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

Ahmet Göncü*
Xian Jiaotong Liverpool University

Mehmet Oğuz Karahan**
Yaşar University

Tolga Umut Kuzubaş***
Boğaziçi University

Abstract

In this paper, we propose a new 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 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 the backtesting method. Our analysis establishes 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 natural gas suppliers via traded weather derivatives.

Keywords: Natural gas demand, heating degree days, temperature modelling, Backtesting, HDD/CDD futures

JEL Classification Numbers: Q47, Q54, C15

Özet

Bu çalışmada, mesken ve ticari doğal gaz tüketimini, bir stokastik sıcaklık modelini temel alan doğal gaz talep tahmini ile birlikte tahmin eden bir yöntem öne sürülmektedir. Talep ve sıcaklık süreçleri ayrı ayrı modellenip, doğal gaz tüketiminin sıcaklığa şartlı dağılımı türetilmektedir. Doğal tüketim ve yerel sıcaklık süreçleri, İstanbul'un günlük doğal gaz tüketim ve sıcaklık verileri kullanılarak tahmin edilmiştir. İlk olarak, doğal gaz tüketiminin şartlı dağılımı kullanılarak nokta tahmincilerinin güven aralıkları oluşturulmuştur. Sonrasında, doğal gaz tüketimi Monte Carlo simülasyonları ile oluşturulan sıcaklık ve tüketim patikaları kullanılarak tahmin edilmiştir. Farklı modellerin tahmin performansları geriye dönük test yöntemiyle gerçek tüketim değerleriyle karşılaştırılmıştır. Ayrıca, analitik çözüm kullanılarak doğal tüketimi ile sıcaklık türev ürünleri arasında bir ilişki ortaya konulmuştur. Bu ilişki doğal gaz tedarikçilerinin sıcaklık türevleri kullanarak talep risklerinden kısmen korunmasına imkan sağlamaktadır.

Anahtar kelimeler: Doğal Gaz Talebi, Isınma Derecesi Günleri, Sıcaklık Modellenmesi, Geriye Dönük Test, IDG/SDG Vadeli Sözleşmeleri

JEL Sınıflaması: Q47, Q54, C15

*Ahmet Göncü Associate Professor Department of Mathematical Sciences Xian Jiaotong Liverpool University Suzhou 215123 China Email: Ahmet.Goncu@xjtlu.edu.cn

**Mehmet Oğuz Karahan Assistant Professor Department of International Trade and Finance Yasar University 35100 Bornova, İzmir, (Turkey) Email: oguz.karahan@yasar.edu.tr

***Tolga Umut Kuzubaş Associate Professor Department of Economics Bogazici University Natuk Birkan Building, 34342, Bebek, İstanbul, (Turkey) Email: 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 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, employs 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) estimates 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) utilizes 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. Gutierrez et al. (2005) uses a Gompertz-type innovation diffusion process as a stochastic growth model of natural gas consumption, which represents the stochastic natural gas consumption process in terms of a deterministic trend plus a stochastic noise component. An excellent survey of the literature on modeling natural gas consumption is given by Soldo (2012).

The majority of studies 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 the 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 estimates expected forecasting errors. However, since they do not model temperature endogenously, their study does not take errors due to temperature

---

[1] For a detailed analysis of the natural gas markets see M.I.T. Energy Initiative (2011). Also, there is a movement towards development of local/regional spot markets. Spot markets allow a transformation of demand risk into price risk. However, cost efficiency of this transformation is subject to liquidity of the local/regional market.

forecasting into account.

In this paper, we propose a new framework to forecast future residential and commercial natural gas consumption which combines natural gas demand estimation with a stochastic temperature model. This new framework offers three significant advantages over existing models in the literature. First, we are able to derive the conditional distribution of natural gas consumption over a given period dependent on the temperature path. Second, by establishing the relationship between the temperature process and natural gas consumption, we represent the conditional expectation of the natural gas consumption in terms of temperature-based weather derivatives. Third, we show that via the theoretical relationship between weather derivatives and natural gas consumption, there emerge partial hedging opportunities for natural gas companies.

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 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) are the main determinant of 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 model forecasts.

Furthermore, we utilize our analytical solution to establish a relationship between 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 natural gas suppliers via traded weather derivatives.

In a nutshell, the framework proposed in this paper, combining residential natural gas demand estimation with a stochastic temperature model, improves forecasting accuracy. Moreover, using the derived analytical relation to weather derivatives provides a hedging strategy or an insurance mechanism for natural gas suppliers against temperature driven demand fluctuations. Our focus is therefore merely on the supply side of the natural gas market and on management of the demand risk faced by natural gas providers.

On the demand-side of the residential natural gas market, efficiency improvements can result in reduced natural gas demand over time. Burlig et al. 2017 analyzes the effects of demand-side efficiency improvements on California schools using a panel fixed effects model. Their study focuses on analyzing realization rates—that is, realized energy savings relative to expected savings of energy investments. It finds that energy upgrades reduced electricity consumption by 2.9 to 4.5 percent on average. However, it finds that on average these reductions constitute approximately 24 percent of expected savings, which are based on engineering estimates. Analyzing realization rates is

important from a microeconomic perspective, since they have immediate implications on the feasibility of undertaking such micro-level energy investments from the perspective of individuals or institutions.

In this paper, we abstract from issues related to the demand side of the natural gas market such as energy efficiency and related improvements in welfare. Improvements in energy efficiency and welfare effects are important considerations but outside the aim of our paper. Of course, demand-side efficiency-improving developments and/or other substitution effects may change demands of individuals who undertake these investments in the long-run. Even if the number of individuals who undertake such investments become significant, since aggregate demand is relevant for the suppliers, we believe that the adaptive nature of our forecasting methodology can incorporate into the model such shifts in demand.

The paper is organized as follows. Section 2 describes the data and discusses the alternative model specifications for demand estimation and the methodology used for modelling daily average temperatures. Section 3 derives the conditional distribution of natural gas consumption and develops 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.

Figure 1: Natural Gas Consumption and Temperature
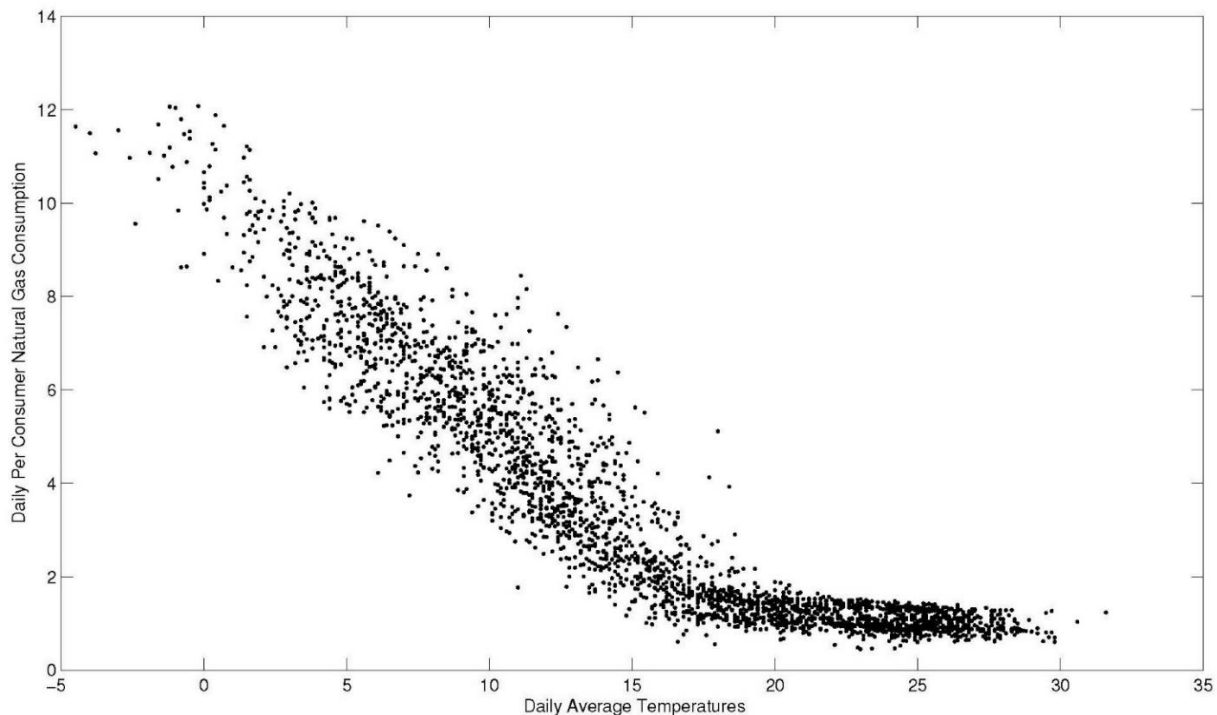
2   Data and Estimation Methodology

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

Natural gas consumption per-consumer is plotted against 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 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.

The 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 a mark-up called "unit service and depreciation charge (USDC)". This mark-up is determined by a public agency, EMRA (Energy Market Regulatory Authority) and differs in each distribution region. Since natural gas prices do not reflect changes in demand conditions, demand estimation can be abstracted from possible supply side simultaneity issues. Similarly, in most of the developing countries (e.g. China) prices are centrally decided either by public or government institutions. Therefore, our framework and methodology apply to a wide range of countries.

We consider different model specifications for our estimation of daily natural gas demand. Table 1 presents model specifications. The level 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. We define the daily average temperature as the average of the

---

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

[3] According to 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).

[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 $T < 18$.

maximum and minimum temperatures observed at a particular measurement station during a given day. Using this definition of daily average temperature, which is denoted 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 $CDD_t = \max(T_t - 18, 0)$.

Table 1: Models for estimating daily per-consumer natural gas demand

Panel A.

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

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

Demand estimation results in Table 2 indicate that HDDs (and therefore temperature) are the driving factor of natural gas demand. Even using only the constant and HDD as explanatory variables, which correspond to Model 1a in Table 1, we obtain an adjusted R^2 of 0.91. By including other explanatory variables such as time and the holiday dummy, the adjusted R^2 improves only marginally.[5]

---

[5] Including the price of natural gas and the holiday dummy as explanatory variables in the regression model improves the fit only slightly.

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

Dependent Variable: c_t

Model	Constant	HDD	CDD	Hol.D.	Time	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)	

Dependent Variable: $\ln(c_t)$

Model	Constant	HDD	CDD	Hol.D.	Time	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)	

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 holiday effects and time trend. It should also be noted that the Turkish natural gas market price is controlled by EMRA and does not fluctuate often throughout time, which reduces the explanatory power of the natural gas price on consumption. To verify the normality of regression residuals the histograms of the residuals together with superimposed normal density are plotted in Figure 3. Histograms show that even though the normality assumption is not perfectly satisfied, the normality assumption is reasonable for analytical tractability.

Figure 2: Fitted Models 1a, 2a, 1b, and 2b.
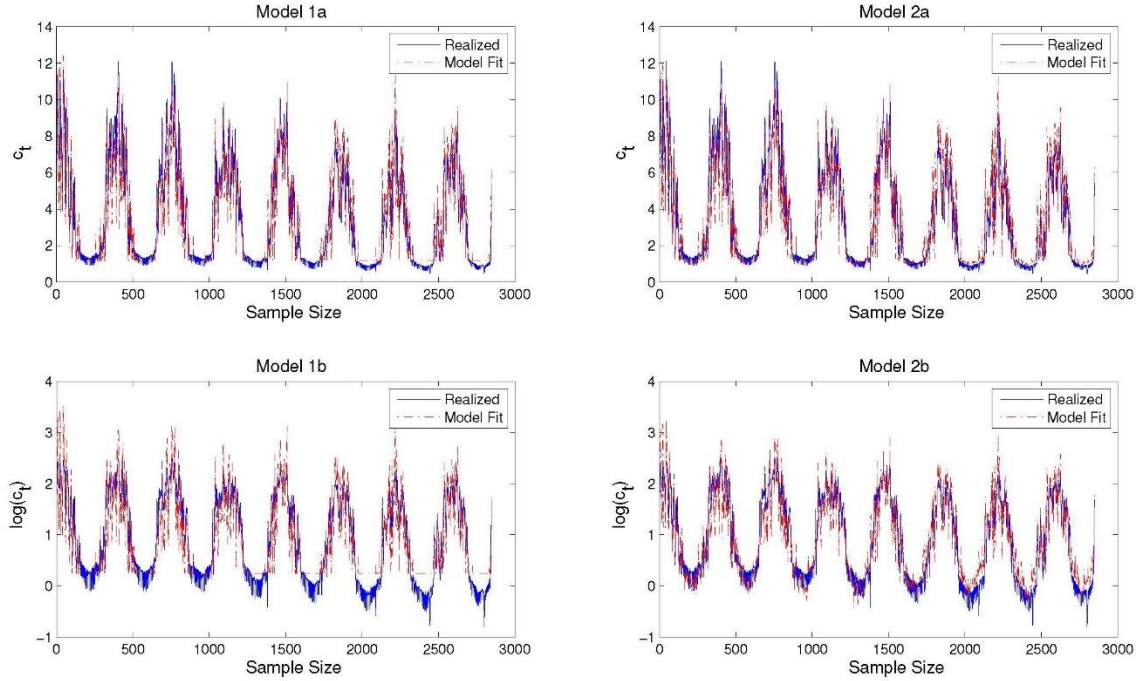
2.3   Temperature Modelling

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 application of this model to daily average temperatures in Istanbul is given in Göncü et al. (2011).
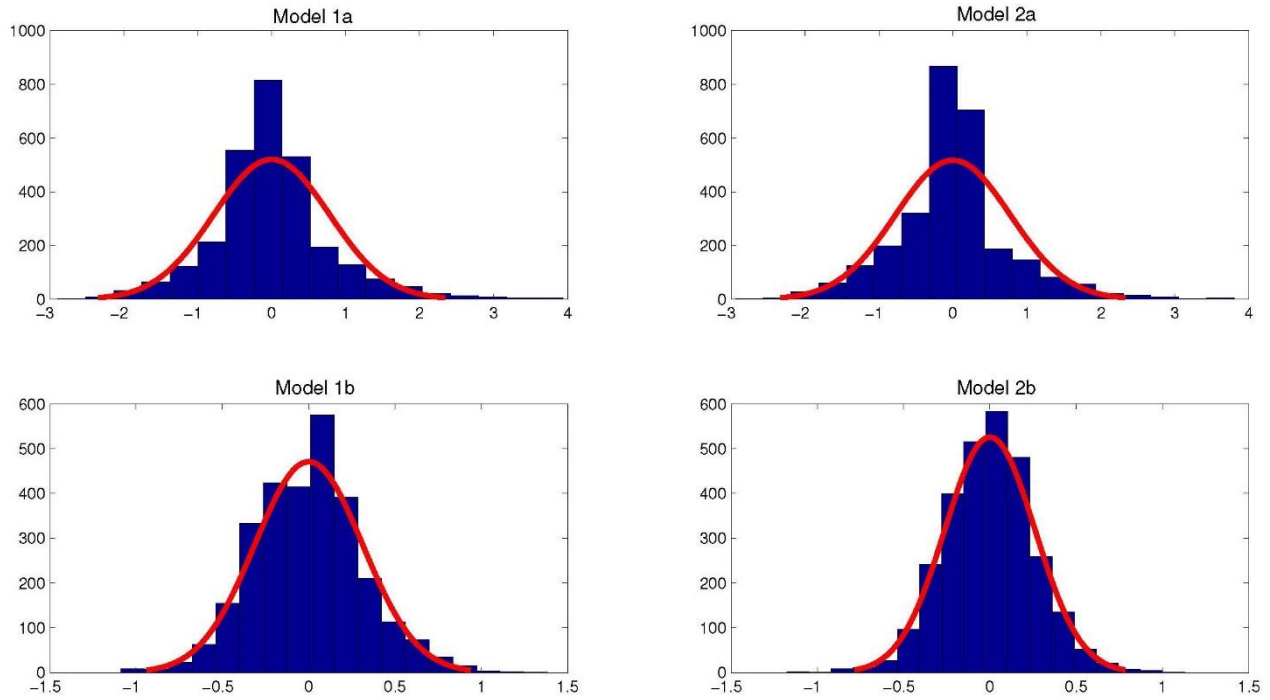
The temperature process is modelled thus:

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

where a is the mean reversion parameter, W_t is P-Brownian motion, and T_t^m represents the long-term mean temperatures.

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

Incorporating market risk, 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 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

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

where $\omega = 2\pi/365$. 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 residuals obtained from the regression model in Equation (3) is provided in Figure 5. Figure 5 confirms that the normality assumption is met.

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

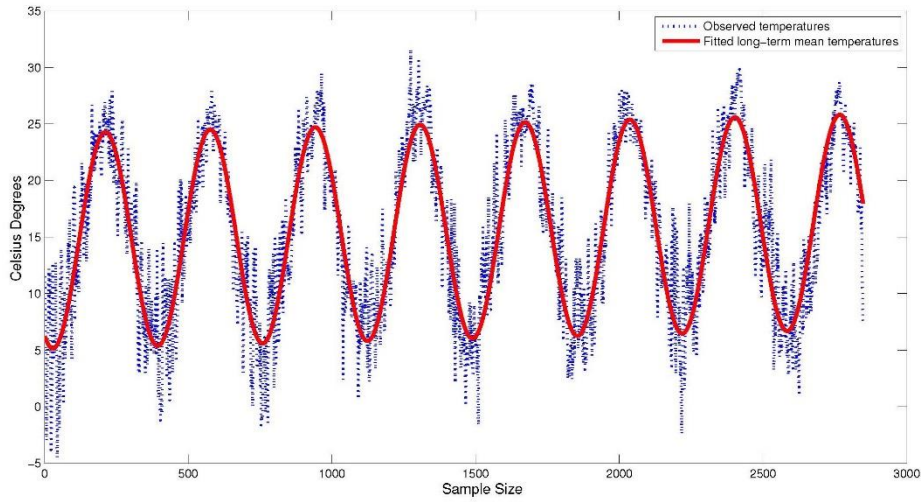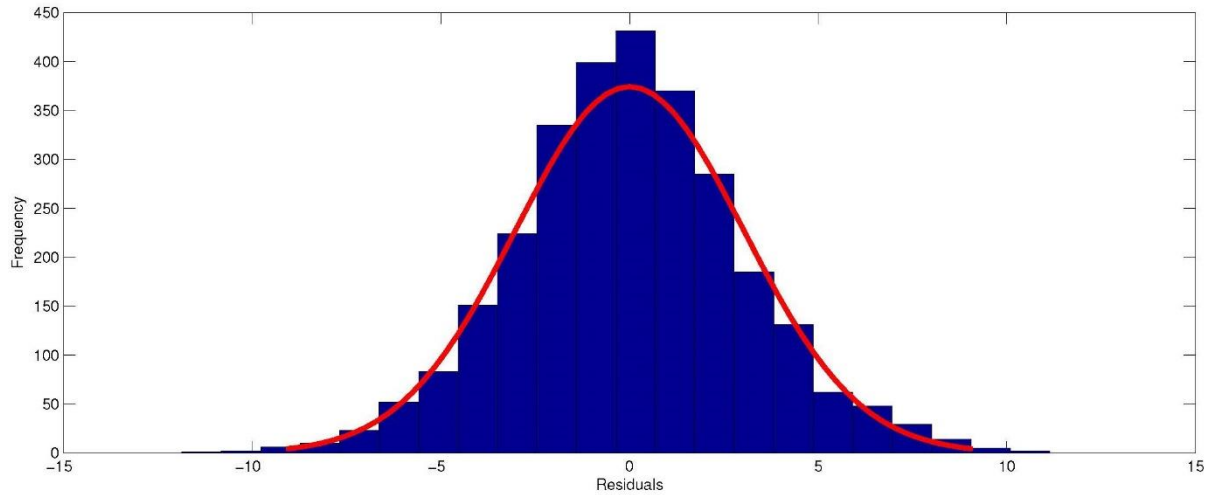|	A	B	C	D	a	R^2
Estimates	14.6463	0.0006	-4.5725	-8.3363	0.2642	0.84
t-statistics	128.4	8.6	56.9	103.1		

Figure 4: Fitted long-term mean temperature function

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

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

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

Then it follows that

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

and

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

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

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

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

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

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 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 natural gas consumption at time t using the results 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 a 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 the previous section. Model 4a is given by:

$$c_t = \beta_0 + \beta_1 HDD_t + \beta_2 CDD_t + + \beta_3 H_t + \beta_4 t + \varepsilon_t, \qquad \varepsilon_t \sim \text{iid } N(0, \sigma_\varepsilon^2) \qquad (9)$$

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

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

and

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

Variance of the conditional natural gas consumption is calculated as follows

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

and

$$\text{var}[c_t | t \notin \chi, \mathcal{F}_s] = \beta_2^2 \text{var}[T_t | \mathcal{F}_s] + \sigma_\varepsilon^2$$
$$(13)$$

denotes the mean and variance of $c_t | t \in \chi$ as $\mu_{c_t,1}$ and $\sigma_{c_t,1}^2$, respectively. Then

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

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 peaking natural gas consumption.

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

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

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

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

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

$$\text{var}[C(t_1, t_M)| \, \mathcal{F}_s] = \sum_{i=1}^{M} \left[\frac{\beta_1^2 \sigma^2}{2a}\left(1 - e^{-2a(t_i - s)}\right)\right] + \sigma_\varepsilon^2 + 2\sum_{i<j} e^{-a(t_j - t_i)} \frac{\sigma^2}{2a}\left(1 - e^{-2a(t_j - t_i)}\right)$$
(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:

$$\log(c_t) = \beta_0 + \beta_1 HDD_t + \beta_2 CDD_t + + \beta_3 H_t + \beta_4 t + \varepsilon_t, \qquad \varepsilon_t \sim \text{iid } N(0, \sigma_\varepsilon^2)$$
(17)

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

$$E_Q[c_t| \, t \in \chi , \mathcal{F}_s] = exp(\mu_{c_t,1} + \sigma_{c_t,1}^2/2)$$
(18)

and

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

Variance is given by

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

and

$$var[c_t| \ t \notin \chi, \ \mathcal{F}_s] = e^{\left(2\mu_{c_t,0}+\sigma^2_{c_t,0}\right)}(e^{\sigma^2_{c_t,0}} - 1)$$
(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 (17) for $t_1 < ... < t_M$ we have

$$E_Q[C(t_1,t_M)| \ \mathcal{F}_s] = \sum_{i=1}^{M}[\beta_0 + \beta_1 E_Q[\text{HDD}_{t_i}|\mathcal{F}_s] + \beta_2 E_Q[\text{CDD}_{t_i}|\mathcal{F}_s] + \beta_3 H_t + \beta_4 t]$$

$$= \sum_{i=1}^{M}[\beta_0 + \beta_3 H_t + \beta_4 t] + \beta_1 F_{HDD}(t_s; t_1, t_M) + \beta_2 F_{CDD}(t_s; t_1, t_M), \quad (22)$$

where $F_{HDD}(t_s; t_1, t_M)$ and $F_{CDD}(t_s; t_1, t_M)$ are the market prices of HDD and CDD futures at time t_s for the measurement period $[t_1, t_M]$, respectively.

For calendar months in which the probability of $CDD > 0$ is negligible (i.e. for $t_1, ..., t_M \in \chi$), then from Equation (10) we obtain the risk neutral expectation of natural gas consumption as:

$$E_Q[C(t_1,t_M)| \ \mathcal{F}_s] = \sum_{i=1}^{M}[\beta_0 + \beta_1 E_Q[\text{HDD}_{t_i}|\mathcal{F}_s] + \beta_2 E_Q[\text{CDD}_{t_i}|\mathcal{F}_s] + \beta_3 H_t + \beta_4 t]$$

$$= \sum_{i=1}^{M}[\beta_0 + \beta_3 H_t + \beta_4 t] + \beta_1 F_{HDD}(t_s; t_1, t_M).$$
(23)

Similarly, from Equation (11) for $t_1, ..., t_M \notin \chi$ where the probability of HDD > 0 is negligible, we obtain

$$E_Q[\text{C}(t_1, t_\text{M})|\,\mathcal{F}_\text{s}] = \sum_{i=1}^{M}[\beta_0 + \beta_1 E_Q[\text{HDD}_{t_i}|\mathcal{F}_s] + \beta_2 E_Q[\text{CDD}_{t_i}|\mathcal{F}_s] + \beta_3 H_t + \beta_4 t]$$

$$= \Sigma_{i=1}^{M}[\beta_0 + \beta_3 H_t + \beta_4 t] + \beta_2 F_{CDD}(t_s; t_1, t_M).$$

(24)

Equation (22) derives the connection between natural gas consumption and HDD/CDD futures for the general case where CDD's and HDD's are observed with positive probabilities, i.e. the transition months. In Equations (23) and (24), we present this relationship for the months where probability of observing positive CDD's and HDD's are negligible, respectively.

3.3   Risk management implications

In the previous section we establish a theoretical link between the expected natural gas consumption and the futures prices of HDD/CDD indices by integrating natural gas demand estimation with a temperature model that is often used for pricing weather derivatives. Using t h e derived relationship in Equation (22), a natural gas company can hedge itself against weather- related changes in natural gas consumption.

Assuming that the price of natural gas is fixed at $\bar{p}$ for $[t_1, t_m]$, say during a given month of the year, revenue of the natural gas company is given by $R(t_1, t_m) = \bar{p}C(t_1, t_m)$. Hence, the expected revenue at time $s \geq t_1$ is given by

$$E[R(t_1, t_m)|\,\mathcal{F}_s] = \bar{p}E[C(t_1, t_m)|\,\mathcal{F}_s] \tag{25}$$

$$E[R(t_1, t_m)|\,\mathcal{F}_s] = \bar{p}(\text{A}(t) + \beta_1 F_{HDD}(t_s; t_1, t_m) + \beta_2 F_{CDD}(t_s; t_1, t_m)), \tag{26}$$

where $A(t) = \Sigma_{i=1}^{M}[\beta_0 + \beta_3 H_t + \beta_4 t]$ is a deterministic function of time.

For example, to hedge the weather-related risk in natural gas consumption, the distributor company can take a short position in $\bar{p}\beta_1/g$ and a long position in $\bar{p}|\beta_2|/g$ units[6] of F_{HDD} and F_{CDD} futures. Note that each futures contract is written on g units of dollars per HDD or CDD.

At maturity $T (= t_m)$, if the realized temperatures were above the risk-neutral expectation levels (say the month was warmer than what the market expected), i.e. $HDD < F_{HDD}$ and $CDD > F_{CDD}$, then the company will obtain a positive payoff from the futures contracts

$$\bar{p}\{\beta_1(F_{HDD} - HDD) + |\beta_2|(CDD - F_{CDD})\}, \tag{27}$$

which is equal to the loss in expected revenues due to warm weather.

---

[6] Note that the coefficient β_2 is negative, so an absolute value is used to represent the number of contracts.

The above example demonstrates that a weather derivatives portfolio can be used to minimize temperature related risks. However, a perfect hedge would require the introduction of new contingent claims on natural gas consumption itself.

4   Backtesting

In this section, we evaluate the forecast performance of these models via the backtesting method at monthly and 10-day forecast horizons. Realized per-consumer natural gas consumption values are compared with the model forecasts from a 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 describe only the backtesting procedure at a 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 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 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 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 the 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 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.

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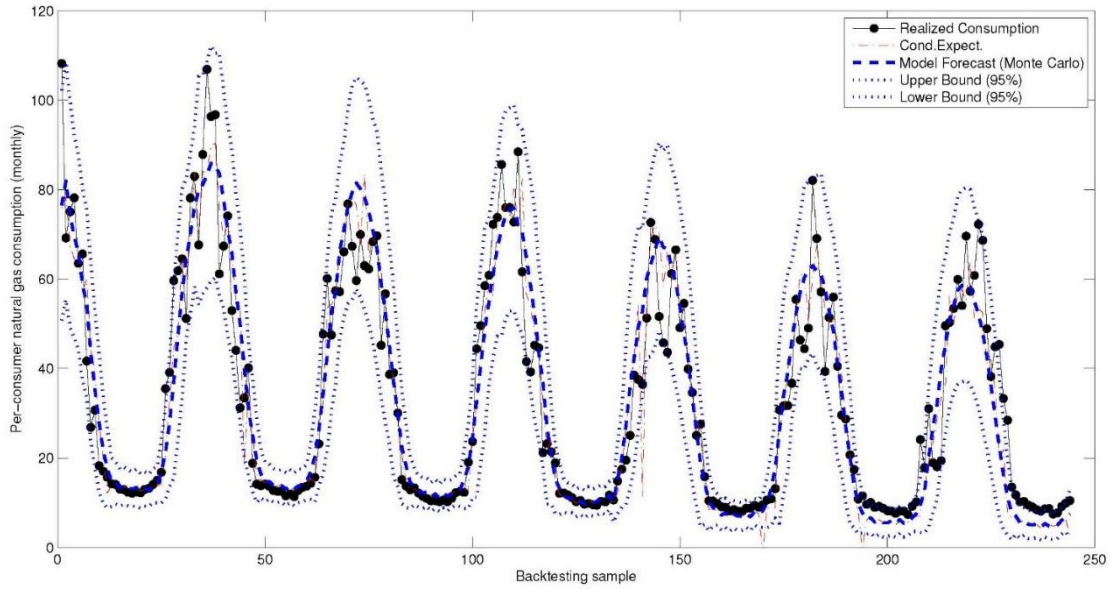
We evaluate the forecast performance of the model by comparing realized and forecasted consumption values for two different forecast horizons, namely 10-day and monthly horizons. We define C_i and $\hat{C_i}$ as

$$C_i = \sum_j c_{i,t_j} \; and \; \hat{C}_i = \sum_j \hat{c}_{i,t_j,}$$
(28)

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

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

$$\frac{1}{N} \sum_{i=1}^{N} \left(\frac{\hat{c}_i - c_i}{c_i} \right)^2,$$
(29)

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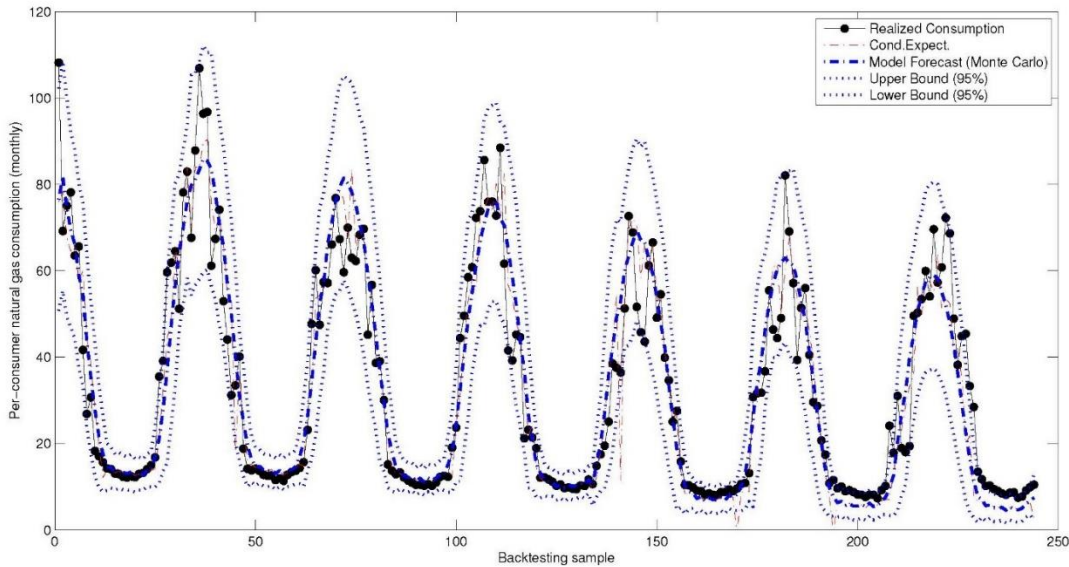
where $i = 1, 2, ..., N$ denotes the number of months used in the backtesting sample, $\hat{C}_i$ represents the model forecast for the ith month backtesting sample, and C_i represents the actual monthly natural gas demand per consumer.

The RMSE results for the considered models are compared in Table 5. Minimum RMSE is achieved in Model 4a for a Monte Carlo simulation and in Model 5b for an analytical solution. A more detailed comparison of Models 4a and 4b is provided in Table 6 at the monthly forecast horizon.[7] As expected, forecasts based on the conditional expectation of natural gas consumption are 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 a Monte Carlo simulation and an analytical solution suggests that - especially for consumption levels (c_t)- forecasting errors are lowest for winter months.

Table 6 suggests that forecasting accuracy of our model is lower for transition months (especially during autumn) compared to winter and summer months. This is mainly due to local climate patterns. In Istanbul, observed summer and winter temperatures are consistently above 18°C and below 18°C, respectively, for each day. However, for transition months, daily temperatures fluctuate around 18°C not only

---

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

within a given month, but also across years, i.e. due to early or late winters. Thus, the use of weather derivatives becomes more critical in locations and/or time periods where temperatures are more volatile.

In Section 3.2, we derive the relationship between natural gas consumption and HDD/CDD futures in Equation (22) which covers the case where we observe both positive HDD's and CDD's with positive probabilities, i.e. the transition months. This suggests that a combination of HDD's and CDD's can be used to hedge the demand risk faced by natural gas suppliers during transition months.

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.

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.

Period	Model 4a		Model 4b	
	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

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. We model daily average temperatures with a mean reverting Ornstein-Uhlenbeck (OU) stochastic process, which is 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 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.

In this paper, we focus on a market structure primarily based on long-term contracts. The presence of a spot market for natural gas enables local suppliers to alleviate the impacts of demand risk by providing an option to readjust their inventories. However, this reduction in demand risk can result in exposures to price and market liquidity risks. Unexpected (negative) temperature shocks are more likely to be region-wide rather than idiosyncratic, which would increase total demand and thus prices. When realized, such risks can be costly to suppliers depending on their ability to transmit price shocks to their customers, which creates incentives to provide demand-side management policies. Furthermore, a natural gas futures market, if it exists, can also be used to hedge this price risk. Overall, based on the climate of the geographical location and costs associated with different risk management options, optimal management of these risks would depend on location-specific characteristics.

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.

Burlig, F., Knittel, C., Rapson, D., Reguant, M. and Wolfram, C., 2017. Machine learning from schools about energy efficiency (No. w23908). National Bureau of Economic Research.

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

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

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.

Gutierrez R., Nafidi A., Sanchez R.G., 2005. Forecasting total natural-gas consumption in Spain by using the stochastic Gompertz innovation diffusion model. Applied Energy, 80(2), 115-124.

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.

M.I.T. Energy Initiative, 2011. The future of natural gas: an interdisciplinary MIT study, http://mitei.mit.edu/system/files/NaturalGas_Report.pdf

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.

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