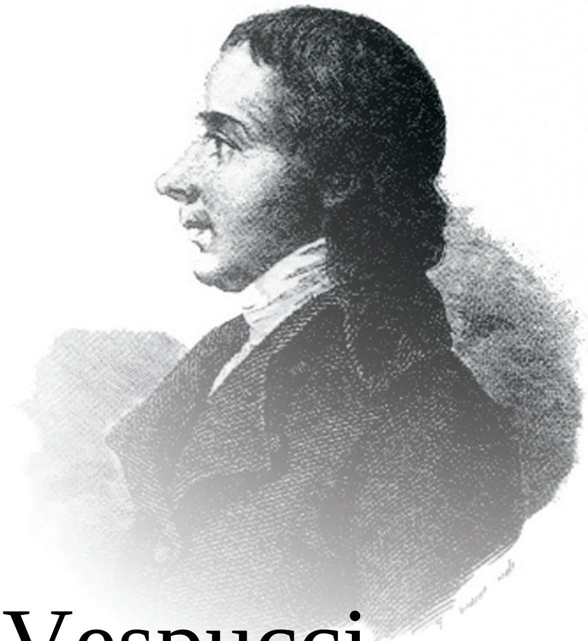
Rapporto n. 8 2007

dmsia unibg.it



F. Maggioni, M.T. Vespucci, S. Gambarini, E. Allevi, M.I. Bertocchi, R. Giacometti and M. Innorta

A stochastic framework for gas retailer based on temperature and oil prices evolution

Serie Ricerca

Dipartimento di Matematica, Statistica, Informatica e Applicazioni "Lorenzo Mascheroni"

UNIVERSITÀ DEGLI STUDI DI BERGAMO

A stochastic framework for gas retailer based on temperature and oil prices evolution

F. Maggioni[1], M.T. Vespucci[2], S. Gambarini, E. Allevi[3], M.I. Bertocchi[1], R. Giacometti[1] and M. Innorta[2].

We consider a new two–stage stochastic optimization model, named OMoGaS–2SV (Optimization Modelling for Gas Seller–Second Stochastic version), to assist companies dealing with gas retail commercialization. Due to nonlinearities present in the objective function, the model can be classified as an NLP mixed integer model, with the profit function depending on the number of contracts with the final consumers, the typology of such consumers and the cost supported to meet the final demand. Constraints related to a maximum daily gas consumption, to yearly maximum and minimum consumption in order to avoid penalties are included. Consumers consumption profiles are also considered. Temperature influences gas consumption of small consumers and is modelled by a mean reverting process. Oil prices influence the energetic indices to which sell and purchase prices are related. Forward curves of energetic indices have been analyzed by econometric models while exchange rates are modelled by a GARCH model. The results obtained by the stochastic version give clear indication of the amount of losses that may appear in the gas seller's budget.

Key Words. Gas sale company, energetic indices, mean reverting process, stochastic programming.

1 Introduction

Starting in 1999 the Italian Natural Gas market has been undergoing a liberalization process aiming at promoting competition and efficiency, while ensuring adequate service quality standards. Timings and methods for the internal gas market liberalization have been introduced following the European Gas Directive; the roles of different segments of the natural gas "chain" have been identified and defined, such as import, production, export, transportation and dispatching, storage, distribution and sale. In particular, the principle has been introduced of unbundling among supply and transport/storage and among distribution and selling. Before liberalization there was a national monopolistic operator, for all activities related to supply, transport, storage and wholesale

[1]Department of Mathematics, Statistic, Computer Science and Applications, Bergamo University, Via dei Caniana 2, Bergamo 24127, Italy.
[2]Department of Informatics and Mathematical Modeling, Bergamo University, Via Marconi 5, Dalmine 24044, Italy.
[3]Department of Quantitative Methods, Brescia University, Contrada S. Chiara, 50, Brescia 25122, Italy.

commercialization, and local monopolistic operators, for distributing and selling to final consumers. After liberalization the following operators run different activities:

- shippers: production/import, re-gasification and wholesale commercialization;
- national distributor: transport on national network and storage;
- local distributors: transport on local network;
- selling companies: purchase gas from shippers and sell it to final consumers.

In 2003 the Italian Regulatory Authority for Electricity and Gas, see [11], defined consumption classes, on the basis of gas consumption in the thermal year, and introduced a new gas tariff with *Delibera* 134_06 in order to guarantee small consumers' protection by applying the transparency principle in the pricing mechanism. The new tariff is based on a detailed splitting in different components, whose values are periodically revised, and represents a maximum price to be applied to small consumers. On the other hand, the gas buying price is influenced by some energetic indices related to the five following oil prices: *Gasoil* 0.2, 1% *Fuel Oil*, 3.5% *Fuel Oil*, *Brent Dated* and a mixture of *Crude Oils*, expressed in \$/(Metric Ton) and whose daily quotations are available on *Platts* [17]. Consequently, in the purchase contract, the gas seller can choose among 17 possible energetic indices formulas given by different linear combinations of five oil prices reported above and by which the gas purchase price is monthly updated. Furthermore, unlike domestic customers, who purchase gas according to the index formula described by Delibera 134_06, industrial customers can choose the index formula by which the gas purchase price is computed, which implies that the gas seller faces the problem of develop a *Risk Management Area for Oil Commodities* to evaluate and control this market risk.

Two optimization models, a deterministic one, see Allevi *et al.*, (2007) [2] and a stochastic one, see Allevi *et al.*, (2007) [4] and Maggioni *et al.*, (2006) [19], have been developed to assist companies dealing with retail commercialization. For each citygate, the gas seller has to decide the customer portfolio, i.e. the number of final customers to supply in each consumption class, and the sell prices to apply to each consumption class. Indeed, different customer portfolios determine different citygate consumption patterns, which shippers refer to when setting the gas price to be paid by the gas seller for the citygate supply. For each thermal year and each citygate there is a contract between shipper and gas seller setting:

- the gas volume required by gas seller in the next thermal year;
- the gas volume required in particular in winter months;
- the maximum daily consumption (capacity) requested by gas seller;
- the purchase price fixed by the shipper;
- energetic index formula.

In the contract it is also specified how to compute penalties, due by gas seller if daily consumption exceeds daily capacity.

The stochasticity considered in Maggioni *et al.*, (2006) [19], is due to the influence of the temperature on consumption whereas sell and purchase prices do not change during the contract. For small customers, using gas either only for cooking or for cooking and heating, or for commercial activities and small industries, gas consumption in winter months strongly depends on the weather conditions: this fact is taken into account in the model, by including a mean reverting process modelling temperature and generating temperature scenarios.

In this paper we consider again temperature as source of stochasticity but we also consider the influence of oil prices on the energetic indices upon which sell and purchase prices are monthly updated. We have modelled oil prices evolution by an endogenous VAR(5) econometric model with five lags, the exchange rates €/$ by an IGARCH model and finally we have built the forward curves of energetic indices by Monte Carlo simulation on the errors.

The models for temperature and energetic indices are presented in sections 2 and 3 respectively. In section 4 the stochastic model, named OMoGaS–2SV, is presented and in section 5 numerical results related to a case study are reported and discussed.

2 The stochastic temperature model

In this section we want to find a stochastic model describing the temperature variations along the months in a year time. We start with some definitions about temperature:

Definition 1 Given a weather station, let T_μ^{max} and T_μ^{min} denote the maximum and the temperatures (in Celsius degrees) measured in day μ, respectively. We define the **mean temperature** of day μ as

$$T_\mu = \frac{T_\mu^{max} + T_\mu^{min}}{2} \quad . \tag{1}$$

Definition 2 Let T_μ denote the mean temperature of day μ. We define **Heating Degree Days** (HDD_μ: measure of cold in winter) and **Cooling Degree Days** (CDD_μ: measure of heat in summer) respectively as

$$HDD_\mu = max\,\{18 - T_\mu, 0\} \quad , \tag{2}$$

$$CDD_\mu = max\,\{T_\mu - 18, 0\} \quad . \tag{3}$$

For a given day HDD and CDD are the numbers of degrees of deviation from a reference temperature level in Bergamo (18° C). The name "heating degree days" refers to the fact that if temperature is below 18° C people tend to use more energy to heat their homes; the name "cooling degree days" refers to the fact that if temperature is above 18° people start turning their air conditioners on. Typically the HDD season is from

November to March, whereas the CDD season is from May to September. April and October are often referred to as "shoulder months".

We have a database of temperatures measured in Bergamo in the last 12 years (1/01/1994–30/11/2005). The database consists of daily minimum and maximum temperatures, from which average daily temperatures are computed using (1). Due to the cyclical nature of the temperature process we find that historical data give a reasonable idea of the temperature level in the future. In Figure 1 we have plotted the daily mean temperatures at Bergamo for the 12 years; it is evident that the temperature process is mean stationary and variance stationary: it could be modeled as a mean reverting process, reverting to some cyclical function. Being the temperature process evidently not deterministic, we must consider the presence of noise. Moreover, the histogram of the daily temperature differences in Bergamo (1994–2005) shows a good fit with the corresponding normal distribution, though the frequency of small differences in daily mean temperature is underestimated. Hence, the temperature process can be modelled as a Brownian Motion.

In order to model the temperature behavior, we consider a Vasicek process with mean reversion through the following stochastic differential equation:

$$\mathrm{d}T_t = a\left(\vartheta - T_t\right)\mathrm{d}t + \sigma \mathrm{d}W_t \ , \tag{4}$$

where T_t is the process to be modelled, $a \in \mathbb{R}$ is the speed of mean reversion, ϑ is the mean (constant) which the process reverts to, σ is the process volatility (constant) and W_t is the Wiener process.
For the temperature process we need a $\vartheta = \vartheta\left(t\right) = \vartheta_t$ computed according to (8); we also need $a = a\left(i\right) = a_i$ and $\sigma = \sigma\left(i\right) = \sigma_i$ as functions changing over the months but constant in each month i.
Then our process becomes

$$\mathrm{d}T_t = a_i\left(\vartheta_t - T_t\right)\mathrm{d}t + \sigma_i \mathrm{d}W_t \ . \tag{5}$$

We need to determine a functional form for ϑ_t and estimates for a_i and σ_i from historical data. Dornier and Queruel, (2000) [13], showed that the process found in (5) is not reverting to ϑ_t; to obtain a process that really reverts to the mean we have to add the term $\vartheta_t^{'}$ to the drift term in (5) so that the equation becomes

$$\mathrm{d}T_t = \left[a_i\left(\vartheta_t - T_t\right) + \frac{\mathrm{d}\vartheta_t}{\mathrm{d}t}\right]\mathrm{d}t + \sigma_i \mathrm{d}W_t \ . \tag{6}$$

The proof of reversion to the mean can be found in the Appendix.

2.1 The mean temperature ϑ_t

By observing the plot of the temperature data measured in Bergamo in the last 12 years, see Figure 1, we note a strong seasonal variation, which can be modelled by the function

$$\sin\left(\omega t + \varphi\right) \ , \tag{7}$$

where t is the time measured in days, $\omega = 2\pi/365$ is the period of oscillation and φ is a phase angle due to the fact that the yearly minimum and maximum mean temperatures do not necessarily occur at January 1 and July 1 respectively. Moreover the mean temperature actually increases each year (the positive trend in the data is weak but it does exist): therefore we assume a linear warming trend. A deterministic model ϑ_t for the mean temperature at time t, is assumed to be given by

$$\vartheta_t = A + Bt + C \sin(\omega t + \varphi) \ , \tag{8}$$

or equivalently by

$$\vartheta_t = A + Bt + C \left[\cos(\varphi) \sin(\omega t) + \sin(\varphi) \cos(\omega t)\right] \ , \tag{9}$$

where we estimate the unknown parameters A, B, C, ω and φ so that the curve given by (9) fits the data.

In order to estimate the parameters in (9), a change of variables is operated and the constants are renamed as follows

$$\begin{cases} A = a_1 \\ B = a_2 \\ C \cos(\varphi) = a_3 \\ C \sin(\varphi) = a_4 \end{cases} \tag{10}$$

or equivalently

$$\begin{cases} A = a_1 \\ B = a_2 \\ C = \sqrt{a_3^2 + a_4^2} \\ \varphi = \arctan\left(\frac{a_4}{a_3}\right) - \pi \end{cases} \tag{11}$$

and we obtain

$$\vartheta_t = a_1 + a_2 t + a_3 \sin(\omega t) + a_4 \cos(\omega t) \ . \tag{12}$$

The numerical values of the parameters in (12) are computed by the least squares method, i.e. the parameter vector $\xi = (a_1, a_2, a_3, a_4)$ is computed that solves

$$min_\xi \left\|\vartheta - \mathbf{X}\right\|^2 \ , \tag{13}$$

where ϑ is the vector whose elements are given by (12) and $\mathbf{X}$ is the data vector. By using the series of 4323 observations of the historic daily temperatures we get

$$\begin{cases} A = 13.33 \\ B = 6.8891 \cdot 10^{-5} \\ C = 10.366 \\ \varphi = -1.7302 \end{cases} \tag{14}$$

In Figure 1 we can see a comparison between the observed temperatures and those estimated by using the deterministic approach given by ϑ_t in the years $1994 - 2005$.

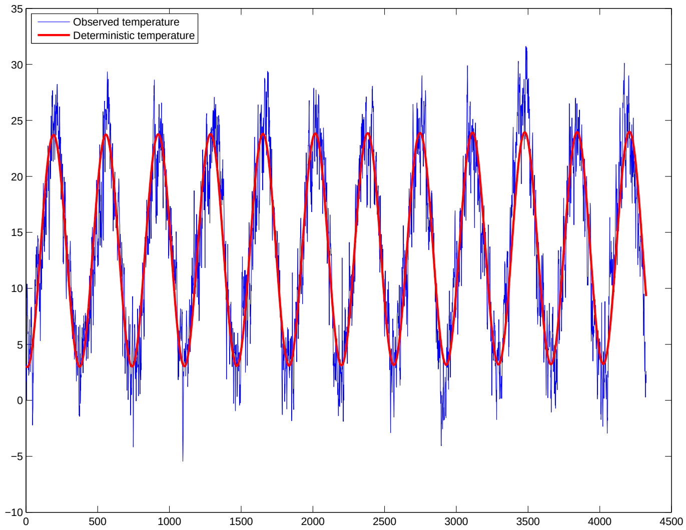


Figure 1: Comparison between measured temperatures and estimated mean $\vartheta(t)$ at Bergamo in the years $1994 - 2005$.

2.2 Estimation of σ_i

For the estimation of the volatility σ_i we follow the same approach as in Alaton *et al.*, (2002) [1], where the quadratic variation σ_i^2 of temperature is assumed to be different along the months in the year, but nearly constant within each month. Since the gas retailer model is based on a monthly discretization, it is only necessary to estimate a volatility value for each month. For this reason σ_i is assumed to be a piece-wise constant function, with a constant value during each month. One possibility is to use an estimator based on the quadratic variation of T_t (see Basawa and Prasaka Rao, (1980) [5])

$$\hat{\sigma}_i^2 = \frac{1}{N_i} \sum_{t=0}^{N_i-1} \left(T_{t+1} - T_t\right)^2 \ , \tag{15}$$

where N_i denotes the number of days of month i and $t = 0$ refers to the last day of the previous month.

Another estimator is derived by discretizing (6) and using the discretized equation as a regression equation. During a given month i, the discretized equation is

$$T_t = \vartheta_t - \vartheta_{t-1} + a_i \vartheta_{t-1} + (1 - a_i)\, T_{t-1} + \sigma_i \epsilon_{t-1} \quad t = 1 \ldots N_i \ , \tag{16}$$

where $\{\epsilon_t\}_{t=1}^{N_i-1}$ are independent standard normally distributed random variables. Thus an efficient estimator of σ_i is (see Brockwell and Davis, (1990) [8]),

$$\hat{\sigma}_i^2 = \frac{1}{N_{i-2}} \sum_{t=1}^{N_i} \left(T_t - (\vartheta_t - \vartheta_{t-1}) - \hat{a}_i \vartheta_{t-1} - (1 - \hat{a}_i)\, T_{t-1}\right)^2 \ , \tag{17}$$

Month	Estimator 1	Estimator 2	Mean Value
January	1.6508	1.6196	1.6352
February	1.5415	1.5515	1.5465
March	1.7455	1.7209	1.7332
April	1.8480	1.8305	1.8393
May	1.8142	1.8013	1.8078
June	1.9871	1.9765	1.9818
July	1.7605	1.7298	1.7452
August	1.6305	1.6402	1.6354
September	1.4805	1.4674	1.4739
October	1.3831	1.3905	1.3868
November	1.5062	1.4933	1.4998
December	1.4912	1.4899	1.4906

Table 1: The estimators of σ_i based on the quadratic variation and the regression approach and their mean value.

where $\hat{a}_i$ is estimated in the following section. In Table 1 for each month i the estimator of σ_i based on the quadratic variation, the one based on the regression approach and their mean value are reported.

2.3 Estimation of Speed of reversion

According to Bibby and Sorensen, (1995) [6], based on observations collected during N_i days of month i, an efficient estimator $\hat{a}_i$ of a_i is the zero of the martingale function given by

$$G(a_i) = \sum_{t=1}^{N_i} \frac{\dot{b}(T_{t-1}; a_i)}{\sigma_{i,t-1}^2} \{T_t - E[T_t | T_{t-1}]\} \ , \tag{18}$$

where $\dot{b}(T_{t-1}; a_i)$ denotes the derivative with respect to a_i of the drift term

$$b(T_t, a_i) = \frac{\mathrm{d}\vartheta_t}{\mathrm{d}t} + a_i(\vartheta_t - T_t) \ . \tag{19}$$

In order to obtain the solution of (18), we have to determine each of the terms $E[T_t|T_{t-1}]$; thus, if we take again the process developed in (6) for a given month i and integrate between day $(t-1)$ and day t in month i, we find

$$T_t = \vartheta_t + e^{-a_i}(T_{t-1} - \vartheta_{t-1}) + e^{-a_i t} \int_{t-1}^{t} \sigma_s e^{a_i s} \mathrm{d}W_s \ , \tag{20}$$

which yields

$$E[T_t | T_{t-1}] = e^{-a_i}(T_{t-1} - \vartheta_{t-1}) + \vartheta_t \ , \tag{21}$$

Month	Estimator $\hat{a}_i$
January	0.2707
February	0.2055
March	0.2017
April	0.1755
May	0.3079
June	0.2364
July	0.3051
August	0.2559
September	0.2666
October	0.1594
November	0.183
December	0.1969

Table 2: The estimator $\hat{a}_i$ based on the formula 23.

because the expected value of an Itô integral is zero.
By substituting (21) in (18) we find

$$G_n(a_i) = \sum_{t=1}^{n} \frac{\dot{b}(T_{t-1}; a_i)}{\sigma_{t-1}^2} \left[T_t - \vartheta_t - e^{-a_i} (T_{t-1} - \vartheta_{t-1}) \right] \ , \tag{22}$$

from which we obtain

$$\hat{a}_i = -\log \left(\frac{\sum_{t=1}^{n} \frac{\vartheta_{t-1} - T_{t-1}}{\sigma_{t-1}^2} (T_t - \vartheta_t)}{\sum_{t=1}^{n} \frac{\vartheta_{t-1} - T_{t-1}}{\sigma_{t-1}^2} (T_{t-1} - \vartheta_{t-1})} \right) \ . \tag{23}$$

Inserting the data of temperatures and the estimator $\hat{\sigma}$ given by (15), we find the estimator $\hat{a}_i$. In Table 2 the values of the estimator $\hat{a}_i$ in the twelve months are reported.

2.4 Generation of temperature scenarios

In this section we consider the problem of generating temperature scenarios. By using Euler approximation scheme, we discretize equation (6) obtaining

$$T_t = \vartheta_t - \vartheta_{t-1} + a_i \vartheta_{t-1} + (1 - a_i) T_{t-1} + \sigma_i \epsilon_{t-1} \ , \tag{24}$$

where $\{\epsilon_t\}_{t=1}^{364}$ are independent standard normally distributed random variables. Figure 2 shows both the evolution of a simulated trajectory of the estimated temperature and its mean ϑ_t, while Figure 3 gives the evolution of 10 scenarios of temperatures.

The following notation is used:

- $\mathbf{T}^s \in \mathbb{R}^{365}$ is the vector of random variables along scenario s, $s = 1, \ldots, N$ which we have obtained using a mean reverting process; the component T_t^s represents the daily average heating degree days for day t, $t = 1, \ldots 365$ along scenario s;

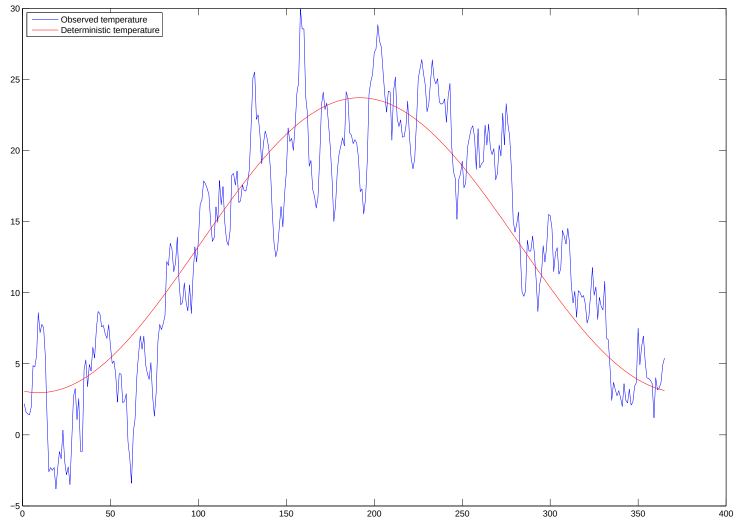


Figure 2: Simulation of sample paths of temperature and the mean estimated by Monte Carlo method.

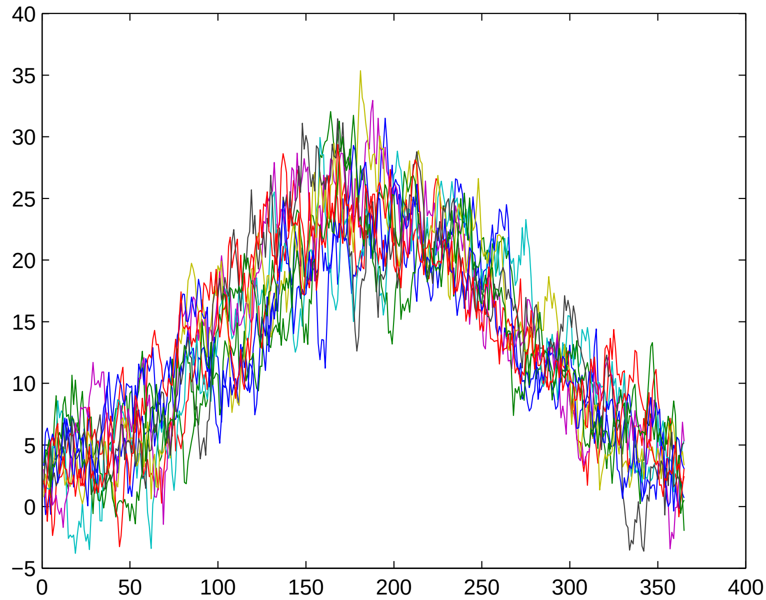


Figure 3: 10 scenarios of temperature estimated by Monte Carlo simulation.

- Due to the fact that the consumption data are monthly data, we generate monthly temperature scenarios from the vector $\mathbf{T}^s$ by averaging. $\mathbf{Tm}^s$ represents the monthly temperature scenario s, whose component Tm_i^s represents the monthly heating degree days for month i, $i = 1, \dots 12$ along scenario s.

- $\bar{T}m_i = \frac{\sum_{s=1}^{N} Tm_i^s}{N}$ for $i = 1, \dots, 12$, is the expected value over all scenarios of the random variable Tm_i^s;

- $\mathbf{\Delta}^s \in \mathbb{R}^{12}$ is the vector of distances of monthly heating degree days from its expected value along scenario s, $s = 1, \dots, N$, i.e. $\Delta_i^s := Tm_i^s - \bar{T}m_i \quad i = 1, \dots, 12, \; s = 1, \dots, N$.

- p^s is the probability related to each scenario s, $s = 1, \dots, N$; we assume equal probability, i.e. $p^s = \frac{1}{N}$, $s = 1, \dots, N$;

3 Forward curves of energetic indices

3.1 Econometric model of the forward curves of energetic indices

In this section we introduce the econometric model which describe oils evolution which the gas price depends from. The considered oils are *Gasoil* 0.2, a primary distillation of crude oil, 1% *Fuel Oil* and 3.5% *Fuel Oil*, respectively a low and high sulphur concentration fuel oils, *Brent Dated*, a crude oil of North Europe and a mixture of *Crude Oils* of Arabian countries. We have analyzed the database of these oils prices expressed in \$/(Metric ton) from January 1998 to June 2005; their behaviors against time are plotted in Figure 4, in particular we have used monthly data because we are interested in long period forecasting.

From the picture we can deduce some characteristics of these oils prices: *positive correlation* (see also the correlation matrix reported in Table 3), *non-stationarity*, *non-trend-stationarity* and an increasing trend in the last 6 observations. In order to test non-stationarity and non-trend-stationarity of the series we have used the *unit-root test* (see Dickey and Fuller (1979), [12]). The test was used also to check the stationarity of returns of prices. In order to be sure to have stationarity on errors of the regression between oil prices, we have to test the *cointegration* (see Engle and Granger (1987), [14]). On this purpose we have checked the prices series cointegration by using the Johansen's procedure (see Johansen, (1988) [18]) based on the *trace test*.

The second step is to estimate the regression model (the PcGive 10.3 package has been used). In order to reduce price volatility, we have considered the logarithm of prices. Moreover, in order to capture the evolution and the interdependencies between the five price time series, we have used a *vectorial autoregressive model* VAR(p):

$$\mathbf{Y}_t = \mathbf{c} + A_1 \mathbf{Y}_{t-1} + \cdots + A_p \mathbf{Y}_{t-p} + \boldsymbol{\epsilon}_t \;, \tag{25}$$

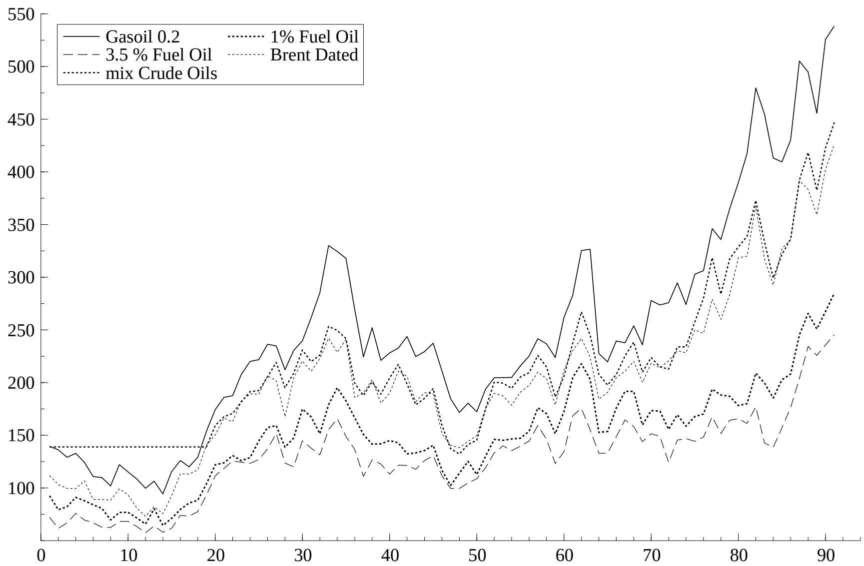


Figure 4: Behaviors of the five oils prices considered from January 1998 to June 2005.

oils prices	*Gasoil 0.2*	1% *Fuel O.*	3.5% *Fuel O.*	*Brent Dated*	mix *Crude O.*
Gasoil 0.2	1.0000	0.92527	0.89595	0.98551	0.97287
Fuel O.	0.92527	1.0000	0.97617	0.93915	0.90788
3.5% *Fuel O.*	0.89595	0.97617	1.0000	0.92699	0.88928
Brent Dated	0.98551	0.93915	0.92699	1.0000	0.9732
mix *Crude O.*	0.97287	0.90788	0.88928	0.9732	1.0000

Table 3: Correlation matrix of the oils prices.

where $\mathbf{Y}_t,\ \ldots,\mathbf{Y}_{t-5}$ are 5×1 vectors, $\mathbf{c}$ is 5×1 vector of constants, $\mathbf{A}_i,\ i = 1, \cdots, p$ are 5×5 matrices and $\boldsymbol{\epsilon}_t$ is a 5×1 vector of error terms with mean equal to zero ($E(\boldsymbol{\epsilon}_t) = 0$), variance–covariance matrix $E(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t')$ given by a 5×5 positive definite matrix and without correlation across time ($E(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_{t-k}') = 0,\ \forall k \neq 0$). By using the Granger's causality technique (see Engle and Granger, (1987) [14]), we have also tested that the VAR model is *endogenous* and finally, through some *tests of correct specification on the errors* and checking the forecasting test reported below in Figure 5, we have concluded that the best model to describe the oil prices evolution is a VAR(5) with five lags given by eq. (25) with $p = 5$ and where:

$$\mathbf{c} = \begin{pmatrix} -0.760758 \\ -0.778114 \\ -0.0353831 \\ -0.426081 \\ -0.910448 \end{pmatrix},$$

$$\mathbf{A}_1 = \begin{pmatrix} 0.0059592 & -0.263299 & -0.126311 & 0.922934 & 0.309547 \\ -0.591126 & 0.74439 & 0.281978 & 1.01744 & -0.426275 \\ -0.669291 & 0.173789 & 0.620519 & 1.32116 & -0.370503 \\ -0.723226 & -0.0943601 & -0.0381978 & 1.16304 & 0.417903 \\ -0.702463 & -0.0299675 & -0.0610118 & 0.843319 & 0.942224 \end{pmatrix},$$

$$\mathbf{A}_2 = \begin{pmatrix} 0.322835 & 0.123564 & 0.0885886 & 0.00826908 & -0.453064 \\ 0.350232 & -0.415353 & -0.225388 & -0.21569 & 0.141177 \\ 0.0514133 & 0.178525 & -0.217244 & -0.0657114 & -0.435402 \\ 0.374726 & 0.0199061 & -0.267055 & -0.0080471 & -0.150649 \\ 0.471634 & -0.122579 & -0.0353559 & -0.0806189 & -0.362299 \end{pmatrix},$$

$$\mathbf{A}_3 = \begin{pmatrix} -0.239912 & 0.00669066 & -0.0871119 & 0.342314 & 0.00302752 \\ -0.21206 & 0.149898 & 0.286594 & -0.0227147 & 0.292289 \\ -0.158735 & -0.266217 & 0.234751 & -0.0354881 & 0.777976 \\ -0.236272 & -0.209425 & 0.366504 & 0.449007 & -0.0507727 \\ -0.455109 & 0.0289447 & 0.212137 & 0.41696 & -0.0137538 \end{pmatrix},$$

$$A_4 = \begin{pmatrix} -0.181235 & 0.409444 & -0.257505 & -0.281725 & 0.223338 \\ -0.269708 & 0.474968 & -0.251569 & 0.3478 & -0.681305 \\ -0.0584977 & 0.784911 & -0.36299 & 0.412928 & -1.05716 \\ -0.0560197 & 0.749746 & -0.474892 & -0.384128 & -0.0478829 \\ 0.0514887 & 0.519774 & -0.481782 & -0.148742 & -0.0688354 \end{pmatrix},$$

$$A_5 = \begin{pmatrix} -0.222403 & -0.379329 & 0.351165 & 0.474257 & 0.0870958 \\ 0.0987397 & -0.451213 & 0.331088 & 0.0164765 & 0.411279 \\ -0.10459 & -0.541229 & 0.0821804 & 0.215786 & 0.504756 \\ 0.0230342 & -0.70368 & 0.528832 & 0.152169 & 0.300377 \\ 0.0919056 & -0.62986 & 0.672984 & 0.161777 & -0.0109883 \end{pmatrix}.$$

3.2 Monte Carlo simulation of the forward curves of energetic indices

Scenarios of (forecasted) oil prices have been generated by Monte Carlo simulation.

The error terms ϵ_t^i $(i = 1, \ldots, 5)$ of the VAR(5) model (see eq. (25)) are correlated normally distributed random variables, thus they can be described by the Brownian motion

$$\epsilon_t^i := Y_t^i - Y^{i,*} = \mathrm{d}Y_t^i = \mu_i \mathrm{d}t + \sigma_i \mathrm{d}W_t^i \ , \qquad i = 1, \ldots, 5 \ , \tag{26}$$

where $Y^{i,*}$ are the observed values, μ_i and σ_i are respectively the mean and the variance of the errors series and W_t^i is the Wiener process. In our particular case eq. (26) is given by:

$$\begin{aligned} \epsilon_t^1 &= -0.80242\mathrm{d}t + 19.862\mathrm{d}W_t^1 \ , \\ \epsilon_t^2 &= -0.47046\mathrm{d}t + 12.42\mathrm{d}W_t^2 \ , \\ \epsilon_t^3 &= -0.39153\mathrm{d}t + 10.399\mathrm{d}W_t^3 \ , \end{aligned}$$

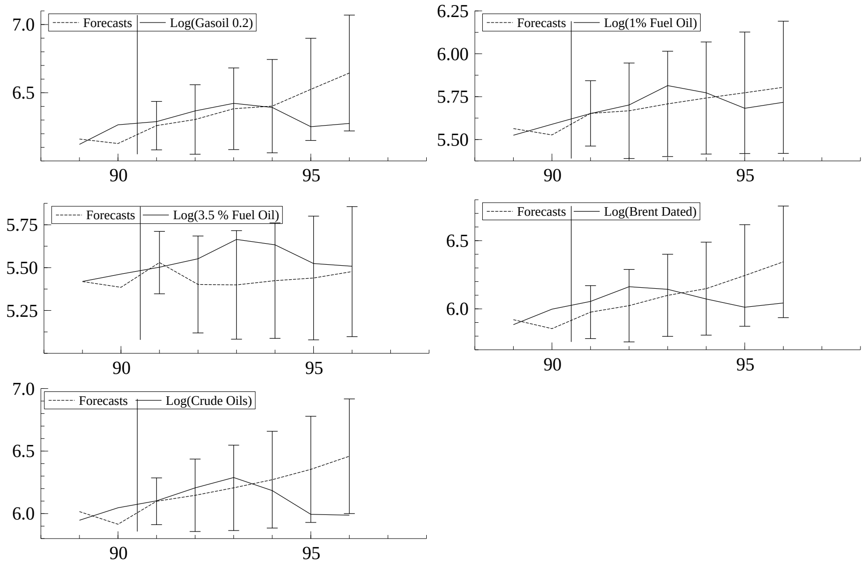


Figure 5: Forecasting analysis of the endogenous model VAR(5) for the oils prices time series, with exclusion of the first 24 observations.

$$\begin{aligned} \epsilon_t^4 &= -0.71279\mathrm{d}t + 16.968\mathrm{d}W_t^4 \ , \\ \epsilon_t^5 &= -0.67033\mathrm{d}t + 17.797\mathrm{d}W_t^5 \ , \end{aligned}$$

where we note that the Wiener processes W_t^i are correlated; we have decomposed the correlation matrix Γ of the errors (see Table 4), by using the *Cholesky decomposition* given by

$$\Gamma = C^T C \ , \tag{27}$$

where $C = [c_{ij}]$ is the lower triangular Cholesky matrix, reported in Table 5, from which eq. (26) can be rewritten as

$$\epsilon_t^i = \mathrm{d}Y_t^i = \mu_i \mathrm{d}t + \sigma_i \sum_{j=1}^{5} c_{ij} \mathrm{d}\widetilde{W}_t^j \ , \qquad i = 1, \ldots, 5 \ , \tag{28}$$

where $\widetilde{W}_t^j$ are independent Wiener processes.

	Gasoil 0.2	1% *Fuel O.*	3.5% *Fuel O.*	*Brent Dated*	mix *Crude O.*
Gasoil 0.2	1.0000	0.77755	0.7091	0.87108	0.8406
Fuel O.	0.77755	1.0000	0.8185	0.75532	0.81416
3.5% *Fuel O.*	0.7091	0.8185	1.0000	0.7581	0.83911
Brent Dated	0.87108	0.75532	0.7581	1.0000	0.93284
mix *Crude O.*	0.8406	0.81416	0.83911	0.93284	1.0000

Table 4: Correlation matrix of the errors ϵ_t^i $i = 1\ldots,5$ of VAR(5) model.

	Gasoil 0.2	1% *Fuel O.*	3.5% *Fuel O.*	*Brent Dated*	mix *Crude O.*
Gasoil 0.2	1.0000	0	0	0	0
Fuel O.	0.77755	0.62882	0	0	0
3.5% *Fuel O.*	0.7091	0.42483	0.56276	0	0
Brent Dated	0.87108	0.12406	0.15586	0.44893	0
mix *Crude O.*	0.8406	0.25532	0.23913	0.29328	0.29155

Table 5: Choleski matrix C obtained by the decomposition of the correlation matrix $\Gamma = C^T C$.

3.3 Forecasting analysis for exchange rates

In this section we consider the problem of simulating exchange rates between dollar and Euro; the oil prices are in fact expressed in dollar, while in the contracts between the shipper and gas retailer seller the price of gas is in Euro. We have analyzed the database of the exchange rates expressed in €/$ from January 1998 to June 2005, their monthly and daily behaviors against time; we have also analyzed their returns. The related plots are shown in Figure 6, in particular we have considered also daily data with the aim to verify possible non-linearity. From the picture as well as and from some tests, such as *unit–root* test and the *Bera–Jarque* test on normality, we can deduce that the exchange rate returns series is stationary whereas the exchange rate series are not; both are not normally distributed. We have estimated an autoregressive model AR(p) varying the number of lags $p = 2, \ldots, 4$. We have obtained that the best process for daily returns of exchange rates is an AR(2) whereas for the monthly case is an AR(1). In both cases the low values of R^2–test (see e.g. Davidson, (2000) [10]) bring us to consider this modelization unreliable.

We have modelled the returns of exchange rates r_t by a *generalized autoregressive conditional heteroskedasticity model* GARCH(p, q) (see Bollerslev, (1986) [7]) as follows

$$r_t = c + \epsilon_t \ , \tag{29}$$

where c is a constant, $\epsilon_t = N(0, \sigma_t^2) = \sigma_t N(0, 1)$ is the error term such that its variance σ_t^2 at time t depends on the squared error terms from p previous periods and on q previous variances:

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j \epsilon_{t-j}^2 + \sum_{k=1}^{q} \beta_k \sigma_{t-k}^2 \ . \tag{30}$$

By using some tests of correct specification we found that the best model is an Integrated Generalized Autoregressive Conditional Heteroskedasticity IGARCH(1,1),

$$\begin{cases} \sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \ , \\ \alpha_1 + \beta_1 = 1 \ , \end{cases} \tag{31}$$

which consists in a restricted version of the GARCH model, where the sum of the persistent parameters α and β sum up to one. The numerical results are given by

$$\begin{cases} r_t = 0.000216 + \epsilon_t \ , \\ \sigma_t^2 = -0.00007 + 0.005748\epsilon_{t-1}^2 + 0.994452\sigma_{t-1}^2 \ . \end{cases} \tag{32}$$

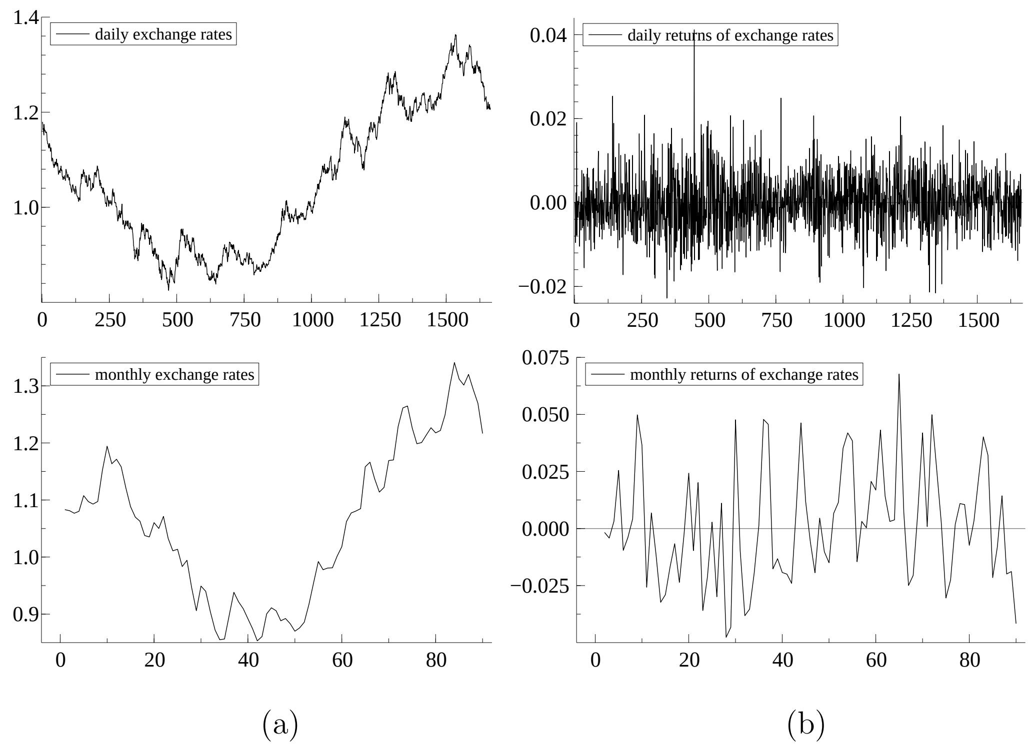


Figure 6: (a) Daily (top row) and monthly (bottom row) exchange rates prices and (b) their relative variation expressed in €/$ from January 1998 to June 2005.

Simulating exchange rates by Monte Carlo method, we built 10000 gas prices scenarios, according to different indices formulas, denoted in the sequel by $\psi = 1, \ldots, 17$.

4 The two–stage stochastic OMoGaS-2SV model

In the literature (see Brooks, (1981) [9], Eydeland and Wolyniec, (2003) [15], Ermoliev and Wets, (1988) [16] and Ruszczynski and Shapiro, (2003) [20]) stochastic approaches in the gas market deal mainly with the scheduling of development of gas fields, the use of gas storage and the gas delivery problem.

The stochastic version of our model, which can be classified as a two-stage stochastic program with recourse, includes a stochastic data process $\boldsymbol{\omega} = (\ \boldsymbol{\Delta},\ \boldsymbol{I_\psi})$ where the first component $\boldsymbol{\Delta}$ represents the temperature and the second component $\boldsymbol{I_\psi}$ represents the oil index price along the year. The consumptions of the first six classes of consumers are considered as dependent on temperature variations along the months. We also consider the stochastic nature of purchase and sell prices monthly updated according to energetic indices related to oil prices.

The following notations are used:

- $I = \{i = 1, \ldots, 12\}$ is the set of month indices, with $i = 1$ corresponding to July and $i = 12$ corresponding to the following June;
- $J = \{j = 1, \ldots, 10\}$ is the set of consumer class indices;
- $\Psi = \{\psi = 1, \ldots, 17\}$ is the set of energetic indices formulas;
- $S = \{s = 1, \ldots, N\}$ is the set of scenario indices;
- c^s_{ij} is the consumption of consumer $j,\ j = 1, \ldots, 6$, in month $i \in I$ along scenario $s \in S$
$$c^s_{ij} = \bar{C}_{ij} + C_{ij}\Delta^s_i, \qquad j = 1, \ldots, 6,\ i \in I,\ s \in S\ , \tag{33}$$
where $\bar{C}_{ij}$ is the average consumption of consumer j in month $i \in I$; for $j = 7, \ldots, 10$ the consumption does not depend on temperature and therefore
$$c_{ij} = \bar{C}_{ij}, \qquad j = 7, \ldots, 10,\ i \in I\ ; \tag{34}$$
- va^s_j is the annual volume of gas for consumer $j,\ j = 1, \ldots, 6$, along scenario $s \in S$
$$va^s_j = \sum_{i=1}^{12} c^s_{ij}, \qquad j = 1, \ldots, 6,\ s \in S\ , \tag{35}$$
for $j = 7, \ldots, 10$ the annual volume of gas is
$$va_j = \sum_{i=1}^{12} \bar{C}_{ij}, \qquad j = 7, \ldots, 10\ ; \tag{36}$$

- vw_j^s is the winter volume of gas for consumer j, $j = 1, \ldots, 6$, along scenario $s \in S$

$$vw_j^s = \sum_{i=5}^{9} c_{ij}^s, \qquad j = 1, \ldots, 6, \; s \in S \;, \tag{37}$$

for $j = 7, \ldots, 10$ the winter volume of gas is

$$vw_j = \sum_{i=5}^{9} \bar{C}_{ij}, \qquad j = 7, \ldots, 10 \;; \tag{38}$$

- r_j^s is the ratio of winter gas consumption with respect to the total annual consumption of consumer j, $j = 1, \ldots, 6$, along scenario $s \in S$

$$r_j^s = \frac{vw_j^s}{va_j^s}, \qquad j = 1, \ldots, 6, \; s \in S \;, \tag{39}$$

for $j = 7, \ldots, 10$ the ratio of winter gas consumption with respect to the total annual consumption is

$$r_j = \frac{vw_j}{va_j}, \qquad j = 7, \ldots, 10 \;; \tag{40}$$

- cd_{ij}^s is the peak consumption per day of customer $j \in J$ in month $i \in I$ for $s \in S$

$$cd_{ij}^s = c_{ij}^s \, \frac{\gamma}{t_i}, \qquad j \in J, \; i \in I, \; s \in S \;, \tag{41}$$

where t_i is the number of days of the month $i \in I$ and γ is a parameter given by the Authority;

- nc_j are the **first stage decision variables** representing the number of consumers of class $j \in J$, restricted to be nonnegative integers, subject to upper bounds, $\overline{nc}_j$,

$$0 \leq nc_j \leq \overline{nc}_j, \quad j \in J \;; \tag{42}$$

- cm_i^s is the citygate consumption of month $i \in I$ along scenario $s \in S$

$$cm_i^s = \sum_{j=1}^{6} c_{ij}^s \cdot nc_j + \sum_{j=7}^{10} c_{ij} \cdot nc_j, \quad i \in I, \; s \in S \;; \tag{43}$$

- ca^s is the gas volume to be purchased for supplying the citygate consumers along scenario $s \in S$

$$ca^s = \sum_{i=1}^{12} cm_i^s, \; s \in S \;; \tag{44}$$

- x^s is the citygate loading factor along scenario $s \in S$ and g is the **first stage decision variable** representing the maximum consumption per day above which the gas seller has to pay a penalty

$$x^s = \frac{ca^s}{365 \cdot g}, \; s \in S \; ; \tag{45}$$

- l_j is the loading factor of consumer class j, $j = 7, \ldots, 10$;

- ${s_{ki}^{+}}^{s}$, $k = 0, 1, 2$ are **second stage decision variables** along scenario $s \in S$ that represent the surplus of consumption in the peak day of winter month i ($i = 5, \ldots, 9$) with respect to gas availability given by the decision variable g. These variables are used in computing the penalties by $\sum_{i=5}^{9} \sum_{k=1}^{2} \mu_{ki} {s_{ki}^{+}}^{s}$ where μ_{ki} is the unitary penalty in month i to be paid on the amount ${s_{ki}^{+}}^{s}$. The unitary penalty μ_{0i} is zero and the surplus variables ${s_{ki}^{+}}^{s}$ must satisfy the relations

$$0 \le {s_{0i}^{+}}^{s} \le \pi_{0i} \cdot g \; , \quad i = 5, \ldots, 9, \quad s \in S, \tag{46}$$

$$\pi_{0i} \cdot g \le {s_{1i}^{+}}^{s} \le \pi_{1i} \cdot g \; , \quad i = 5, \ldots, 9, \quad s \in S, \tag{47}$$

$$\pi_{2i} \cdot g \le {s_{2i}^{+}}^{s} \; , \quad i = 5, \ldots, 9, \quad s \in S, \tag{48}$$

  where π_{ki} represents the width of penalizations classes $k = 0, 1$ (no upper bound for class $k = 2$);

- cw^s is the citygate consumption in winter months along scenario $s \in S$

$$cw^s = \sum_{i=5}^{9} cm_i^s, \; s \in S \; ; \tag{49}$$

- h^s is the ratio of winter gas consumption with respect to total annual consumption along scenario $s \in S$

$$h^s = \frac{cw^s}{ca^s}, \; s \in S \; ; \tag{50}$$

- ϕ_ψ, ($\psi \in \Psi$) are first stage decision binary variables which identify the more appropriate indexation formula to be included (0 when a specific index formula ψ, $\psi \in \Psi$, is not chosen, 1 when ψ, $\psi \in \Psi$, is chosen). They have to satisfy the relation

$$\sum_{\psi=1}^{17} \phi_\psi = 1 \; ; \tag{51}$$

- $P_{i\,\psi_b}^{;s}$ is the purchase price to be paid by the gas seller to the shipper along the scenario $s \in S$, at month $i \in I$: it is expressed as a linear function of x^s and of indices formulas, and is defined as

$$P_{i\,\psi_b}^{s} = P^s + k_{\psi_b}(I_{i\,\psi_b}^{s} - I_{0\,\psi_b}) = QT + QS + q + m \cdot x^s + k_{\psi_b}(I_{i\,\psi_b}^{s} - I_{0\,\psi_b}) \; , \tag{52}$$

where q is the intercept and m is the slope; QT and QS are fixed by the Italian Regulatory Authority, k_{ψ_b} is a constant depending by the index buying formula $\psi \in \Psi$, $I^s_{i\,\psi_b}$ is a energetic index given with monthly, bimonthly or quarterly cadence according to the following general formula

$$I^s_{i\,\psi} = A_\psi(\frac{GasO^s_i}{a_\psi}) + B_\psi(\frac{1\%\,FO^s_i}{b_\psi}) + C_\psi(\frac{3.5\%\,FO^s_i}{c_\psi}) + D_\psi(\frac{Brent^s_i}{d_\psi}) + F_\psi(\frac{Crude\,O^s_i}{f_\psi}) \ ,$$

where A_ψ, B_ψ, C_ψ, D_ψ, F_ψ and a_ψ, b_ψ, c_ψ, d_ψ, f_ψ are constant depending on the choice of ψ ($\psi = \psi_b$ for buying and $\psi = \psi_s$ for selling) such that

$$A_\psi + B_\psi + C_\psi + D_\psi + F_\psi = 1, \qquad \psi \in \Psi \ ,$$

and finally $I_{0\,\psi}$ is the energetic index at the moment of the drawing up of the contract and it is fixed during all the length of the contract.

- $P'^{\,s}_{ij\,\{\psi=1\}}$ is the price to be paid by the first 6 classes of consumers along the scenario $s \in S$ and is defined as

$$P'^{\,s}_{ij\,\{\psi_s=1\}} = (CMP + k_{\{\psi_s=1\}}(I^s_{i\{\psi_s=1\}} - I_{0\,\{\psi_s=1\}}) + QVD) \cdot (1 - \alpha_j) \ , \qquad (53)$$

  where the value of CMP cover raw material costs (production, importation and transport), it is decided by the Italian Regulatory Authority and it is monthly updated by the term $k_{\{\psi_s=1\}}(I^s_{i\{\psi_s=1\}} - I_{0\,\{\psi_s=1\}})$ defined by the Authority with Decree 134/06; QVD too is fixed by the Authority and covers retail commercialization costs and finally α_j ($0 \leq \alpha_j < 1$) is a parameter representing possible discount fixed by the gas seller to be applied to consumer j;

- $P''^{\,s}_{i\,j\,\psi_s}$ is the price applied by the gas seller to consumer class j, $j = 7, \ldots, 10$, along the scenario $s \in S$

$$P''^{\,s}_{ij\,\psi_s} = \sum_{\psi_b=1}^{17} \phi_{\psi_b} P^s_{i^*,\psi_b} - \beta_j \cdot (1 - \frac{x^s}{l_j}) + \delta_j \cdot (r_j - h^s) + \lambda_j + k_{\psi_s}\left(I^s_{i \geq i^*,\,\psi_s} - I_{0\,\psi_s}\right) \ , \quad (54)$$

  where β_j and δ_j are constant values, λ_j is a possible recharge which can be applied to the industrial consumer class j and i^* is the month when the industrial clients draw up the contract with the gas retailer.

We choose as objective function the expected value of the gas seller profit:

$$w \ = \ E[R\,(nc_j, g) - C\,(nc_j, g) - Pt\left(s^{+\ s}_{ki}, g\right)] \ , \qquad (55)$$

where

$$R\left(nc_j,g\right)=\sum_{j=1}^{6}\sum_{i=1}^{12}P_{ij\{\psi_s=1\}}^{'\ s}\cdot c_{ij}^{s}\cdot nc_j+\sum_{j=7}^{10}\sum_{i=1}^{12}\sum_{\psi_s=1}^{17}\phi_{\psi_s}P_{ij\,\psi_s}^{''\ s}\cdot\bar{C}_{ij}\cdot nc_j\ , \tag{56}$$

represents the revenue,

$$C\left(nc_j,g\right)=\sum_{i=1}^{12}\sum_{\psi_b=1}^{17}\phi_{\psi_b}P_{i\psi_b}^{s}\cdot cm_i^{s}\ , \tag{57}$$

represents the costs and

$$Pt\left(s_{ki}^{+\ s},g\right)=\sum_{i=5}^{9}\sum_{k=0}^{2}\mu_{ki}s_{ki}^{+\ s}\ . \tag{58}$$

represents the penalties of gas retailer.

Notice that

- the expected value of revenues from the first six classes of consumers is

$$E\left[\sum_{j=1}^{6}\sum_{i=1}^{12}(P_{ij\{\psi_s=1\}}^{'\ s}\cdot c_{ij}^{s}\cdot nc_j)\right]=$$

$$=\sum_{j=1}^{6}\left(1-\alpha_j\right)nc_j\sum_{i=1}^{12}\bar{C}_{ij}\left(CMP+QVD+k_{\{\psi_s=1\}}\left(E[I_{i\{\psi_s=1\}}^{s}]-I_{0\{\psi_s=1\}}\right)\right)\ , \tag{59}$$

  where we have taken into account that

$$E\left[\Delta_i^s\right]=0\ , \tag{60}$$

  and

$$E\left[\Delta_i^s I_{i\{\psi_s=1\}}^{s}\right]=E\left[\Delta_i^s\right]E\left[I_{i\{\psi_s=1\}}^{s}\right]=0\ ,$$

  because of the independence of Δ_i^s on $I_{i\{\psi_s=1\}}^{s}$ and of eq. (60) .

- the expected value of revenues from the last four consumer classes is

$$E\left[\sum_{j=7}^{10}\sum_{i=1}^{12}\sum_{\psi_s=1}^{17}\phi_{\psi_s}P_{ij\,\psi_s}^{''\ s}\cdot\bar{C}_{ij}\cdot nc_j\right]=$$

$$\sum_{j=7}^{10}nc_j\sum_{i=1}^{12}\bar{C}_{ij}\left\{E[P_j^{''\ s}]+\sum_{\psi_b=1}^{17}\phi_{\psi_b}k_{\psi_b}\left(E[I_{i*\psi_b}^{s}]-I_{0\psi_b}\right)\right\}+$$

$$+\sum_{j=7}^{10} nc_j \sum_{i=i^*}^{12} \bar{C}_{ij} \sum_{\psi_s=1}^{17} \phi_{\psi_s} k_{\psi_s} \left(E[I^s_{i\psi_s}] - I_{0\psi_s} \right) \ , \quad (61)$$

where

$$P_j^{''\ s} = P^s - \beta_j \cdot (1 - \frac{x^s}{l_j}) + \delta_j \cdot (r_j - h^s) + \lambda_j \ , \quad (62)$$

and where we have considered that the industrial consumptions are independent of temperature.
Notice that

$$\begin{aligned} E\left(P_j^{''\ s}\right) &= E\left(P^s\right) - \beta_j \cdot \left(1 - E\left(\frac{x^s}{l_j}\right)\right) + \delta_j \cdot \left(r_j - E\left(h^s\right)\right) + \lambda_j = \quad (63) \\ &= QT + QS + q - \beta_j + \delta_j r_j + \lambda_j + \left(m + \frac{\beta_j}{l_j}\right) \sum_{s=1}^{N} x^s p^s - \delta_j \sum_{s=1}^{N} h^s p^s \ . \end{aligned}$$

- the expected value of the costs is

$$E\left[\sum_{i=1}^{12} \sum_{\psi_b=1}^{17} \phi_{\psi_b} P^s_{i\psi_b} \cdot cm^s_i\right] = (QT + QS + q) \sum_{i=1}^{12} E\left[cm^s_i\right] +$$

$$+\frac{m}{365 \cdot g} \left(E\left[\sum_{i=1}^{12} \left(cm^s_i\right)^2\right] + 2E\left[\sum_{\substack{i,k=1 \\ k>i}}^{12} \left(cm^s_i\right)\left(cm^s_k\right)\right]\right) +$$

$$+\sum_{\psi_b=1}^{17} \phi_{\psi_b} \sum_{i=1}^{12} k_{\psi_b} \left(E[I^s_{i\psi_b}] - I_{0\psi_b} \right) E[cm^s_i]; \quad (64)$$

where

$$E\left[\left(cm^s_i\right)^2\right] = \sum_{s=1}^{N} \left(cm^s_i\right)^2 \ p^s \ ,$$

and

$$E\left[\sum_{\substack{i,k=1 \\ k>i}}^{12} \left(cm^s_i\right)\left(cm^s_k\right)\right] = \sum_{s=1}^{N} \left(\sum_{\substack{i,k=1 \\ k>i}}^{12} \left(cm^s_i\right)\left(cm^s_k\right)\right) p^s \ ;$$

- the expected value of the penalties is

$$E\left[\sum_{i=5}^{9} \sum_{k=1}^{2} \mu_{ki} s_{ki}^{+\ s}\right] = \sum_{s=1}^{N} \left(\sum_{i=5}^{9} \sum_{k=0}^{2} \mu_{ki} s_{ki}^{+\ s}\right) p^s \ . \quad (65)$$

The constraints of our stochastic problem are the following:

$$0 \leq nc_j \leq \overline{nc}_j \ , \quad j \in J \ , \tag{66}$$

$$\sum_{j=1}^{6} cd_{ij}^{s} \cdot nc_j + \sum_{j=7}^{10} cd_{ij} \cdot nc_j - g \leq \sum_{k=0}^{2} s_{ki}^{+^s} \ , \quad i = 5, \ldots, 9, \ s \in S \ , \tag{67}$$

$$0 \leq s_{0i}^{+^s} \leq \pi_{0i} \cdot g \ , \quad i = 5, \ldots, 9, \ s \in S \ , \tag{68}$$

$$\pi_{0i} \cdot g \leq s_{1i}^{+^s} \leq \pi_{1i} \cdot g \ , \quad i = 5, \ldots, 9, \ s \in S \ , \tag{69}$$

$$\pi_{2i} \cdot g \leq s_{2i}^{+^s} \ , \quad i = 5, \ldots, 9, \ s \in S \ , \tag{70}$$

$$\sum_{\psi_s=1}^{17} \phi_{\psi_s} = 1 \ , \qquad \sum_{\psi_b=1}^{17} \phi_{\psi_b} = 1 \ . \tag{71}$$

We note that in our model, as shown in the previous simplifying formulas (59), (61) and (64), we take only the temperature as unique source of stochasticity because the indices formulas on oil prices appear only as expected value in the objective function and not in constraints; consequently we can drop oil prices as random variable and consider these as parameters which are monthly updated.
Notice that the problem may also be formulated as a *2-stage stochastic model with recourse* as follow:

$$\max \ E_{\boldsymbol{\xi}} \left[f\left(x, y(\ \boldsymbol{\Delta})\right) \right] \ , \tag{72}$$

$$Ax = b \ , \tag{73}$$

$$T(\ \boldsymbol{\Delta})x + Wy(\ \boldsymbol{\Delta}) = h(\ \boldsymbol{\Delta}) \ , \tag{74}$$

$$x \geq 0, \ y(\ \boldsymbol{\Delta}) \geq 0 \ , \tag{75}$$

where $\boldsymbol{\xi} = \left(h\left(\ \boldsymbol{\Delta}\right), T\left(\ \boldsymbol{\Delta}\right)\right)$ is a random vector influenced by random temperature data. In our problem the first stage decision variables x involves:

- the number of customers nc_j of class $j \in J$;
- the daily capacity g above which the gas seller has to pay a penalty;
- binary variables ϕ_ψ, $\psi \in \Psi$;

whereas the second stage decision variable $y(\ \boldsymbol{\Delta})$ involves the surplus in consumption in the peak day $s_{ki}^{+^s}$ in winter month i. Furthermore the first stage constraint (73) is represented by equations (66) and (71) and the second stage constraint (74) by equations (67), (68), (69) and (70).

5 Results and model validations

In this section, we show the results of our stochastic model for a local gas seller who has to decide the customer portfolio structure in a village in Northern Italy (Sotto il Monte). The simulation is based on the data of thermal year 2004-2005 (for these data see Allevi *et al.*, (2005) [3]). We have developed a simulation framework based on ACCESS 97, for database management, on MATLAB release 12, for data visualization, and on GAMS release 21.5, for optimization. In the GAMS framework the DICOPT solver has been used for the nonlinear mixed integer optimization problem. DICOPT solves a series of NLP subproblems by CONOPT2 and MIP subproblems by CPLEX. The relation between the part of the purchase price P^s, which does not depend by oil prices index formulas, and x^s is estimated by the gas seller through a linear regression using the data related to year 2004-2005 for all citygates managed by the gas seller. The regression of P^s values has also been tried on the annual volume ca^s, h^s and g but it has been found not significant. Indeed, the value of R^2-test (see e.g. Davidson, (2000) [10]) with the regression on x^s is 0.603, therefore not highly significant. However, the introduction of non parametric regression, would introduce a more complicated function in the model. On the other side, linear regression is currently used by the gas seller in their simulations. In our case we use:

$$P^s(x^s) = QT + QS + 18.348 - 3.866 \cdot x^s \ , \tag{76}$$

where the intercept value 18.348 and the slope value -3.866; the values QT and QS are given by the Italian Regulatory Authority: in our numerical experiments $QT =$ 2.4953171 Eurocent/Stm^3 and $QS = 0.63882$ Eurocent/Stm^3.

The relation between the consumption of consumer j, $j = 1 \ldots, 6$, in month $i \in I$ along scenario $s \in S$, c^s_{ij} and the deviation from mean value over scenarios, Δ^s_i, is supposed to be linear with intercept equal to $\bar{C}_{ij}$ and the other coefficient computed via a linear regression. The regression results to be significative for all the consumers.
In particular, we choose to work under the assumption that all the contracts on sale and on purchase are stipulated in the same month (i.e. $i^* = 1$), obviously we can modify the objective function to take into account of possible diversified effective dates.
The model has been validated by running several tests both in the deterministic (see Allevi *et al.*, (2007) [2]) and in the stochastic case. For the stochastic model, we have solved 10000 times the problem, each time with $N = 80$ scenarios randomly chosen with the procedure described in Section 2.4. The optimal values both in the function and in the decision variables are stable. We report in Table 6 their average over 10000 trials. The first column refers to the solution obtain by a first stochastic model described in a previous paper (see Maggioni *et al.* (2006) [19]) in which the source of stochasticity was again given by the dependence of consumption by the temperature but the purchase and sell prices did not change during the year, whereas the second column refers to the model described here in which the dependence of gas prices by energetic indices is considered; in Table 7 are also reported the optimal values of purchase prices along the months in a year obtained as solutions of the second stochastic

	First stochastic	Second stochastic	
$Profit$	152208	197264	Euro
P	19.67	23.85	Eurocent/Stm^3
P'	25.00	29.37	Eurocent/Stm^3
P''	19.88	26.52	Eurocent/Stm^3
ca	4484407	4484406	Stm^3
g	26399	26251	Stm^3
x	0.4654	0.4678	
ψ_b		$\psi_b = 11$: 1% Fuel Oil 12,1,1	
ψ_s		$\psi_s = 13$: C 6,1,1	

Table 6: Optimal values for citygate Sotto il Monte respectively in the first (see [19]) and second stochastic case.

model.
By looking Table 6 we can note that, even if the two models are different and consequently the correspondent results not comparable, the optimal profit value of the second model, is much higher than the one of the first model; this is due to the fact that gas retailer can choose to buy according to the lowest oil prices index formula ψ_b and to sell to industrial customers at highest one ψ_s.

In Figure 7 are reported the forward mean values along 1000 simulations of the energetic indices $I_\psi - I_{0\psi}$, with $\psi \in \Psi$, in the period between July 2005 and June 2006. As expected, the optimal purchase index formula is given by $\psi_b = 11$ which corresponds to the lowest index formula 1% Fuel Oil 12,1,1, and the optimal sell formula is $\psi_s = 13$, that is the highest index formula, C 6,1,1.

Another test has also been considered with the aim to quantify the losses in the gas seller budget in the case of non optimal purchase and sell index formulas. In this case we have fixed $\psi_b = \psi_s =$ 1% 3.5% Fuel Oil 9,1,1 (see Figure 7) and we have reported in the first column of Table 8 the relative optimal profit value compared to the case of optimal index formulas $\psi_b = 11$ and $\psi_s = 13$.

We have also solved the problem (55) with the related constrained (66) to (71) without using the structure of two-stage stochastic optimization according to the following special sampling strategy:

- we assume that temperatures along the different months are independent;
- for each month $i = 1, \ldots, 12$ we compute

$$M_i = \max_{ns \in [1,\ldots,NS]} Tm_i^{ns} \qquad \text{and} \qquad m_i = \min_{ns \in [1,\ldots,NS]} Tm_i^{ns};$$

  where NS is the number of simulations;
- N is the number of intervals in which we divide the difference $M_i - m_i$ for $i = 1, \ldots, 12$;

Month	$P_{i,\{\psi_b=11\}}$ Eurocent/Stm^3
January	20.80
February	21.39
March	22.01
April	22.63
May	23.10
June	23.69
July	24.36
August	24.92
September	25.43
October	25.77
November	25.96
December	26.21

Table 7: Optimal values of purchase prices in the different months in a year as solutions of second stochastic model.

	non-optimal indices formula	Second stochastic	
$Profit$	135561	197264	Euro
P	24.96	23.86	Eurocent/Stm^3
P'	29.38	29.38	Eurocent/Stm^3
P''	25.17	26.53	Eurocent/Stm^3
ca	4484656	4484406	Stm^3
g	26264	26251	Stm^3
x	0.4678	0.4678	
ψ_b	1% 3.5% Fuel Oil 9,1,1	1% Fuel Oil 12,1,1	
ψ_s	1% 3.5% Fuel Oil 9,1,1	C 6,1,1	

Table 8: Optimal values for citygate Sotto il Monte respectively in the case of choose of non-optimal indices formula $\psi_b = \psi_s =$ 1% 3.5% Fuel Oil 9,1,1 C 6,1,1 and second stochastic case.

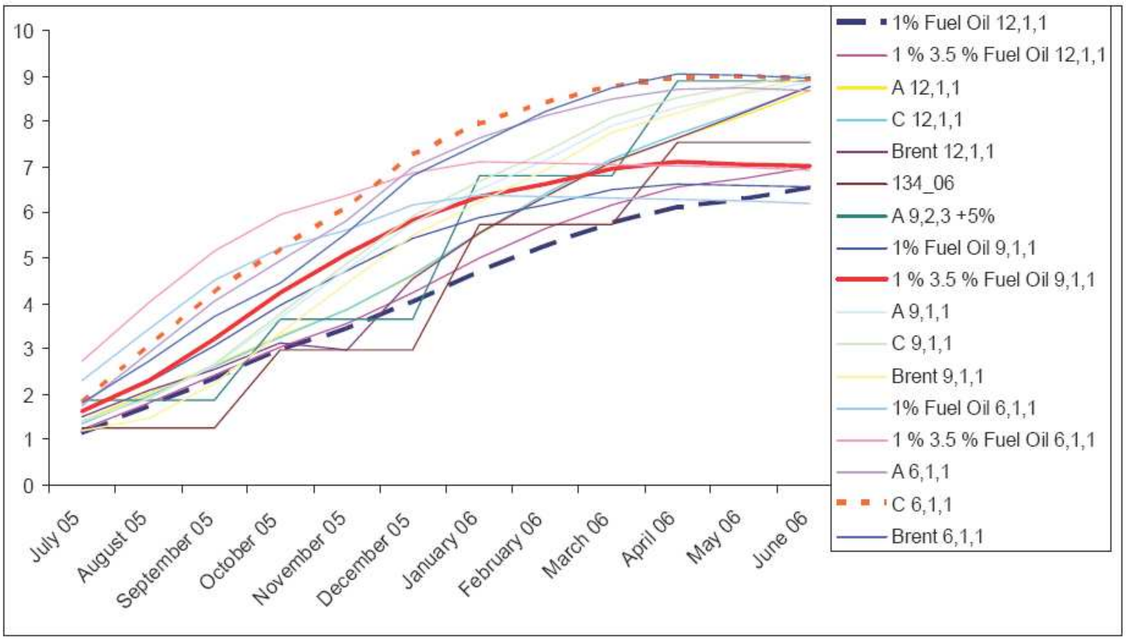


Figure 7: Forward mean values along 1000 simulations of the energetic indices $I_\psi - I_{0\psi}$ in the period between July 2005 and June 2006. Ax expected the optimal solutions for purchase index formula is given by $\psi_b = 11$ which corresponds to the lowest index formula 1% Fuel Oil 12,1,1 and for sell by $\psi_s = 13$ which corresponds to the highest index formula C 6,1,1.

- $int\,(i) := \dfrac{M_i - m_i}{N}, \quad i = 1, \ldots, 12$, is the width of the interval i;
- $\mathbf{pt}_i \in \mathbb{R}^N$ are the vectors of the middle points of the intervals of temperature in each month i; the component

  $$pt_{i,k} = m_i + \left(k - \frac{1}{2}\right) int\,(i)\,, \qquad k = 1, \ldots, N, \quad i = 1, \ldots, 12\;,$$

  is the middle point of each interval $k = 1, \ldots, N$ for each month $i = 1, \ldots, 12$;
- $\mathbf{p}_i \in \mathbb{R}^N$ are the vectors of probabilities along the intervals in month i; the component p_i^r is the probability associated to the middle point of the interval $r = 1, \ldots, N$ related to month i; it is given by the frequency of the corresponding interval.
- The sample of temperatures in each months is ordered from the coldest to the warmest temperature in that month.

For this sampling technique, we analyze the sensitivity of solution as the number of intervals in each month increases. Therefore, we tested the case of $NS = 1000$ temperature simulations as shown in Figures 8 and 9. In Figure 8 we can see the behavior of

		Special sampling technique
$Profit$	196728	Euro
P	23.85	Eurocent/Stm^3
P'	29.37	Eurocent/Stm^3
P''	26.52	Eurocent/Stm^3
ca	4484656	Stm^3
g	26264	Stm^3
x	0.4678	
ψ_b	$\psi_b = 11$: 1% Fuel Oil 12,1,1	
ψ_s	$\psi_s = 13$: C 6,1,1	

Table 9: Optimal values for citygate Sotto il Monte respectively in case of special sampling technique.

the optimal profit value for an increasing number of temperature intervals: in this situation we observe that the optimal profit value is decreasing until a convergence around a value between 196470 and 196435; the decreasing behavior is due to the fact that, for an increasing number of temperature intervals, colder scenarios are taking into account with the consequent reduction of profit due to penalties. Figure 9 represents the convergence in the range between 26246 and 26318 of maximum daily consumption g as the number of temperature intervals increases. Furthermore the optimal values of the variables nc_j ($j \in J$) and ϕ_ψ ($\psi \in \Psi$) are the same for all the number of temperature intervals considered.

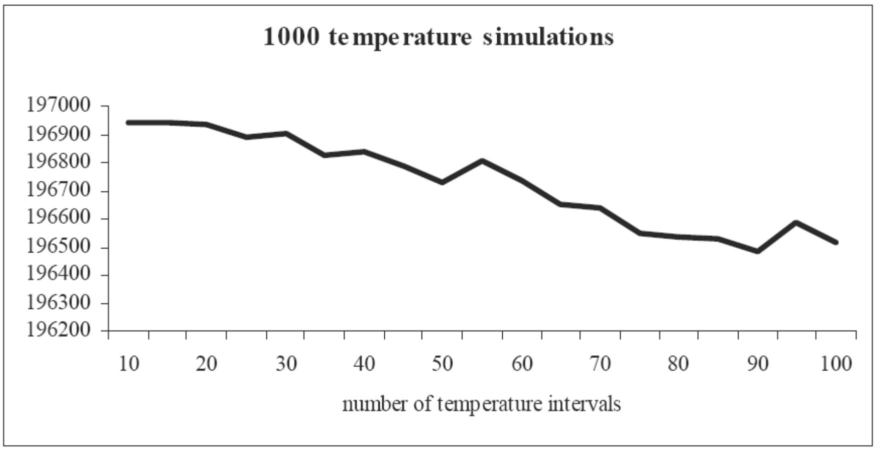


Figure 8: Case of $NS = 1000$ temperature simulations: optimal profit value as the number of temperature intervals increases.

Finally, in order to relate clients number to the best selling price on the market $\bar{P}'$ if small customers, or $\bar{P}''$ if industrial ones, we have also thought to introduce in the

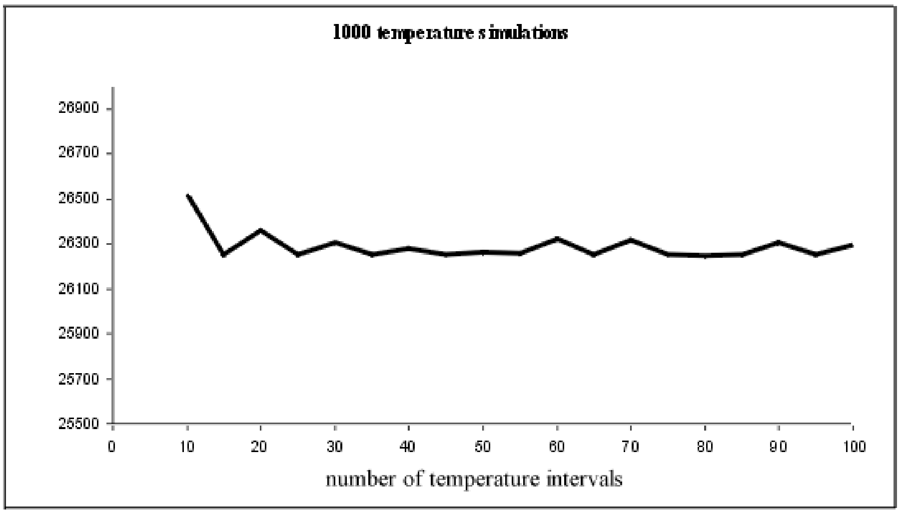


Figure 9: Case of $NS = 1000$ temperature simulations: solution of the first stage decision variable maximum consumption per day g, as the number of temperature intervals increases.

model the following non linear constraints

$$(\bar{P}'_j - P'^{\,s}_{ij\{\psi_s=1\}})\vartheta_j + (\bar{P}'_j - P'^{\,s}_{ij\{\psi_s=1\}} - m)(1-\vartheta_j) \geq 0 \quad j=1,\ldots,6,\ i\in I,\ \forall s \quad (77)$$

$$(\bar{P}''_j - P''^{\,s}_{ij\,\psi_s})\vartheta_j + (\bar{P}''_j - P''^{\,s}_{ij\,\psi_s} - m)(1-\vartheta_j) \geq 0 \quad j=7,\ldots,10,\ i\in I,\ \forall s \quad (78)$$

$$nc_j \leq \vartheta_j \overline{nc}_j \quad j \in J\ , \quad (79)$$

where ϑ_j $(j \in J)$ is a binary variable, $\overline{nc}_j$ is the maximal number of customer class $j \in J$ and m a negative number such that $|m| \gg (\bar{P}_j - P_{ij})$, $i \in I$, $j \in J$. By the way we have not taken into account in the model these constraints because they add a further computational complexity which is not solvable by standard algorithm.

6 Conclusions

We have proposed a stochastic model for the management of a gas retail company where we have considered temperature as source of stochasticity, but we have taken into account also information on oil prices to which gas purchase and sell prices are related and monthly updated. In order to allow the gas seller to evaluate the possibility of a client to migrate to other providers, a possible extension of the model consists in including linear constraints among prices and number of clients. Finally, there are various possible extensions of the model, for instance including different type of contracts with the shipper (portfolio management of gas contracts) or considering storage facilities for a gas shipper-seller.

Acknowledgements

The authors would like to thank Dr. Michal Kaut and Prof. Jitka Dupačová for helpful suggestions on a preliminary draft of this paper.

Appendix

Denoting with N_i the number of days of a specific month i, in order to prove that the process found in (6) is mean reverting we set

$$Y_t = \left[e^{\left(\sum_{k=1}^{i-1} \int_0^{N_k} a_k \mathrm{d}s\right) + \int_0^{t-\sum_{k=1}^{i-1} N_k} a_i \mathrm{d}s} \right] (\vartheta_t - T_t) \tag{80}$$

$$= \left[e^{\left(\sum_{k=1}^{i-1} N_k a_k\right) + a_i \left(t - \sum_{k=1}^{i-1}\right) N_k} \right] (\vartheta_t - T_t) \ ,$$

then Itô's formula implies

$$\mathrm{d}Y_t = \left[e^{\left(\sum_{k=1}^{i-1} N_k a_k\right) + a_i \left(t - \sum_{k=1}^{i-1}\right) N_k} \right] [\left(\frac{\mathrm{d}\vartheta_t}{\mathrm{d}t} + a_i (\vartheta_t - T_t) \right) \mathrm{d}t + \tag{81}$$

$$- \left(a_i (\vartheta_t - T_t) + \frac{\mathrm{d}\vartheta_t}{\mathrm{d}t} \right) \mathrm{d}t - \sigma_t \mathrm{d}W_t] \ ,$$

hence

$$Y_t - Y_0 = - \int_0^t \sigma_s \left[e^{\left(\sum_{k=1}^{i-1} N_k a_k\right) + a_i \left(s - \sum_{k=1}^{i-1}\right) N_k} \right] \mathrm{d}W_s \ , \tag{82}$$

that is

$$\left[e^{\left(\sum_{k=1}^{i-1} N_k a_k\right) + a_i \left(t - \sum_{k=1}^{i-1}\right) N_k} \right] (\vartheta_t - T_t) = \vartheta_0 - T_0 - \int_0^t \left[e^{\left(\sum_{k=1}^{i-1} N_k a_k\right) + a_i \left(s - \sum_{k=1}^{i-1}\right) N_k} \right] \sigma_s \mathrm{d}W_s \ , \tag{83}$$

but $\vartheta_0 = T_0 = C$ and thus

$$T_t = \vartheta_t + e^{-\left[\left(\sum_{k=1}^{i-1} N_k a_k\right) + a_i \left(t - \sum_{k=1}^{i-1}\right) N_k\right]} \int_0^t \left[e^{\left(\sum_{k=1}^{i-1} N_k a_k\right) + a_i \left(s - \sum_{k=1}^{i-1}\right) N_k} \right] \sigma_s \mathrm{d}W_s \ , \tag{84}$$

from which we can see that the process reverts to its mean ϑ_t because the expected value of an Itô Integral is zero.

References

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

[2] Allevi, E., Bertocchi, M.I., Innorta, M., Vespucci, M.T. 2007 *A mixed integer nonlinear optimization model for gas sale company*, Optimization Letters, **1**(1), 61-69, Springer, Berlin.

[3] Allevi, E., Bertocchi, M.I., Innorta, M., Vespucci, M.T. 2005 *A stochastic optimization model for gas sales companies*, , Department of Mathematics, Statistic, Computer science and Applications, Working Paper 11, University of Bergamo.

[4] Allevi, E., Bertocchi, M.I., Innorta, M., Vespucci, M.T. 2007 *A stochastic optimization model for a gas sale company*, IMA Journal of Management Mathematics, 1-14, DOI:10.1093/imaman/dpm004.

[5] Basawa, I.V. and Prakasa Rao B.L.S. 1980 *Statistical Inference for Stochastic Processes*, Academic Press.

[6] Bibby, B. M. and Sorenses, M. 1995 *Martingale Estimation Functions for Discretely Observed Diffusion Processes*, Bernoulli, **I** numbers I/II.

[7] Bollerslev, T. 1986 *Generalized Autoregressive Conditional Heteroskedasticity*, Journal of Econometrics, **31**, 307–327.

[8] Brockwell, P.J. and Davis, R.A. 1990 *Time Series: Theory and Methods*, Springer, Second edition.

[9] Brooks, R. E. 1981 *Using generalized networks to forecast natural gas distribution and allocation during periods of shortage*, Mathematical Programming Study, **15**, 23–42.

[10] Davidson J. 2000 Econometric Theory, Blackwell, ISBN 0-631-21584-0.

[11] *Deliberazione n.138, 4/12/2003 Criteri per la determinazione delle condizioni economiche di fornitura del gas naturale ai clienti finali e disposizioni in materia di tariffe per l'attività di distribuzione.*

[12] Dickey, D.A. & W.A. Fuller 1979 *Distribution of the Estimators for Autoregressive Time Series with a Unit Root*, Journal of the American Statistical Association, **74**, 427-431.

[13] Dornier, F. and Queruel, M. 2000 *Pricing weather derivatives by marginal value*, Quantitative Finance, V.1, Institute of Physics Publishing.

[14] Engle, R. F. and Granger, C. W. J. 1987 *Co-integration and error-correction: Representation, estimation and testing.* Econometrica **55**, 251–276.

[15] Eydeland, A., Wolyniec, K. 2003 *Energy and power risk management*, Wiley.

[16] Ermoliev, Y., Wets, J.-B. 1988 *Numerical techniques for stochastic optimization*, Springer Verlag.

[17] *Oilgram Price Report*, www.platts.com.

[18] S. Johansen, 1988 *Statistical Analysis of Cointegration Vectors*, Journal of Economic Dynamics and Control **12**, 231-254. Reprinted in R.F. Engle and C.W.J. Granger (eds.), Long-run Economic Relationships, Readings in Cointegration, Oxford University Press (1991).

[19] Maggioni, F., Allevi, E., Bertocchi, M.I., Innorta, M., Vespucci, M.T. 2006 *A gas retail stochastic optimization model by mean reverting temperature scenarios*, to appear in Communications to SIMAI Congress, ISSN 1827-9015 .

[20] Ruszczynski, A., Shapiro, A. 2003 *Stochastic programming*, Elsevier.

Redazione

Dipartimento di Matematica, Statistica, Informatica ed Applicazioni
Università degli Studi di Bergamo
Via dei Caniana, 2
24127 Bergamo
Tel. 0039-035-2052536
Fax 0039-035-2052549

La Redazione ottempera agli obblighi previsti dall'art. 1 del D.L.L. 31.8.1945, n. 660 e successive modifiche

Stampato nel 2007
presso la Cooperativa
Studium Bergomense a r.l.
di Bergamo