Forecasting Daily Residential Natural Gas Consumption: A Dynamic Temperature Modelling Approach

Ahmet Göncü* Mehmet Oğuz Karahan† Tolga Umut Kuzubaş ‡

April 30, 2013

Abstract

In this paper, we propose a methodology to forecast residential and commercial natural gas consumption which combines natural gas demand estimation with a stochastic temperature model. We model demand and temperature processes separately and derive the distribution of natural gas consumption conditional on temperature. Natural gas consumption and local temperature processes are estimated using daily data on natural gas consumption and temperature for Istanbul, Turkey. First, using the derived conditional distribution of the natural gas consumption we obtain confidence intervals of point forecasts. Second, we forecast natural gas consumption by using temperature and consumption paths generated by Monte Carlo simulations. We evaluate the forecast performance of different model specifications by comparing the realized consumption values with the model forecasts by backtesting method. We utilize our analytical solution to establish a relationship between the traded temperature-based weather derivatives, i.e. HDD/CDD futures, and expected natural gas consumption. This relationship allows for partial hedging of the demand risk faced by the natural gas suppliers via traded weather derivatives.

Keywords: Natural gas demand, Temperature modelling, Monte Carlo Simulation

JEL Classification Numbers: Q47, Q54, C15

*. Address: Xian Jiaotong Liverpool University, Department of Mathematical Sciences, Suzhou 215123 China. e-mail: Ahmet.Goncu@xjtlu.edu.cn

†. Address: Bogazici University, Center for Economics and Econometrics 34342 Bebek, Istanbul, (Turkey). e-mail: oguzkarahan@gmail.com.

‡. Address: Bogazici University, Department of Economics, Natuk Birkan Building, 34342 Bebek, Istanbul, (Turkey). e-mail: umut.kuzubas@boun.edu.tr

1 Introduction

Natural gas is a widely used energy source in industrial, commercial and residential sectors. While other conventional energy sources, such as, oil or coal, have relatively lower transportation costs, in most cases, natural gas transportation requires higher initial investments. As a result, local and international natural gas markets are historically based on long-term contracts[1]. Given this market structure, one of the risk factors for natural gas distributors is the demand uncertainty. Therefore, accurate forecasting of the demand for natural gas is critical for an efficient management of energy resources.

Estimation and forecasting of residential natural gas consumption has drawn significant attention from the literature. [Liu and Lin (1991)], using monthly and quarterly data for Taiwan, employ multiple-input transfer function models to study the relationship between natural gas consumption, temperature and price. [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)] utilizes a Bayesian vector autoregressive methodology to forecast energy demand for China including demand for natural gas, which predicts a significant increase in natural gas consumption.

[Ediger and Akar (2007)] uses autoregressive integrated moving average and seasonal moving average models to forecast future overall energy demand in Turkey, including natural gas. [Aras and Aras (2004)] estimate aggregate natural gas demand in residential areas of Eskisehir, 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. [Erdogdu (2010)] employs an ARIMA model to forecast natural gas demand using quarterly data spanning the period 1988 to 2005.

[1]For a detailed analysis of the natural gas markets see [M.I.T. Energy Initiative (2011)]

Majority of studies in the literature use monthly or quarterly data to estimate natural gas demand. This aggregation is likely to result in an information loss. The use of daily data also enables us to conduct a more efficient statistical analysis due to large sample properties of the estimators. Moreover, one of the main determinants of residential natural gas consumption is temperature and therefore, in order to obtain reliable forecasts, one should embed temperature forecasting into the forecasting procedure. Potocnik et al. (2007) incorporates weather forecast data in their model and estimate expected forecasting errors. Although their forecasting methodology is different than ours, their objective is close to our paper. However, since they do not model temperature endogenously, their study do not take errors due to temperature forecasting into account.

In this paper, we propose a framework to forecast future residential and commercial natural gas consumption which combines natural gas demand estimation with a stochastic temperature model. We model demand and temperature processes separately and derive the distribution of natural gas consumption conditional on temperature. First, using the derived conditional distribution of the natural gas consumption we obtain confidence intervals of point forecasts. Second, we forecast natural gas consumption by using temperature and consumption paths generated by Monte Carlo simulations. Then, we compare the performance of these forecasting procedures.

We apply our framework using daily natural gas consumption and temperature data from Istanbul, Turkey. We estimate different model specifications for a robust analysis of the natural gas demand. Estimation results indicate that heating degree days (daily HDDs) is the main determinant of the natural gas demand. We model temperature using a mean-reverting Ornstein-Uhlenbeck process based on the model by [Alaton et al. (2002)].

Starting from the initial one year period of our sample and iteratively expanding the estimation window, we obtain forecasts based on both analytical solution and Monte Carlo simulations. We evaluate relative forecast performances of these procedures by comparing realized consumption values with the model forecasts.

Furthermore, we utilize our analytical solution to establish a relationship between the traded temperature-based weather derivatives, i.e. /CDD futures, and expected natural gas consumption. This relationship allows for partial hedging of the demand risk faced by the natural gas suppliers via traded weather derivatives.

The paper is organized as follows. Section 2.1 describes the dataset used in the analysis. Section 2.2 provides the alternative model specifications for demand estimation and presents the estimation results. Section 2.3 describes the methodology used for modelling daily average temperatures. Section 3 derives the conditional distribution of the natural gas consumption and derives a theoretical relationship between the risk neutral expectation of the natural gas consumption and the heating/cooling degree days (HDD/CDD) futures. Section 4 evaluates the forecast performance of the considered models using Monte Carlo simulation and the conditional expectation derived. Section 5 concludes the paper.

2 Data and Estimation Methodology

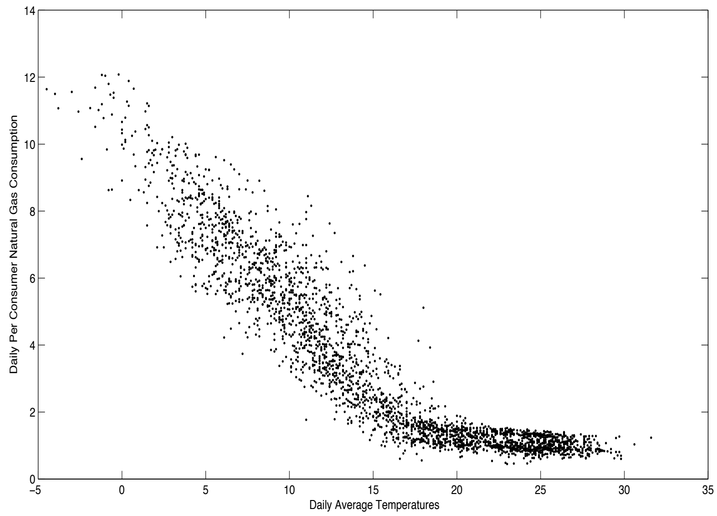
In this section, we describe our dataset and models for natural gas demand and temperature. We also present estimation results for these models for the whole sample. In order to obtain forecasts, we re-estimate these parameters for each backtesting sample.

2.1 Data

The data for natural gas consumption is obtained from IGDAS, the only natural gas distributor in Istanbul, Turkey. The dataset contains 2848 daily observations of residential and commercial natural gas consumption in urban areas[2] and the number of consumers for the time period between January 1, 2004 to October 18, 2011. The dataset of daily average

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

Figure 1: Natural Gas Consumption and Temperature



temperatures for Istanbul is obtained from Turkish State Meteorological Service for the same time period.

Natural gas consumption per-consumer is plotted against the daily average temperatures in Figure 1. It can be observed that approximately below 18°Celsius degrees the relationship between natural gas consumption and temperature is linear and natural gas consumption responds to temperature changes. This observation is consistent with the energy industry threshold of 18°Celsius degrees[3] for defining the heating and cooling degree days.

2.2 Time Series Modeling of Natural Gas Demand

In this section, we introduce demand model specifications and present estimation results for the whole sample.

Natural gas market in Turkey is not competitive (natural gas markets are not fully liberalized in most of the emerging economies) in the sense that distributors set prices using

[3]According to the energy industry practice in the US market, the reference temperature for defining heating/cooling degree days is equal to 65 Fahrenheit ($\sim$ 18°Celsius degrees).

a constant mark-up which is determined by a governmental agency, EMRA (Energy Market Regulatory Authority). Since natural gas prices do not reflect changes in demand conditions, demand estimation can be abstracted from possible supply side simultaneity issues.

We consider different model specifications for estimation of daily natural gas demand. In Table 1 model specifications are presented. The levels of natural gas consumption per-consumer, (c_t) is the dependent variable in "Panel A", whereas in "Panel B", the natural logarithm of consumption per-consumer ($\ln(c_t)$) is used. The explanatory variables and the notation used are as follows: daily heating degree days (HDD_t); daily cooling degree days (CDD_t)[4] time trend (t); natural gas prices (p_t); and the holiday dummy (H_t), which is used to capture possible effects of holidays on consumption. In Model 5b, the price is also in logarithmic form to estimate the price elasticity of demand directly. We define the daily average temperature as the average of the maximum and minimum temperatures observed at a particular measurement station during a given day. Using this definition of daily average temperature, which is denoted by T_t (measured in Celsius degrees), we define the heating degree day (HDD) as $HDD_t = \max(18 - T_t, 0)$. For a given day, if the temperature is below 18°Celsius degrees then this is considered as a heating degree day ($HDD > 0$). We define CDD similarly as $CDD_t = \max(T_t - 18, 0)$.

Demand estimation results in Table 2 indicate that HDDs (and therefore temperature) is the driving factor of the natural gas demand. Even only using the constant and HDD as explanatory variables, which corresponds to Model 1a in Table 1, we obtain an adjusted R^2 of 0.91. By including other explanatory variables such as the price and holiday dummy, the adjusted R^2 only improves marginally.

Figure 2 plots the actual natural gas consumption (response variable) versus the fitted values of the Models 1a, 1b, 2a, and 2b. Estimations are robust to price levels, possible

[4]Note that including HDD and CDD in the regression equation is equivalent to regressing natural gas consumption on temperature and an interaction dummy variable which is equal to 1 whenever $T < 18$.

Figure 2: Fitted Models 1a, 2a, 1b, and 2b.

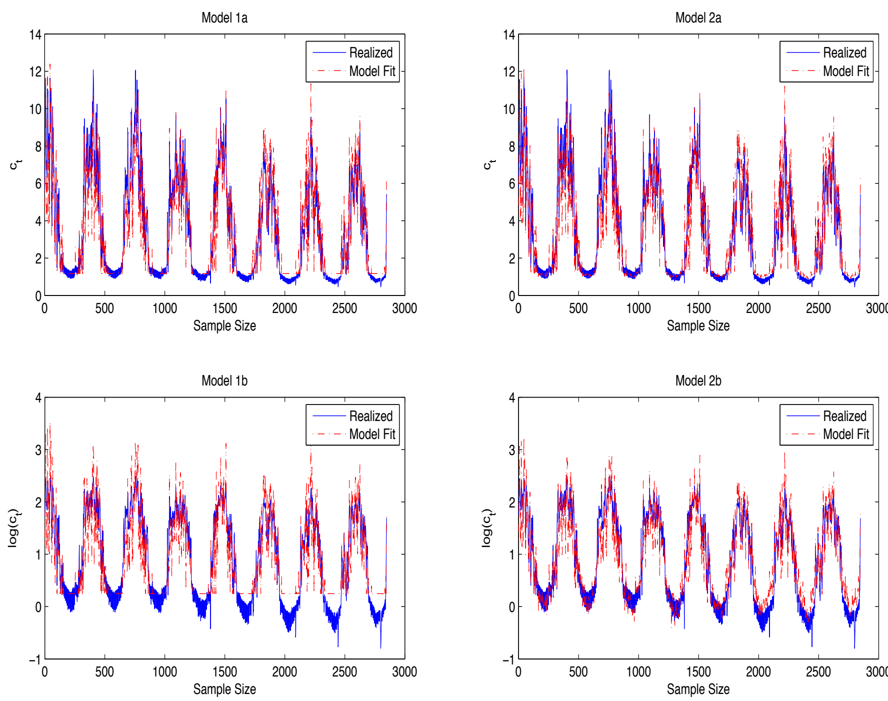


Table 1: Models considered for the estimation of daily per-consumer natural gas demand

Panel A.	
Model 1a	$c_t = \beta_0 + \beta_1 HDD_t + \epsilon_t$
Model 2a	$c_t = \beta_0 + \beta_1 HDD_t + \beta_2 CDD_t + \epsilon_t$
Model 3a	$c_t = \beta_0 + \beta_1 HDD_t + \beta_2 CDD_t + \beta_3 H_t + \epsilon_t$
Model 4a	$c_t = \beta_0 + \beta_1 HDD_t + \beta_2 CDD_t + \beta_3 H_t + \beta_4 t + \epsilon_t$
Model 5a	$c_t = \beta_0 + \beta_1 HDD_t + \beta_2 CDD_t + \beta_3 H_t + \beta_4 t + \beta_5 p_t + \epsilon_t$
Panel B.	
Model 1b	$\ln(c_t) = \beta_0 + \beta_1 HDD_t + \epsilon_t$
Model 2b	$\ln(c_t) = \beta_0 + \beta_1 HDD_t + \beta_2 CDD_t + \epsilon_t$
Model 3b	$\ln(c_t) = \beta_0 + \beta_1 HDD_t + \beta_2 CDD_t + \beta_3 H_t + \epsilon_t$
Model 4b	$\ln(c_t) = \beta_0 + \beta_1 HDD_t + \beta_2 CDD_t + \beta_3 H_t + \beta_4 t + \epsilon_t$
Model 5b	$\ln(c_t) = \beta_0 + \beta_1 HDD_t + \beta_2 CDD_t + \beta_3 H_t + \beta_4 t + \beta_5 \ln(p_t) + \epsilon_t$

For simplicity we denote all the regression residuals with ϵ_t, where ϵ_t i.i.d. $\sim N(0, \sigma_\epsilon^2)$.

holiday effects and time trend. It should also be noted that in the Turkish natural gas market price is controlled by the government and does not fluctuate often throughout time, which reduces the explanatory power of natural gas price on consumption.

To verify the normality of regression residuals the histograms of the residuals together with the superimposed normal density are plotted in Figure 3. Histograms show that even though the normality assumption is not perfectly satisfied, for analytical tractability (which is utilized in Section 3) the normality assumption is reasonable.

2.3 Temperature Modelling

Our results in Section 2.2 indicate that heating degree days (HDDs), and thus temperature, is the principal variable that explains most of the variation in daily natural gas demand per-consumer. For an accurate forecasting of natural gas demand, we need to be able to capture the dynamic behavior of daily average temperatures. We model daily average temperatures using an Ornstein-Uhlenbeck stochastic process a la [Alaton et al. (2002)]. An applica-

Table 2: Estimated demand equations. t-statistics are given in parenthesis.

Dependent Variable: c_t

Model	Constant	HDD	CDD	Hol.D.	Time	Price	Adj.R^2
1a	1.1775	0.4982					0.9154
	(60.4)	(175.5)					
2a	1.3123	0.4853	-0.0360				0.9165
	(45.2)	(138.6)	(-6.2)				
3a	1.4206	0.4855	-0.0368	-0.3379			0.9199
	(47.2)	(141.6)	(-6.5)	(-11.0)			
4a	2.0195	0.4810	-0.0274	-0.3326	-0.0004		0.9362
	(57.9)	(156.9)	(-5.4)	(-12.2)	(-27.0)		
5a	2.2974	0.4816	-0.0290	-0.3311	-0.0003	-0.8795	0.9369
	(37.7)	(157.8)	(-5.8)	(-12.2)	(-10.1)	(-5.3)	

Dependent Variable: $\ln(c_t)$

Model	Constant	HDD	CDD	Hol.D.	Time	ln(Price)	Adj.R^2
1b	0.2474	0.1464					0.8537
	(31.7)	(128.9)					
2b	0.5027	0.1219	-0.0681				0.8979
	(51.5)	(103.5)	(-35.1)				
3b	0.5514	0.1220	-0.0685	-0.1519			0.9054
	(55.4)	(107.5)	(-36.7)	(-14.9)			
4b	0.7297	0.1206	-0.0657	-0.1503	-0.0001		0.9209
	(61.8)	(116.2)	(-38.4)	(-16.2)	(-23.7)		
5b	0.9148	0.1204	-0.0654	-0.1507	-0.0002	0.1637	0.9218
	(26.5)	(116.5)	(-38.4)	(-16.4)	(-16.8)	(5.69)	

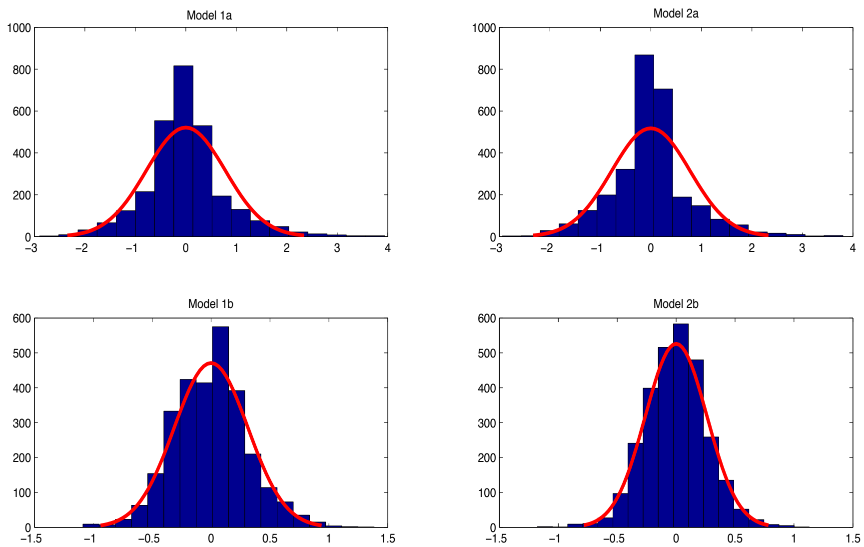
tion of this model to the Istanbul daily average temperatures is given in [Göncü et al. (2011)].

The temperature process is modelled as:

$$dT_t = \left(\frac{dT_t^m}{dt} + a(T_t^m - T_t) \right) dt + \sigma_t dW_t \qquad (1)$$

where a is the mean reversion parameter, W_t is $\mathbf{P}$-Brownian motion, and T_t^m represents the long-term mean temperatures. If we incorporate the market price of risk into the same model

Figure 3: Histogram of residuals with superimposed standard normal density for models 1a, 2a, 1b, and 2b.



then Equation (1) can be rewritten as

$$dT_t = \left(\frac{dT_t^m}{dt} + a(T_t^m - T_t) - \lambda\sigma_t \right) dt + \sigma_t dW_t^*, \qquad (2)$$

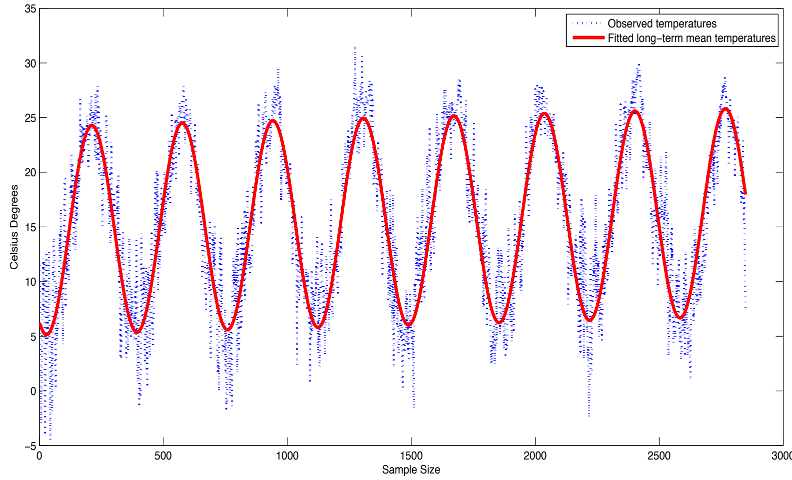
where W_t^* is $\mathbf{Q}$-Wiener process under the risk neutral probability measure. The market price of risk λ can be inferred from the prices of weather derivatives.

The long-term mean temperatures can be modelled by

$$T_t^m = A + Bt + C\sin(wt) + D\cos(wt) \qquad (3)$$

where $w = 2\pi/365$. The parameters A, B, C and D are estimated by the least-squares method to maximize the goodness-of-fit. The estimated parameters for the whole sample along with the t-statistics are provided in Table 3. In Figure 4 the actual daily average temperatures and the fitted long-term mean temperatures are plotted. The histogram of the

Figure 4: Fitted long-term mean temperature function



residuals obtained from the regression model in Equation (3) is provided in Figure 5. Figure 5 confirms that the normality assumption is met.

For a starting point T_s, the solution of Equation (2) is given by

$$T_t = (T_s - T_s^m)e^{-a(t-s)} + T_t^m + \int_s^t e^{-a(t-s)} \sigma_\tau dW_\tau - \int_s^t \lambda \sigma_u e^{-a(t-u)} du. \tag{4}$$

Then it follows that

$$E_{\mathbf{Q}}[T_t | \mathcal{F}_s] = (T_s - T_s^m)e^{-a(t-s)} + T_t^m - \int_s^t \lambda \sigma_u e^{-a(t-u)} du \tag{5}$$

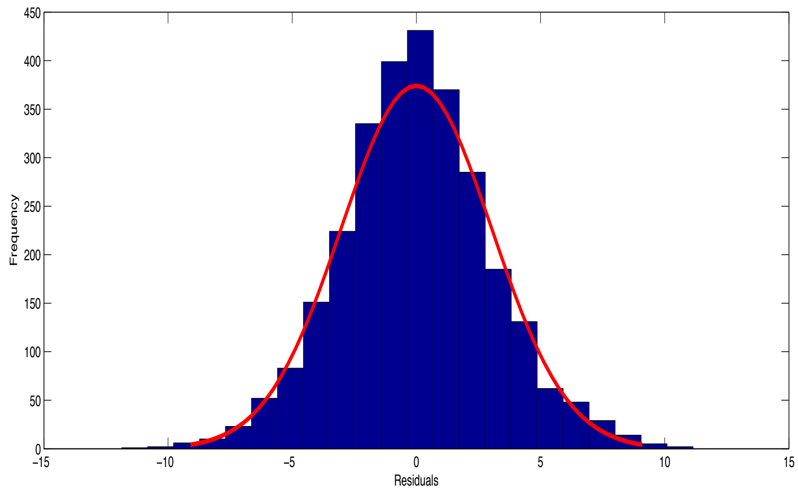
and

$$var(T_t | \mathcal{F}_s) = \int_s^t \sigma_u^2 e^{-2a(t-u)} du. \tag{6}$$

If the volatility is constant over the interval $[s, t]$ then the variance is

$$var(T_t | \mathcal{F}_s) = \frac{\sigma^2}{2a}(1 - e^{-2a(t-s)}). \tag{7}$$

Figure 5: Histogram of the residuals obtained from the fitted long-term mean temperatures



The covariance of daily temperatures for $0 \leq s \leq t \leq u$ is derived as

$$cov(T_t, T_u|\mathcal{F}_s) = e^{-a(u-t)} var(T_t|\mathcal{F}_s), \tag{8}$$

which will be needed in the derivation of conditional expectation for natural gas consumption.

In Table 4, we provide volatility estimates of the daily temperatures using quadratic variation and regression methods for each month of the year (see [Alaton et al. (2002)] for details of these two methods).

Table 3: Fitted parameters of the temperature model given by Equation (1)

	A	B	C	D	a	R^2
Estimates	14.6463	0.0006	-4.5725	-8.3363	0.2642	0.84
t-statistics	128.4	8.6	56.9	103.1		

Table 4: Monthly volatility estimates of daily average temperatures

Month	Regression	Quadratic Variation
Jan	2.43	2.50
Feb	2.59	2.49
Mar	2.52	2.59
Apr	2.21	2.22
May	1.71	1.75
Jun	1.67	1.74
Jul	1.26	1.33
Aug	1.23	1.15
Sep	1.47	1.45
Oct	1.70	1.73
Nov	2.56	2.53
Dec	2.48	2.45

3 Conditional Distribution of Natural Gas Consumption

In this section we derive the conditional distribution of the natural gas consumption at time t using the solution of the dynamic model for the daily temperatures. Hence, we obtain analytical expressions for the conditional expectation and the variance of the natural gas consumption, which we use for obtaining forecasts and confidence intervals.

Furthermore, given traded temperature based HDD/CDD futures, our approach establishes the analytical relationship between the consumption model and futures prices of HDD/CDD contracts. Therefore, we derive the risk neutral expectation of natural gas consumption given traded HDD/CDD futures. However, it is also possible to set the market price of risk parameter to be equal to zero, i.e. $\mathbf{Q} = \mathbf{P}$, and work with the physical probability measure.

3.1 Derivation of the Conditional Distribution of Natural Gas Consumption

We will consider two model specifications, Model 4a and 4b, respectively. However, the same approach can be applied to other model specifications considered in previous section. Model 4a is given by:

$$c_t = \beta_0 + \beta_1 HDD_t + \beta_2 CDD_t + \beta_3 H_t + \beta_4 t + \epsilon_t, \quad \epsilon_t \sim \text{ iid } N(0, \sigma_\epsilon^2). \tag{9}$$

T_t is normally distributed with the given mean and variance in Equations (5) and (6), respectively. For any day t, we have either heating or cooling degree day, i.e. $HDD \geq 0$ or $CDD > 0$. We set $\chi := \{t | HDD \geq 0\}$, then the conditional expectation of natural gas consumption for day t at day s ($s < t$) follows as

$$E_{\mathbf{Q}}[c_t | t \in \chi, \mathcal{F}_s] = \beta_0 + \beta_1(18 - E_{\mathbf{Q}}[T_t|\mathcal{F}_s]) + \beta_3 H_t + \beta_4 t \tag{10}$$

and

$$E_{\mathbf{Q}}[c_t | t \notin \chi, \mathcal{F}_s] = \beta_0 + \beta_2(E_{\mathbf{Q}}[T_t|\mathcal{F}_s] - 18) + \beta_3 H_t + \beta_4 t. \tag{11}$$

Variance of the conditional natural gas consumption is calculated as follows

$$var(c_t | t \in \chi, \mathcal{F}_s) = \beta_1^2 var(T_t|\mathcal{F}_s) + \sigma_\epsilon^2 \tag{12}$$

and

$$var(c_t | t \notin \chi, \mathcal{F}_s) = \beta_2^2 var(T_t|\mathcal{F}_s) + \sigma_\epsilon^2. \tag{13}$$

Denote the mean and variance of $c_t | t \in \chi$ as $\mu_{c_t,1}$ and $\sigma^2_{c_t,1}$, respectively. Then $c_t | t \in \chi \sim N(\mu_{c_t,1}, \sigma^2_{c_t,1})$, whereas the mean and variance of $c_t | t \notin \chi$ is denoted as $\mu_{c_t,0}$ and $\sigma^2_{c_t,0}$. Then

$c_t | t \notin \chi \sim N(\mu_{c_t,0}, \sigma^2_{c_t,0})$.

For simplicity suppose we want to obtain a forecast of natural gas consumption during a winter month, where HDD_t is always positive. We should also note that forecasting the natural gas during winter months is particularly important due to the peaking natural gas consumption.

Suppose given the information at time s we want to forecast $C(t_1, t_M|\mathcal{F}_s) := c_{t_1} + c_{t_2} + ... + c_{t_M}$, where c_{t_i} denotes the per-consumer natural gas consumption for day i and $t_1, ..., t_M \in \chi$. Then the conditional expectation and variance of $C(t_1, t_M|\mathcal{F}_s)$ is given by

$$E_{\mathbf{Q}}[C(t_1, t_M)|\mathcal{F}_s] = \sum_{i=1}^{M} [\beta_0 + \beta_1(18 - E_{\mathbf{Q}}[T_{t_i}|\mathcal{F}_s]) + \beta_3 H_t + \beta_4 t], \tag{14}$$

where $E_{\mathbf{Q}}[T_{t_i}|\mathcal{F}_s]$ is given in Equation (5). The variance is given by $(s \leq t_1)$

$$var(C(t_1, t_M|\mathcal{F}_s)) = \sum_{i=1}^{M} \left[\beta_1^2 var(T_{t_i}|\mathcal{F}_s) + \sigma_\epsilon^2\right] + 2\sum_{i<j} cov(T_{t_i}, T_{t_j}|\mathcal{F}_s), \tag{15}$$

where the variance and covariance of temperatures are given in Equations (7) and (8), respectively. By substitution we obtain

$$\begin{aligned} var(C(t_1, t_M|\mathcal{F}_s)) &= \sum_{i=1}^{M} \left[\frac{\beta_1^2\sigma^2}{2a}(1 - e^{-2a(t_i - s)}) + \sigma_\epsilon^2\right] \\ &+ 2\sum_{i<j} e^{-a(t_j - t_i)} \frac{\sigma^2}{2a}(1 - e^{-2a(t_j - t_i)}). \end{aligned} \tag{16}$$

Next, we consider Model 4b where the dependent variable is the logarithm of the natural gas consumption. Similarly, we derive the conditional expectation under this alternative model specification. Model 4b is given by:

$$\log(c_t) = \beta_0 + \beta_1 HDD_t + \beta_2 CDD_t + \beta_3 H_t + \beta_4 t + \epsilon_t, \quad \epsilon_t \text{ iid } \sim N(0, \sigma_\epsilon^2). \tag{17}$$

Similarly, due to the normality of temperatures the natural gas consumption is log-normally distributed. Then $c_t|t \in \chi \sim$ Log-Normal$(\mu_{c_t,1}, \sigma^2_{c_t,1})$, whereas $c_t|t \notin \chi \sim$ Log-Normal$(\mu_{c_t,0}, \sigma^2_{c_t,0})$. Then the conditional expectation of natural gas consumption for $s < t$ is given by

$$E_{\mathbf{Q}}[c_t|t \in \chi, \mathcal{F}_s] = \exp(\mu_{c_t,1} + \sigma^2_{c_t,1}/2) \tag{18}$$

and

$$E_{\mathbf{Q}}[c_t|t \notin \chi, \mathcal{F}_s] = \exp(\mu_{c_t,0} + \sigma^2_{c_t,0}/2). \tag{19}$$

The variance is given by

$$var(c_t| \in \chi, \mathcal{F}_s) = e^{(2\mu_{c_t,1}+\sigma^2_{c_t,1})}(e^{\sigma^2_{c_t,1}} - 1) \tag{20}$$

and

$$var(c_t| \notin \chi, \mathcal{F}_s) = e^{(2\mu_{c_t,0}+\sigma^2_{c_t,0})}(e^{\sigma^2_{c_t,0}} - 1). \tag{21}$$

3.2 Risk-neutral expectation of natural gas consumption in the existence of temperature-based futures

In this section, we derive a relationship between the conditional expectation of natural gas consumption and temperature based futures contracts. For example, various types of temperature futures such as, heating degree days (HDDs), cooling degree days (CDDs) and cumulative average temperatures (CAT), are traded in the Chicago Mercantile Exchange (CME). Next, we establish a connection between futures prices of HDD/CDD contracts and natural gas consumption.

We consider Model 4a. From Equation (10) for $t_1, ..., t_M \in \chi$ we obtain the risk neutral expectation of natural gas consumption as:

$$
\begin{aligned}
E_{\mathbf{Q}}[C(t_1,t_M)|\mathcal{F}_s] &= \sum_{i=1}^{M}[\beta_0+\beta_1 E_{\mathbf{Q}}[HDD_{t_i}|\mathcal{F}_s]+\beta_3 H_t+\beta_4 t] \qquad (22)\\
&= \sum_{i=1}^{M}[\beta_0+\beta_3 H_t+\beta_4 t]+\beta_1 F_{HDD}(t_1,t_M),
\end{aligned}
$$

where $F_{HDD}(t_1, t_M)$ is the market price of HDD futures contract for the measurement period $[t_1, t_M]$.

Similarly, from Equation (11) for $t_1, ..., t_M \notin \chi$ we obtain

$$
\begin{aligned}
E_{\mathbf{Q}}[C(t_1,t_M)|\mathcal{F}_s] &= \sum_{i=1}^{M}[\beta_0+\beta_2 E_{\mathbf{Q}}[CDD_{t_i}|\mathcal{F}_s]+\beta_3 H_t+\beta_4 t] \qquad (23)\\
&= \sum_{i=1}^{M}[\beta_0+\beta_3 H_t+\beta_4 t]+\beta_2 F_{CDD}(t_1,t_M),
\end{aligned}
$$

where $F_{CDD}(t_1, t_M)$ is the market price of CDD futures contract for the measurement period $[t_1, t_M]$.

4 Backtesting

In this section, we evaluate the forecast performance of the considered models via backtesting method at monthly and 10-day forecast horizons. Realized per-consumer natural gas consumption values are compared with the model forecasts from Monte Carlo simulation and the conditional expectation derived in the previous section. Below, we describe our methodology which relies on an iterative process for obtaining model forecasts at different time horizons. Without loss of generality, we only describe the backtesting procedure at monthly forecast horizon.

1. We estimate the demand and temperature models using the first 365 days of our sample

i.e. we use data from January 1, 2004 to December 31, 2004. This part of the sample is used for model estimation only.

2a. (*Monte Carlo Simulation*) For the the first consequent month (for the period January 1, 2005 to January 31, 2005), we simulate temperature and natural gas consumption paths for each day using the simulated temperatures as an input in the demand model.

2b. (*Conditional Expectation - Analytical Solution*) Using the results in Section 3 we calculate the conditional expectation of the natural gas consumption for the next consequent month (for the period January 1, 2005 to January 31, 2005).

3. At the next step, we expand estimation window by including the consequent month and repeat previous steps, i.e. we re-estimate both the demand and temperature models, which allows us to use all the realized information to obtain the forecast of the consequent month.

4. We repeat this iterative procedure until all the forecasts are obtained within the back-testing sample.

In Figure 6 and 7 we plot the realized values for monthly and 10-day periods of the natural gas consumption versus the forecasts obtained using Model 4a. The model performs well in forecasting the level and the seasonality of natural gas consumption.

We evaluate forecast performance of the model by comparing realized and forecasted consumption values for two different forecast horizons, namely 10-day and monthly horizons. We define C_i and $\hat{C}_i$ as

$$C_i = \sum_j c_{i,t_j} \quad and \quad \hat{C}_i = \sum_j \hat{c}_{i,t_j}, \tag{24}$$

Figure 6: Backtesting of the natural gas demand forecasts using *Model 4a* at the 10-day forecast horizon. Forecasts are obtained via two methods: Monte Carlo simulation and the analytical conditional expectation derived. Confidence intervals are plotted at the 95% confidence level.

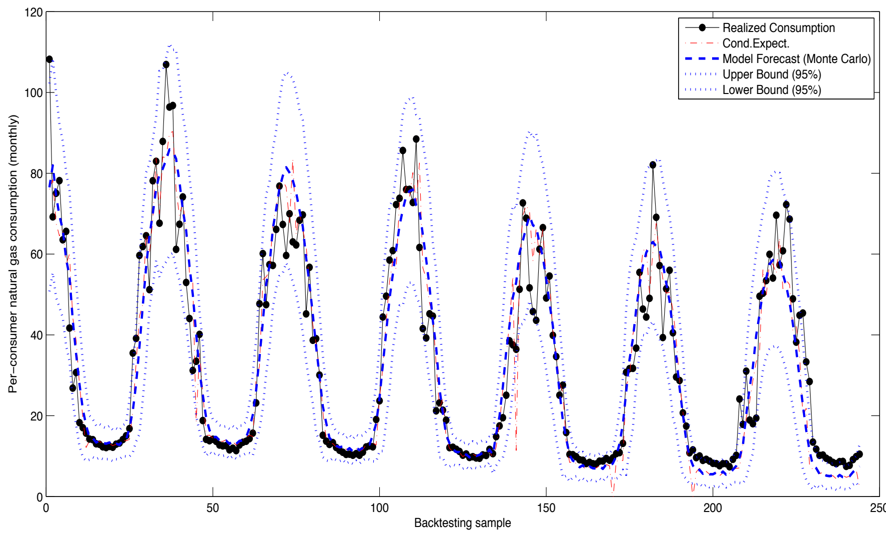
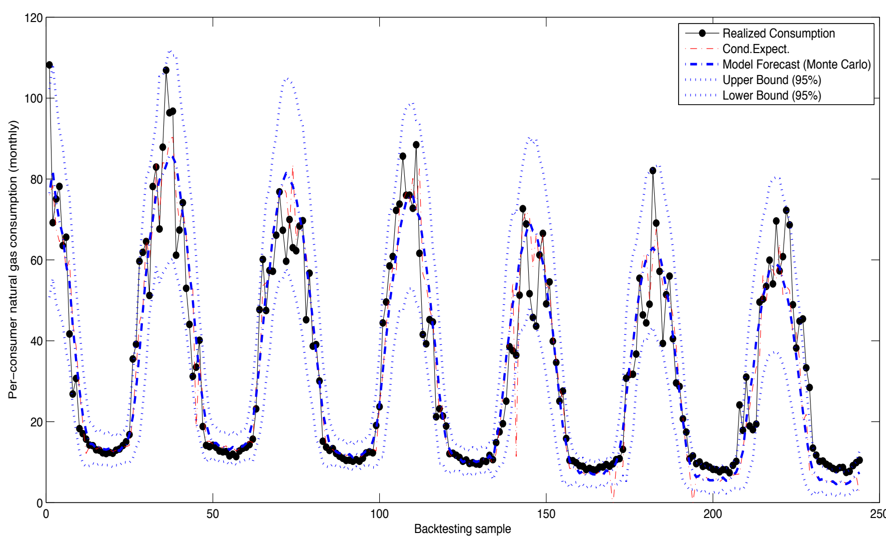


Figure 7: Backtesting of the monthly natural gas demand forecasts using *Model 4a* at the monthly forecast horizon. Forecasts are obtained via two methods: Monte Carlo simulation and the analytical conditional expectation derived. Confidence intervals are plotted at the 95% confidence level.



where t_j denotes the day of a given month i, c_{ij} and $\hat{c}_{ij}$ are the realized and forecasted consumption on jth day of month i, respectively.

For a quantitative comparison of different models, the relative mean square errors (RMSE) are calculated by the following formula

$$\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\hat{C}_i - C_i}{C_i}\right)^2, \tag{25}$$

where $i = 1, 2, ..., N$ denotes the number of months used in the backtesting sample, $\hat{C}_i$ represents the model forecast for the ith month backtesting sample, whereas C_i represents the actual monthly natural gas demand per-consumer.

The RMSE results for the considered models are compared in Table 5. Minimum RMSE is achieved in Model 4a for Monte Carlo simulation and in Model 5b for analytical solution.

A more detailed comparison of Models 4a and 4b are provided in Table 6 at the monthly forecast horizon.[5] As expected, forecasts based on the conditional expectation of natural gas consumption is particularly more accurate for winter or summer months, which reduces the need for a Monte Carlo simulation. However, the accuracy of the analytical result decays during transition months when it is more likely that temperature fluctuates around the threshold value 18°C.

Since natural gas demand for space heating purposes peaks during winter months, forecasting accuracy for these periods is more critical. Monthly RMSE of forecasts based on both Monte Carlo simulation and analytical solution suggest that -especially for consumption levels (c_t)- forecasting errors are lowest for winter months.

[5]For space considerations we only present RMSE results at the monthly forecast horizon.

Table 5: Comparison of different models with respect to the Relative Mean Square Error (RMSE): all months of the year are included in the backtesting sample. The sample size is fixed at 5000 for Monte Carlo simulations.

Model	Monte Carlo Simulation	Analytical Solution
1a	0.0680	0.0660
2a	0.0570	0.0594
3a	0.0546	0.0497
4a	0.0302	0.0428
5a	0.0584	0.0702

Model	Monte Carlo Simulation	Analytical Solution
1b	0.0896	0.1078
2b	0.0532	0.0588
3b	0.0538	0.0579
4b	0.0369	0.0377
5b	0.0327	0.0321

5 Conclusion

In this paper, we propose a framework to forecast future residential and commercial natural gas consumption which combines natural gas demand estimation with a stochastic temperature model. Our framework incorporates a model for temperature process into a demand model which allows us to derive the distribution of natural gas consumption conditional on temperature. We obtain point forecasts and confidence intervals using (1) the derived conditional distribution and (2) temperature and consumption paths generated by Monte Carlo simulations and evaluate relative forecast performances.

We apply our framework using daily natural gas consumption and temperature data from Istanbul, Turkey. For a robust analysis of natural gas demand, we estimate different model specifications using data on heating and cooling degree days (HDDs/CDDs), natural gas price, and holiday dummy variables. Our estimations identify heating degree days and thus daily temperatures, as the main determinant of natural gas demand. Including the price of natural gas and holiday dummy as explanatory variables in the regression model improves

Table 6: Comparison of different models with respect to the Relative Mean Square Error (RMSE) for different periods of the year. The forecasting performance of the considered models are reported. The sample size is fixed at 5000 for Monte Carlo simulations.

	Model 4a		Model 4b	
Period	Monte Carlo	Analytical Solution	Monte Carlo	Analytical Solution
Jan	0.0179	0.0222	0.0262	0.0464
Feb	0.0132	0.0116	0.0168	0.0285
Mar	0.0328	0.0253	0.0304	0.0267
Apr	0.0334	0.0325	0.0103	0.0156
May	0.0345	0.0783	0.0535	0.0372
Jun	0.0279	0.0283	0.0434	0.0443
Jul	0.0293	0.0274	0.0222	0.0248
Aug	0.0269	0.0256	0.0307	0.0361
Sep	0.0159	0.0194	0.0546	0.0520
Oct	0.0614	0.1652	0.0777	0.0740
Nov	0.0559	0.0662	0.0641	0.0614
Dec	0.0130	0.0120	0.0071	0.0050
Dec-Jan-Feb	0.0440	0.0458	0.0501	0.0799
Jun-Jul-Aug	0.0841	0.0813	0.0963	0.1052

the fit only slightly. We model daily average temperatures with a mean reverting Ornstein-Uhlenbeck (OU) stochastic process which is a a commonly used approach in modelling daily average temperatures (see [Alaton et al. (2002)] and [Göncü et al. (2011)]).

In order to evaluate the forecast performance, we rely on the backtesting method. We start from the initial year of our sample, and we iteratively expand the estimation window to obtain backtesting samples based on the conditional distribution and Monte Carlo simulations. We evaluate relative forecast performances of these procedures by comparing realized consumption values with the model forecasts by comparing relative mean square errors (RMSE). Backtesting of model forecasts at monthly and 10-days time increments shows that modelling natural gas demand using heating/cooling degree days, a constant and a time trend, yields reliable forecasts. The analytical solution derived for the conditional expectation of natural gas consumption is particularly more useful for winter and summer months, which reduces the need for a Monte Carlo simulation. However, during transition months we observe that the accuracy of the analytical result decays.

Furthermore, we establish a relationship between traded temperature-based weather derivatives and expected natural gas consumption. Using our approach, if a weather derivatives market exists for the considered location, then the implied natural gas demand per-consumer with respect to the risk neutral probability measure can be derived. As a result, this relationship allows for partial hedging of the demand risk of natural gas suppliers via weather derivatives.

References

Alaton, P.,Djehiche, B., and Stillberger, D. 2002. On Modelling and Pricing Weather Derivatives. Applied Mathematical Finance, **9**, 1-20.

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.

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

Göncü, A., Karahan, M. O., and Kuzubaş, T. U. 2011. Pricing of Temperature-based Weather Options for Turkey. Iktisat Isletme ve Finans, **26**, 33-50.

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.

Ediger, V., 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.

Potocnik, P., Thaler,M., Govekar, E., Grabec,I., Poredos,A., 2007. Forecasting risks of natural gas consumption in Slovenia. Energy Policy **35** 4271-4282

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.

M.I.T. Energy Initiative 2011. The future of natural gas: an interdisciplinary MIT study,