# A COMPARISON OF STOCHASTIC MODELS OF NATURAL GAS CONSUMPTION: AN APPLICATION FOR TURKEY

SERLİ KİREMİTÇİYAN

BOĞAZİÇİ UNIVERSITY

2014

# A COMPARISON OF STOCHASTIC MODELS OF NATURAL GAS CONSUMPTION: AN APPLICATION FOR TURKEY

Thesis submitted to the

Institute for Graduate Studies in the Social Sciences

in partial fulfillment of the requirements for the degree of

Master of Arts

in

Economics

by

Serli Kiremitçiyan

Boğaziçi University

2014

Thesis Abstract

Serli Kiremitçiyan, ”A Comparison of Stochastic Models of Natural Gas Consumption: An Application for Turkey”

In this thesis, we model the dynamic behaviour of natural gas consumption using continuous-time stochastic models to incorporate their significant advantages over the discrete-time models into the modeling process. In addition to offering a wide set of choices for the drift and volatility terms and yielding analytical solutions for any forecast horizon, continuous-time models can also be used in the pricing of contingent claims depending on natural gas consumption since they enable more reliable forecasts at high-frequency levels. Here, we also document that the per consumer natural gas consumption data exhibit stationarity, strong seasonality, mean reversion, and serial correlation. Hence, we study the application of a One-factor mean-reverting process and stochastic Gompertz diffusion model on the modeling of daily natural gas consumption in Istanbul, Turkey that will also incorporate the empirical observations. In the comparison of their forecasting performances which are tested via the backtesting method at different forecast horizons, we find out that the One-factor mean-reverting process is more advantageous than the Gompertz diffusion process. To illustrate the pricing implications of these models, we price two hypothetical contracts and find out that the results vary from one model to another and hence, the choice of model becomes crucial in the real world.

Tez Özeti

Serli Kiremitçiyan, ”Stokastik Doğal Gaz Tüketim Modellerinin Kıyaslanması: Türkiye Üzerine Bir Uygulama”

Bu tezde, sürekli stokastik modeller aracılığıyla, doğal gaz tüketiminin dinamik davranışı modellenmektedir. Burada, kesikli modellerin aksine, sürekli modellerin sahip olduğu belirli avantajların modelleme sürecine dahil edilmesi hedeflenmektedir. Sürüklenme ve volatilite terimleri için çok sayıda seçenek sunmaları ve herhangi bir tahmin periyodu için geçerli olabilecek olan analitik çözümler önermelerinin yanı sıra, sürekli modeller, yüksek frekanslarda daha güvenilir tahmin yapılmasını kolaylaştırdıklarından, koşullu yükümlülüklerin fiyatlanmasında da kullanılabilmektedirler. Ayrıca, tüketici başına düşen doğal gaz tüketim serisinin durağanlık, ortalamaya dönme, yüksek mevsimsellik etkisi ve serisel korelasyon içerdiği de gösterilmiştir. Gözlemlenen empirik özellikleri dahil edecek şekilde, ortalamaya dönmeli bir stokastik süreç ile stokastik Gompertz difüzyon modelleri önerilmekte ve İstanbul’a ait olan günlük doğal gaz tüketim verilerine uygulanmaktadırlar. Farklı tahmin periyotları için gerçekleştirilen geriye dönük testler aracılığıyla yapılan kıyaslamalarda, ortalamaya dönmeli stokastik sürecin, stokastik Gompertz difüzyon modeline göre, daha avantajlı olduğu sonucuna varılmıştır. Burada incelenen modellerin, fiyatlama aşamasındaki olası sonuçlarını gözlemleyebilmek adına, iki örnek sözleşmenin fiyatlaması gerçekleştirilmiş ve modellere göre sonuçların çok değişken olduğu görülmüştür. Buna dayanarak, fiyatlama aşamasındaki model seçiminin çok önemli olacağı sonucuna varılmıştır.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor, Assist. Prof. Tolga Umut Kuzubaş for his support and guidance in every step of my study. I would also like to thank to Assist. Prof. Ahmet Göncü for his valuable contributions to the thesis from the very beginning. I dedicate this thesis of mine to my family who have always encouraged me in all my life. This thesis would not have been possible without their supports.

CONTENTS

FIGURES

TABLES

# CHAPTER I

# INTRODUCTION

Natural gas is an energy source that constitutes an increasing share in the world's energy resources in terms of both production and consumption figures. Since it is an environmentally friendly fuel that can also act as a viable substitute to other fossil-based energy sources, it is subject to major international trade flows. The recent developments in the conventional natural gas production technologies provide even more incentives that increase the share of natural gas in the total energy mix. As an emerging country with growing energy needs, Turkey's natural gas consumption has increased rapidly over the last two decades. (EPDK, 2013).

The Turkish natural gas industry has undergone a process of reconstruction to create a competitive natural gas market. For this purpose, a new law, the Natural Gas Market Law, was enacted in 2001 which aimed to change the monopolistic structure of the market and form a more competitive one by establishing an independent regulatory authority, the Energy Market Regulatory Authority (EMRA or EPDK in Turkish). However, it is believed that the reform has not worked out as expected so far (Erdoğdu, 2010b). Nevertheless, as the liberalization of the energy market continues, short term forecasting and demand management of natural gas also gain importance. Moreover, following the introduction of competitiveness, it is also very plausible to assume that new contingent claims on the consumption amounts or prices of natural gas will probably become available to meet the needs of market participants.

Accurate modeling and forecasting of natural gas consumption are vital for efficient management of resources which is also suggested by the significant amount of research that has been conducted. A review of the literature on the modeling 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 temperatures. 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, Katırcıoğlu, Aykan, Okumuş and Kılınçer (2001) and Sarak and Satman (2003) utilize degree days to explain the relation between natural gas demand and temperature levels. Erdoğdu (2010a) employs an ARIMA model to forecast natural gas demand using quarterly data over the period of 1988 to 2005.

Other than the time series models, different forecasting tools such as artificial neural networks, genetic algorithms and grey prediction models are also used. For instance, Taşpınar, Çelebi and Tutkun (2013) uses various computational methods, such as a seasonal ARIMA model with exogenous variables (SARIMAX), artificial neural networks and ordinary least squares, to forecast daily natural gas consumption in Turkey. Sabo, Scitovski, Vazler and Zekic-Susac (2011) present mathematical models of natural gas consumption that depend on exponential, Gompertz and logistic models that aim to forecast hourly natural gas consumption on the basis of the past natural gas consumption and temperature data. 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 that aims to capture demand patterns by estimating the seasonality and transitory components as well as the general trend itself. Crompton and Wu (2005) utilize a Bayesian vector autoregressive methodology to forecast the energy demand for China, including the demand for natural gas, using real fuel prices, real GDP and population data.

In the literature, only the studies by Göncü, Karahan and Kuzubaş (2013) and Gutiérrez, Nafidi and Gutiérrez Sánchez (2005b) (similarly, Gutiérrez, Nafidi and Gutiérrez Sánchez (2006)) consider the use of stochastic processes driven by Brownian motion to model natural gas consumption. Using continuous-time stochastic models have 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 and any contingent claim dependent on the path of the natural gas consumption can be priced. Third, the empirical characteristics of the natural gas consumption data can be described via a large choice of drift or volatility terms. Additionally, the forecasts obtained from time series models with high-frequency data would not be reliable since explanatory variables such as the macroeconomic variables or temperatures are difficult to predict at high-frequency levels. All these points indicate that the current lack of focus on continuous-time models in this framework might change and continuous time models might dominate the future of energy modelling.

In this study, we consider the only continuous-time models in the natural gas modeling literature, which belong to Göncü et al. (2013), that adopts the model in Lucia and Schwartz (2002), Gutiérrez et al. (2005b) and Gutiérrez et al.(2006). Lucia and Schwartz (2002) use a One-factor Mean-Reverting process to model the electricity prices and power derivatives in the Nordic electricity market and power exchange. On the other hand, Gutiérrez et al. (2005b) and Gutiérrez et al. (2006) use a Gompertz diffusion process to model the total natural-gas consumption in Spain and the electricity consumption in Morocco, respectively. The models that we use are also driven by Brownian motion noise terms which are represented as stochastic differential equations. The analytical formulas for the conditional expectations and variances are derived which are used in the comparison of the models in terms of their forecasting powers and capacities in capturing the empirical properties of consumption. Finally, they are used in the pricing of new derivative contracts.

The remaining of this paper is structured as follows. In Chapter II, we present the empirical properties of the existing natural gas consumption data and analyze the systematic patterns that evidently exist in the data. From Chapter III through VI, we present details of the proposed models, the One-Factor Mean-Reverting Process and the Stochastic Gompertz Diffusion Model in light of the empirical results, and the pricing of contingent claims that rely on these models. Chapter VII presents the estimation results and compare these models in terms of their forecasting powers and capacities in capturing the empirical properties of consumption. Finally, Chapter VIII concludes.

CHAPTER II

DATA ANALYSIS

The data used in this study is obtained from İGDAŞ, the only natural gas distributor in Istanbul, which spans the time period between January 1, 2004 to October 18, 2011. It contains 2848 daily observations of residential and commercial natural gas consumption in urban areas, excluding the industrial use. The dataset includes the consumption amounts in addition to the number of consumers in this time period. Moreover, natural gas consumption per consumer series is created to account for the effects of the growing population in Istanbul.

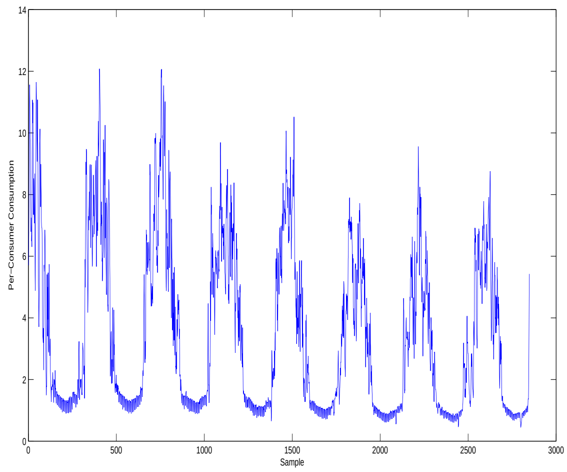


Figure 1. Per consumer daily natural gas consumption

In Figure 1, per consumer natural gas consumption is plotted for the entire sampling period. In contrast, data points specific to warm and cold seasons are labelled in the figure below.

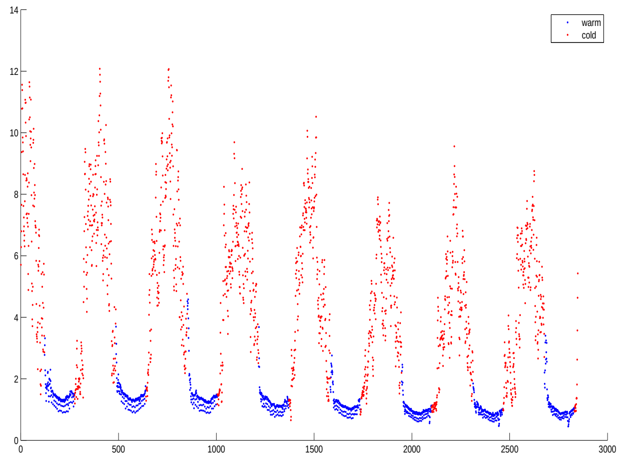


Figure 2. Per consumer daily natural gas consumption (warm and cold seasons)

Here, Figure 2 shows a seasonal pattern and mean reversion to the seasonal means such that these effects seem to be stronger during warm seasons where the natural gas consumption is low and deviations seem to be larger during cold seasons where the consumption is high. Therefore, a suitable model needs to incorporate a clear seasonality with slow mean reversion. Furthermore, as the figure below also suggests, a holiday dummy variable should be included to observe the specific effect of holidays on consumption.

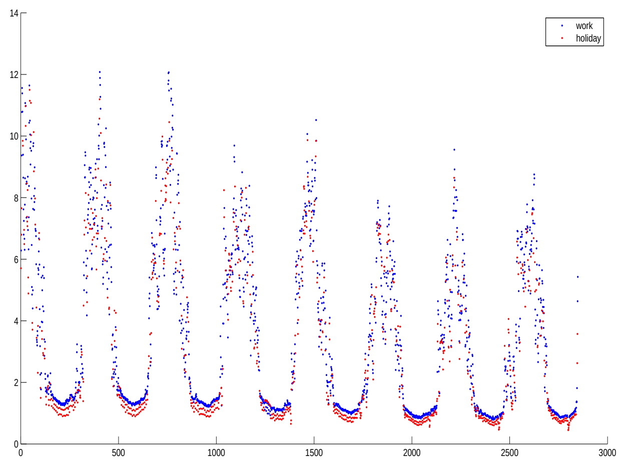


Figure 3. Per consumer daily natural gas consumption (workdays and holidays)

Even though the logarithm of per consumer consumption series will be our variable of interest in the following sections, Table 1 reports the descriptive statistics for both the consumption per consumer series and the logarithm of this.

Table 1. Descriptive Statistics

Panel A: The Entire Dataset

| Series | N.Obs | Mean | Min | Max | Std.Dev. | Skew. | Kurt. |
|---|---|---|---|---|---|---|---|
| $c_t$ | 2848 | 3.4303 | 0.4501 | 12.078 | 2.6887 | 0.8820 | 2.6884 |
| $\ln(c_t)$ | 2848 | 0.9094 | -0.798 | 2.4914 | 0.8181 | 0.1614 | 1.5729 |

Panel B: Warm Seasons

| Series | N.Obs | Mean | Min | Max | Std.Dev. | Skew. | Kurt. |
|---|---|---|---|---|---|---|---|
| $c_t$ | 1224 | 1.2262 | 0.4501 | 4.5795 | 0.4584 | 2.9729 | 17.3297 |
| $\ln(c_t)$ | 1224 | 0.1527 | -0.7981 | 1.5216 | 0.3047 | 0.8224 | 5.3581 |

Panel C: Cold Seasons

| Series | N.Obs | Mean | Min | Max | Std.Dev. | Skew. | Kurt. |
|---|---|---|---|---|---|---|---|
| $c_t$ | 1624 | 5.0915 | 0.6546 | 12.0780 | 2.4694 | 0.2656 | 2.4254 |
| $\ln(c_t)$ | 1624 | 1.4796 | -0.4237 | 2.4914 | 0.5890 | -0.7640 | 2.8225 |

Panel D: Holidays

| Series | N.Obs | Mean | Min | Max | Std.Dev. | Skew. | Kurt. |
|---|---|---|---|---|---|---|---|
| $c_t$ | 904 | 3.2600 | 0.4501 | 11.4998 | 2.6433 | 0.8606 | 2.5968 |
| $\ln(c_t)$ | 904 | 0.8263 | -0.7981 | 2.4423 | 0.8639 | 0.1350 | 1.5499 |

Panel E: Working Days

| Series | N.Obs | Mean | Min | Max | Std.Dev. | Skew. | Kurt. |
|---|---|---|---|---|---|---|---|
| $c_t$ | 1944 | 3.5095 | 0.7865 | 12.0780 | 2.7066 | 0.8905 | 2.7133 |
| $\ln(c_t)$ | 1944 | 0.9480 | -0.2401 | 2.4914 | 0.7932 | 0.2060 | 1.5390 |

In Table 1, the dataset is divided into sections as warm/cold seasons and holidays/working days to form an idea about the empirical properties of consumption and log-consumption series which can be variant within different sampling periods. Similar to the convention followed by Lucia and Schwartz (2002), warm seasons include observations from May 1 to September 30 for every year in the sampling period.

Both the per consumer consumption itself and the logarithm of this series show higher daily means in cold seasons than in warm seasons. Similarly, the standard deviations for the warm and cold seasons show significant differences such that they are more volatile in cold seasons than in warm seasons. Since the logarithm of per consumer consumption will be considered to be normally distributed in the forthcoming models, the kurtosis estimates for the log-consumption series are critical. The kurtosis estimates for the log-consumption series are lower than three under the null hypothesis of normality, except in warm seasons which indicate that extremely low and high consumptions are more probable in warm seasons. The skewness estimates for the consumption series are all positive which indicate that high extreme values are more probable than low extreme values. When the estimates for the entire dataset are compared with the estimates from holidays/working days categories, we observe that the moments of the entire dataset remain in between those regarding holidays (being the lower ones) and working days (being the higher ones).

Nonstationarity tests are conducted for the consumption itself and the logarithm of this series using Augmented Dickey-Fuller (ADF) test and Phillips-Perron (PP) unit root test with drift. The table below shows the values of the t-statistics under the null hypothesis of a unit root using a certain number of lags. Both tests reject the presence of a unit root at 99% confidence level. Hence, these series are stationary which also supports the use of a mean reverting process.

Table 2. Results of Statistical Tests for the Entire Dataset

ADF test statistics

| Series | 1 lag | 8 lags | 15 lags | 22 lags | 29 lags | Critical Value |
|---|---|---|---|---|---|---|
| $c_t$ | -6.8094 | -4.5912 | -3.5862 | -3.6698 | -3.7445 | -3.4363 |
| $\ln(c_t)$ | -5.4374 | -3.5713 | -3.1443 | -3.3158 | -3.7762 | -3.4363 |

PP unit root with drift test statistics

| Series | 1 lag | 8 lags | 15 lags | 22 lags | 29 lags | Critical Value |
|---|---|---|---|---|---|---|
| $c_t$ | -6.0807 | -5.3601 | -4.7877 | -4.5946 | -4.7394 | -3.4363 |
| $\ln(c_t)$ | -5.1369 | -4.2604 | -3.8949 | -3.8996 | -4.0718 | -3.4363 |

The entire dataset of per consumer log-consumption series displays some predictability in the autocorrelation and partial autocorrelation functions (ACF and

PACF, respectively).

In Figure 4, the sample partial autocorrelation function of the log-natural gas consumption shows the existence of strong serial correlation with a highly significant first order lag.

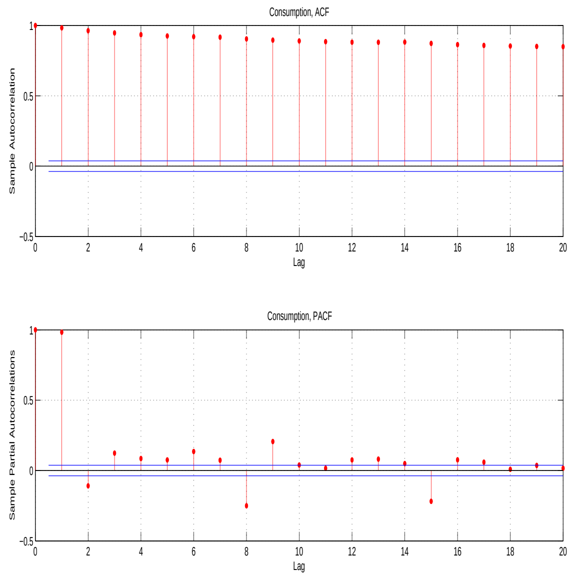


Figure 4. ACF and PACF of daily, per consumer log-consumption series

In the following sections, these empirical observations will be incorporated into continuous-time models.

# CHAPTER III

# PRELIMINARIES

## Brownian Motion

Brownian Motion or the Wiener process,$\{W_t, t \geq 0\}$, is a one-dimensional Gaussian process with continuous paths and independent increments following the necessary and sufficient conditions that are listed below:

i) $W_t$ is continuous and $W_0 = 0$ with probability one,

ii) $W_t \sim N(0, t)$ under $\mathbb{P}$ measure ($\mathbb{P}$-Brownian Motion),

iii) $W_t - W_s \sim N(0, t - s)$ under $\mathbb{P}$ and is independent of the filtration $\{\mathcal{F}_s\}$,

i.e., the increments $W_t - W_s$ are stationary and independent for $t \geq s$ (Baxter and Rennie, 2003, Chapter 3). When a fixed discrete time interval $\Delta t > 0$ is considered, a trajectory for the Wiener process can be simulated in the following way:

$$W(\Delta t) = W(\Delta t) - W(0) \sim N(0, \Delta t) \sim \sqrt{\Delta t} N(0, 1)$$

and for any other increment,

$$W(t + \Delta t) - W_t \sim N(0, \Delta t) \sim \sqrt{\Delta t} N(0, 1)$$

## Stochastic Processes

A stochastic process $X$ is a continuous process $\{X_t, t \geq 0\}$ such that it can be written as

$$X_t = x_0 + \int_0^t \mu_u du + \int_0^t \sigma_u dW_u \tag{1}$$

which can also be expressed in the differential form as

$$dX_t = \mu_t dt + \sigma_t dW_t \tag{2}$$

where $\mu$ and $\sigma$ are random $\mathcal{F}$-previsible processes such that $\int_0^t (\sigma_u^2 + |\mu_u|)du$ is finite for all times t with probability one (Baxter and Rennie, 2003, Chapter 3).

Diffusion Processes

Diffusion processes $\{X_t, t \geq 0\}$ solve stochastic differential equations of the form:

$$dX_t = a(t, X_t)dt + b(t, X_t)^{1/2} dW_t, \quad X_0 = x_0 \tag{3}$$

with the non-random initial condition $x_0$. The solution of this equation is expressed as:

$$X_t = x_0 + \int_0^T a(u, X(u))du + \int_0^T b(u, X(u))^{1/2} dW_u \tag{4}$$

The two deterministic functions $a(t, X_t)$ and $b(t, X_t)$ represent drift and diffusion coefficients of the stochastic differential equation in Equation (3), respectively. Since Equation (4) is an Itô drift-diffusion process, these coefficients should be measurable such that:

$$P\left\{\int_0^T \sup_{|x| \leq R} (|a(t, x)| + b(t, x))dt < \infty\right\} = 1 \tag{5}$$

for all $T, R \in [0, \infty)$.

Assumption (Global Lipschitz): For all $x, y \in \mathbb{R}$ and $t \in [0, T]$, there exists a constant $K < +\infty$ such that

$$|a(t,x) - a(t,y)| + |b(t,x)^{1/2} - b(t,y)^{1/2}| < K|x-y| \tag{6}$$

The following linear growth condition ensures that the solution $X_t$ does not explode in a finite time.

Assumption (Linear Growth): For all $x, y \in \mathbb{R}$ and $t \in [0, T]$, there exists a constant $C < +\infty$ such that

$$|a(t,x)| + |b(t,x)| < C(1 + |x|) \tag{7}$$

Under these assumptions (Equations (6) and (7)), the stochastic differential equation in Equation (3) will have a unique, continuous, and adapted strong solution such that

$$\mathbb{E}\left\{\int_0^T |X_t|^2 dt\right\} < \infty$$

The fact that the solution $X$ is of strong type implies the pathwise uniqueness of the result. The weak solution obtained under the local versions of the same conditions imply that any two weak solutions $X(1)$ and $X(2)$ are not necessarily pathwise identical while their distributions are, which is still enough for likelihood inference. The following equations are the local versions of the same conditions.

Assumption (Local Lipschitz): For any $N < \infty$, $|x|, |y| \leq N$ and $t \in [0, T]$, there exist a constant $L_N > 0$ such that

$$|a(t,x) - a(t,y)| + |b(t,x) - b(t,y)| < L_N|x - y|$$

and

$$2xa(t,x) + b(t,x) < A(1 + x^2)$$

which is only a condition that limits the growth of the diffusion coefficient. With the initial condition $X_0 = x_0$, the following condition holds for some $A$:

$$xa(t,x_0) + b(t,x_0) \leq A(1 + x_0^2)$$

where $a(.)$ is locally bounded and $b(.)$ is continuous and positive. Under this assumption, the stochastic differential equation in Equation (3) will have a unique weak solution (Iacus, 2008, Chapter 1).

## Change of Measure and the Radon-Nikodým Derivative

Given $\mathbb{P}$ and $\mathbb{Q}$ equivalent measures and a time horizon $T$, a random variable $\frac{d\mathbb{Q}}{d\mathbb{P}}$ defined on $\mathbb{P}$-possible paths that takes positive real values can be defined such that:

(i) $\mathbb{E}_{\mathbb{Q}}(X_T) = \mathbb{E}_{\mathbb{P}}(\frac{d\mathbb{Q}}{d\mathbb{P}}X_T)$, for all claims knowable by time T

(ii) $\mathbb{E}_{\mathbb{Q}}(X_T|\mathcal{F}_s) = \zeta_s^{-1}\mathbb{E}_{\mathbb{P}}(\zeta_t X_t|\mathcal{F}_s), s \leq t \leq T$

where $\zeta_t$ is the process $\mathbb{E}_{\mathbb{P}}(\frac{d\mathbb{Q}}{d\mathbb{P}}|\mathcal{F}_s)$, the Radon-Nikodým derivative (Baxter and Rennie, 2003, Chapter 3).

## Girsanov's Theorem

Girsanov's Theorem is a change-of-measure theorem for stochastic processes where the original measure is changed to an equivalent probability measure. When the following two stochastic differential equations are considered and the probability measures for these equations are denoted by $\mathbb{P}$ and $\mathbb{Q}$, respectively:

$$dX_t = a_1(t, X_t)dt + b(t, X_t)^{1/2}dW_t$$
$$dX_t = a_2(t, X_t)dt + b(t, X_t)^{1/2}dW_t$$

Assuming that the coefficients satisfy the assumptions in Equations (6) and (7) with non-random initial values that equal to the same constant, the two probability measures $\mathbb{P}$ and $\mathbb{Q}$ are equivalent and the corresponding Radon-Nikodým derivative becomes:

$$\frac{d\mathbb{Q}}{d\mathbb{P}}(X) = \exp\left\{\int_0^T \frac{a_2(s, X_s) - a_1(s, X_s)}{b(s, X_s)}dX_s - \frac{1}{2}\int_0^T \frac{a_2(s, X_s)^2 - a_1(s, X_s)^2}{b(s, X_s)}ds\right\} \tag{8}$$

The significance of this derivative is that inferences from diffusion processes will be carried out after obtaining the above likelihood ratio from the corresponding Radon-Nikodým derivative (Baxter and Rennie, 2003, Chapter 3).

## Kolmogorov Equations

Transformation of diffusion processes via Kolmogorov equations is of great interest in stochastic calculus. Especially, the transformation to a Wiener process is particularly important since it is used to derive transition probability density functions. Let $X_t$ be a one-dimensional diffusion process defined on a finite or infinite interval $I$ where the transition density is defined from the value $y$ at time $s$ to value $x$ at time $t$ by $f(y, s|x, t)$ where

$$f(y, s|x, t) = \frac{\partial}{\partial y} P\{X_s \leq y, X_t = x\} \tag{9}$$

From the Markovian property of diffusion, we can define the transition density function that satisfies the two Kolmogorov equations, which are known as the Kolmogorov forward:

$$\frac{\partial}{\partial t} f(y, s|x, t) = -\frac{\partial(a(t, x) f(y, s|x, t))}{\partial x} + \frac{1}{2} \frac{\partial^2 (b(t, x) f(y, s|x, t))}{\partial x^2} \tag{10}$$

and the Kolmogorov backward equations:

$$-\frac{\partial}{\partial s} f(y, s|x, t) = a(s, y) \frac{\partial f(y, s|x, t)}{\partial y} + \frac{1}{2} b(s, y) \frac{\partial^2 f(y, s|x, t)}{\partial y^2} \tag{11}$$

where $a(., .)$ and $b(., .)$ are the drift and diffusion coefficients (Iacus, 2008, Chapter 1).

One-dimensional diffusion processes whose transition probability density function satisfies the Kolmogorov equations can be derived from the Wiener process. The necessary and sufficient conditions as well as the conditions that ensure the existence of such transformations are derived by Ricciardi(1976).

The required transformation that will change Equation (3) into the Wiener process $\{W(t'), t' \geq t'_0\}$, which is defined on the interval obtained by applying the following transformations on interval $I$, is defined as:

$$x' = \psi(t, x) \tag{12}$$

$$t' = \varphi(t) \tag{13}$$

Setting,

$$f(y, s|x, t) = \frac{\partial \psi(s, y)}{\partial y} f'(y', s'|x', t') = f'(\psi(s, y), \varphi(s)|\psi(t, x), \varphi(t)) \frac{\partial \psi(s, y)}{\partial y} \tag{14}$$

in which case Equation (10) becomes:

$$\frac{\partial}{\partial t'} f'(y, s|x, t) + \frac{\partial^2}{\partial y'^2} f'(y, s|x, t) = 0 \tag{15}$$

The most general form of transformation is expressed in Ricciardi's Theorem. If and only if

$$a(t, x) = \frac{1}{4} \frac{\partial b(t, x)}{\partial x} + \frac{b(t, x)^{1/2}}{2} \left\{ c_1(t) + \int_z^x \frac{c_2(t) b(t, y) + \frac{\partial b(t, y)}{\partial y}}{[b(t, y)]^{3/2}} dy \right\} \tag{16}$$

where $c_1(t)$ and $c_2(t)$ are arbitrary functions with only time dependence, then there exists a transformation as in Equation (12). This transformation is then given by:

$$\begin{aligned} \psi(t, x) = (k_1)^{1/2} \exp\left[ -\frac{1}{2} \int_0^t d\tau c_2(\tau) \right] \int_z^x \frac{dy}{[b(t, y)]^{1/2}} \\ -\frac{(k_1)^{1/2}}{2} \int_{t_2}^t d\tau c_1(\tau) \exp\left[ -\frac{1}{2} \int_0^\tau d\theta c_2(\theta) \right] + k_2 \end{aligned} \tag{17}$$

and

$$\varphi(t) = k_1 \int_{t_1}^{t} d\tau \exp\left[-\int_0^{\tau} d\theta c_2(\theta)\right] + k_3 \qquad (18)$$

where $z \in I$, $t_i \in [0, \infty)$ and $k_i$'s are arbitrary constants with only the restriction that $k_1 > 0$.

The expression of the density function for the Wiener process, i.e. the normal distribution function, provides the final equation (Ricciardi, 1977) :

$$f(y, s|x, t) = (2\pi[\varphi(s) - \varphi(t)])^{1/2} \exp\left\{-\frac{[\psi(s, y) - \psi(t, x)]^2}{2[\varphi(s) - \varphi(t)]}\right\} \frac{\partial\psi(s, y)}{\partial y} \qquad (19)$$

CHAPTER IV

# ONE-FACTOR MEAN-REVERTING PROCESS

The Ornstein-Uhlenbeck process incorporates a mean reverting behaviour into a stochastic model which is driven by Brownian motion. The one-dimensional process $\{Y_t, t \in [0, T]\}$ is defined as the solution of the following stochastic differential equation:

$$dY_t = \kappa(\theta - Y_t)dt + \sigma dW_t, \quad Y_0 = y_0 \tag{20}$$

where $\kappa$, $\theta$ and $\sigma \geq 0$ are real constants, and $\{W_t, t \in [0, T]\}$ is the standard Brownian motion defined on the probability space $(\Omega, \mathcal{F}, P)$ with filtration $\{\mathcal{F}_t\}_0^\infty$. The parameter $\theta$ shows the long run mean reversion level which will equal to zero in our framework. Thus, deviations from the long run level are corrected at level $\kappa$ and subject to random deviations of magnitude $\sigma$ (Maller et al., 2009).

$$dY_t = -\kappa Y_t dt + \sigma dW_t, \quad Y_0 = y_0 \tag{21}$$

where $\kappa$ is the speed of mean reversion and $\sigma$ is the volatility of the process. Contrary to Brownian motion, this process has a finite variance for all $t \geq 0$. By applying the Itô's formula to the transformation $Y_t e^{\kappa t}$:

$$d(Y_t e^{\kappa t}) = Y_t \kappa e^{\kappa t} dt + e^{\kappa t}\, dY_t = \sigma e^{\kappa t} dW_t.$$

Integrating both sides:

$$Y_t e^{\kappa t} = y_0 + \int_0^t \sigma e^{\kappa s} W_s$$

From this, the solution of Equation (21) for any $Y_s = y_s$ is obtained as:

$$Y_t = y_s e^{-\kappa(t-s)} + \int_s^t e^{-\kappa(t-u)} \sigma dW_u \tag{22}$$

and thus, $Y_t \sim N\left(y_s e^{-\kappa(t-s)}, \int_s^t e^{-2\kappa(t-u)} \sigma^2 du\right)$. Since the volatility of the process is assumed to be constant (Iacus, 2008, Chapter 1) :

$$\int_s^t e^{-\kappa(t-u)} \sigma^2 dW_u = \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa(t-s)})$$

When $\kappa > 0$, $\lim_{t\to\infty} Y_t$ will be finite and equal to $\frac{\sigma^2}{2\kappa}$. Since $W_t$ has stationary independent increments, $Y_t$ is a stationary Markovian process. When $\kappa \leq 0$, Equation (22) will not be stationary since $\int_s^t e^{-\kappa(t-u)} dW_u$ will tend to infinity in probability as $t \to \infty$ (Maller, Müller and Szimayer, 2009)

One-Factor Mean-Reverting Process with a Deterministic Function

Following the one-factor model in Göncü et al. (2013), the per consumer natural gas consumption $\{X_t, t \in [0, T]\}$ is modelled as a combination of a seasonal deterministic component and a mean reverting stochastic process which is specified in the exponential form as:

$$X_t = \exp(f(t) + Y_t) \tag{23}$$

where $f(t)$ is a bounded deterministic function of time and $Y_t$ follows Equation (22). From Equation (23), we can rewrite Equation (21) as:

$$d(\ln(X_t) - f(t)) = -\kappa(\ln(X_t) - f(t))dt + \sigma dW_t \tag{24}$$

which shows that when $\ln(X_t)$ deviates from the deterministic term $f(t)$, it tends to move back to it at a rate proportional to its deviation. Applying Itô's formula, the stochastic differential equation for the natural gas consumption is derived as:

$$dX_t = \kappa(d(t) - \ln(X_t))X_t dt + \sigma X_t dW_t, \tag{25}$$

where $d(t) = \frac{1}{\kappa}\left(\frac{\sigma^2}{2} + \frac{df(t)}{dt}\right) + f(t)$. From this, $X_t$ can be obtained as:

$$X_t = f(t) + y_s e^{-\kappa(t-s)} + \int_s^t e^{-\kappa(t-u)} \sigma dW_u \tag{26}$$

Since $\ln(X_t)$ is is normally distributed because of the normality of $Y_t$, $X_t$ is log-normally distributed. Then, the conditional expectation and variance of Equation (26) with respect to the filtration $\mathcal{F}_s$ at time $0 < s < t$ become equal to:

$$
\begin{aligned}
E[X_t|\mathcal{F}_s] &= \exp\left(E[\ln(X_t)|\mathcal{F}_s] + \frac{1}{2}var(\ln(X_t)|\mathcal{F}_s)\right) \\
&= \exp\left(f(t) + (\ln X_s - f(s))e^{-\kappa(t-s)} + \frac{\sigma^2}{4\kappa}(1 - e^{-2\kappa(t-s)})\right)
\end{aligned} \tag{27}
$$

and

$$
\begin{aligned}
var(X_t|\mathcal{F}_s) &= E[X_t|\mathcal{F}_s]^2 \times \exp\left(var(\ln(X_t)|\mathcal{F}_s) - 1\right) \\
&= \exp\left(2f(t) + 2(\ln X_s - f(s))e^{-\kappa(t-s)} + \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa(t-s)})\right) \times \\
&\left[\exp\left(\frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa(t-s)})\right) - 1\right]
\end{aligned} \tag{28}
$$

respectively. From these, we observe that the conditional mean converges to $f(\infty) + \frac{\sigma^2}{4\kappa}$ in the long run and the conditional variance converges to a function of the value of the deterministic component at infinity.

## The Deterministic Function

The deterministic function $f(t)$ aims to capture the predictable behaviour depending on the nature and characteristics of per consumer natural gas consumption time series. Following Lucia and Schwartz (2002), we capture the seasonality in natural gas consumption with the specification below:

$$
f(t) = \beta_0 + \beta_1 H_t + \sum_{i=1}^{p} \alpha_i \sin(iwt) + \gamma_i \cos(iwt), \tag{29}
$$

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 $p$ shows the number of sine and cosine terms in the Fourier expansion which are expected to reflect the seasonal pattern prevalent in the data. $\beta_0$, $\beta_1$, $\alpha_i$ and $\gamma_i$'s are all constant parameters.

## Estimation of the One-Factor Model

Since discrete observations are used in reality, we need to express this model in the discrete form. When we discretize Equation (21), substitute $f(t)$ from Equation (29) and take the number of sine and cosine terms as $p = 2$, we obtain:

$$
\begin{aligned}
\ln(X_t) &= \beta_0 + \beta_1 H_t + \sum_{i=1}^{2} \alpha_i \sin(iwt) + \gamma_i \cos(iwt) + Y_t \\
Y_t &= \phi Y_{t-1} + u_t, \qquad u_t \sim N(0, \sigma^2)
\end{aligned}
\tag{30}
$$

for $t = 0, 1, 2, ..., T$ with i.i.d. innovations $u_t$ that are normal random variables with mean zero and variance $\sigma^2$. Equivalently, this model can be written as:

$$
\begin{aligned}
\ln X_t &= z_t = \Theta(\Phi, \nu_t) + \epsilon_t \\
\epsilon_t &= \phi \epsilon_{t-1} + u_t, \qquad u_t \sim N(0, \sigma^2)
\end{aligned}
\tag{31}
$$

where $\Theta$ is a function of the vector of explanatory variables $\nu_t$ and vector of parameters $\Phi$, and $z_t$ is the dependent variable. By rearranging, we obtain the following Autoregressive Distributed Lag (ADL) model with lags of the deterministic component and the first order lag of consumption:

$$z_t = \phi z_{t-1} + \Theta(\Phi, \nu_t) - \Theta(\Phi, \nu_{t-1}) + u_t, \tag{32}$$

where the parameters are estimated simultaneously by using Non-Linear Least Squares procedure and the mean reversion parameter will be given as $\hat{\kappa} = 1 - \hat{\phi}$ (Göncü et al., 2013).

# CHAPTER V

# GOMPERTZ DIFFUSION PROCESS

## Deterministic Gompertz Model

The deterministic Gompertz model is a continuous-time model that is used to describe diverse growth phenomena in nature. The process $\{X_t, t \in [0, T]\}$ that is known to follow this model can be expressed as:

$$dX_t = (\alpha - \beta \ln X_t)X_t dt, \quad X_0 = x_0 \tag{33}$$

where the parameter $\alpha$ shows the intrinsic growth rate and $\beta$ shows the growth deceleration factor that need to be estimated from the realizations of $X_t$ for $t \in [0, T]$. For any $X_s = x_s$ at any given point in time with predetermined values of $\alpha$ and $\beta$, the solution of Equation (33) is:

$$X_t = \exp\left(\ln(X_s)e^{-\beta(t-s)} + \frac{\alpha}{\beta}(1 - e^{-\beta(t-s)})\right). \tag{34}$$

which is also named as the "Gompertz Curve".

The problem with the deterministic Gompertz model is that it does not take into account random fluctuations that almost surely exist in the realizations of $X_t$ and does not represent the heteroscedasticity prevalent in the process. Thus, an extension of this model is necessary such that a stochastic model is obtained (Ferrante, Bompadre, Possati and Leone, 2000).

Randomness can be taken into account in a number of ways, either by assuming a time-variant intrinsic growth rate or growth deceleration factor. We incorporate randomness into Equation (33) by assuming a time-variant intrinsic growth rate, while keeping the growth deceleration factor constant.

Time-variance is introduced into the intrinsic growth rate in the following way:

$$\theta_t = a + \sigma\eta_t \tag{35}$$

where $a$ is the constant mean of $\theta_t$, $\sigma > 0$ is the diffusion coefficient and $\eta_t$ is a Gaussian white noise process (Ferrante et al., 2000). Accordingly, the stochastic Gompertz diffusion model can be described by the following one-dimensional stochastic differential equation:

$$dX_t = a(t, X_t)dt + b(t, X_t)^{1/2}dW_t, \quad X_0 = x_0 \tag{36}$$

where $\{W_t, t \in [0, T]\}$ is a standard Brownian motion process defined on the probability space $(\Omega, \mathcal{F}, P)$ with filtration $\{\mathcal{F}_t\}_0^\infty$. We assume that, for all $t \in [0, T]$, each observation of $X_t$, $x_t \in (0, \infty)$ (Gutiérrez, Nafidi and Gutiérrez Sánchez, 2004).

The drift coefficient $a(t, x)$ can be expressed with or without exogenous factors that may also affect the behaviour of the process. The case where we do not include exogenous factors will be referred as the stochastic 'Homogenous' Gompertz Diffusion Model and the opposite case will be referred as the stochastic

'Non-Homogenous' Gompertz Diffusion Model.

## Stochastic Homogenous Gompertz Diffusion Model

In this section, the One-dimensional, Homogeneous Gompertz diffusion process is introduced by the infinitesimal moments and the corresponding Kolmogorov equations.

In this model, the drift coefficient $a(t,x)$ and the diffusion coefficient $b(t,x)$ in Equation (36) are real-valued functions which are given as:

$$a(t,x) = \alpha x - \beta x \ln(x) \tag{37}$$

$$b(t,x) = \sigma^2 x^2 \tag{38}$$

where $\alpha, \beta$ and $\sigma$ are real-valued parameters. After substitution, we obtain the following stochastic differential equation:

$$dX_t = (\alpha - \beta \ln X_t)X_t dt + \sigma X_t dW_t, \quad X_0 = x_0 \tag{39}$$

Equation (39) has a unique solution $\{X_t, t \in [0,T]\}$ that is obtained by applying the Itô's formula to the transformation $e^{\beta t} \ln X_t$ (Gutiérrez et al., 2004).

$$
\begin{aligned}
de^{\beta t}\ln X_t &= \beta e^{\beta t}\ln(X_t)dt + e^{\beta t}\frac{1}{X_t}dX_t - \frac{1}{2}e^{\beta t}\frac{1}{X_t^2}(dX_t)^2 \\
&= e^{\beta t}\left(\alpha - \sigma^2/2\right)dt + e^{\beta t}\sigma dW_t
\end{aligned}
$$

Integrating both sides:

$$
e^{\beta t}\ln X_t = e^{\beta s}\ln(x_s) + \frac{(\alpha - \sigma^2/2)}{\beta}(e^{\beta t} - e^{\beta s}) + \sigma\int_s^t e^{b\tau}dW_\tau
$$

From this, the solution of Equation (39) for any $X_s = x_s$ can be obtained as:

$$
X_t = \exp\left(\ln(x_s)e^{-\beta(t-s)} + \frac{(\alpha - \sigma^2/2)}{\beta}(1 - e^{-\beta(t-s)}) + \sigma\int_s^t e^{-\beta(t-\tau)}dW_\tau\right). \tag{40}
$$

The conditional expectation under this process is given by:

$$
\begin{aligned}
\mathbb{E}[X_t|\mathcal{F}_s] = \exp&\left(\ln x_s e^{-\beta(t-s)} + \frac{(\alpha - \sigma^2/2)}{\beta}(1 - e^{-\beta(t-s)})\right) \\
&.\mathbb{E}\left[\exp\left(\sigma\int_s^t e^{-\beta(t-\tau)}dW_\tau\right)\right]
\end{aligned}
$$

Since $W_t - W_s \sim N(0, t-s)$,

$$
\sigma\int_s^t e^{-\beta(t-\tau)}dW_\tau \sim N\left(0, \sigma^2\int_s^t e^{-2\beta(t-\tau)}d\tau\right)
$$

Hence,

$$\exp\left(\sigma\int_s^t e^{-\beta(t-\tau)}dW_\tau\right) \sim \Lambda\left(0, \exp\left(\frac{\sigma^2}{2}\int_s^t e^{-2\beta(t-\tau)}d\tau\right)\right)$$

where $\Lambda$ indicates the log-normal distribution. We obtain the final form of the conditional expectation function as:

$$\mathbb{E}[X_t|\mathcal{F}_s] = \exp\left(\ln x_s e^{-\beta(t-s)} + \frac{(\alpha-\sigma^2/2)}{\beta}(1-e^{-\beta(t-s)}) + \frac{\sigma^2}{4\beta}(1-e^{-2\beta(t-s)})\right) \tag{41}$$

Analysis through Kolmogorov Equations

In this section, the Kolmogorov equations are used to derive the transition probability density function. Hence, the One-dimensional Homogeneous Gompertz diffusion process is introduced by the corresponding Kolmogorov equations.

The diffusion process $\{X_t, t \in [0,T]\}$ can be expressed by the following density function:

$$f(y,s|x,t) = \frac{\partial}{\partial y}P\{X_s \leq y, X_t = x\} \tag{42}$$

with infinitesimal moments that are defined as in Equations (37) and (38). The Kolmogorov equations corresponding to the defined process are expressed in the following form:

$$\frac{\partial}{\partial t}f(y,s|x,t) = -\frac{\partial((\alpha x - \beta x\ln(x))f(y,s|x,t))}{\partial x} + \frac{1}{2}\frac{\partial^2(\sigma^2 x^2 f(y,s|x,t))}{\partial x^2} \tag{43}$$

and

$$-\frac{\partial}{\partial s} f(y,s|x,t) = (\alpha y - \beta y \ln(y)) \frac{\partial f(y,s|x,t)}{\partial y} + \frac{1}{2}\sigma^2 y^2 \frac{\partial^2 f(y,s|x,t)}{\partial y^2} \tag{44}$$

Next, the transformation in the Ricciardi's Theorem is applied to determine the transition density function. The condition of Ricciardi's Theorem is verified using the infinitesimal moments which implies the existence of a transformation. From Equation (16):

$$\begin{aligned} a(t,x) &= \frac{2\sigma^2 x}{4} + \frac{\sigma x}{2}\left\{c_1(t) + \int_z^x \frac{c_2(t)\sigma^2 x^2}{(\sigma^2 x^2)^{3/2}} dy\right\} \\ &= \frac{\sigma^2 x}{2} + \frac{c_1(t)\sigma x}{2} + c_2(t)\frac{x}{2}\left(\ln(x) - \ln(z)\right) \end{aligned} \tag{45}$$

The Gompertz process verifies the necessary and sufficient condition of this theorem when $c_1$ and $c_2$ are given as:

$$\begin{aligned} c_1(t) &= \frac{2}{\sigma}(\alpha - \frac{\sigma^2}{2} - \beta \ln(z)) \\ c_2(t) &= -2\beta \end{aligned} \tag{46}$$

Then, Equation (45) becomes:

$$\begin{aligned} a(t,x) = \frac{\sigma^2 x}{2} + \frac{\frac{2}{\sigma}(\alpha - \frac{\sigma^2}{2} - \beta \ln(z))\sigma x}{2} - 2\beta\frac{x}{2}\left(\ln(x) - \ln(z)\right) \\ = \alpha x - \beta \ln(x) \end{aligned} \tag{47}$$

The appropriate transformations become:

$$
\begin{aligned}
\psi(t,x) = (k_1)^{1/2}\exp\left[\int_0^t d\tau\beta\right]\int_z^x \frac{dy}{\sigma y} - \frac{(k_1)^{1/2}}{2}\int_{t_2}^t d\tau\frac{2}{\sigma} \\
(\alpha - \frac{\sigma^2}{2} - \beta\ln(z))\exp\left[\int_0^\tau d\theta\beta\right] + k_2 \\
= (k_1)^{1/2}\frac{e^{\beta(t-t_0)}}{\sigma}[\ln(x)-\ln(z)] - \frac{(k_1)^{1/2}}{\sigma\beta}(\alpha - \frac{\sigma^2}{2} - \beta\ln(z)) \\
[e^{\beta(t-t_0)} - e^{\beta(t_2-t_0)}] + k_2 \\
= (k_1)^{1/2}\frac{e^{\beta(t-t_0)}}{\sigma}\ln(x) - (k_1)^{1/2}\frac{e^{\beta(t-t_0)}}{\sigma}\ln(z) \\
-\frac{(k_1)^{1/2}}{\sigma\beta}(\alpha - \frac{\sigma^2}{2})[e^{\beta(t-t_0)} - e^{\beta(t_2-t_0)}] \\
+\frac{(k_1)^{1/2}}{\sigma}\ln(z)e^{\beta(t-t_0)} - \frac{(k_1)^{1/2}}{\sigma}(\beta\ln(z))e^{\beta(t_2-t_0)} + k_2
\end{aligned} \tag{48}
$$

$$
\begin{aligned}
\varphi(t) &= k_1\int_{t_1}^t d\tau\exp(\int_0^\tau 2\beta d\theta) + k_3 \\
&= \frac{k_1}{2\beta}[e^{2\beta(t-t_0)} - e^{2\beta(t_1-t_0)}] + k_3
\end{aligned} \tag{49}
$$

Since $k_i$'s are arbitrary constants with only one restriction, i.e. $k_1 > 0$, we can take $k_1 = 1$, $t_0 = t_1 = t_2 = s$ and equate:

$$
\begin{aligned}
k_2' = k_2 + \frac{1}{\sigma\beta}(\alpha - \frac{\sigma^2}{2}) - \frac{1}{\sigma}(\beta\ln(z)) + k_2 = 0 \\
k_3' = k_3 - \frac{1}{2\beta} = 0
\end{aligned} \tag{50}
$$

Hence, the transformations are obtained to be:

$$\begin{aligned} \psi(t,x) &= \frac{e^{\beta(t-s)}}{\sigma}\ln(x) - \frac{\alpha-\sigma^2}{\sigma\beta}e^{\beta(t-s)} \\ \varphi(t) &= \frac{1}{2\beta}e^{2\beta(t-s)} \end{aligned} \tag{51}$$

Therefore, we can write the transition density function using Equation (19) in the following way:

$$\begin{aligned} f(y,s|x,t) &= \left(2\pi\left[\frac{1}{2\beta}(1-e^{-2\beta(t-s)})\right]\right)^{-1/2} \times \frac{1}{y\sigma} \\ \exp&\left\{-\frac{\left[\frac{\ln(y)}{\sigma} - \frac{e^{\beta(t-s)}}{\sigma}\ln(x) - \frac{\alpha-\sigma^2/2}{\sigma\beta}(1-e^{\beta(t-s)})\right]^2}{2\left[\frac{1}{2\beta}(1-e^{2\beta(t-s)})\right]}\right\} \\ &= \left(2\pi\left[\frac{\sigma^2}{2\beta}(1-e^{-2\beta(t-s)})\right]\right)^{-1/2} \times \frac{1}{y} \\ \exp&\left\{-\frac{\left[\ln(y) - e^{\beta(t-s)}\ln(x) - \frac{\alpha-\sigma^2/2}{\beta}(1-e^{\beta(t-s)})\right]^2}{\left[\frac{\sigma^2}{\beta}(1-e^{2\beta(t-s)})\right]}\right\} \end{aligned} \tag{52}$$

This transformation changes the state-space $\mathbb{R}^+$ into $\mathbb{R}$ and the transition density function corresponds to the density of a univariate log-normal distribution that depends on $(t-s)$ which is the temporal parameter of the process (Gutiérrez, Gutiérrez Sánchez, Nafidi, Román and Torres, 2005a). Since $X_t$ is log-normally distributed, $\ln(X_t)$ is normally distributed with $N(\mu(s,t), \sigma^2\nu^2(s,t))$. Then, the expectation and variance of $X_t$ equal to:

$$\mathbb{E}(X_t) = \exp(\mu(s,t) + \frac{1}{2}\sigma^2\nu^2(s,t)) \tag{53}$$

and

$$Var(X_t) = \left[\exp(\mu(s,t) + \frac{1}{2}\sigma^2\nu^2(s,t))\right]^2 \exp(\sigma^2\nu^2(s,t) - 1) \tag{54}$$

We define a standard normal random variable $Q$ where $\varrho(s,t) = X_t|X_s = x_s$

$$Q = \frac{\ln(\varrho(s,t)) - \mu(s,t)}{\sigma\nu(s,t)} \sim N(0,1) \tag{55}$$

with mean $\mu(s,t)$ and variance $\sigma^2\nu^2(s,t)$ which are expressed as:

$$\begin{aligned} \mu(s,t) &= \ln x_s e^{-\beta(t-s)} + \frac{\alpha - \sigma^2/2}{\beta}(1 - e^{-\beta(t-s)}) \\ \nu^2(s,t) &= \frac{1}{2\beta}(1 - e^{-2\beta(t-s)}) \end{aligned} \tag{56}$$

From these, we can obtain a $100p\%$ confidence interval for $\varrho(s,t)$ with the following form:

$$\begin{aligned} \varrho_{lower}(s,t) &= \exp\{\mu(s,t) - \xi\sigma\nu(s,t)\} \\ \varrho_{upper}(s,t) &= \exp\{\mu(s,t) + \xi\sigma\nu(s,t)\} \end{aligned}$$

with $\xi = F^{-1}_{N(0,1)}(\frac{1+p}{2})$ where $F^{-1}_{N(0,1)}$ is the inverse cumulative normal standard distribution (Gutiérrez et al., 2005b).

Moments

As the random variable $X_t|X_s = x_s$ follows log-normal distribution, we have:

$$\mathbb{E}[X_t^r|\mathcal{F}_s] = \exp\left\{r\mu(s,t) + \frac{r^2}{2}\sigma^2\nu^2(s,t)\right\} \tag{57}$$

$$= \exp\left\{r \ln x_s e^{-\beta(t-s)} + \frac{r(\alpha - \sigma^2/2)}{\beta}(1 - e^{-\beta(t-s)}) + \frac{r^2\sigma^2}{4\beta}(1 - e^{-2\beta(t-s)})\right\}$$
$$= \exp\{r \ln x_s e^{-\beta(t-s)}\} \exp\left\{\frac{r\alpha}{\beta}(1 - e^{-\beta(t-s)})\right\}$$
$$\exp\left\{\frac{r\sigma^2}{4\beta}(1 - e^{-\beta(t-s)})[r(1 + e^{-\beta(t-s)}) - 2]\right\}$$

The conditional expectation of $X_t$ is given in Equation (41) and the conditional variance is expressed in the following way:

$$Var[X_t|\mathcal{F}_s] = \mathbb{E}[X_t^2|\mathcal{F}_s] - \mathbb{E}[X_t|\mathcal{F}_s]^2 = \exp\{2 \ln x_s e^{-\beta(t-s)}\} \exp\left\{\frac{2\alpha}{\beta}(1 - e^{-\beta(t-s)})\right\}$$
$$\exp\left\{\frac{-\sigma^2}{\beta}(1 - e^{-\beta(t-s)}) + \frac{\sigma^2}{2\beta}(1 - e^{-2\beta(t-s)})\right\}\left(\exp\left\{\frac{\sigma^2}{2\beta}(1 - e^{-2\beta(t-s)})\right\} - 1\right)$$
$$= \exp\{2 \ln x_s e^{-\beta(t-s)}\} \exp\left\{\frac{2\alpha}{\beta}(1 - e^{-\beta(t-s)})\right\} \exp\left\{\frac{-\sigma^2}{2\beta}(1 - e^{-\beta(t-s)})^2\right\}$$
$$\left(\exp\left\{\frac{\sigma^2}{2\beta}(1 - e^{-2\beta(t-s)})\right\} - 1\right) \tag{58}$$

Inference on the Model

Inference on this process aims to obtain estimates of the moments. Since we do not have continuous samples in practice, we consider approximations based on discrete observations. The conditional likelihood is obtained based on the transition density function of the Gompertz diffusion process in the following way:

$$\mathcal{L}(x_0, x_1, x_2, ..., x_n, \alpha, \beta, \sigma^2) = \prod_{i=1}^{n} f(x_i, t_i | x_{i-1}, t_{i-1})$$

$$\begin{aligned} \mathcal{L} = \prod_{i=1}^{n} \left\{ 2\pi \left[ \frac{\sigma^2}{2\beta} (1 - e^{-2\beta\tau_j}) \right] \right\}^{-1/2} \frac{1}{x_j} \times \\ \exp \left\{ - \frac{\left[ \ln(x_j) - e^{-\beta\tau_j} \ln(x_{j-1}) - \frac{\gamma}{\beta}(1 - e^{-\beta\tau_j}) \right]^2}{\frac{\sigma^2}{\beta}(1 - e^{-2\beta\tau_j})} \right\} \end{aligned} \tag{59}$$

where $\gamma = \alpha - \sigma^2/2$ and $\tau_j = t_j - t_{j-1}$.

$$\begin{aligned} \ln(\mathcal{L}) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\frac{\sigma^2}{2\beta}) - \frac{1}{2} \sum_{j=1}^{n} \ln(1 - e^{-2\beta\tau_j}) - \sum_{j=1}^{n} \ln(x_j) \\ -\frac{\beta}{\sigma^2} \sum_{j=1}^{n} \left\{ \frac{[\ln(x_j) - e^{-\beta\tau_j} \ln(x_{j-1}) - \frac{\gamma}{\beta}(1 - e^{-\beta\tau_j})]^2}{(1 - e^{-2\beta\tau_j})} \right\} \end{aligned} \tag{60}$$

By taking the partial derivatives with respect to $\gamma$, $\beta$ and $\sigma^2$ and solving $\frac{\partial \mathcal{L}}{\partial \gamma} = 0$, $\frac{\partial \mathcal{L}}{\partial \beta} = 0$ and $\frac{\partial \mathcal{L}}{\partial \sigma^2} = 0$ simultaneously, we find the likelihood estimators of the parameters.

$$\frac{\partial \ln(\mathcal{L})}{\partial \gamma} = \frac{2}{\sigma^2} \sum_{j=1}^{n} \left\{ \frac{[\ln(x_j) - e^{-\beta\tau_j} \ln(x_{j-1}) - \frac{\gamma}{\beta}(1 - e^{-\beta\tau_j})]}{(1 + e^{-\beta\tau_j})} \right\} \tag{61}$$

$$\frac{\partial \ln(\mathcal{L})}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{\beta}{\sigma^4}\sum_{j=1}^{n}\left\{\frac{[\ln(x_j) - e^{-\beta\tau_j}\ln(x_{j-1}) - \frac{\gamma}{\beta}(1-e^{-\beta\tau_j})]^2}{(1-e^{-2\beta\tau_j})}\right\} \tag{62}$$

$$\begin{aligned}\frac{\partial \ln(\mathcal{L})}{\partial \beta} = -\frac{2\beta}{\sigma^2}\sum_{j=1}^{n}\left\{\frac{[\ln(x_j) - e^{-\beta\tau_j}\ln(x_{j-1}) - \frac{\gamma}{\beta}(1-e^{-\beta\tau_j})]}{(1-e^{-2\beta\tau_j})}\right\} \times \\ \left[e^{-\beta\tau_j}\tau_j\ln(x_{j-1}) - \gamma\left(\frac{e^{-\beta\tau_j}\tau_j}{\beta} - \frac{1-e^{-\beta\tau_j}}{\beta^2}\right)\right] + \\ \frac{2\beta}{\sigma^2}\sum_{j=1}^{n} e^{-\beta\tau_j}\tau_j\left\{\frac{[\ln(x_j) - e^{-\beta\tau_j}\ln(x_{j-1}) - \frac{\gamma}{\beta}(1-e^{-\beta\tau_j})]^2}{(1-e^{-2\beta\tau_j})^2}\right\} + \\ \frac{n}{2\beta} - \sum_{j=1}^{n}\frac{e^{-2\beta\tau_j}\tau_j}{1-e^{-2\beta\tau_j}} - \frac{1}{\sigma^2}\sum_{j=1}^{n}\left\{\frac{[\ln(x_j) - e^{-\beta\tau_j}\ln(x_{j-1}) - \frac{\gamma}{\beta}(1-e^{-\beta\tau_j})]^2}{(1-e^{-2\beta\tau_j})}\right\}\end{aligned} \tag{63}$$

From Equations (61) and (63), we solve the estimates for $\gamma$ and $\sigma^2$ where $\beta$ is to be found numerically.

$$\hat{\gamma} = \hat{\beta}\left(\sum_{j=1}^{n}\frac{1-e^{-\hat{\beta}\tau_j}}{1+e^{-\hat{\beta}\tau_j}}\right)^{-1}\sum_{j=1}^{n}\left\{\frac{[\ln(x_j) - e^{-\hat{\beta}\tau_j}\ln(x_{j-1})]}{(1+e^{-\hat{\beta}\tau_j})}\right\} \tag{64}$$

$$\hat{\sigma^2} = \frac{2\hat{\beta}}{n}\sum_{j=1}^{n}\left\{\frac{[\ln(x_j) - e^{-\hat{\beta}\tau_j}\ln(x_{j-1}) - \frac{\hat{\gamma}}{\hat{\beta}}(1-e^{-\hat{\beta}\tau_j})]^2}{(1-e^{-2\hat{\beta}\tau_j})}\right\} \tag{65}$$

$$-\frac{2\hat{\beta}}{\hat{\sigma^2}}\sum_{j=1}^{n}\left\{\frac{e^{-\hat{\beta}\tau_j}\tau_j[\ln(x_j)-e^{-\hat{\beta}\tau_j}\ln(x_{j-1})-\frac{\hat{\gamma}}{\hat{\beta}}(1-e^{-\hat{\beta}\tau_j})]}{(1-e^{-2\hat{\beta}\tau_j})}\right\}\left[\ln(x_{j-1})-\frac{\hat{\gamma}}{\hat{\beta}}\right]-$$

$$\sum_{j=1}^{n}\frac{e^{-2\hat{\beta}\tau_j}}{(1-e^{-2\hat{\beta}\tau_j})}+\frac{2\hat{\beta}}{\hat{\sigma^2}}\sum_{j=1}^{n}e^{-2\hat{\beta}\tau_j}\tau_j\frac{[\ln(x_j)-e^{-\hat{\beta}\tau_j}\ln(x_{j-1})-\frac{\hat{\gamma}}{\hat{\beta}}(1-e^{-\hat{\beta}\tau_j})]^2}{(1-e^{-2\hat{\beta}\tau_j})^2}=0$$

(66)

$\hat{\beta}$ has to be solved from the above equation using numerical methods such as the Newton-Raphson Method. For simplicity, assuming $\tau_j = t_j - t_{j-1} = 1$ for $j = 1, \ldots, n$, the equations for $\hat{\gamma}$ and $\hat{\sigma^2}$ can be simplified as:

$$\hat{\gamma}=\frac{\hat{\beta}}{(n)(1-e^{-\hat{\beta}})}\sum_{j=1}^{n}\left\{\ln(x_j)-e^{-\hat{\beta}}\ln(x_{j-1})\right\} \tag{67}$$

$$\hat{\sigma^2}=\frac{2\hat{\beta}}{(n)(1-e^{-2\hat{\beta}})}\sum_{j=1}^{n}\left[\ln(x_j)-e^{-\hat{\beta}}\ln(x_{j-1})-\frac{\hat{\gamma}}{\hat{\beta}}(1-e^{-\hat{\beta}})\right]^2 \tag{68}$$

Note that, as $\hat{\beta} \to 0$ , $\hat{\sigma^2}$ converges to:

$$\lim_{\hat{\beta}\to 0}\hat{\sigma^2}=\frac{1}{n}\sum_{j=1}^{n}[\ln(x_j)-\ln(x_{j-1})-\hat{\gamma}]^2 \tag{69}$$

As an alternative method to finding the transition probability density function using the Kolmogorov equations, the likelihood function that will be obtained from the Radon-Nikodým derivative that converts Equation (36) into a Wiener process can be used to obtain the parameter estimates. In this section, the drift parameters $\alpha$ and $\beta$ are estimated from maximum likelihood based on continuous sampling while the diffusion coefficient $\sigma$ is estimated approximately from an observed sample path $\{X_t,t \in [0,T]\}$. Since we cannot have continuous sampling in practice, we have to consider discrete sampling as an approximation based on discrete observations of the process at times $t_0 = 0, \ldots, t_n = T$ (Gutierrez et al., 2004).

Consider the following two stochastic differential equations:

$$dX_t = a(t, X_t)dt + b(t, X_t)^{1/2} dW_t$$
$$dX_t = b(t, X_t)^{1/2} dW_t$$

where $a$ and $b$ are defined as in Equations (37) and (38). If we denote the probability measures for these equations by $\mathbb{P}$ and $\mathbb{Q}$, respectively, the likelihood function or the Radon-Nikodým derivative becomes:

$$\frac{d\mathbb{Q}}{d\mathbb{P}}(X) = \exp\left\{\int_0^T \frac{-a(s, X_s)}{b(s, X_s)} dX_s + \frac{1}{2}\int_0^T \frac{a(s, X_s)^2}{b(s, X_s)} ds\right\}$$

$$\begin{aligned}\mathcal{L} = \exp\Bigg\{\int_0^T \frac{-\alpha X_s + \beta X_s \ln(X_s)}{\sigma^2 X_s^2} dX_s + \\ \frac{1}{2}\int_0^T \frac{\alpha^2 X_s^2 - 2\alpha\beta X_s^2 \ln(X_s) + \beta^2 X_s^2 \ln(X_s)^2}{\sigma^2 X_s^2} ds\Bigg\} \\ = \exp\Bigg\{\frac{-\alpha}{\sigma^2}\int_0^T \frac{1}{X_s} dX_s + \frac{\beta}{\sigma^2}\frac{\ln(X_s)}{X_s} dX_s + \frac{\alpha}{2\sigma^2}\int_0^T ds - \\ \frac{\alpha\beta}{\sigma^2}\int_0^T \ln(X_s) ds + \frac{\beta^2}{2\sigma^2}\int_0^T \ln(X_s)^2 ds\Bigg\}\end{aligned}$$

In order to find the estimates, we solve the equations $\frac{\partial \mathcal{L}}{\partial \alpha} = 0$ and $\frac{\partial \mathcal{L}}{\partial \beta} = 0$ simultaneously and find the likelihood estimators of the drift parameters, $\alpha$ and $\beta$ in

Equation (39) as (Gutierrez et al., 2005b):

$$\hat{\alpha} = \frac{\left(\int_0^T \frac{dX_t}{X_t}\right)\left(\int_0^T \ln(X_t)^2 dt\right) - \left(\int_0^T \frac{\ln(X_t)}{X_t} dX_t\right)\left(\int_0^T \ln(X_t) dt\right)}{T\left(\int_0^T \ln(X_t)^2 dt\right) - \left(\int_0^T \ln(X_t) dt\right)^2} \tag{70}$$

and

$$\hat{\beta} = \frac{\left(\int_0^T \frac{dX_t}{X_t}\right)\left(\int_0^T \ln(X_t) dt\right) - T\left(\int_0^T \frac{\ln(X_t)}{X_t} dX_t\right)}{T\left(\int_0^T \ln(X_t)^2 dt\right) - \left(\int_0^T \ln(X_t) dt\right)^2} \tag{71}$$

respectively. As stated in Gutierrez et al. (2005b), the above 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 $\alpha$ and $\beta$.

The Itô integrals in expressions (70) and (71) are calculated using the Itô's formula:

$$\int_0^T \frac{dX_t}{X_t} = \ln(\frac{X_T}{X_0}) + \frac{\sigma^2}{2}T$$

$$\int_0^T \ln(X_t)^2 dt = \sum_{t=1}^{T-1} \ln(X_t)^2 + \frac{\ln(X_0)^2 + \ln(X_T)^2}{2}$$

$$\int_0^T \frac{\ln(X_t)}{X_t} dX_t = \frac{\ln^2(X_T) - \ln^2(X_0)}{2} - \frac{\sigma^2 T}{2} + \frac{\sigma^2}{2}\left(\int_0^T \ln(X_t) dt\right)$$

$$\int_0^T \ln(X_t) dt = \sum_{t=1}^{T-1} \ln(X_t) + \frac{\ln(X_0) + \ln(X_T)}{2}$$

The diffusion coefficient $\sigma$ is estimated through the application of Itô's Lemma (Katsamaki and Skiadas, 1995).

$$d\left(\frac{1}{X_t}\right) = -\frac{1}{X_t^2}dX_t + \frac{1}{2}\frac{2}{X_t^3}(dX_t)^2 = \frac{-dX_t}{X_t^2}$$
$$+\frac{\sigma^2 dt}{X_t} = X_t d\left(\frac{1}{X_t}\right) = \frac{-dX_t}{X_t} + \sigma^2 dt$$

Subtituting $dX_t \cong X_t - X(t-1)$ and $d(\frac{1}{X_t}) \cong \frac{1}{X_t} - \frac{1}{X(t-1)}$ and for an approximate value of $\sigma$, the following is satisfied:

$$X_t\left(\frac{1}{X_t} - \frac{1}{X(t-1)}\right) = \frac{X(t-1) - X_t}{X_t} + \sigma^2 t \tag{72}$$

Solving for $\sigma$:

$$\hat{\sigma} = \frac{|X_t - X(t-1)|}{\sqrt{tX_tX(t-1)}} \tag{73}$$

Appling the same value of $\sigma$ to a set of real data from $t = 2$ to $t = T$, the volatility $\sigma$ is estimated as:

$$\hat{\sigma} = \frac{1}{T-1}\sum_{t=2}^{T}\frac{|X_t - X(t-1)|}{\sqrt{tX_tX(t-1)}}. \tag{74}$$

An Alternative Estimation Method

We discretize the stochastic differential equation given in Equation (39) as:

$$\ln X_{t+1} = \omega_0 + \omega_1 \ln X_t + \eta_t, \qquad \eta_t \sim N(0, \sigma^2 \Delta t), \tag{75}$$

where $\eta_t \sim N(0, \sigma^2 \Delta t)$ denotes the noise component. We can consider Equation (75) as a least-squares fitting problem where $\omega_0 = \alpha \Delta t$ and $\omega_1 = 1 - \beta \Delta t$. Rewriting:

$$X_{t+1} = e^{\omega_0 + \eta_t} X_t^{\omega_1} \tag{76}$$

and

$$\mathbb{E}[X_{t+1} | \mathcal{F}_t] = X_t^{\omega_1} \exp\left(\omega_0 + \frac{\sigma^2 \Delta t}{2}\right) \tag{77}$$

Similarly, for any forecast horizon $h$:

$$\mathbb{E}[X_{t+h} | \mathcal{F}_t] = X_t^{\omega_1^h} \exp\left(\omega_0 \sum_{i=1}^{h} \omega_1^{i-1} + \frac{\sigma^2 \Delta t}{2} \sum_{i=1}^{h} \omega_1^{i-1}\right) \tag{78}$$

<u>Stochastic Gompertz Diffusion Model: Non-Homogenous Case</u>

In this section, the Non-Homogeneous Gompertz diffusion process is introduced by the infinitesimal moments and the corresponding Kolmogorov equations.

We consider an extension of the stochastic Gompertz homogeneous diffusion process by introducing exogenous factors which are time-variant functions that affect the drift of the process. Accordingly, the stochastic Non-homogenous Gompertz diffusion model can be described by the following one-dimensional stochastic differential equation for $\{X_t, t \in [t_0, T]\ , t_0 \geq 0\}$ which takes on positive real values and has finite infinitesimal moments:

$$dX_t = a(t, X_t)dt + b(t, X_t)^{1/2}dW_t, \quad X_0 = x_0$$

where $\{W_t, t \in [0, T]\}$ is the one-dimensional Wiener process and $x_0$ is a fixed real number belonging to $\mathbb{R}^+$. Furthermore, the drift coefficient $a(t, x)$ and the diffusion coefficient $b(t, x)$ for each observation of $X_t$, $x \in (0, \infty)$, are real-valued functions which are given as:

$$a(t, x) = h(t)x - \beta x \ln(x) \tag{79}$$

$$b(t, x) = \sigma^2 x^2 \tag{80}$$

where $\beta \in \mathbb{R}$, $\sigma > 0$ and $h(t) = \alpha_0 + \sum_{i=1}^{q} \alpha_i g_i(t)$ with $g_i(t)$ as the exogenous factors that are continuous functions in $[t_0, T]$. The parameters $\alpha$, $\beta$ and $\sigma$ are time-independent and real-valued parameters. After substitution, we obtain the final form:

$$dX_t = (h(t) - \beta \ln X_t)X_t dt + \sigma X_t dW_t, \quad X_0 = x_0 \tag{81}$$

Equation (81) has a unique solution which is obtained by applying the Itô's formula to the transformation $e^{\beta t} \ln X_t$.

$$de^{\beta t} \ln X_t = e^{\beta t} \left(h(t) - \sigma^2/2\right) dt + e^{\beta t} \sigma dW_t$$

Integrating both sides:

$$e^{\beta t} \ln X_t = e^{\beta s} \ln(x_s) + \int_s^t \left(h(t) - \sigma^2/2\right) e^{\beta \tau} d\tau + \sigma \int_s^t e^{\beta \tau} dW_\tau$$

From this, the solution of Equation (81) for any $X_s = x_s$ is obtained as (Gutierrez et al., 2006):

$$X_t = \exp\left(\ln(x_s) e^{-\beta(t-s)} + \int_s^t \left(h(t) - \sigma^2/2\right) e^{-\beta(t-\tau)} d\tau + \sigma \int_s^t e^{-\beta(t-\tau)} dW_\tau\right). \quad (82)$$

Since $\int_s^t e^{-\beta(t-\tau)} dW_\tau \sim N(0, \int_s^t e^{-2\beta(t-\tau)} d\tau)$, the conditional expectation under this process is given by:

$$\mathbb{E}[X_t | \mathcal{F}_s] = \exp\left(\ln(x_s) e^{-\beta(t-s)} + \frac{\alpha_0 - \frac{\sigma^2}{2}}{\beta}(1 - e^{-\beta(t-s)}) + \frac{\sigma^2}{4\beta}(1 - e^{-2\beta(t-s)})\right) \cdot \exp\left(\sum_{i=1}^q \alpha_i \int_s^t g_i(\tau) e^{-\beta(t-\tau)} d\tau\right) \quad (83)$$

<u>Analysis through Kolmogorov Equations</u>

In this section, the One-dimensional Non-Homogeneous Gompertz diffusion process is introduced by the forward and backward Kolmogorov equations and the transition probability density function is derived using Ricciardi's Theorem in the same manner as in the Homogeneous Gompertz diffusion process.

With infinitesimal moments defined as in Equations (79) and (80), the Kolmogorov equations corresponding to the defined process are expressed in the following form:

$$\frac{\partial}{\partial t}f(y,s|x,t) = -\frac{\partial((h(t)x - \beta x\ln(x))f(y,s|x,t))}{\partial x} + \frac{1}{2}\frac{\partial^2(\sigma^2 x^2 f(y,s|x,t))}{\partial x^2} \tag{84}$$

and

$$-\frac{\partial}{\partial s}f(y,s|x,t) = (h(s)y - \beta y\ln(y))\frac{\partial f(y,s|x,t)}{\partial y} + \frac{1}{2}\sigma^2 y^2\frac{\partial^2 f(y,s|x,t)}{\partial y^2} \tag{85}$$

Next, the condition of Ricciardi's Theorem is verified using the infinitesimal moments. From Equation (16):

$$\begin{aligned} a(t,x) &= \frac{2\sigma^2 x}{4} + \frac{\sigma x}{2}\left\{c_1(t) + \int_z^x \frac{c_2(t)\sigma^2 x^2}{(\sigma^2 x^2)^{3/2}}dy\right\} \\ &= \frac{\sigma^2 x}{2} + \frac{c_1(t)\sigma x}{2} + c_2(t)\frac{x}{2}\left(\ln(x) - \ln(z)\right) \end{aligned} \tag{86}$$

The Non-Homogenous Gompertz diffusion process verifies the necessary and sufficient condition with $c_1$ and $c_2$ given as:

$$
\begin{aligned}
c_1(t) &= \frac{2}{\sigma}(h(t) - \frac{\sigma^2}{2} - \beta \ln(z) \\
c_2(t) &= -2\beta
\end{aligned} \tag{87}
$$

Then, Equation (86) becomes:

$$
\begin{aligned}
a(t,x) &= \frac{\sigma^2 x}{2} + \frac{\frac{2}{\sigma}(h(t) - \frac{\sigma^2}{2} - \beta \ln(z))\sigma x}{2} - 2\beta \frac{x}{2}\left(\ln(x) - \ln(z)\right) \\
&= h(t)x - \beta \ln(x)
\end{aligned} \tag{88}
$$

The appropriate transformations become:

$$
\begin{aligned}
\psi(t,x) = (k_1)^{1/2} \exp\left[\int_0^t d\tau \beta\right] \int_z^x \frac{dy}{\sigma y} - \frac{(k_1)^{1/2}}{2} \int_{t_2}^t d\tau . \frac{2}{\sigma}(h(\tau) - \frac{\sigma^2}{2} - \beta \ln(z)) \\
\exp\left[\int_0^\tau d\theta \beta\right] + k_2 \\
= (k_1)^{1/2} \frac{e^{\beta(t-t_0)}}{\sigma}[\ln(x) - \ln(z)] + \frac{(k_1)^{1/2}}{\sigma\beta}(\frac{\sigma^2}{2} + \beta \ln(z))[e^{\beta(t-t_0)} - e^{\beta(t_2-t_0)}] \\
- \frac{(k_1)^{1/2}}{\sigma} \int_{t_2}^t d\tau h(\tau) e^{\beta(\tau - t_0)} + k_2
\end{aligned}
$$

$$
\begin{aligned}
= (k_1)^{1/2} \frac{e^{\beta(t-t_0)}}{\sigma} \ln(x) + \frac{(k_1)^{1/2}}{\sigma\beta} \frac{\sigma^2}{2}[e^{\beta(t-t_0)} - e^{\beta(t_2-t_0)}] - \frac{(k_1)^{1/2}}{\sigma\beta} \beta \ln(z) e^{\beta(t_2-t_0)} \\
- \frac{(k_1)^{1/2}}{\sigma} \int_{t_2}^t h(\tau) e^{\beta(\tau-t_0)} d\tau + k_2
\end{aligned} \tag{89}
$$

$$\begin{aligned}\varphi(t) &= k_1 \int_{t_1}^{t} d\tau \exp(\int_0^{\tau} 2\beta d\theta) + k_3 \\ &= \frac{k_1}{2\beta}[e^{2\beta(t-t_0)} - e^{2\beta(t_1-t_0)}] + k_3\end{aligned} \tag{90}$$

Since $k_i$'s are arbitrary constants with only one restriction, i.e. $k_1 > 0$, we can take $k_1 = 1$, $t_0 = t_1 = t_2 = s$ and equate

$$\begin{aligned}k_2' &= k_2 - \frac{\sigma}{2\beta} - \frac{1}{\sigma}\ln(z) = 0 \\ k_3' &= k_3 - \frac{1}{2\beta} = 0\end{aligned} \tag{91}$$

Hence, the transformations are obtained to be:

$$\begin{aligned}\psi(t,x) &= \frac{e^{\beta(t-s)}}{\sigma}\ln(x) + \frac{\sigma}{2\beta}e^{\beta(t-s)} - \frac{1}{\sigma}\int_s^t h(\tau)e^{\beta(\tau-s)}d\tau \\ \varphi(t) &= \frac{1}{2\beta}e^{2\beta(t-s)}\end{aligned} \tag{92}$$

We find the transition density function of the process which happens to be the density of a univariate log-normal distribution (Gutierrez et al., 2006):

$$\begin{aligned}f(y,s|x,t) &= \left(2\pi\left[\frac{1}{2\beta}(1-e^{-2\beta(t-s)})\right]\right)^{-1/2} \times \frac{1}{y\sigma} \\ &\exp\left\{-\frac{\left[\frac{\ln(y)}{\sigma} - \frac{e^{\beta(t-s)}}{\sigma}\ln(x) + \frac{\sigma^2}{2\beta}(1-e^{\beta(t-s)}) - \frac{1}{\sigma}\int_s^t h(\tau)e^{\beta(\tau-s)}d\tau\right]^2}{2\left[\frac{1}{2\beta}(1-e^{2\beta(t-s)})\right]}\right\}\end{aligned} \tag{93}$$

$$= \left(2\pi\left[\frac{\sigma^2}{2\beta}(1-e^{-2\beta(t-s)})\right]\right)^{-1/2} \times \frac{1}{y}$$
$$\exp\left\{-\frac{\left[\ln(y) - e^{\beta(t-s)}\ln(x) + \frac{\sigma^2}{2\beta}(1-e^{\beta(t-s)}) - \int_s^t h(\tau)e^{\beta(\tau-s)}d\tau\right]^2}{\left[\frac{\sigma^2}{\beta}(1-e^{2\beta(t-s)})\right]}\right\}$$

Since $X_t$ is log-normally distributed, $\ln(X_t)$ is normally distributed with $N(\mu(s,t), \sigma^2\nu^2(s,t))$ which are equal to:

$$\mu(s,t) = \ln(x_s)e^{\beta(t-s)} - \frac{\sigma^2}{2\beta}(1-e^{\beta(t-s)}) + \int_s^t h(\tau)e^{\beta(\tau-s)}d\tau$$
$$= \ln(x_s)e^{-\beta(t-s)} + \frac{\alpha_0 - \frac{\sigma^2}{2}}{\beta}(1-e^{-\beta(t-s)}) + \sum_{i=1}^{q}\alpha_i\int_s^t g_i(\tau)e^{-\beta(t-\tau)}d\tau$$
$$\nu^2(s,t) = \frac{1}{2\beta}(1-e^{-2\beta(t-s)})$$

where we substitute $h(t) = \alpha_0 + \sum_{i=1}^{q}\alpha_i g_i(t)$.

Moments

As the random variable $X_t|X_s = x_s$ follows a log-normal distribution with $\Lambda(\mu(s,t), \sigma^2\nu^2(s,t))$, we have:

$$\begin{aligned} \mathbb{E}[X_t^r|\mathcal{F}_s] &= \exp\left\{r\mu(s,t) + \frac{r^2}{2}\sigma^2\nu^2(s,t)\right\} = \exp\left\{r\ln x_s e^{-\beta(t-s)}\right. \\ &\left. -\frac{r\sigma^2/2}{\beta}(1-e^{-\beta(t-s)}) + r\int_s^t h(\tau)e^{\beta(\tau-s)}d\tau + \frac{r^2\sigma^2}{4\beta}(1-e^{-2\beta(t-s)})\right\} \\ &= \exp\{r\ln x_s e^{-\beta(t-s)}\}\exp\left\{r\int_s^t h(\tau)e^{\beta(\tau-s)}d\tau\right\} \\ &\exp\left\{\frac{r\sigma^2}{4\beta}(1-e^{-\beta(t-s)})[r(1+e^{-\beta(t-s)})-2]\right\} \end{aligned} \tag{94}$$

The conditional expectation of $X_t$ is given in Equation (83) and the conditional variance is expressed in the following way:

$$Var[X_t|\mathcal{F}_s] = \mathbb{E}[X_t^2|\mathcal{F}_s] - \mathbb{E}[X_t|\mathcal{F}_s]^2 = \exp\{2\ln x_s e^{-\beta(t-s)}\} \exp\left\{2\int_s^t h(\tau)e^{\beta(\tau-s)}d\tau\right\}$$
$$\exp\left\{\frac{-\sigma^2}{\beta}(1-e^{-\beta(t-s)}) + \frac{\sigma^2}{2\beta}(1-e^{-2\beta(t-s)})\right\}\left(\exp\left\{\frac{\sigma^2}{2\beta}(1-e^{-2\beta(t-s)})\right\}-1\right)$$
$$= \exp\{2\ln x_s e^{-\beta(t-s)}\} \exp\left\{2\int_s^t h(\tau)e^{\beta(\tau-s)}d\tau\right\}$$
$$\exp\left\{\frac{-\sigma^2}{2\beta}(1-e^{-\beta(t-s)})^2\right\}\left(\exp\left\{\frac{\sigma^2}{2\beta}(1-e^{-2\beta(t-s)})\right\}-1\right) \tag{95}$$

Inference on the Model

The parameters of the Non-Homogeneous Gompertz diffusion process are estimated by the maximum likelihood method using the following discrete sampling of the process(as an approximation of continuous sampling):

$$\mathcal{L}(x_0, x_1, x_2, ..., x_n, \alpha, \beta, \sigma^2) = \prod_{i=1}^{n} f(x_i, t_i | x_{i-1}, t_{i-1})$$

$$\mathcal{L} = \prod_{i=1}^{n}\left(2\pi\left[\frac{\sigma^2}{2\beta}(1-e^{-2\beta\tau_j})\right]\right)^{-1/2} \times \frac{1}{x_j}$$
$$\exp\left\{-\frac{\left[\ln(x_j) - e^{\beta\tau_j}\ln(x_{j-1}) + \frac{\sigma^2}{2\beta}(1-e^{\beta\tau_j}) - \int_{t_{j-1}}^{t_j} h(\theta)e^{\beta(\theta-t_{j-1})}d\theta\right]^2}{\left[\frac{\sigma^2}{\beta}(1-e^{2\beta\tau_j})\right]}\right\} \tag{96}$$

where $\tau_j = t_j - t_{j-1}$.

$$\ln(\mathcal{L}) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\left(\frac{\sigma^2}{2\beta}\right) - \frac{1}{2}\sum_{j=1}^{n}\ln(1-e^{-2\beta\tau_j}) - \sum_{j=1}^{n}\ln(x_j)$$
$$-\frac{\beta}{\sigma^2}\sum_{j=1}^{n}\left\{\frac{\left[\ln(x_j) - e^{\beta\tau_j}\ln(x_{j-1}) + \frac{\sigma^2}{2\beta}(1-e^{\beta\tau_j}) - \int_{t_{j-1}}^{t_j} h(\theta)e^{\beta(\theta-t_{j-1})}d\theta\right]^2}{(1-e^{-2\beta\tau_j})}\right\} \tag{97}$$

In this section, it is difficult to address the problem of estimation in the same way as in the Homogenous Gompertz diffusion process. However, it is possible to simplify the likelihood function so that it is analytically tractable which is accomplished using $h(t) = \alpha_0 + \sum_{i=1}^{q}\alpha_i g_i(t)$ form for the exogenous factors.

Parameter Estimation

Consider a discrete sampling of the process $(x_0, \ldots, x_n)$ with the initial condition $P(X(t_0) = x_0) = 1$. The associated conditional likelihood function has the following form:

$$\mathcal{L}(x_0, x_1, ..., x_n, a, \sigma^2) = \prod_{i=1}^{n} f(x_i, t_i | x_{i-1}, t_{i-1})$$

where $a = \left(\alpha_0 - \frac{\sigma^2}{2}, \alpha_1, ..., \alpha_q\right)$ if we have $q$ exogenous factors and the probability density function of the process follows:

$$f(x, t|x_s, s) = [2\pi\sigma^2\nu^2(s,t)]^{-\frac{1}{2}}x^{-1}\exp\left\{-\frac{[\ln(x) - \mu(s,t)]^2}{2\sigma^2\nu^2(s,t)}\right\}$$

with $\mu(s,t)$ and $\nu^2(s,t)$ are taken as in Equation (94) where equally-spaced time intervals will be assumed $t_j - t_{j-1} = h$ for $j = 1, \ldots, n$ with $h = 1$.

We will express the likelihood function in a condensed form using the matrix notation by introducing the following variables.

$$
\begin{gathered}
\gamma = \frac{(1-e^{-\beta})}{\beta} \\
a_{(q+1)\times 1} = \left(\alpha_0 - \frac{\sigma^2}{2}, \alpha_1, ..., \alpha_q\right) \\
\nu = \nu^2(t_{j-1}, t_j) = \frac{1}{2\beta}(1 - e^{-2\beta}) \\
\iota_0 = x_0, \iota_i = \nu^{-1}(\ln(x_i) - e^{-\beta}\ln(x_{i-1})) \\
v_{n\times 1} = (\iota_1, ..., \iota_n)' \\
u_i = \nu^{-1}\left(\gamma, \int_{t_{i-1}}^{t_i} g_1(\tau)e^{-\beta(t_i-\tau)}d\tau, ..., \int_{t_{i-1}}^{t_i} g_q(\tau)e^{-\beta(t_i-\tau)}d\tau\right)' \\
U_{(q+1)\times n} = (u_1, ..., u_n)
\end{gathered}
$$

for $i = 1, \ldots, n$. With this notation, the likelihood function becomes:

$$
\mathcal{L}(\iota_1, ..., \iota_n, a, \beta, \sigma^2) = [2\pi\sigma^2\nu]^{-\frac{n}{2}} \exp\left\{-\frac{(v - U'a)'(v - U'a)}{2\sigma^2}\right\} \tag{98}
$$

We obtain the log-likelihood function as:

$$
\ln(\mathcal{L}) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2\nu) - \frac{(v - U'a)'(v - U'a)}{2\sigma^2} \tag{99}
$$

By differentiating the log-likelihood function with respect to $a$ and $\sigma^2$, we obtain the following expressions:

$$Uv = UU'a$$
$$n\sigma^2 = (v - U'a)'(v - U'a)$$

and a third equation is obtained from the derivative with respect to $\beta$, which leads to the following equation when the previous two equations are used:

$$\left(\nu^{-1}e^{-\beta}I_x' - a'\frac{\partial U}{\partial \beta}\right)(v - U'a) = 0$$

where $I_x = (\ln(x_0), \ln(x_1) \ldots \ln(x_n))'$ and $\frac{\partial U}{\partial \beta}$ is derivative of the elements in $U$.

We find the estimates of $a$ and $\sigma^2$ as:

$$\begin{aligned}
\hat{a} &= (UU')^{-1}Uv \\
n\hat{\sigma^2} &= (v - U'a)'(v - U'a) = v'v - v'U'a - av'v + av'U'(UU')^{-1}Uv \\
&= v'(I_n - U'(UU')^{-1}U)v = v'Hv \\
&\left(\nu^{-1}e^{-\beta}I_x' - v'U'(U'U)^{-1}\frac{\partial U}{\partial \beta}\right)Hv = 0
\end{aligned} \tag{100}$$

where $H = I_n - U'(UU')^{-1}U$ is an idempotent and symmetric matrix.

It is not possible to explicitly calculate the estimator of $\beta$ from the last equation since the components in $U$ and $H$ matrices are dependent on the integrals of expressions $g_i(t)$ $(i = 1, \ldots, q)$ which are not known continuously. Hence, they are taken as polygonal functions which are only known from the discrete observations of data. Letting these discrete observations be denoted by $x_{i,j}$ for $i = 0, \ldots, n$ and $j = 1, \ldots, q$, the polygonal functions are typically constructed from the observed values of the exogenous factors (Gutierrez et al., 2006).

$$g_j(t) = x_{i-1,j} + (x_{i,j} - x_{i-1,j})(t_i - t_{i-1}) \tag{101}$$

Hence, the integrals in matrix $U$ become:

$$\int_{t_{i-1}}^{t_i} g_j(\tau)e^{-\beta(t_i-\tau)}d\tau = \gamma z_{i,j}(\beta) \tag{102}$$

where

$$z_{i,j}(\beta) = x_{i-1,j} + (x_{i,j} - x_{i-1,j})\frac{\beta - 1 + e^{-\beta}}{\beta(1-e^{-\beta})} \tag{103}$$

<u>The Case with No Exogenous Factors - Homogenous Gompertz Diffusion Model</u>

The parameters for the Homogenous Gompertz Diffusion Model can be obtained as a particular case of the Non-Homogenous Gompertz Diffusion Model. When there are no exogenous factors, $g_i(t) = 0$ for all $i = 1, \ldots, q$ and the matrix $U$ reduces to a row vector $U = \gamma\nu^{-1}(1, \ldots, 1)$ and $v = \nu^{-1}(I_x - e^{-\beta}J_x)$ where
$I_x = (\ln(x_0), \ln(x_1) \ldots \ln(x_n - 1))'$ and $J_x = (\ln(x_1), \ln(x_2) \ldots \ln(x_n))'$.
Furthermore, the nonlinear equation that is numerically solved to obtain $\beta$, Equation (100), converts into $\nu^{-1}e^{-\beta}I_x' = 0$. Solving it explicitly, we find the estimate of $\beta$ (Gutierrez et al., 2006):

$$\hat{\beta} = \ln\left(\frac{J_x' H J_x}{J_x' H I_x}\right) \tag{104}$$

We also obtain the estimates of other two parameters as:

$$\begin{aligned} \hat{a} &= \gamma^{-1}\nu(UU')^{-1}Uv + \frac{\hat{\sigma^2}}{2} \\ \hat{\sigma^2} &= \frac{1}{n}v'Hv \end{aligned} \tag{105}$$

An Alternative Estimation Method

We discretize the stochastic differential equation given in Equation (81) as:

$$\ln X_{t+1} = a_0 + \sum_{i=1}^{q} a_i g_i(t) + \beta \ln X_t + \eta_t, \qquad \eta_t \sim N(0, \sigma^2 \Delta t) \tag{106}$$

where $a_0 = \alpha_0 \Delta t$, $a_i = \alpha_i \Delta t$, $b = (1 - \beta \Delta t)$. Equation (106) can be written as

$$X_{t+1} = e^{h(t)\Delta t + \eta_t} X_t^b \tag{107}$$

and

$$E[X_{t+1}|\mathcal{F}_t] = X_t^{\alpha_1} \exp\left(h(t) + \frac{\sigma^2 \Delta t}{2}\right) \tag{108}$$

Similarly, for any forecast horizon $h$:

$$E[X_{t+h}|\mathcal{F}_t] = X_t^{(1-\beta\Delta t)^h} \exp\left( h(t)\Delta t \sum_{i=1}^{h} (1-\beta\Delta t)^{i-1} + \frac{\sigma^2\Delta t}{2} \sum_{i=1}^{h} (1-\beta\Delta t)^{i-1} \right) \tag{109}$$

# CHAPTER VI

# CONTINGENT CLAIMS ON NATURAL GAS CONSUMPTION

In this section, we consider the pricing of contingent claims (futures and call/put options) that are defined on natural gas consumption. The price of a futures contract which depends on the future natural gas consumption level at maturity $T$ and the current consumption level $X_s$ is given by:

$$F(X_T, X_s) = E_{\mathbb{Q}}[X_T | \mathcal{F}_s] \tag{110}$$

where the expectation is taken with respect to a risk neutral measure $\mathbb{Q}$ and filtration $\{\mathcal{F}_t\}$. Similarly, the prices of European type call/put options with payoffs that are defined as $\max(X_T - K, 0)$ or $\max(K - X_T, 0)$ given the strike consumption level $K$, respectively are expressed as:

$$Call(X_T, X_s) = E_{\mathbb{Q}}[e^{-r_f(T-s)} \max(X_T - K, 0)] \tag{111}$$

and

$$Put(X_T, X_s) = E_{\mathbb{Q}}[e^{-r_f(T-s)} \max(K - X_T, 0)] \tag{112}$$

where the prices at time $0 < s < T$ are obtained by considering the risk neutral expectations given the risk-free rate $r_f$.

The expectations in Equations (110), (111) and (112) are taken with respect to the risk neutral measure. Hence, we need to convert from the physical measure, which is shown in Equation (21), to the risk-neutral measure. We introduce the market price of risk parameter $\lambda$ so that the drift rate in this equation becomes equal to the risk-free interest rate. Now, the stochastic part of the model converts into:

$$dY_t = -\kappa(Y_t + \lambda\sigma/\kappa)dt + \sigma dW_t^*, \tag{113}$$

where $dW_t^* = dW_t + \lambda dt$ with the process $W_t^*$ following a $\mathbb{Q}$-Brownian Motion. Hence, the expected growth rate of $Y_t$ becomes $-\kappa(Y_t + \lambda\sigma/\kappa)$ which should be equal to the risk free rate $r_f$. Then, the futures price will be given by:

$$F(X_T, X_s) = E_{\mathbb{Q}}[X_T|\mathcal{F}_s]$$
$$= \exp\left(f(T) + (\ln X_s - f(s))e^{-\kappa(T-s)} - \lambda\frac{\sigma}{\kappa}(1 - e^{-\kappa(T-s)}) + \frac{\sigma^2}{4\kappa}(1 - e^{-2\kappa(T-s)})\right) \tag{114}$$

which is similar to Equation (27).

In the pricing of European type call options, we need to derive the expression for the expectation in Equation (111). We start with

$$E_{\mathbb{Q}}[\max(X_T - K, 0)] = \int_K^{\infty} (X_T - K)g(X_T)dX_T \tag{115}$$

where $g(X_T)$ is the probability density function of $X_T$. Since $X_T$ is log-normally distributed, $ln(X_T)$ is normally distributed with $N(\mu_{ln(X_T)}, \sigma_{ln(X_T)})$. From the relation between normal and log-normal distributions, the expectation and variance of $X_T$ will equal to:

$$E_{\mathbb{Q}}(X_T) = \exp\left(\mu_{\ln(X_T)} + \frac{1}{2}\sigma^2_{\ln(X_T)}\right) \tag{116}$$

and

$$Var_{\mathbb{Q}}(X_T) = \left[\exp\left(\mu_{\ln(X_T)} + \frac{1}{2}\sigma^2_{\ln(X_T)}\right)\right]^2 \exp(\sigma^2_{\ln(X_T)} - 1) \tag{117}$$

From Equation (116), we find:

$$\mu_{\ln(X_T)} = \ln(E_{\mathbb{Q}}(X_T)) - \frac{1}{2}\sigma^2_{\ln(X_T)} \tag{118}$$

Using $Q = \frac{\ln(X_T) - \mu_{\ln(X_T)}}{\sigma_{\ln(X_T)}}$ which is standard normal $N(0,1)$ with probability density function $\phi(Q)$, we can rewrite Equation (115) as:

$$\begin{aligned} E_{\mathbb{Q}}[\max(X_T - K, 0)] = &\int_{\frac{\ln(X_T) - \mu_{\ln(X_T)}}{\sigma_{\ln(X_T)}}}^{\infty} \exp(Q.\sigma_{\ln(X_T)} + \mu_{\ln(X_T)})\phi(Q)dQ \\ &-K\int_{\frac{\ln(X_T) - \mu_{\ln(X_T)}}{\sigma_{\ln(X_T)}}}^{\infty} \phi(Q)dQ \end{aligned} \tag{119}$$

where $X_T = \exp(Q.\sigma_{\ln(X_T)} + \mu_{\ln(X_T)})$ from the definition of $Q$. Since $\phi(Q) = \frac{1}{\sqrt{2\pi}}\exp(\frac{-Q^2}{2})$,

$$\exp(Q.\sigma_{\ln(X_T)} + \mu_{\ln(X_T)})\phi(Q) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-Q^2 + 2Q\sigma_{\ln(X_T)} + 2\mu_{\ln(X_T)}}{2}\right)$$
$$= \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-Q^2 + 2Q\sigma_{\ln(X_T)} - \sigma^2_{\ln(X_T)}}{2}\right) \exp\left(\mu_{\ln(X_T)} + \frac{\sigma^2_{\ln(X_T)}}{2}\right) \tag{120}$$
$$= \exp\left(\mu_{\ln(X_T)} + \frac{\sigma^2_{\ln(X_T)}}{2}\right) \phi(Q - \sigma_{\ln(X_T)})$$

Hence, Equation (115) becomes:

$$E_{\mathbb{Q}}[\max(X_T - K, 0)] = \exp\left(\mu_{\ln(X_T)} + \frac{\sigma^2_{\ln(X_T)}}{2}\right) \int_{\frac{\ln(X_T) - \mu_{\ln(X_T)}}{\sigma_{\ln(X_T)}}}^{\infty} \phi(Q - \sigma_{\ln(X_T)}) dQ -$$
$$K \int_{\frac{\ln(X_T) - \mu_{\ln(X_T)}}{\sigma_{\ln(X_T)}}}^{\infty} \phi(Q) dQ \tag{121}$$

As a result, the call price is found to be:

$$Call(X_T, X_s) = e^{-r_f(T-s)} \left[\exp\left(\mu_{\ln(X_T)} + \frac{\sigma^2_{\ln(X_T)}}{2}\right) \Phi\left(\frac{-\ln(K) + \mu_{\ln(X_T)}}{\sigma_{\ln(X_T)}} + \sigma_{\ln(X_T)}\right)\right.$$
$$\left. -K\Phi\left(\frac{-\ln(K) + \mu_{\ln(X_T)}}{\sigma_{\ln(X_T)}}\right)\right] \tag{122}$$

where $\Phi(.)$ denotes the cumulative distribution function. Similarly, the put price $Put(X_T, X_s)$ is found to be:

$$Put(X_T, X_s) = e^{-r_f(T-s)} \left[K\Phi\left(\frac{\ln(K) - \mu_{\ln(X_T)}}{\sigma_{\ln(X_T)}}\right) - \right.$$
$$\left. \exp\left(\mu_{\ln(X_T)} + \frac{\sigma^2_{\ln(X_T)}}{2}\right) \Phi\left(\frac{\ln(K) - \mu_{\ln(X_T)}}{\sigma_{\ln(X_T)}} - \sigma_{\ln(X_T)}\right)\right] \tag{123}$$

where the variables $\mu_{\ln(X_T)}$ and $\sigma_{\ln(X_T)}$ are known to equal:

$$\mu_{\ln(X_T)} = f(t) + (\ln X_s - f(s))e^{-\kappa(t-s)} - \frac{\lambda\sigma}{\kappa}(1 - e^{-\kappa(t-s)}) \tag{124}$$

and

$$\sigma^2_{\ln(X_T)} = \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa(t-s)}) \tag{125}$$

Pricing under Stochastic Gompertz Diffusion Model

Homogenous Gompertz Diffusion Model

Dividing the stochastic Homogenous Gompertz diffusion model given in Equation (39) by $X_t$, we obtain:

$$\frac{dX_t}{X_t} = (\alpha - \beta \ln X_t)dt + \sigma dW_t, \quad X_s = x_s \tag{126}$$

where $\alpha - \beta \ln X_t$ shows the expected growth rate in $X_t$. In the risk-neutral world, the drift rate in Equation (126) is reduced by $\lambda\sigma X_t$ factor and the expected growth rate becomes $\alpha - \beta \ln X_t - \lambda\sigma$ which should be equal to the risk free rate $r_f$. Now, Equation (126) converts into the following version:

$$\frac{dX_t}{X_t} = (\alpha - \beta \ln X_t - \lambda\sigma)dt + \sigma dW^*_t \tag{127}$$

where $dW_t^* = dW_t + \lambda dt$. Then, the futures price will be given by:

$$F(X_T, X_s) = E_{\mathbb{Q}}[X_T|\mathcal{F}_s] = \exp\left(\ln X_s e^{-\beta(T-s)} + \frac{\gamma}{\beta}(1 - e^{-\beta(T-s)}) + \frac{\sigma^2}{4\beta}(1 - e^{-2\beta(T-s)})\right) \tag{128}$$

where $\gamma = \alpha - \sigma^2/2 - \lambda\sigma$ . Remembering Equations (111) and (115), $\ln(X_T)$ is again normally distributed with $N(\mu_{\ln(X_T)}, \sigma_{\ln(X_T)})$ and the expectation and variance of $X_T$ are expressed the same way as in Equations (116) and (117). We use the same standard normal variable $Q$ and all the rest follows in the same way until we obtain the same call and put option pricing equations as in Equations (122) and (123). The only differences are in the formulation of $\mu_{\ln(X_T)}$ and $\sigma_{\ln(X_T)}$ which are known to equal:

$$\mu_{ln(X_T)} = \ln(X_s)e^{-\beta(T-s)} + \frac{\gamma}{\beta}(1 - e^{-\beta(T-s)}) \tag{129}$$

and

$$\sigma^2_{ln(X_T)} = \frac{\sigma^2}{2\beta}(1 - e^{-2\beta(T-s)}) \tag{130}$$

where $\gamma = \alpha - \sigma^2/2 - \lambda\sigma$ .

Non-Homogenous Gompertz Diffusion Model

Introducing the market price of risk parameter $\lambda$, similar to the Homogenous

Gompertz model, into the stochastic Non-Homogenous Gompertz diffusion model which is given as in Equation (81), we obtain:

$$\frac{dX_t}{X_t} = (h(t) - \lambda\sigma - \beta \ln X_t)dt + \sigma dW_t, \quad X_s = x_s \tag{131}$$

where $dW_t^* = dW_t + \lambda dt$. In this case, $h(t) - \lambda\sigma - \beta \ln X(t) = r_f$. Substituting $h(t) = \alpha_0 + \sum_{i=1}^{q} \alpha_i g_i(t)$, the futures price will then be given by:

$$\begin{gathered} F(X_T, X_s) = E_{\mathbb{Q}}[X_T|\mathcal{F}_s] = \exp\Big(\ln(X_s)e^{-\beta(T-s)} + \\ \frac{\alpha_0 - \lambda\sigma - \frac{\sigma^2}{2}}{\beta}(1 - e^{-\beta(T-s)}) + \frac{\sigma^2}{4\beta}(1 - e^{-2\beta(T-s)})\Big) \cdot \\ \exp\left(\sum_{i=1}^{q} \alpha_i \int_s^T g_i(\tau)e^{-\beta(T-\tau)}d\tau\right) \end{gathered} \tag{132}$$

where $\gamma = \alpha_0 - \lambda\sigma - \sigma^2/2$. Remembering Equations (111) and (115), $\ln(X_T)$ is again normally distributed with $N(\mu_{\ln(X_T)}, \sigma_{\ln(X_T)})$ and the expectation and variance of $X_T$ are expressed the same way as in Equations (116) and (117). We use the same standard normal variable $Q$ and all the rest follows in the same way until we obtain the same call and put option pricing equations as in Equations (122) and (123). The only differences are in the formulation of $\mu_{\ln(X_T)}$ and $\sigma_{\ln(X_T)}$ which are now expressed as:

$$\mu_{\ln(X_T)} = \ln(X_s)e^{-\beta(T-s)} + \frac{\gamma}{\beta}(1 - e^{-\beta(T-s)}) + \sum_{i=1}^{q} \alpha_i \int_s^T g_i(\tau)e^{-\beta(T-\tau)}d\tau \tag{133}$$

and

$$\sigma^2_{\ln(X_T)} = \frac{\sigma^2}{2\beta}(1 - e^{-2\beta(T-s)}) \tag{134}$$

Pricing of General Contingent Claims

A wide range of claims can be priced using the described stochastic systems in the Monte Carlo simulation framework. When we consider a general payoff function that depends on the path of natural gas consumption $H(x_s, \ldots, x_T)$ in the interval $[s, T]$, the price of a claim becomes $E_{\mathbb{Q}}[e^{-r_f(T-s)}H(X_s, \ldots, X_T)|\mathcal{F}_s]$ which will be estimated by the Monte Carlo simulation in the absence of analytical formulas in the following way:

$$e^{-r_f(T-s)}\frac{1}{N}\sum_{i}^{N} H(X_{s,i}, \ldots, X_{T,i})$$

where $H(X_{s,i}, \ldots, X_{T,i})$ shows the consumption process simulated over the interval $[s, T]$ for any path $i$.

# CHAPTER VII

# RESULTS

## Estimation of Model Parameters in the One-Factor Mean-Reverting Process

In this part, the results from the One-Factor Mean-Reverting process and its implications for the valuation of derivative securities are discussed. In this model, the logarithm of per consumer natural gas consumption is described in terms of two components: a deterministic trend including any kind of predictable behaviour and a mean reverting stochastic component. The functional form of the deterministic component in Equation (29) can also be modified to include natural gas prices. However, since the prices are centrally determined such that they do not change frequently in the short run, it is also possible to leave them in the stochastic part.

The estimation results for the entire sampling period (all 2848 daily observations) are reported in the table below.

Table 3. Nonlinear Regression Results for the One-Factor Mean Reverting Process

| Parameters | $\beta_0$ | $\beta_1$ | $\alpha_1$ | $\alpha_2$ | $\gamma_1$ | $\gamma_2$ |
|---|---|---|---|---|---|---|
| Estimates | 0.9604 | -0.1149 | 0.5083 | 0.0282 | 0.9338 | 0.0460 |
| Upper CI | 1.0112 | -0.1056 | 0.5785 | 0.0960 | 1.0050 | 0.1137 |
| Lower CI | 0.9096 | -0.1243 | 0.4382 | -0.0397 | 0.8627 | -0.0217 |

Table 3. continued.

| Parameters | $\phi$ | $\kappa$ | $\sigma$ |
|---|---|---|---|
| Estimates | 0.9043 | 0.0957 | 0.1339 |
| Upper CI | 0.8885 | 0.1115 | 0.1444 |
| Lower CI | 0.9202 | 0.0798 | 0.1226 |

The lag structure representing the number of sine and cosine terms in the deterministic function is determined based on Akaike and Schwarz Information criteria, which results in $p = 2$ in this case. In this model, the sign of the coefficient of the holiday dummy variable $\beta_1$ is negative since, as expected from the results in Figure 3, per consumer consumptions are lower on holidays. The parameter $\phi$ that represents the AR(1) term in Equation (32) is significant which suggests that the reversion coefficient $\kappa$ is also significant.

The proposed deterministic function $f(t)$ and the actual consumption values are plotted in the figure below from which we see that the function is able to capture most of the seasonality.

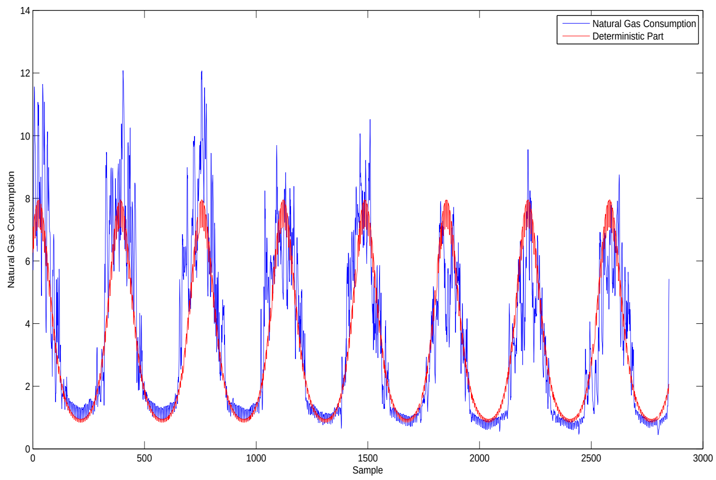


Figure 5. Realized per consumer natural gas consumption and the fitted deterministic function

The residuals that are obtained from the differences between these two series follow an autoregressive process of at least order one, as the sample autocorrelation function suggests. Since the mean reverting stochastic process proposed in Equation (21) corresponds to an AR(1) process in the discrete form, it seems to be a suitable choice, as it is also suggested by Figure 6. Finally, the residuals are plotted in Figure 7 which shows no significant serial correlation.

We need to evaluate the forecasting performance of the model by the

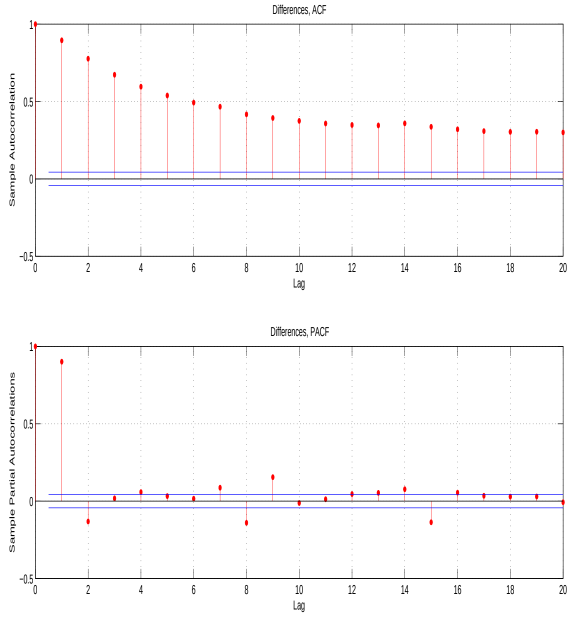


Figure 6. ACF and PACF for the differences between realized per consumer consumption and the fitted deterministic function

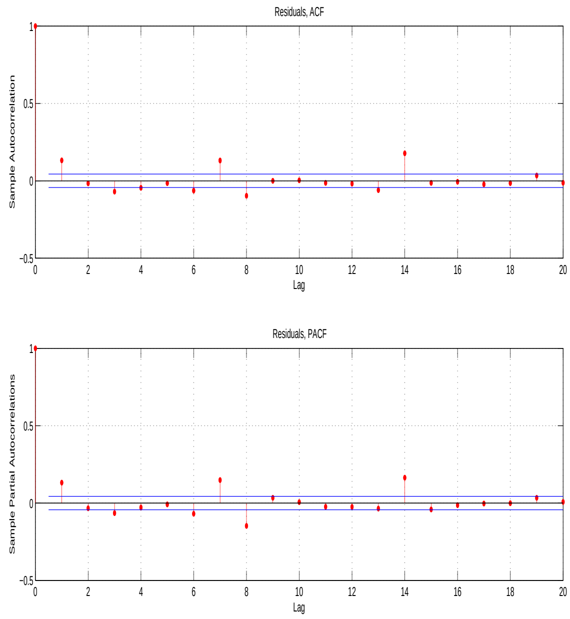


Figure 7. ACF and PACF for the residuals of the One-Factor Mean-Reverting Process

backtesting method at daily, weekly and monthly forecast horizons. In this part, we estimate the model using only the observations from the initial two-years period and apply the backtesting methodology for the remaining data points. The iterative process is as follows:

(i) We estimate the model using the initial two-years period, i.e. only the data from January 1, 2004 to December 31, 2005 (730 data points)

(ii) We calculate the conditional expectations of the natural gas consumption at each forecast horizon ($h = 1, 7, 30$), i.e. $\hat{X}_{t+h}$, such that Equation (27) converts to the following version:

$$\hat{X}_{t+h} = E[X_{t+h}|\mathcal{F}_t] = \exp\left(f(t+h) + (\ln X_t - f(t))e^{-\kappa h} + \frac{\sigma^2}{4\kappa}(1 - e^{-2\kappa h})\right) \tag{135}$$

where $t + h > 730$. We compare these conditional expectations with the realized per consumer consumptions at time $t + h$.

(iii) At time t+1, we expand the estimation window by one more day and repeat previous steps, i.e. we re-estimate the model parameters and with the updated parameters, we again calculate the forecasts $\hat{X}_{t+h+1}$.

(iv) We repeat this procedure until all the forecasts are calculated within the backtesting sample (all 2118 data points).

We plot the realized natural gas consumptions and the forecasted values for different horizons in the figure below. We also obtain the confidence intervals for the point forecasts using the expressions for the conditional expectations $\mu(t, t + h)$ and variances $\nu^2(t, t + h)$:

$$E[X_{t+h}|\mathcal{F}_t] = \exp\left(\mu(t, t+h) + \frac{1}{2}\sigma^2\nu^2(t, t+h)\right)$$

$$\mu(t, t+h) = \exp\left(f(t+h) + (\ln X_t - f(t))e^{-\kappa h}\right)$$
$$\nu^2(t, t+h) = \frac{\sigma^2}{4\kappa}(1 - e^{-2\kappa h})$$

where $t + h > 730$ and $h = 1, 7, 30$.

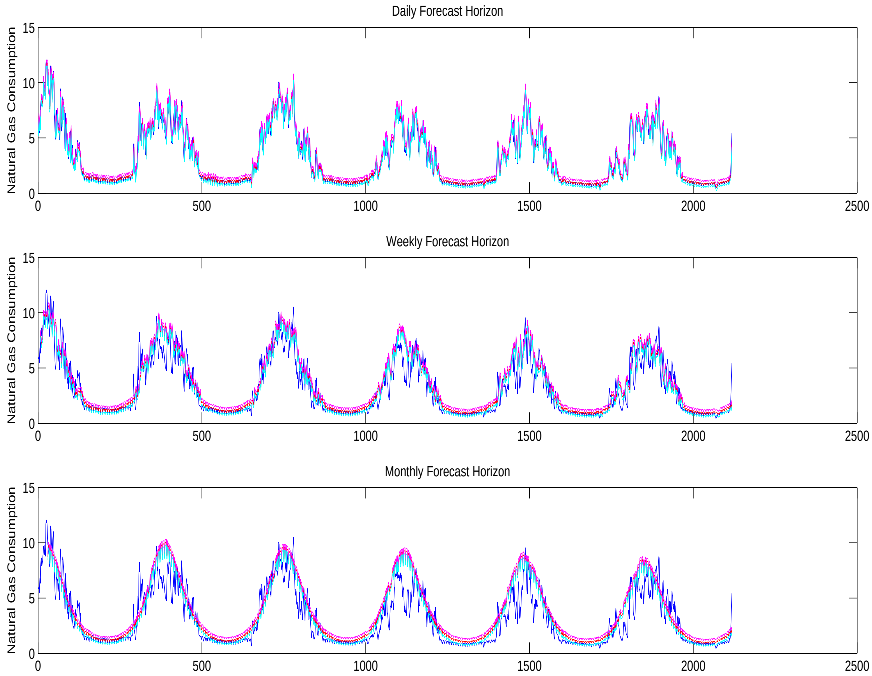


Figure 8. Backtesting with One-Factor Mean-Reverting process

From Figure 8, we see that the confidence intervals get wider as the forecast horizon increases. We calculate the Relative Mean Square Errors (RMSE) as:

$$RMSE = \frac{1}{N}\sum_{t=1}^{N}\left(\frac{\hat{X}_{t+h} - X_{t+h}}{X_{t+h}}\right)^2 \tag{136}$$

where $t$ denotes the number of observations used in the backtesting sample and $h$ is the number of days in each forecast horizon.

Table 4. Comparison of RMSE's for the One-Factor Mean-Reverting Process

| Forecast horizon | | |
|---|---|---|
| Daily | Weekly | Monthly |
| 0.0176 | 0.1230 | 0.2194 |

As expected, the best forecast performance is achieved at daily forecast horizon and the forecasting power reduces significantly at the monthly level. However, the confidence intervals successfully represent the range of variation in natural gas consumption.

### Estimation of Model Parameters in the Homogenous Gompertz Diffusion Model

In the Homogenous Gompertz Diffusion model, the logarithm of per consumer natural gas consumption is described by the stochastic differential equation in Equation (39) with the functional forms for the drift and diffusion coefficients given as in Equations (37) and (38), respectively. The likelihood estimators of these parameters are obtained in Equations (70), (71) and (74). Alternatively, a least squares estimation is proposed by discretizing the same stochastic differential equation. Finally, the same parameters can also be obtained as a particular case of the Non-Homogenous Gompertz Diffusion Model under the assumption that there are no exogenous factors.

The estimation results for the whole sampling period (all of $2848$ daily observations) are reported in the table below for all three approaches. In addition, the sample autocorrelation and partial autocorrelation functions are drawn for the residuals that are obtained from the differences between the logarithm of the daily predicted and realized per consumer consumption series.

Table 5. Estimated Parameters for the Homogenous Gompertz Diffusion Model

| Likelihood Estimators | | | |
|---|---|---|---|
| Parameters | $\alpha$ | $\beta$ | $\sigma$ |
| Estimates | $2.0765 \times 10^{-5}$ | $3.3486 \times 10^{-5}$ | 0.0040 |
| Least Squares Estimates | | | |
| Parameters | $\alpha$ | $\beta$ | $\sigma$ |
| Estimates | 0.0151 | 0.0166 | 0.0040 |
| Upper CI | 0.0233 | 0.0233 | |
| Lower CI | 0.0069 | 0.0100 | |
| Non-Homogenous Gompertz, Special Case | | | |
| Parameters | $\alpha$ | $\beta$ | $\sigma$ |
| Estimates | 0.0155 | 0.0168 | 0.0224 |

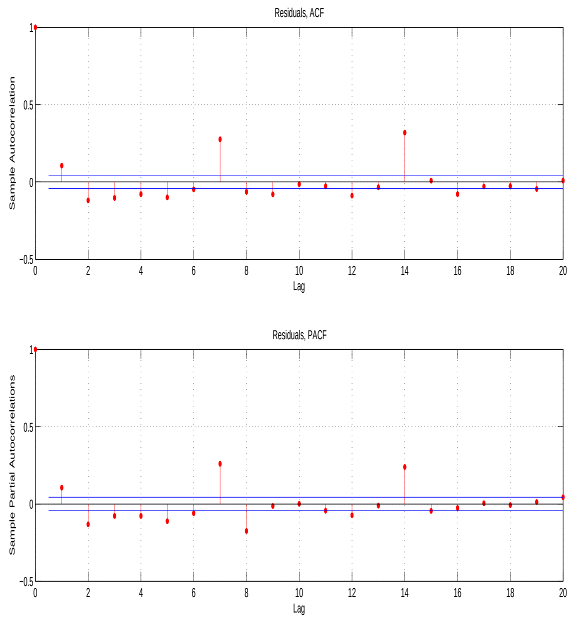


Figure 9. ACF and PACF for the residuals of the Homogenous Gompertz diffusion model (estimated via likelihood estimators)

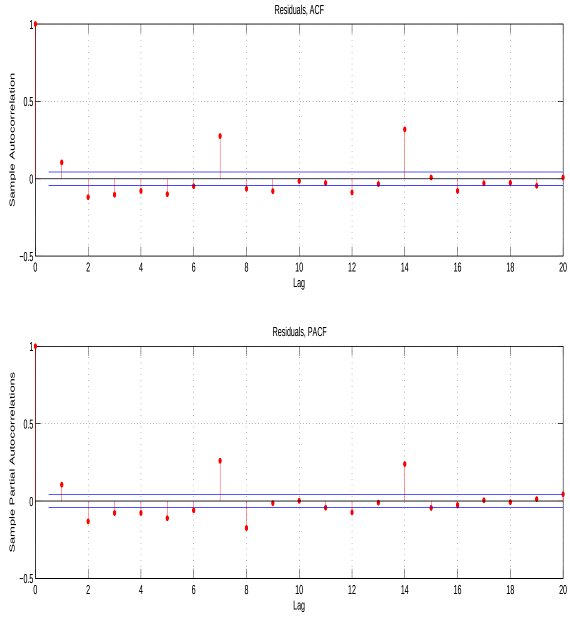


Figure 10. ACF and PACF for the residuals of the Homogenous Gompertz diffusion model (estimated with least squares)

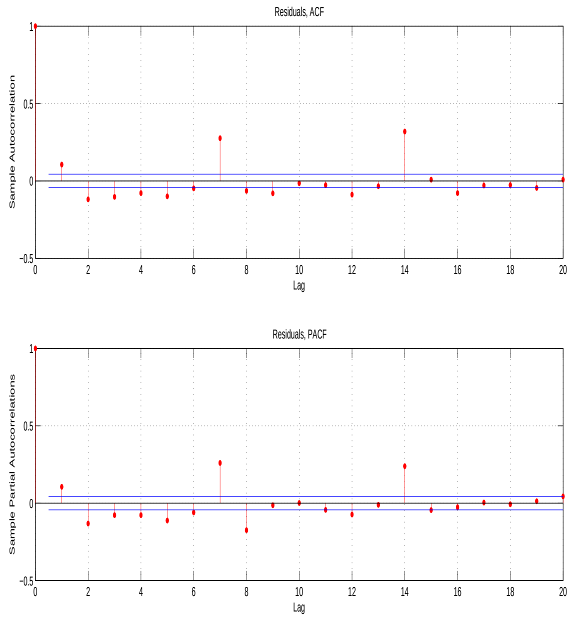


Figure 11. ACF and PACF for the residuals of the Homogenous Gompertz diffusion model (as a special case of Non-Homogenous Gompertz)

We evaluate the forecasting performance of these models by the backtesting method at daily, weekly and monthly forecast horizons. The methodology that is used in this part is explained in the previous section where Equation (41) is used to calculate the conditional expectations at each forecast horizon ($h = 1, 7, 30$). Remembering Equations (53) and (56):

$$\mathbb{E}(X_t) = \exp\left(\mu(s,t) + \frac{1}{2}\sigma^2\nu^2(s,t)\right)$$

$$\mu(s,t) = \ln x_s e^{-\beta(t-s)} + \frac{\alpha - \sigma^2/2}{\beta}(1 - e^{-\beta(t-s)})$$

$$\nu^2(s,t) = \frac{1}{2\beta}(1 - e^{-2\beta(t-s)})$$

We calculate the Relative Mean Square Errors (RMSE) as:

Table 6. Comparison of RMSE's for the 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 |
| Non-Homogenous Gompertz, Special Case | | |
| Daily | Weekly | Monthly |
| 0.0202 | 0.1121 | 0.4551 |

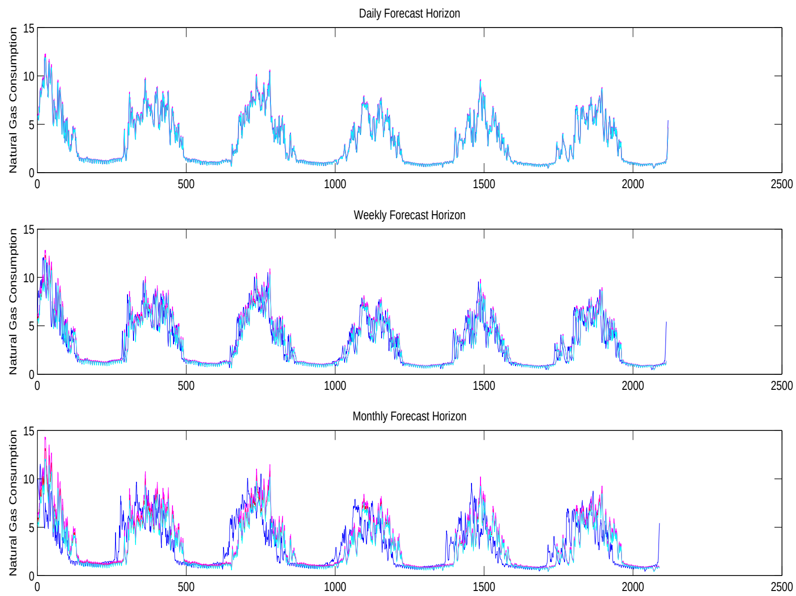


Figure 12. Backtesting with Homogenous Gompertz Diffusion model (estimated via likelihood estimators)

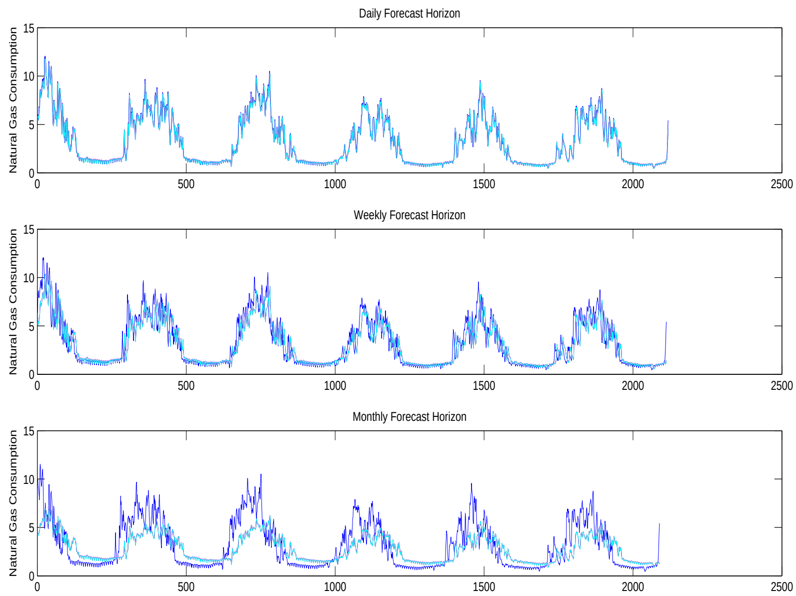


Figure 13. Backtesting with Homogenous Gompertz Diffusion model (estimated with least squares)

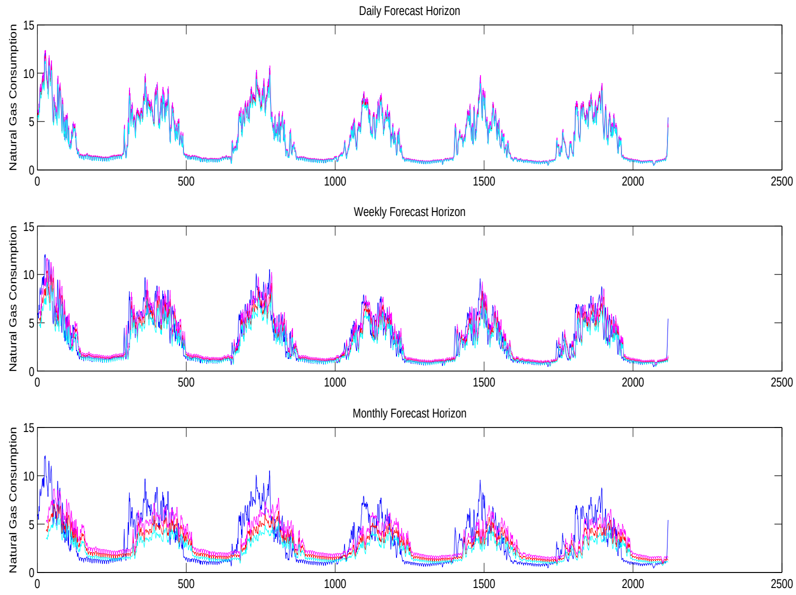


Figure 14. Backtesting with Homogenous Gompertz Diffusion model (as a special case of Non-Homogenous Gompertz)

Estimation of Model Parameters in the Non-Homogenous Gompertz Diffusion Model

In the Non-Homogenous Gompertz Diffusion model, the logarithm of per consumer natural gas consumption is described by the stochastic differential equation in Equation (81) with the functional forms for the drift and diffusion coefficients given as in Equations (79) and (80), respectively. As opposed to the Homogenous Gompertz Diffusion model, we use $\sin(wt)$, $\sin(2wt)$, $\cos(wt)$ and $\cos(2wt)$ as exogenous factors to represent the seasonality in the data.

The likelihood estimators of these parameters are obtained in Equation (100). Alternatively, a least squares method is proposed by discretizing the same stochastic differential equation as in Equation (106). The estimation results from both approaches for the entire sampling period (all $2848$ daily observations) are reported in the table below.

Table 7. Estimated Parameters for the Non-Homogenous Gompertz Diffusion Model

| Likelihood Estimators | | | | | | | |
|---|---|---|---|---|---|---|---|
| Parameters | $\alpha_0$ | $\alpha_1$ | $\alpha_2$ | $\alpha_3$ | $\alpha_4$ | $\beta$ | $\sigma$ |
| Estimates | 0.1094 | 6.9415 | -2.5431 | 0.1965 | -0.0614 | 0.1185 | 0.0234 |
| Least Squares Estimates | | | | | | | |
| Parameters | $\alpha_0$ | $\alpha_1$ | $\alpha_2$ | $\alpha_3$ | $\alpha_4$ | $\beta$ | $\sigma$ |
| Estimates | 0.1028 | 6.5412 | -2.3281 | 0.1843 | -0.0541 | 0.1114 | 0.1445 |
| Upper CI | 0.1187 | 7.5225 | -1.6315 | 0.3985 | 0.1693 | 0.1276 | |
| Lower CI | 0.0857 | 5.4858 | -2.9616 | -0.0398 | -0.2700 | 0.0936 | |

The sample autocorrelation and partial autocorrelation functions are drawn for the residuals that are obtained from the differences between the logarithm of the daily realized and predicted per consumer consumption series.

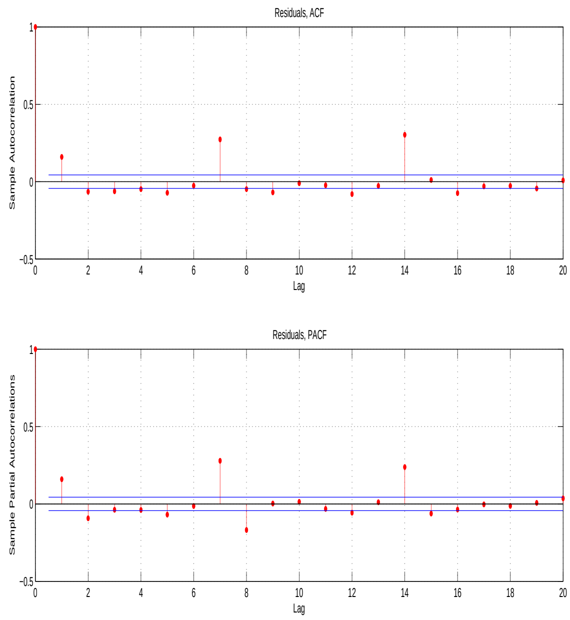


Figure 15. ACF and PACF for the residuals of the Non-Homogenous Gompertz diffusion model

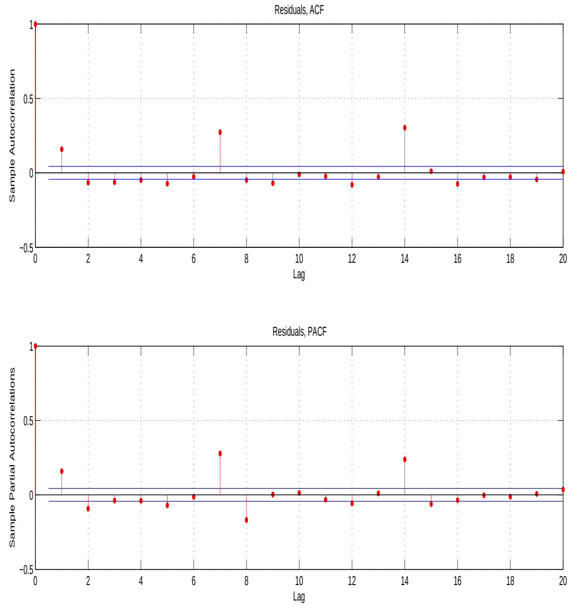


Figure 16. ACF and PACF for the residuals of the Non-Homogenous Gompertz diffusion model (estimated with least squares)

We evaluate the forecasting performance of these models by the backtesting method at daily, weekly and monthly forecast horizons. The methodology that is used in this part is explained in the previous section where Equation (83) is used to calculate the conditional expectations at each forecast horizon ($h = 1, 7, 30$). Remembering Equations (53) and (56):

$$\mathbb{E}(X_t) = \exp\left(\mu(s,t) + \frac{1}{2}\sigma^2\nu^2(s,t)\right)$$

$$\begin{aligned}\mu(s,t) = \mu(s,t) &= \ln(x_s)e^{\beta(t-s)} - \frac{\sigma^2}{2\beta}(1 - e^{\beta(t-s)}) + \int_s^t h(\tau)e^{\beta(\tau-s)}d\tau \\ &= \ln(x_s)e^{-\beta(t-s)} + \frac{\alpha_0 - \frac{\sigma^2}{2}}{\beta}(1 - e^{-\beta(t-s)}) + \sum_{i=1}^{q}\alpha_i\int_s^t g_i(\tau)e^{-\beta(t-\tau)}d\tau\end{aligned}$$

$$\nu^2(s,t) = \frac{1}{2\beta}(1 - e^{-2\beta(t-s)})$$

We calculate the Relative Mean Square Errors (RMSE) as:

Table 8. Comparison of RMSE's for the 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 |

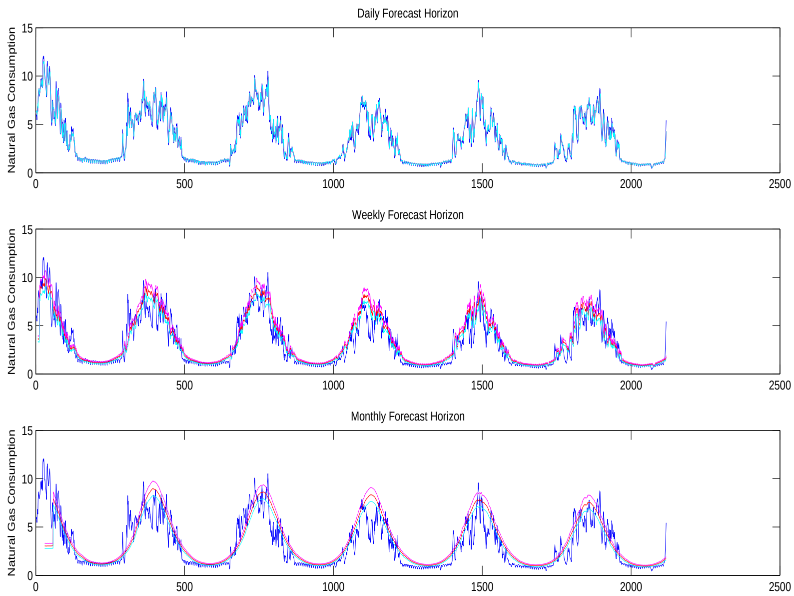


Figure 17. Backtesting with Non-Homogenous Gompertz Diffusion model

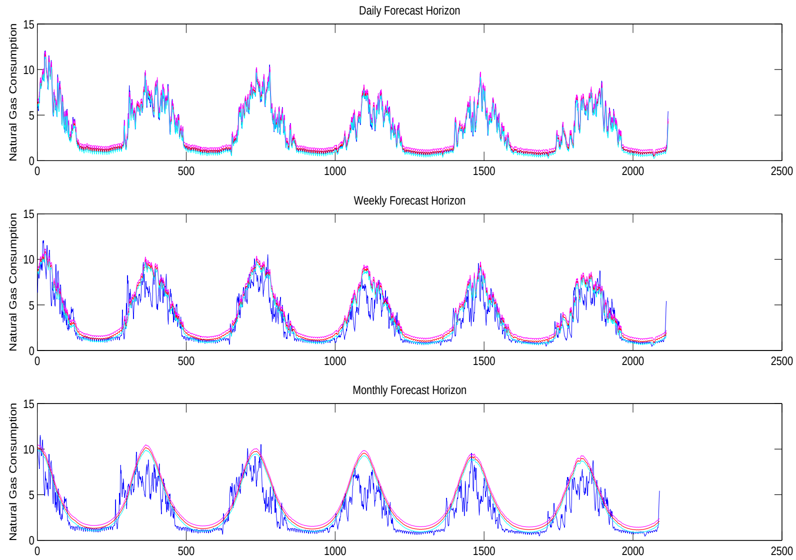


Figure 18. Backtesting with Non-Homogenous Gompertz Diffusion model (estimated with least squares)

## Pricing of Contingent Claims, Numerical Examples

In this section, we consider two option pricing examples. In the first numerical example, we assume a European option with the contract period of January 1-31, 2011 and find the price as of December 31, 2010, just before the contract period starts. Here, the payoff of the call option is given as $\max(c_T - K, 0)$, where $c_T$ is the per consumer natural gas consumption at the maturity date, January 31, 2011. First, we estimate the parameters using the dataset from the beginning until December 31, 2010. Then, we compare the option prices that are obtained from the pricing formulas derived.

As a second example, we consider an arithmetic option where the underlying variable is the arithmetic average of the natural gas consumption within the period January 1-31, 2011. Since there is no known analytical solution for this option, we use Monte Carlo simulation by generating 20,000 sample paths of consumption during the same period and calculate the average discounted payoffs as estimates for option prices.

Table 9. Comparison of European Option Prices under Different Models (Risk Free Rate 5%)

One-Factor Mean-Reverting Process

| Strike | 4.5 | 5 | 5.5 | 6 | 6.5 | 7 | 7.5 | 8 | 8.5 |
|---|---|---|---|---|---|---|---|---|---|
| Call | 2.2066 | 1.7966 | 1.4345 | 1.1249 | 0.8679 | 0.6603 | 0.4962 | 0.3690 | 0.2721 |
| Put | 0.0822 | 0.1702 | 0.3060 | 0.4944 | 0.7354 | 1.0257 | 1.3595 | 1.7303 | 2.1313 |

Homogenous Gompertz Diffusion Model

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Call | 1.8356 | 1.3377 | 0.8397 | 0.3422 | 0.0115 | 0 | 0 | 0 | 0 |
| Put | 0 | 0 | 0 | 0.0004 | 0.1677 | 0.6541 | 1.1521 | 1.6500 | 2.1480 |

Homogenous Gompertz Diffusion Model (Least Squares)

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Call | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Put | 0.1978 | 0.6957 | 1.1937 | 1.6916 | 2.1896 | 2.6875 | 3.1855 | 3.6834 | 4.1814 |

Homogenous Gompertz Diffusion Model (Non-Homogenous Gompertz, Special Case)

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Call | 1.8241 | 1.3261 | 0.8284 | 0.3509 | 0.0585 | 0.0024 | 0 | 0 | 0 |
| Put | 0 | 0 | 0.0002 | 0.0207 | 0.2262 | 0.6681 | 1.1636 | 1.6615 | 2.1595 |

Non-Homogenous Gompertz Diffusion Model

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Call | 1.8217 | 1.3254 | 0.8469 | 0.4422 | 0.1780 | 0.0540 | 0.0124 | 0.0022 | 0.0003 |
| Put | 0 | 0.0017 | 0.0211 | 0.1144 | 0.3481 | 0.7221 | 1.1785 | 1.6662 | 2.1623 |

Non-Homogenous Gompertz(Least Squares)

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Call | 3.6815 | 3.1835 | 2.6856 | 2.1876 | 1.6897 | 1.1917 | 0.6938 | 0.1958 | 0 |
| Put | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.3021 |

Table 10. Comparison of Option Prices for an Arithmetic Option under Different Models (Risk Free Rate 5%)

One-Factor Mean-Reverting Process

| Strike | 4.5 | 5 | 5.5 | 6 | 6.5 | 7 | 7.5 | 8 | 8.5 |
|---|---|---|---|---|---|---|---|---|---|
| Call | 3.1203 | 2.6223 | 2.1244 | 1.6264 | 1.1285 | 0.6305 | 0.1326 | 0 | 0 |
| Put | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.3654 | 0.8633 |
| Homogenous Gompertz Diffusion Model | | | | | | | | | |
| Call | 1.8179 | 1.3199 | 0.8220 | 0.3240 | 0 | 0 | 0 | 0 | 0 |
| Put | 0 | 0 | 0 | 0 | 0.1739 | 0.6719 | 1.1698 | 1.6678 | 2.1657 |
| Homogenous Gompertz Diffusion Model (Least Squares) | | | | | | | | | |
| Call | 0.6470 | 0.1491 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Put | 0 | 0 | 0.3489 | 0.8468 | 1.3448 | 1.8427 | 2.3407 | 2.8386 | 3.3366 |
| Homogenous Gompertz Diffusion Model (Non-Homogenous Gompertz, Special Case) | | | | | | | | | |
| Call | 2.8074 | 2.3094 | 1.8115 | 1.3135 | 0.8156 | 0.3176 | 0 | 0 | 0 |
| Put | 0 | 0 | 0 | 0 | 0 | 0 | 0.1803 | 0.6783 | 1.1762 |
| Non-Homogenous Gompertz Diffusion Model | | | | | | | | | |
| Call | 0.6565 | 0.1585 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Put | 0 | 0 | 0.3394 | 0.8374 | 1.3353 | 1.8333 | 2.3312 | 2.8292 | 3.3271 |
| Non-Homogenous Gompertz(Least Squares) | | | | | | | | | |
| Call | 2.7956 | 2.2977 | 1.7997 | 1.3018 | 0.8038 | 0.3059 | 0 | 0 | 0 |
| Put | 0 | 0 | 0 | 0 | 0 | 0 | 0.1921 | 0.6900 | 1.1880 |

# CHAPTER VIII
# CONCLUSION

In this thesis, we aim to model the dynamic behaviour of natural gas consumption while we capture the empirical properties that the consumption data shows and based on these results, we want to forecast and price various contingent claims. For this purpose, we have used continuous-time stochastic models to take advantage of the analytical solutions that can be derived for any forecast horizon which can also be used to make reliable forecasts at high-frequency levels. Hence, we have studied the application of One-factor mean-reverting process and stochastic Gompertz diffusion model in this context.

We apply our methodology to model and forecast daily natural gas consumption in Istanbul, Turkey. A time-varying deterministic function has also been incorporated into these models to account for the seasonal pattern that has been derived after a thorough analysis of the data. In the specific structure of this component, a cyclical trend and a holiday dummy variable have been used. Yet, natural gas prices have been left out since they are centrally determined and they vary infrequently. As the autocorrelation and partial autocorrelation functions for the residuals suggest, the effect of leaving prices in the error terms does not cause major problems since we do not see any striking pattern. In fact, such disturbing patterns are not observed in any of the residuals even though they are obtained from very different models.

When we compare these various models in terms of their forecasting powers, it appears that the One-factor mean-reverting process is more advantageous than the Gompertz diffusion process. From the Relative Mean Square Errors (RMSE) that are listed in Tables 4, 6 and 8, it is visible that including sinusoidal variables as exogenous factors improves the prediction powers of the models, especially for longer horizons (at the monthly horizon, 0.2194 in the One-factor mean-reverting process and 0.2003 in the Non-Homogenous Gompertz Diffusion model). The RMSE results for the Homogenous Gompertz Diffusion model are close to the Non-Homogenous case, excluding only the monthly horizon. As expected, the role of deterministic trends become more pronounced at longer horizons. Moreover, the models that are estimated from the likelihood functions perform better in the backtesting procedure since this method is more sophisticated than the least squares method.

The estimates that are obtained from the models are then used in the valuation of contracts that depend on natural gas consumption. Since there are no contracts that are actually traded in the Turkish Derivatives Exchange (TurkDex), hypothetical contracts are priced to see the implications of the models. Nevertheless, such derivative contracts on natural gas consumption or prices will probably become available in the future as the efficient management of resources gains importance. In both pricing examples, we see the expected trends: the call option prices decrease as the strike prices increase and the put option prices increase as the strike prices increase. The results that are estimated with least squares method show that they usually predict lower option prices for European calls and higher option prices for European puts than their counterparts. Moreover, the prices that are found in both examples are very different from one model to another. Hence, the choice of model that will be used in pricing becomes crucial.

Finally, the question of which model to use can be answered when the specific area of application is considered. In the context of pricing, obtaining accurate daily consumption predictions becomes highly important since daily settlement amounts are usually used in futures contracts or options. 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.

# BIBLIOGRAPHY

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

Baxter, M. & Rennie, A. (2003). *Financial calculus: An introduction to derivative pricing*. Cambridge: Cambridge University Press.

Crompton, P. & Wu, Y. (2005). Energy consumption in China: past trends and future directions. *Energy Economics*, 27(1), 195-208.

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

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

Erdoğdu, E. (2010). A Review Of Turkish Natural Gas Distribution Market. *Renewable and Sustainable Energy Reviews*, 14, 806–813.

Ferrante, L., Bompadre, S., Possati, L. & Leone, L. (2000). Parameter estimation in a gompertzian stochastic model for tumor growth. *Biometrics*, 56(4), 1076-1081.

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

Gutiérrez, R., Nafidi, A. & Gutiérrez Sánchez, R. (2004). Inference in the stochastic gompertz diffusion model with continuous sampling. *Monografías del Seminario Matemático García de Galdeano*, 31, 347–353.

Gutiérrez, R., Gutiérrez Sánchez, R., Nafidi, A., Román, P. & Torres, F. (2005). Inference in Gompertz-type Nonhomogeneous stochastic systems by means of discrete sampling. *Cybernetics and Systems: An International Journal*, 36, 203-216.

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

Gutiérrez, R., Nafidi, A. & Gutiérrez Sánchez, R. (2006). Electricity consumption in Morocco: Stochastic Gompertz diffusion analysis with exogenous factors. *Applied Energy*, 83, 1139–51.

Gümrah, F., Katırcıoğlu, D., Aykan, Y., Okumuş, S. & Kılınçer, N. (2001). Modeling of gas demand using degree-day concept: Case study for Ankara. *Energy Sources*, 23, 101-114.

Iacus, S. M. (2008). *Simulation and inference for stochastic differential equations: With R examples*. New York: Springer-Verlag.

Katsamaki, A. & Skiadas, C. H. (1995). Analytic solution and estimation of parameters on a stochastic exponential model for a technological diffusion process. *Applied Stochastic Models And Data Analysis*, 11, 59-75.

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

Maller, R. A., Müller, G. & Szimayer, A. (2009). Ornstein–Uhlenbeck processes and extensions. In T. G. Andersen, R. A. Davis, J. P. Kreiß, T. V. Mikosch (Ed.), *Handbook of financial time series* (pp. 421-437). Berlin: Springer-Verlag Berlin Heidelberg.

Ricciardi, L. M. (1976). On the transformation of diffusion processes into the Wiener process. *Journal Of Mathematical Analysis And Applications*, 54, 185-199.

Sabo, K., Scitovski, R., Vazler, I. & Zekic-Susac, M. (2011). Mathematical models of natural gas consumption. *Energy Conversion and Management*, 52, 1721–1727.

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

Sarak, H. & 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.

Taşpınar, F., Çelebi, N. & Tutkun, N. (2013). Forecasting of daily natural gas consumption on regional basis in Turkey using various computational methods. *Energy and Buildings*, 56, 23–31.

T.C. Enerji Piyasasi Düzenleme Kurumu (EPDK) Doğalgaz Piyasası Dairesi Başkanlığı (2013). *Doğalgaz Piyasasi Sektör Raporu* Retrieved June 29, 2014 from http://www.epdk.gov.tr/documents/dogalgaz/rapor_yayin/Ddp_yayin_rapor_2012.pdf