obtaining accurate daily consumption predictions is more important since daily settlement amounts are usually used in futures contracts or options. On the other hand,in the context of demand estimation which represent the aspect of gas suppliers or governmental institutions, monthly predictions can be more important in which case the Non-Homogenous Gompertz Diffusion model seems to perform better than the other models, especially when likelihood estimators are used for the sample period that we have chosen.

The proposed approach can be generalized to include a noise term that is driven by more general processes such as Levy process. Alternatively, the fit of the seasonality function can be improved by the use of non-parametric estimation techniques. The mean reverting stochastic process used in this paper leads to an AR(1) process in residuals. However, our modelling approach can be generalized to include higher order serial correlation in the residuals, which can be modelled via continuous autoregressive processes. Another interesting implication of our model is that contingent claims can be defined with respect to the natural gas consumption and priced within the same framework.

References

Aras, H. and Aras, N. 2004. Forecasting residential natural gas demand. Energy Sources, **26**, 463-472.

Crompton, P. and Wu, Y. 2005. Energy consumption in China: Past trends and future directions. Energy Economics, **27**, 195-208.

T.C. Enerji Piyasasi Düzenleme Kurumu (EPDK), Doğalgaz Piyasası Dairesi Başkanlığı 2013. Doğalgaz Piyasasi Sektör Raporu, `http://www.epdk.gov.tr/documents/dogalgaz/rapor_yayin/Ddp_yayin_rapor_2012.pdf`

Erdogdu, E. 2010. Natural gas demand in Turkey. Applied Energy, **87**, 211-219.

Erdogdu, E. 2010. A review of Turkish Natural gas distribution market. Renewable and Sustainable Energy Reviews, **14**, 806–813.

Göncü, A., Karahan, M. O., and Kuzubaş, T. U. 2011. A stochastic model for natural gas consumption: An application for Turkey. Iktisat Isletme ve Finans, **28**(332), 33-56.

Gümrah, F., Katircioglu, D., Aykan, Y., Okumus, S., and Kilinçer, N. 2001. Modeling of gas demand using degree-day concept: Case study for Ankara. Energy Sources, **23**, 101-114.

Gutierrez, R., Nafidi, A., Gutierrez Sanchez R. 2005. Forecasting total natural-gas consumption in Spain by using the stochastic Gompertz innovation diffusion model. Applied Energy, **80**, 115-24.

Gutierrez, R., Nafidi, A., Gutierrez Sanchez R. 2006. Electricity consumption in Morocco: Stochastic Gompertz diffusion analysis with exogenous factors. Applied Energy, **83**, 1139–51.

Ediger, V. and Akar, S. 2007. ARIMA forecasting of primary energy demand by fuel in Turkey. Energy Policy, **35**, 1701-1708.

Liu, L. M. and Lin, M. W. 1991. Forecasting residential consumption of natural gas using monthly and quarterly time series. International Journal of Forecasting, **7**, 3-16.

Lucia, J.J. and Schwartz, E.S. 2002. Electricity prices and power derivatives: Evidence from the Nordic power exchange. Review of Derivatives Research, **5**(1), 5-50.

Sanchez-Ubeda, E. and Berzosa, A. 2007. Modeling and forecasting industrial end-use natural gas consumption. Energy Economics, **29**, 710-742.

Sarak, H. and Satman, A. 2003. The degree-day method to estimate the residential heating natural gas consumption in Turkey: A case study. Energy, **28**, 929-939.

Soldo, B. 2012. Forecasting natural gas consumption. Applied Energy, **93**, 26-37.