

A TWO-STAGE STOCHASTIC OPTIMIZATION MODEL FOR A GAS SALE RETAILER

F. MAGGIONI, M. T. VESPUCCI, E. ALLEVI, M. I. BERTOCCHI
AND M. INNORTA

The paper deals with a new stochastic optimization model, named OMoGaS-SV (Optimization Modelling for Gas Seller-Stochastic Version), to assist companies dealing with gas retail commercialization. Stochasticity is due to the dependence of consumptions on temperature uncertainty. 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 and to consumption profiles are included. The results obtained by the stochastic version give clear indication of the amount of losses that may appear in the gas seller's budget and are compared with the results obtained by the deterministic version (see Allevi et al. [2]).

Keywords: gas sale company, mean reverting process, stochastic programming

AMS Subject Classification: 90C15, 90B50, 46N10

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 between supply and transport/storage and between distribution and selling. Before liberalization there was a national monopolistic operator, for all activities related to supply, transport, storage and wholesale 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 to final consumers.

In 2003 the Italian Regulatory Authority for Electricity and Gas, see [9], defined consumption classes, on the basis of gas consumption in the thermal year, and introduced a new gas tariff, 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.

In a previous paper, see Allevi et al. [2], a deterministic optimization model has 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 for next thermal year;
- the gas volume required in winter months;
- the maximum daily consumption (capacity) requested by the gas seller;
- the purchase price fixed by the shipper.

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

In this paper we introduce stochasticity in the model due to the influence of temperature on consumptions. For domestic customers, using gas either only for cooking or for cooking and heating, and for commercial activities and small industries, gas consumption in winter months strongly depends on weather conditions. A first simple model to take into account this fact has been introduced in Allevi et al. [3]; here we propose a mean reverting process modelling temperature. This model is presented in Section 2. In Section 3 the two-stage stochastic optimization model, named OMoGaS-SV, is described and in Section 4 numerical results related to a case study are reported and discussed.

2. THE STOCHASTIC TEMPERATURE MODEL

In this section we introduce a stochastic model describing the temperature variations along the months in a year time. We strictly refer to the model introduced by Alaton et al. [1]. We start with some definitions about temperature:

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

$$T_t = \frac{T_t^{\max} + T_t^{\min}}{2}. \tag{1}$$

Definition 2. Let T_t denote the mean temperature of day t . We define *Heating Days Degree* (HDD_t : measure of cold in winter) and *Cooling Days Degree* (CDD_t : measure of heat in summer) respectively as

$$HDD_t = \max\{18 - T_t, 0\}, \tag{2}$$

$$CDD_t = \max\{T_t - 18, 0\}. \tag{3}$$

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 Days Degree” refers to the fact that if temperature is below 18°C people tend to use more energy to heat their homes; the name “Cooling Days Degree” 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. We have plotted the daily mean temperatures at Bergamo for the 12 years (see Figure 1); it is evident that the temperature process is mean stationary and variance stationary so it should be a mean reverting process, reverting to some cyclical function. However, because temperature process is evidently not deterministic, we must consider the presence of noise. Looking at the histogram of the daily temperature differences in Bergamo (1994–2005), it shows a good fit with the corresponding normal distribution, though the frequency of small differences in daily mean temperature is underestimated. This is the reason which brings us to choose the Brownian Motion as model of temperature process.

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

$$dT_t = a(\vartheta - T_t) dt + \sigma dW_t, \tag{4}$$

where T_t is the process to be modelled, $a \in \mathbb{R}$ is the speed of mean reversion, ϑ the mean to which the process reverts to (constant), σ the volatility of the process (constant) and W_t is the Wiener process.

For the temperature process we need a $\vartheta = \vartheta(t) = \vartheta_t$, $a = a(i) = a_i$ and $\sigma = \sigma(i) = \sigma_i$ as functions changing over the months but constant in each month i .

Then our process becomes

$$dT_t = a_i (\vartheta_t - T_t) dt + \sigma_i dW_t . \tag{5}$$

Thus, we need to determine a functional form for ϑ_t and estimates for a_i and σ_i from historical data. In Dornier and Queruel [10], it is shown 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 ϑ_t' to the drift term in (5) so that the equation becomes

$$dT_t = \left[a_i (\vartheta_t - T_t) + \frac{d\vartheta_t}{dt} \right] dt + \sigma_i dW_t . \tag{6}$$

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. Following Alaton et al. [1] we first assume a deterministic model ϑ_t for the mean temperature at time t , given by the function

$$\vartheta_t = A + Bt + C \sin(\omega t + \varphi) , \tag{7}$$

where t is the time measured in days, 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, $\omega = 2\pi/365$ is the frequency of oscillation and the linear part corresponds to a linear warming trend. Equation (7) can be rewritten as

$$\vartheta_t = A + Bt + C [\cos(\varphi) \sin(\omega t) + \sin(\varphi) \cos(\omega t)] . \tag{8}$$

In order to estimate the parameters in (8), a change of variables is operated and the constants are renamed as follows

$$\begin{cases} A = a_1 \\ B = a_2 \\ C \cos(\varphi) = a_3 \\ C \sin(\varphi) = a_4 \end{cases} \tag{9}$$

or equivalently

$$\begin{cases} A = a_1 \\ B = a_2 \\ C = \sqrt{a_3^2 + a_4^2} \\ \varphi = \arctan\left(\frac{a_4}{a_3}\right) - \pi \end{cases} \tag{10}$$

and we obtain

$$\vartheta_t = a_1 + a_2 t + a_3 \sin(\omega t) + a_4 \cos(\omega t) . \tag{11}$$

The numerical values of the parameters in (11) are computed by the least squares method, i. e. is computed the parameter vector $\xi = (a_1, a_2, a_3, a_4)$ that solves

$$\min_{\xi} \|\vartheta - \mathbf{X}\|^2 , \tag{12}$$

where ϑ is the vector whose elements are given by (11) and \mathbf{X} is the data vector. By using the series of 4323 observations of the historical daily temperatures we get

$$\begin{cases} A = 13.33 \\ B = 6.8891 \cdot 10^{-5} \\ C = 10.366 \\ \varphi = -1.7302. \end{cases} \tag{13}$$

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.

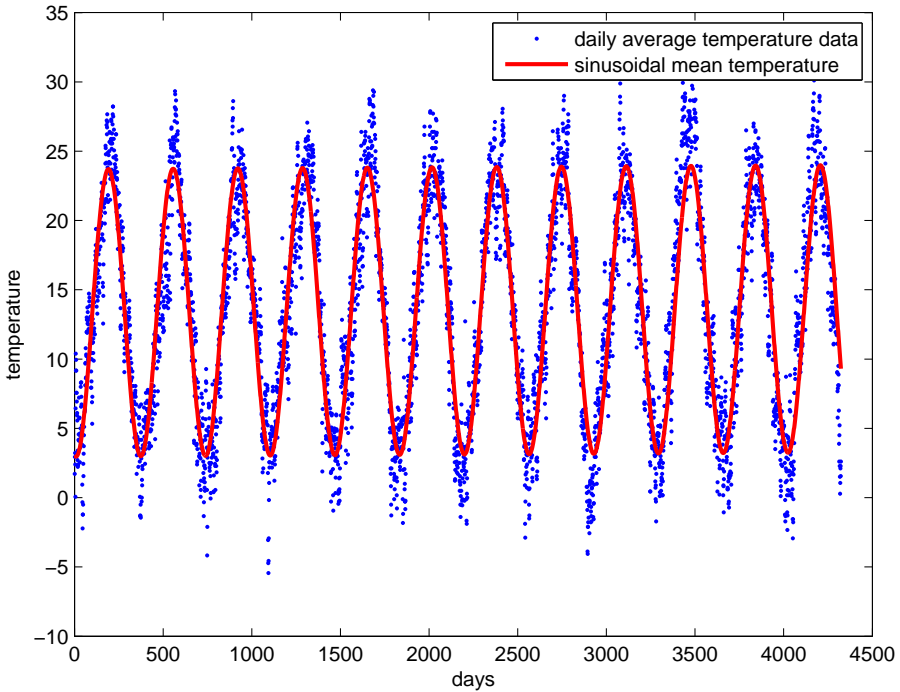


Fig. 1. Comparison between daily average temperature and estimated mean temperature ϑ_t according to formula (7) at Bergamo in the years 1994–2005.

There may be some criticism towards the sinusoidal model (7) as the real annual cycle of temperature shows some signs of asymmetry. An alternative deterministic model ϑ_t can be considered: according to (7) and (13), the temperature increase is about 0.025°C per year and can be thus ignored (only one year scenarios are generated). Hence the deterministic model ϑ_t (t are days) can be considered as a 365-periodic function. The following simple model can be chosen for ϑ_t : subscript $i = 0, 1, \dots, 12, 13$ refers to past December, present January, February, \dots , Decem-

ber, next January; d_i is the middle point of month i , $\bar{T}_i, i = 1, \dots, 12$, is the average of month i temperature obtained by averaging the real data over years 1994–2005; define $\bar{T}_0 = \bar{T}_{12} = \bar{T}_{13} = \bar{T}_1$; connect points $[d_i, \bar{T}_i], i = 0, \dots, 13$ by cubic interpolation splines, getting thus (one period of) ϑ_t . This ϑ_t may be handled in the same way as ϑ_t given by (7), i. e., may be plugged into (15) (model ϑ_t may be adjusted so that \bar{T}_i and monthly mean values based on the model coincide).

Looking at Figure 2 we observe that the cubic interpolation spline and the sinusoidal function based on averaged month $i, i = 1, \dots, 12$, temperatures show quite similar graphs, which, in fact, represent the average behaviour of the empirical data. The Figure gives a clear evidence that the rather different (monthly or daily) input data make a sensible difference in the output. Nevertheless, being the improvement using splines limited, we prefer to use the sinusoidal model (7) based on daily data suggested by Alaton et al. [1].

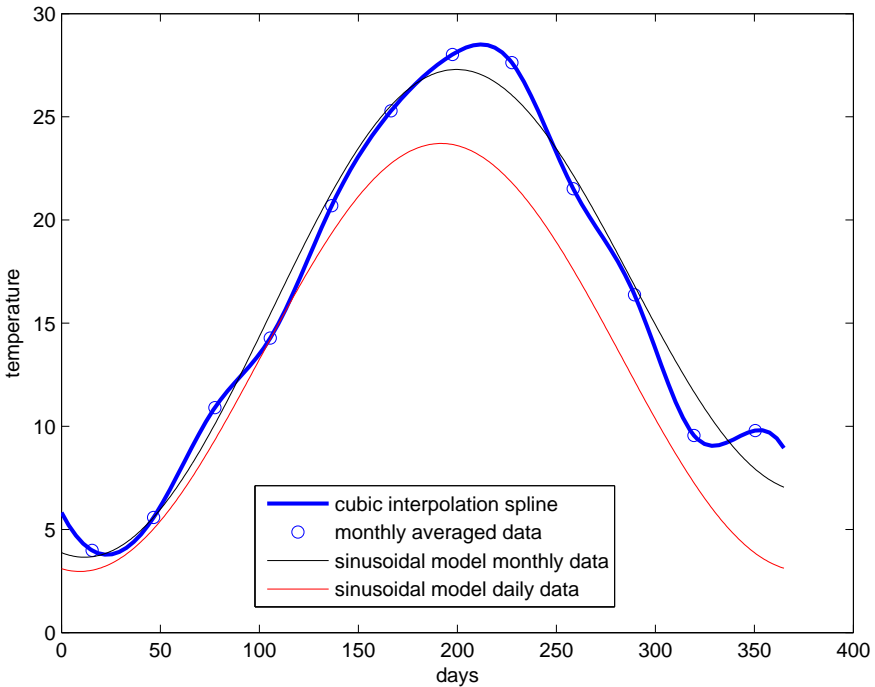


Fig. 2. Comparison over one year among cubic interpolation spline, sinusoidal function, both based on averaged month $i, i = 1, \dots, 12$; temperatures obtained by averaging the real data over years 1994–2005, and sinusoidal function based on daily data.

2.2. Estimation of σ_i

For the estimation of the volatility σ_i we follow the same approach as in Alaton et al. [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. Because only the mean value of temperature in each month is needed, it is not necessary to use a more elaborate model. For this reason σ_i is assumed to be a piece-wise constant function, with a constant value during each month. A crude estimator is based on the quadratic variation of T_t (see Basawa and Prasaka Rao [4])

$$\hat{\sigma}_i^2 = \frac{1}{N_i} \sum_{t=0}^{N_i-1} (T_{t+1} - T_t)^2, \tag{14}$$

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

The estimator that will be used in the sequel is derived by discretizing (6) and using the discretized equation as a regression equation. During a given month i , the discretized equation is

$$T_t = \vartheta_t - \vartheta_{t-1} + a_i \vartheta_{t-1} + (1 - a_i) T_{t-1} + \sigma_i \epsilon_{t-1} \quad t = 1 \dots N_i, \tag{15}$$

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 [6]),

$$\hat{\sigma}_i^2 = \frac{1}{N_i - 2} \sum_{t=1}^{N_i} (T_t - (\vartheta_t - \vartheta_{t-1}) - \hat{a}_i \vartheta_{t-1} - (1 - \hat{a}_i) T_{t-1})^2, \tag{16}$$

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 [5], 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

$$G(a_i) = \sum_{t=1}^{N_i} \frac{\dot{b}(T_{t-1}; a_i)}{\sigma_{i,t-1}^2} \{T_t - E[T_t | T_{t-1}]\}, \tag{17}$$

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

$$b(T_t, a_i) = \frac{d\vartheta_t}{dt} + a_i (\vartheta_t - T_t). \tag{18}$$

In order to obtain the solution of (17), 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

$$T_t = \vartheta_t + e^{-a_i} (T_{t-1} - \vartheta_{t-1}) + e^{-a_i t} \int_{t-1}^t \sigma_s e^{a_i s} dW_s, \tag{19}$$

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

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

which yields

$$E [T_t|T_{t-1}] = e^{-a_i} (T_{t-1} - \vartheta_{t-1}) + \vartheta_t, \tag{20}$$

because the expected value of an Itô integral is zero.

By substituting (20) in (17) we find

$$G_{N_i}(a_i) = \sum_{t=1}^{N_i} \frac{\dot{b}(T_{t-1}; a_i)}{\sigma_{t-1}^2} [T_t - \vartheta_t - e^{-a_i} (T_{t-1} - \vartheta_{t-1})], \tag{21}$$

from which we obtain

$$\hat{a}_i = -\log \left(\frac{\sum_{t=1}^{N_i} \frac{\vartheta_{t-1} - T_{t-1}}{\sigma_{t-1}^2} (T_t - \vartheta_t)}{\sum_{t=1}^{N_i} \frac{\vartheta_{t-1} - T_{t-1}}{\sigma_{t-1}^2} (T_{t-1} - \vartheta_{t-1})} \right). \tag{22}$$

Inserting the data of temperatures and the estimator $\hat{\sigma}$ given by (14), 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. Using Euler approximation scheme, we discretize equation (6) obtaining

$$T_t = \vartheta_t - \vartheta_{t-1} + a_i \vartheta_{t-1} + (1 - a_i) T_{t-1} + \sigma_i \epsilon_{t-1}, \tag{23}$$

where $\{\epsilon_t\}_{t=1}^{364}$ are independent standard normally distributed random variables. Figure 3 shows both the evolution of a simulated trajectory of the estimated temperature and its mean ϑ_t , while Figure 4 gives the evolution of 10 scenarios of temperatures.

Table 2. The estimator \hat{a}_i based on the formula (22).

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

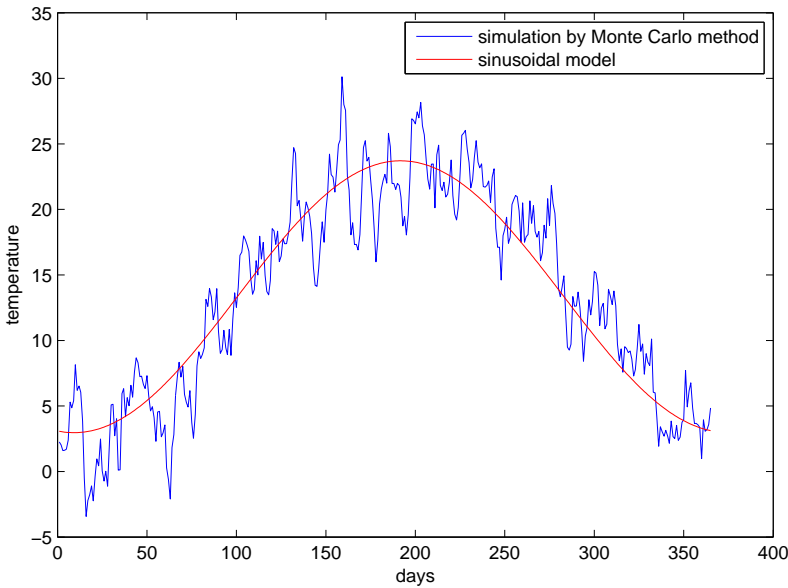


Fig. 3. Simulation of sample path of temperature estimated by Monte Carlo method according to eq. (23) and estimated mean temperature ϑ_t according to the sinusoidal model (7).

The following notation is used:

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

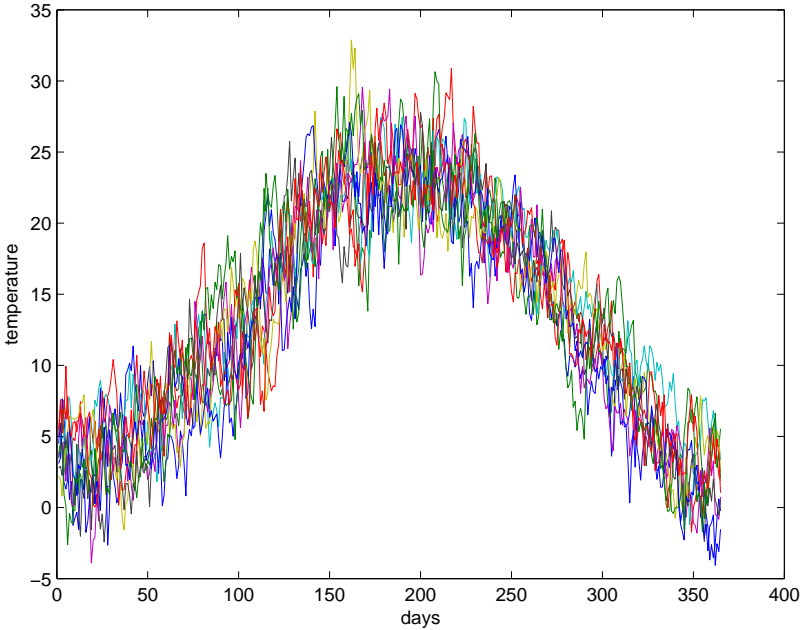


Fig. 4. 10 scenarios of temperature estimated by Monte Carlo simulation according to eq. (23).

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

In the literature (see Brooks [7], Eydeland and Wolyniec [11], Ermoliev and Wets [12] and Ruszczyński and Shapiro [13]) 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, uses the temperature Δ as source of uncertainty. The consumptions of the first six classes of consumers are considered as dependent on temperature variations along the months.

The following notations are used:

- $I = \{i = 1, \dots, 12\}$ is the set of month indices, with $i = 1$ corresponding to July and $i = 12$ corresponding to the following June;
- $J = \{j = 1, \dots, 10\}$ is the set of consumer class indices;
- $S = \{s = 1, \dots, N\}$ is the set of scenario indices;
- c_{ij}^s is the consumption of consumer j , $j = 1, \dots, 6$, in month $i \in I$ along scenario $s \in S$

$$c_{ij}^s = \bar{C}_{ij} + C_{ij}\Delta_i^s, \quad j = 1, \dots, 6, i \in I, s \in S, \quad (24)$$

where \bar{C}_{ij} is the average consumption of consumer j in month $i \in I$; for $j = 7, \dots, 10$ the consumption does not depend on temperature and therefore

$$c_{ij} = \bar{C}_{ij}, \quad j = 7, \dots, 10, i \in I; \quad (25)$$

- va_j^s is the annual volume of gas for consumer j , $j = 1, \dots, 6$, along scenario $s \in S$

$$va_j^s = \sum_{i=1}^{12} c_{ij}^s, \quad j = 1, \dots, 6, s \in S, \quad (26)$$

for $j = 7, \dots, 10$ the annual volume of gas is

$$va_j = \sum_{i=1}^{12} \bar{C}_{ij}, \quad j = 7, \dots, 10; \quad (27)$$

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

$$vw_j^s = \sum_{i=5}^9 c_{ij}^s, \quad j = 1, \dots, 6, s \in S, \quad (28)$$

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

$$vw_j = \sum_{i=5}^9 \bar{C}_{ij}, \quad j = 7, \dots, 10; \quad (29)$$

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

$$r_j^s = \frac{vw_j^s}{va_j^s}, \quad j = 1, \dots, 6, \quad s \in S, \quad (30)$$

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

$$r_j = \frac{vw_j}{va_j}, \quad j = 7, \dots, 10; \quad (31)$$

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

$$cd_{ij}^s = c_{ij}^s \frac{\gamma}{t_i}, \quad j = 1, \dots, 6, \quad i \in I, \quad s \in S, \quad (32)$$

where t_i is the number of days of the month $i \in I$ and γ is a parameter given by the Authority; we note that for industrial customers $j = 7, \dots, 10$ the formula simplifies in

$$cd_{ij} = c_{ij} \frac{\gamma}{t_i}, \quad j = 7, \dots, 10, \quad i \in I; \quad (33)$$

- 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, \bar{nc}_j ,

$$0 \leq nc_j \leq \bar{nc}_j, \quad j \in J; \quad (34)$$

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

$$cm_i^s = \sum_{j=1}^6 c_{ij}^s \cdot nc_j + \sum_{j=7}^{10} c_{ij} \cdot nc_j, \quad i \in I, \quad s \in S; \quad (35)$$

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

$$ca^s = \sum_{i=1}^{12} cm_i^s, \quad s \in S; \quad (36)$$

- 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

$$x^s = \frac{ca^s}{365 \cdot g}, \quad s \in S; \quad (37)$$

- l_j is the loading factor of consumer $j = 7, \dots, 10$, $l_j = \frac{va_j}{cd_j \cdot 365}$, where cd_j is the maximum daily capacity of consumer j , $cd_j = \max_i \{cd_{ij}\}$;

- 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 ($i = 5, \dots, 9$) 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

$$0 \leq s_{0i}^{+s} \leq \pi_{0i} \cdot g, \quad i = 5, \dots, 9, \quad s \in S, \quad (38)$$

$$\pi_{0i} \cdot g \leq s_{1i}^{+s} \leq \pi_{1i} \cdot g, \quad i = 5, \dots, 9, \quad s \in S, \quad (39)$$

$$\pi_{2i} \cdot g \leq s_{2i}^{+s}, \quad i = 5, \dots, 9, \quad s \in S, \quad (40)$$

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 $s \in S$

$$cw^s = \sum_{i=5}^9 cm_i^s, \quad s \in S; \quad (41)$$

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

$$h^s = \frac{cw^s}{ca^s}, \quad s \in S; \quad (42)$$

- P^s is the purchase price to be paid by the gas seller to the shipper along scenario $s \in S$: it is expressed as a linear function of x^s , and is defined as

$$P^s = QT + QS + q + m \cdot x^s, \quad s \in S; \quad (43)$$

where q is the intercept and m is the slope; QT and QS are fixed by the Italian Regulatory Authority;

- P'_j is the price to be paid by the first 6 classes of consumers and is defined as

$$P'_j = (CMP + QVD) \cdot (1 - \alpha_j), \quad (44)$$

where the values of CMP and QVD are fixed by the Italian Regulatory Authority and cover raw material costs (production, importation and transport) and retail commercialization costs respectively; α_j is a parameter representing possible discount fixed by the gas seller to be applied to consumer j ;

- P''_j^s is the price applied by the gas seller to consumer class j , $j = 7, \dots, 10$ along scenario $s \in S$

$$P''_j^s = P^s - \beta_j \cdot \left(1 - \frac{x^s}{l_j}\right) + \delta_j \cdot (r_j - h^s) + \lambda_j, \quad (45)$$

where β_j and δ_j are constant values and λ_j is a possible recharge which can be applied to the industrial consumer class j .

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

$$w = E \left[\sum_{j=1}^6 (P'_j \cdot va_j^s \cdot nc_j) + \sum_{j=7}^{10} (P_j''^s \cdot va_j \cdot nc_j) - P^s \cdot ca^s - \sum_{i=5}^9 \sum_{k=1}^2 \mu_{ki} s_{ki}^{+s} \right]. \tag{46}$$

Notice that

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

$$E \left[\sum_{j=1}^6 (P'_j \cdot va_j^s \cdot nc_j) \right] = \sum_{j=1}^6 \left(P'_j \cdot \sum_{i=1}^{12} \bar{C}_{ij} \cdot nc_j \right); \tag{47}$$

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

$$E \left[\sum_{j=7}^{10} (P_j''^s \cdot va_j \cdot nc_j) \right] = \sum_{j=7}^{10} \left(nc_j E[P_j''^s] \cdot \sum_{i=1}^{12} c_{ij} \right), \tag{48}$$

being the industrial consumptions independent of temperature. Notice that

$$\begin{aligned} E(P_j''^s) &= E(P^s) - \beta_j \cdot \left(1 - E\left(\frac{x^s}{l_j}\right) \right) + \delta_j \cdot (r_j - E(h^s)) + \lambda_j \\ &= QT + QS + q - \beta_j + \delta_j r_j + \lambda_j + \left(m + \frac{\beta_j}{l_j} \right) \sum_{s=1}^N x^s p^s - \delta_j \sum_{s=1}^N h^s p^s; \end{aligned} \tag{49}$$

- the expected value of the costs is

$$\begin{aligned} &E[P^s \cdot ca^s] (QT + QS + q) \sum_{i=1}^{12} E[cm_i^s] \\ &+ \frac{m}{365 \cdot g} \left(E \left[\sum_{i=1}^{12} (cm_i^s)^2 \right] + 2E \left[\sum_{\substack{i,k=1 \\ k>i}}^{12} (cm_i^s) (cm_k^s) \right] \right), \end{aligned} \tag{50}$$

where

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

and

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

- the expected value of the penalties is

$$E \left[\sum_{i=5}^9 \sum_{k=1}^2 \mu_{ki} s_{ki}^{+s} \right] = \sum_{s=1}^N \left(\sum_{i=5}^9 \sum_{k=0}^2 \mu_{ki} s_{ki}^{+s} \right) p^s . \tag{51}$$

The constraints of our stochastic problem are the following:

$$0 \leq nc_j \leq \bar{nc}_j, \quad j \in J, \tag{52}$$

$$\sum_{j=1}^6 cd_{ij}^s \cdot nc_j + \sum_{j=7}^{10} cd_{ij} \cdot nc_j - g \leq \sum_{k=0}^2 s_{ki}^{+s}, \quad i = 5, \dots, 9, s \in S, \tag{53}$$

$$0 \leq s_{0i}^{+s} \leq \pi_{0i} \cdot g, \quad i = 5, \dots, 9, s \in S, \tag{54}$$

$$\pi_{0i} \cdot g \leq s_{1i}^{+s} \leq \pi_{1i} \cdot g, \quad i = 5, \dots, 9, s \in S, \tag{55}$$

$$\pi_{2i} \cdot g \leq s_{2i}^{+s}, \quad i = 5, \dots, 9, s \in S. \tag{56}$$

Notice that the problem may also be formulated as a *two-stage stochastic model with recourse* as follows:

$$\max E_{\xi} [f(x, y(\Delta))] , \tag{57}$$

$$Ax = b, \tag{58}$$

$$T(\Delta)x + Wy(\Delta) = h(\Delta), \tag{59}$$

$$x \geq 0, y(\Delta) \geq 0, \tag{60}$$

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

- 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;

whereas the second stage decision variable $y(\Delta)$ involves the surplus in consumption in the peak day s_{ki}^{+s} in winter month i . Furthermore the first stage constraint (58) is represented by equation (52) and the second stage constraint (59) by equations (53), (54), (55) and (56).

4. 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. [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 purchase price P^s 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 [8]) 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 his simulations. In our case we use:

$$P^s(x^s) = QT + QS + 18.348 - 3.866 \cdot x^s, \quad (61)$$

where the intercept value is 18.348 and the slope value is -3.866 ; the values QT and QS are given by the Italian Regulatory Authority: in our numerical experiments $QT = 2.4953171$ Eurocent/Stm³ and $QS = 0.63882$ Eurocent/Stm³.

The relation between the consumption of consumer j , $j = 1, \dots, 6$, in month $i \in I$ along scenario $s \in S$, c_{ij}^s , and the deviation from mean value over scenarios, Δ_i^s , is supposed to be linear, with intercept equal to \bar{C}_{ij} and the other coefficient computed via a linear regression. The regression results seem to be significant for all the consumers.

The model has been validated by running several tests both in the deterministic (see Allevi et al. [2]) and in the stochastic case. The deterministic results are reported in Table 3. For the stochastic model, we report the result obtained by solving 10000 times the problem, each time with $N = 50$ 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 4 their average over 10000 trials.

Table 3. Optimal values for citygate Sotto il Monte in the deterministic case.

Profit	154265	Euro
P	19.67	Eurocent/Stm ³
ca	4484406	Stm ³
g	26399	Stm ³
x	0.4654	

Table 4. Average optimal values over 10000 simulations for citygate Sotto il Monte in the stochastic case with $N = 50$.

Profit	152219	Euro
P	19.67	Eurocent/Stm ³
ca	4484525	Stm ³
g	26309	Stm ³
x	0.4669	

While in the deterministic case, the consumption surplus in January and February is under 10% and therefore no penalization has to be paid, in the stochastic case a nonzero penalization is applied in scenarios with high variance in consumptions. In fact the stochastic approach gives indications to the gas seller that in scenarios with colder temperatures, he could face the possibility of a reduced profit due to penalties. This solution, though, allows gas seller to have the same purchase price of the deterministic case and therefore the same selling price for the industrial customer; this means that industrial consumer is still very important and worthwhile to belong to the retail seller’s portfolio.

To validate the model we analyze the sensitivity of solutions to different number of scenarios. We have run 1000 and 10000 simulations with increasing number of scenarios. In Figures 5 and 6 we report for each number of scenarios the average optimal value over the corresponding number of simulations; we observe that the optimal profit converges to a value between 152200 and 152210.

5. CONCLUSIONS

We have proposed a stochastic model for the management of a gas sale company where the uncertainty is based on a mean reversion stochastic process for the evolution of temperature; as the number of scenarios increases, the complexity of the problem also increases: one further possibility is to devise a new algorithm that decouples computation of g from all other decision variables so that the problem becomes linear. Moreover, there exists a relation between purchase price p and international price indices, since gas seller must choose the index of reference among a certain number of admitted choices: it is possible to investigate the influence on P of future variations of these indices to help gas seller in taking his decision.

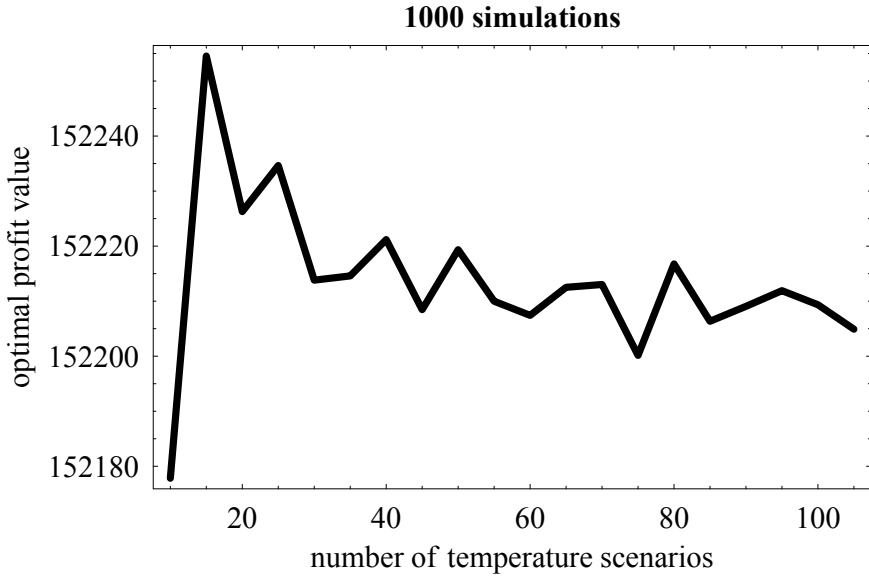


Fig. 5. Case of 1000 simulations: optimal profit value as the number of scenarios increases.

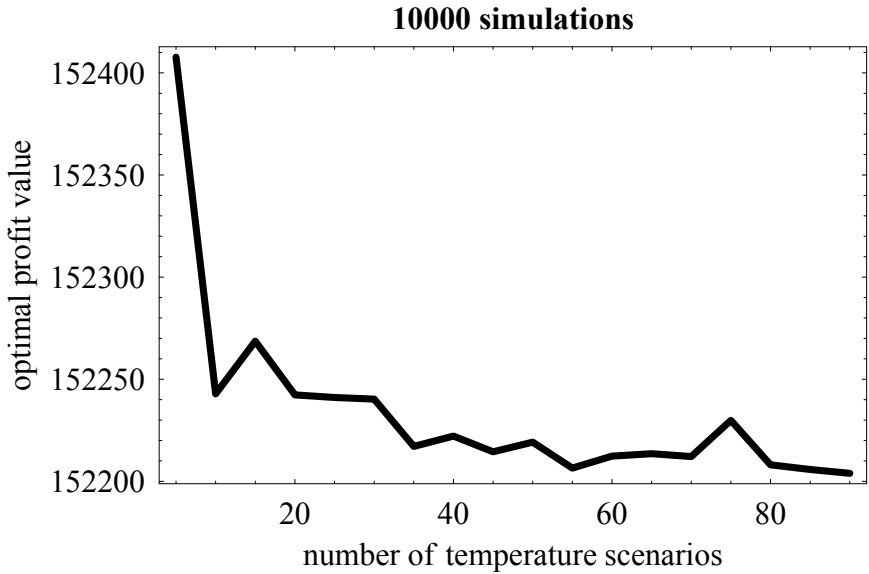


Fig. 6. Case of 10000 simulations: optimal profit value as the number of scenarios increases.

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

$$\begin{aligned}
 Y_t &= \left[e^{(\sum_{k=1}^{i-1} \int_0^{N_k} a_k ds) + \int_0^{t - \sum_{k=1}^{i-1} N_k} a_i ds} \right] (\vartheta_t - T_t) \\
 &= \left[e^{(\sum_{k=1}^{i-1} N_k a_k) + a_i (t - \sum_{k=1}^{i-1} N_k)} \right] (\vartheta_t - T_t),
 \end{aligned}
 \tag{A.1}$$

then Itô's formula implies

$$\begin{aligned}
 dY_t &= \left[e^{(\sum_{k=1}^{i-1} N_k a_k) + a_i (t - \sum_{k=1}^{i-1} N_k)} \right] \left[\left(\frac{d\vartheta_t}{dt} + a_i (\vartheta_t - T_t) \right) dt \right. \\
 &\quad \left. - \left(a_i (\vartheta_t - T_t) + \frac{d\vartheta_t}{dt} \right) dt - \sigma_t dW_t \right],
 \end{aligned}
 \tag{A.2}$$

hence

$$Y_t - Y_0 = - \int_0^t \sigma_s \left[e^{(\sum_{k=1}^{i-1} N_k a_k) + a_i (s - \sum_{k=1}^{i-1} N_k)} \right] dW_s,
 \tag{A.3}$$

that is

$$\begin{aligned}
 &\left[e^{(\sum_{k=1}^{i-1} N_k a_k) + a_i (t - \sum_{k=1}^{i-1} N_k)} \right] (\vartheta_t - T_t) \\
 &= \vartheta_0 - T_0 - \int_0^t \left[e^{(\sum_{k=1}^{i-1} N_k a_k) + a_i (s - \sum_{k=1}^{i-1} N_k)} \right] \sigma_s dW_s,
 \end{aligned}$$

but $\vartheta_0 = T_0 = C$ and thus

$$T_t = \vartheta_t + e^{-[(\sum_{k=1}^{i-1} N_k a_k) + a_i (t - \sum_{k=1}^{i-1} N_k)]} \int_0^t \left[e^{(\sum_{k=1}^{i-1} N_k a_k) + a_i (s - \sum_{k=1}^{i-1} N_k)} \right] \sigma_s dW_s,
 \tag{A.4}$$

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

ACKNOWLEDGEMENT

Earlier versions of this paper were presented at the VIII SIMAI Congress, Ragusa (Italy, 2006), at the XXXVII AIRO Conference, Cesena (Italy, 2006). The authors greatly acknowledge the useful comments and helpful suggestions of an anonymous referee.

(Received October 23, 2007.)

REFERENCES

[1] P. Alaton, B. Djehiche, and D. Stillberger: On modelling and pricing weather derivatives. *Appl. Math. Finance* 9 (2002), 1, 1–20.

- [2] E. Allevi, M. I. Bertocchi, M. Innorta, and M. T. Vespucci: A mixed integer nonlinear optimization model for gas sale company. *Optim. Lett.* 1 (2007), 1, 61–69.
- [3] E. Allevi, M. I. Bertocchi, M. Innorta, and M. T. Vespucci: A stochastic optimization model for gas sales companies. *IMA J. Management Math.* (2007), 1–14.
- [4] I. V. Basawa and B. L. S. Prakasa Rao: *Statistical Inference for Stochastic Processes*. Academic Press, London 1980.
- [5] B. M. Bibby and M. Sorensen: Martingale estimation functions for discretely observed diffusion processes. *Bernoulli* 1 (1995), 1/2, 17–39.
- [6] P. J. Brockwell and R. A. Davis: *Time Series: Theory and Methods*. Second edition. Springer, Berlin 1996.
- [7] R. E. Brooks: Using generalized networks to forecast natural gas distribution and allocation during periods of shortage. *Math. Programming Stud.* 15 (1981), 23–42.
- [8] J. Davidson: *Econometric Theory*. Blackwell Publishing, 2000.
- [9] *Deliberazione 138 (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.*
- [10] F. Dornier and M. Queruel: *Pricing weather derivatives by marginal value*. Quantitative Finance 1, Institute of Physics Publishing 2000.
- [11] A. Eydeland and K. Wolyniec: *Energy and Power Risk Management*. Wiley, New York 2003.
- [12] Y. Ermoliev and J.-B. Wets: *Numerical Techniques for Stochastic Optimization*. Springer-Verlag, Berlin 1988.
- [13] A. Ruszczyński and A. Shapiro: *Stochastic Programming*. Elsevier, Amsterdam 2003.

Francesca Maggioni and Marida Bertocchi, Department of Mathematics, Statistic, Computer Science and Applications, Bergamo University, Via dei Caniana 2, Bergamo 24127. Italy.

Maria Teresa Vespucci and Mario Innorta, Department of Informatics and Mathematical Modeling, Bergamo University, Via Marconi 5, Dalmine 24044. Italy.

Elisabetta Allevi Department of Quantitative Methods, Brescia University, Contrada S. Chiara, 50, Brescia 25122. Italy.

e-mails: francesca.maggioni@unibg.it, marida.bertocchi@unibg.it, mtvespucci@tin.it, mario.innorta@fastwebnet.it, allevi@eco.unibs.it