Applied Energy xxx (2008) xxx–xxx

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

The trend of the total stock of the private car-petrol in Spain: Stochastic modelling using a new gamma diffusion process

R. Gutiérrez *, R. Gutiérrez-Sánchez, A. Nafidi

Department of Statistics and Operational Research, University of Granada, Facultad de Ciencias, Campus de Fuentenueva 18071 Granada, Spain

ARTICLE INFO

Article history:
Received 6 February 2008
Received in revised form 25 March 2008
Accepted 26 March 2008
Available online xxxx

Keywords:
Stochastic gamma diffusion process
Statistical inference with discrete sampling
Estimated trend functions
Forecasting
Total stock of private cars-petrol in Spain

ABSTRACT

The main aim of this study is to model the trend of the evolution of the total stock of private petrol-driven cars. In Spain, as in other EU countries, this trend between 2000 and 2005 differed significantly from that observed from 1986 to 1999. Moreover, it varies greatly from that corresponding to the stock of diesel-driven cars, which consistently presents an exponential Gompertz-type increase. Spain constitutes a typical example of a failure to observe the maximum CO_2 emission levels assigned to it by 2012 under the Kyoto Protocol (1992); a significant percentage of these excess emissions is accounted for by the land transport sector, in general, and by the private cars subsector, in particular. This paper proposes a stochastic model based on a new non homogeneous stochastic gamma-type diffusion process which it is a stochastic version of a Gamma function type deterministic growth model considered in Skiadas [Skiadas CH. Methods of growth functions formulation. In: Gutiérrez R, Valderrama M, editors. Selected topics on stochastic modelling. World Scientific; 1994. p. 296–310]. We describe its main probabilistic characteristics and establish a statistical methodology by which it can be fitted to real data and obtain medium-term forecasts that, in statistical terms, are quite accurate.

1. Introduction

In recent decades important advances have been made in modelling based on stochastic diffusion processes; these have been applied in many scientific fields such as the environment, energy, stochastic economics, tumour growth and that of populations in general. Various authors have studied these diffusion processes from the standpoint of the corresponding Ito SDEs, and most successful predictive models have been based on forms of the deterministic diffusion process extended to the corresponding stochastic version, for example Katsamaki and Skiadas [2] in the case of the exponential model, Skiadas and Giovanis [3] in the case of the Bass model and Giovanis and Skiadas [4] in the case of the logistic model.

The question of statistical inference and the problem of parameter estimation in these processes have received considerable attention in recent years, in situations in which the process is observed continuously or discretely. In most cases, the parameter estimation is based on approximating the maximum likelihood (ML) methodology. A large body of literature has addressed this question; important research on the topic, both in general and in particular cases includes, for example, Bibby and Sorensen [5]; Ait Sahalia [6], Durham and Gallant [7], Eugene [8] and the extensive review given in Prakasa-Rao [9] and Kutoyants [10], without forgetting the early works focused on other methodologies, such as the generalized method of moments by Chan et al. [11]; the non parametric method described by Arapis and Gao [12], and a method based on Bayesian analysis reported by Elerian et al. [13].

Another way of defining and studying stochastic diffusion processes is based on Kolmogorov equations (backwards and forwards, or Fokker–Plank) associated with the corresponding infinitesimal moments (drift and diffusion coefficients). This method of approaching and studying the topic is particularly interesting when we wish to construct non homogeneous versions of a diffusion by introducing just time functions (exogenous factors) into the infinitesimal moments. These exogenous factors or variables evolve in time externally to the evolution that corresponds to the endogenous variable that is being modelled in each case. Thus, we may obtain various non homogeneous extensions of the homogeneous stochastic diffusion processes, which are based on the incorporation of certain exogenous factors that affect the drift of the basic homogeneous diffusion under consideration. The exogenous factors that are considered may be of diverse characteristics, with or without external information.

In the case in which external information is present, the exogenous factors are completely defined by their observed values, such as monthly or annual data. This case has been considered by

* Corresponding author. Tel./fax: +34 958243267.
E-mail address: rgjaimez@ugr.es (R. Gutiérrez).

0306-2619/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.apenergy.2008.03.016

Gutiérrez et al. [14,15] for the lognormal process, by Gutiérrez et al. [16,17] for the Gompertz process, and by Gutiérrez et al. [18] for the Vasicek process. These latter studies consider non homogenous models whose drifts are linear combinations of exogenous factors of the present type.

In the second case (lacking external information), exogenous factors are functions of time and no observed values are available for these factors. This case has been studied by Gutiérrez et al. [19] for the lognormal process, considering an exogenous factor given by a time polynomial; Ferrante et al. [20] considered a Gompertz process with an exogenous factor given by the sum of two exponential functions. Another case involving the Gompertz process was examined by Albano and Giorno [21], who took an exogenous factor that was defined by a logarithmic function. Finally, Gutiérrez et al. [22] considered an exogenous factor of an exponential type within the homogeneous lognormal process; the resultant process is a Gompertz diffusion type process that depends on the initial value of the process.

This method for obtaining non homogeneous versions of univariate diffusion processes by introducing exogenous factors has also been studied in some cases of multivariate diffusion processes, for example by Gutiérrez et al. [23] in the case of multivariate lognormal diffusion and by Gutiérrez et al. [24] in the case of the bivariate Gompertz diffusion process.

The statistically adjusted trend of the stochastic diffusion process used in modelling sometimes presents divergences from the behaviour of the real data, which in the first place leads to the search for exogenous variables, i.e. those external to the process, to achieve a suitable correction of the modelling. In other words, the model is corrected in order to consider a non homogeneous version of the model such that the new non homogeneous trend, which is affected by such exogenous variables, provides a better description of the real behaviour of the phenomenon.

However, there are sometimes found to be discrepancies between the values adjusted by the diffusion that is being used for modelling and the observed values, in such a way that it is not possible to correct these discrepancies by the mere fact of introducing exogenous factors into the baseline diffusion. This statement is based on the fact that, structurally, the real phenomenon that is being studied does not correspond to the model that is utilized even when the latter is extended to a non homogeneous version. Hence, we must find another model, one that is fundamentally different from the one initially taken, one that provides an appropriate modelling of the phenomenon and that better fits the behaviour of the real data. One example of such a situation is that described in the study and prediction of the evolution of the total stock of private cars-petrol in Spain, taking as a basis for analysis the real data for this total stock of private cars-petrol, for the period 1986–2005, as shown in Fig. 1. Obviously, the exponential trend of the lognormal stochastic diffusion process is not suitable; neither are the trends of the logistic or the Gompertz processes, or the trends of other models that have been studied for example, of Rayleigh-type diffusion process (see Gutiérrez et al. [25]), cubic model diffusion (see Gutiérrez et al. [26]) or inverse CIR model (Gutiérrez et al. [27]). A rapid examination of Fig. 1 suggests that a gamma function might be appropriate for describing the trend of the behaviour of the real data concerning the total stock of petrol-driven vehicles. For this reason, defining a diffusion process whose trend was proportional to a Gamma-type function was set as the technical objective of the present study.

In this paper, therefore, we propose a new stochastic type gamma diffusion process (SGDP). The paper is structured as follows: in Section 2, we identify the main characteristics of the proposed process, namely the analytical expression, the probability transition density function (ptdf) and the trend functions. In Section 3, the parameter estimators are derived by the maximum likelihood method, based on the discrete sampling of the process, after which the distributions of the likelihood estimators are obtained. Thus, we obtain the confidence intervals of the parameters and of the trend functions. In the final section, the process is applied to time series data on the total stock of petrol-driven private cars in Spain.

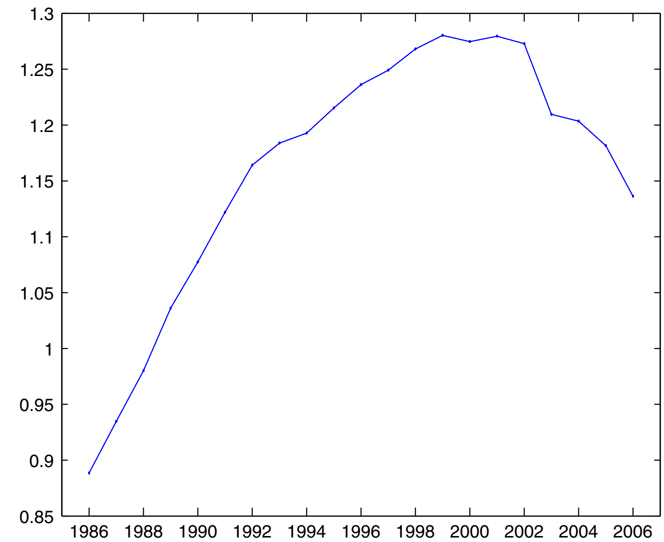

Fig. 1. Real data.

2. The model and its characteristics

2.1. Formulation of the model

Stochastic diffusion processes may be generated in various ways; one is to begin with a discrete probabilistic model and, by limit steps, then obtain a continuous time-series diffusion model. Following this approach, Ricciardi [28] for example, derived a lognormal, logistic diffusion process, while Albano and Giorno [21] obtained a Gompertz diffusion process.

Another way of arriving at stochastic diffusion processes is to use the mechanism of the regulation function introduced by Ricciardi [28], and which consists in including a regulatory random function that depends on a stationary Gaussian process $\Lambda(t)$ with a zero mean and a delta-type correlation function: $\mathbb{E}[\Lambda(t)] = 0$, $\mathbb{E}[\Lambda(t_1) \cdot \Lambda(t_2)] = \sigma^2 \delta(t_2 - t_1)$, in the Malthusian model; we then obtain the well known regulated growth process that is given by the following equation

$$\frac{dx(t)}{dt} = \alpha x(t)[1 - \Phi(t, x)],$$

where $\Phi(t, x)$ is the regulation function that involves a change in the growth rate with the population size and with time.

By using this mechanism, various diffusion processes have been obtained, considering particular cases of $\Phi(t, x)$. Thus, for example, Ricciardi [28] studied the case in which $\Phi(t, x) = \frac{\beta}{\alpha} x - \frac{1}{\alpha}\Lambda(t)$ and so the resulting process is one of logistic diffusion; by assuming $\Phi(t, x) = \frac{\beta}{\alpha}\log(x) - \frac{1}{\alpha}\Lambda(t)$, we obtain the Gompertz diffusion process in which the bounds do not depend on the initial value; while for $\Phi(t, x) = 1 - e^{-\beta(t-t_1)} - \frac{1}{\alpha}\Lambda(t)$ we obtain the Gompertz diffusion process that depends on the initial value (see Gutiérrez et al. [22]).

In this study, the above methodology has been adapted to define a new gamma-type stochastic diffusion process. To do so, we took the function $\Phi(t, x)$ in the following form:

$$\Phi(t, x) = 1 - \frac{1}{t} + \frac{\beta}{\alpha} - \frac{1}{\alpha}\Lambda(t).$$

Please cite this article in press as: Gutiérrez R et al., The trend of the total stock of the private car-petrol in Spain: ..., Appl Energ (2008), doi:10.1016/j.apenergy.2008.03.016

Thus, the following equation was obtained:

$$\frac{dx(t)}{dt} = \left(\frac{\alpha}{t} - \beta\right)x(t) + x(t)\Lambda(t),$$

given that $\Lambda(t)dt = \sigma dw(t)$, where $w(t)$ is a standard Wiener process. This equation can be expressed in the normal SDE form as follows:

$$dx(t) = \left(\frac{\alpha}{t} - \beta\right)x(t)dt + \sigma x(t)dw(t). \quad (1)$$

We shall now consider the process $\{x(t); t \in [t_1, T]\}$ defined in Eq. (1), with value in $(0, \infty)$ and the initial condition $P[x(t_1) = x_{t_1}] = 1$, where $t_1 > 0$ and x_{t_1} is a positive real value.

2.2. Analytical expression, ptdf and moments of the SGDP

It can be proved that the functionals coefficients in Eq. (1) are non-anticipative and satisfy the Lipschitz and the growth conditions and consequently, that there exists a unique, strong solution to Eq. (1) [see, for example, Liptser and Shiryayev [29], Theorem 4.6].

Furthermore, it is straightforward to show that these functionals are Borel measurable and satisfy the uniform Lipschitz condition and the c-Holder, in particular order 1 Holder, conditions (see, for example, Wong and Hajek [30], Propositions 4.1 and 7.1]. Consequently, there exists a separable, measurable and almost surely (as) sample continuous diffusion process $\{x(t); t \in [t_1, T]\}$ which is the unique (as) solution to Ito's differential equation Eq. (1) with infinitesimal moments (drift and diffusion coefficients) given, respectively, by

$$a(x,t) = \left(\frac{\alpha}{t} - \beta\right)x \quad \text{and} \quad b(x,t) = \sigma^2 x^2.$$

The strong solution to Eq. (1) can be obtained by Ito's formula, transforming the latter using the function $y(t) = \log(x(t))$ to the following SDE

$$dy(t) = \left(\frac{\alpha}{t} - \beta - \frac{\sigma^2}{2}\right)dt + \sigma dw(t); \quad y(t_1) = \log(x_{t_1}).$$

From which we deduce that the analytical expression of the solution to the SDE Eq. (1) is

$$x(t) = x_{t_1}\left(\frac{t}{t_1}\right)^{\alpha} e^{-(\beta+\sigma^2/2)(t-t_1)} \exp[\sigma(w(t) - w(t_1))]$$

then, the random variable $x(t)|x(t_1) = x_{t_1}$ has a one-dimensional lognormal distribution $\Lambda_1[\mu(t_1, t, x_{t_1}), \sigma^2(t - t_1)]$, where $\mu(s, t, x)$ is given by

$$\mu(s,t,x) = \log(x) + \alpha\log(t/s) - (\beta + \sigma^2/2)(t - s).$$

From which, we deduce that the tpdf of the process is

$$f(y,t|x,s) = [2\pi\sigma^2(t-s)]^{-1/2} \exp\left(-\frac{[\log(y) - \mu(s,t,x)]^2}{2\sigma^2(t-s)}\right). \quad (2)$$

Taking into account that the random variable $x(t) \mid x(s) = x_s$ is distributed as $\Lambda_1[\mu(s,t,x_s), \sigma^2(t-s)]$ and bearing in mind the properties of this distribution, the rth conditional moment of the process is expressed by

$$\mathbb{E}[x^r(t)|x(s) = x_s] = \exp\left(r\mu(s,t,x_s) + \frac{r\sigma^2}{2}(t-s)\right)$$
$$= x_s\left(\frac{t}{s}\right)^{r\alpha} e^{-r\beta(t-s)} \exp\left(\frac{r(r-1)}{2}\sigma^2(t-s)\right).$$

Then, the conditional trend function of the process is

$$\mathbb{E}[x(t)|x(s) = x_s] = x_s\left(\frac{t}{s}\right)^{\alpha} e^{-\beta(t-s)}. \quad (3)$$

Assuming the initial condition $P(x(t_1) = x_1) = 1$, the trend function of the process is

$$\mathbb{E}[x(t)] = \frac{x_{t_1} e^{\beta t_1}}{t_1^{\alpha}} t^{\alpha} e^{-\beta t}. \quad (4)$$

Remarks

- Note that in the absence of white noise (i.e. $\sigma = 0$), the solution of the ordinary differential equation associated with the SDE Eq. (1) is $x(t) = kt^{\alpha}e^{-\beta t}$, which is proportional to the Gamma density function. We can see that the trend function given in Eq. (4) is also proportional to the density function of the gamma distribution.
- For $\alpha = 0$ and $\beta < 0$, we find the stochastic homogeneous lognormal diffusion process studied by Tintner and Sengupta [31].

3. Statistical inference on the model

3.1. Maximum likelihood parameter estimation

As the tpdf of the process is known (lognormal distribution), we can estimate the parameters involved in the process, making use of discrete sampling, based on the conditioned likelihood function obtained as the product of the corresponding process transitions (given by Eq. (2)). Let us examine the following process sample, $x_1, \ldots, x_n$ at the instants $t_1, t_2, \ldots, t_n$. Assuming the initial distribution $P[x(t_1) = x_1] = 1$, the conditioned likelihood function associated with the process and with the above sample is given as follows:

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

Let us now perform a change of variable, in order to work with a known likelihood function and to calculate the maximum likelihood estimators in a simpler way. Consider the following transform: $v_i = (t_i - t_{i-1})^{-1/2}(\log(x_i) - \log(x_{i-1}))$, $i = 2, \ldots, n$, then, with the reparametrization $\mathbf{a} = (\alpha, -(\beta + \sigma^2/2))'$ and if we denote by $u_i = (t_i - t_{i-1})^{-1/2}(\log(t_i/t_{t-1}), t_i - t_{i-1})'$, the likelihood function for the transformed sample is

$$\mathbb{L}_{v_2,\ldots,v_n}(\mathbf{a}, \sigma^2) = \left[2\pi\sigma^2\right]^{-(n-1)/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=2}^{n}(v_i - u_i'\mathbf{a})^2\right).$$

Let be $\mathbf{V} = (v_2, \ldots, v_n)'$ and $\mathbf{U}$ is the $2 \times (n-1)$ matrix, whose rank is 2, and given by $\mathbf{U} = (\mathbf{u}_2, \ldots, \mathbf{u}_n)$, then, the likelihood function can be rewrite in the following form

$$\mathbb{L}_{\mathbf{V}}(\mathbf{a}, \sigma^2) = [2\pi\sigma^2]^{-(n-1)/2} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{V} - \mathbf{U}'\mathbf{a})'(\mathbf{V} - \mathbf{U}'\mathbf{a})\right). \quad (5)$$

After calculating the derivatives of the log-likelihood function with respect to $\mathbf{a}$ and σ^2, the likelihood equations are

$$\mathbf{U}(\mathbf{V} - \mathbf{U}'\hat{\mathbf{a}}) = 0,$$
$$(n-1)\hat{\sigma}^2 = (\mathbf{V} - \mathbf{U}'\hat{\mathbf{a}})'(\mathbf{V} - \mathbf{U}'\hat{\mathbf{a}}).$$

From which, the likelihood estimators of the parameters are

$$\hat{\mathbf{a}} = (\mathbf{U}\mathbf{U}')^{-1}\mathbf{U}\mathbf{V}, \quad (6)$$
$$(n-1)\hat{\sigma}^2 = \mathbf{V}'H_{\mathbf{U}}\mathbf{V}, \quad (7)$$

where $H_{\mathbf{U}} = I_{n-1} - \mathbf{U}'(\mathbf{U}\mathbf{U}')^{-1}\mathbf{U}$

3.2. Properties of maximum likelihood estimators

3.2.1. Distribution of maximum likelihood estimators

Noting that the Eq. (5) is also the density function of the random vector $\mathbf{V}$, this latter density can be rewritten in the following form

Please cite this article in press as: Gutiérrez R et al., The trend of the total stock of the private car-petrol in Spain: ..., Appl Energ (2008), doi:10.1016/j.apenergy.2008.03.016

$$\mathbb{L}_{\mathbf{V}}(\mathbf{a},\sigma^2)=\frac{1}{(2\pi)^{(n-1)/2}|\sigma^2 I_{n-1}|^{1/2}}\times\exp\left(-\frac{1}{2}(\mathbf{V}-\mathbf{U}'\mathbf{a})'(\sigma^2 I_{n-1})^{-1}(\mathbf{V}-\mathbf{U}'\mathbf{a})\right).$$

From which, we deduce that $\mathbf{V}\sim\mathcal{N}_{n-1}[\mathbf{U}'\mathbf{a},\sigma^2\mathbf{I}_{n-1}]$.

The rank of $\mathbf{U}$ is 2, Then, $(\mathbf{U}\mathbf{U}')^{-1}\mathbf{U}$ has the same rank, and we have

$$(\mathbf{U}\mathbf{U}')^{-1}\mathbf{U}\mathbf{V}\sim\mathcal{N}_2[(\mathbf{U}\mathbf{U}')^{-1}\mathbf{U}\mathbf{U}'\mathbf{a},\sigma^2(\mathbf{U}\mathbf{U}')^{-1}(\mathbf{U}\mathbf{U}')\mathbf{U}\mathbf{U}')^{-1})$$

and therefore, we have

$$\hat{\mathbf{a}}\sim\mathcal{N}_2[\mathbf{a},\sigma^2(\mathbf{U}\mathbf{U}')^{-1}].$$

To obtain the distribution of $\hat{\sigma}^2$, we make use of the following result (see for example [32], corollary 2.11.2):

Corollary 1. *If $Z\sim\mathcal{N}_p[\mu,\Sigma],\Sigma$ non singular and $B_{p\times p}$ symmetric, then, $Z'BZ\sim\chi^2_k(\delta)$, where $k=\operatorname{rank}(B)$ and $\delta=\mu'B\mu$ if and only if $B\Sigma$ is idempotent.*

As $H_{\mathbf{U}}$ is symmetric and idempotent, then, $\operatorname{rank}(H_{\mathbf{U}})=\operatorname{tr}(H_{\mathbf{U}})=n-3$, then using the last result in the particular case: $Z=\sigma^{-1}\mathbf{V},\Sigma=I_{n-1},B=H_{\mathbf{U}}$ and $\mu=\mathbf{U}'\mathbf{a}$, we have

$$\frac{\mathbf{V}'}{\sigma}H_{\mathbf{U}}\frac{\mathbf{V}}{\sigma}\sim\chi^2_{n-3}(\delta),\quad\text{with}\quad\delta=\mathbf{a}'\mathbf{U}H_{\mathbf{U}}\mathbf{U}'\mathbf{a}=0$$

and therefore

$$\frac{(n-1)\hat{\sigma}^2}{\sigma^2}\sim\chi^2_{(n-3)}.$$

3.2.2. Independence of maximum likelihood estimators

To show that $\hat{\mathbf{a}}$ and $\hat{\sigma}^2$ are independently distributed, we may make use of the following independence result between linear and quadratic forms (see for example [32] corollary 2.11.4):

Corollary 2. *Let $Z\sim\mathcal{N}_p[\mu,\Sigma]$, with $\Sigma>0$. Then, $y_j=Z'A_jZ+2b_j'Z+c_j, j=1,2$ are independently distributed if and only if $A_1\Sigma A_2=0, A_2\Sigma b_1=0, A_1\Sigma b_2=0$, and $b_1'\Sigma b_2=0$.*

Let $B=(\mathbf{U}\mathbf{U}\prime)^{-1}\mathbf{U}$. Then, by decomposing the matrices B and $H_{\mathbf{U}}$ as follows: $B=\begin{pmatrix}B_1\\B_2\end{pmatrix}$ and $H_{\mathbf{U}}=(H_{\mathbf{U},1},H_{\mathbf{U},2},\dots,H_{\mathbf{U},n-1})$, where B_1 and B_2 are $1\times(n-1)$-vector and $H_{\mathbf{U},i}$ is the column of $H_{\mathbf{U}}$ ($(n-1)\times 1$-vector for $i=1,\dots,n-1$). We have: $\hat{\mathbf{a}}=B\mathbf{V}=(B_1\mathbf{V},B_2\mathbf{V})'$. Then, to show that $\hat{\sigma}^2$ and $\hat{\mathbf{a}}$ are independently distributed we need only show that $\hat{\sigma}^2$ is distributed independently of each component of the vector $\hat{\mathbf{a}}$, that is, of $B_1\mathbf{V}$ and of $B_2\mathbf{V}$.

On the one hand, we have: $BH_{\mathbf{U}}=(\mathbf{U}\mathbf{U}')^{-1}\mathbf{U}(I_{n-1}-\mathbf{U}'(\mathbf{U}\mathbf{U}')^{-1}\mathbf{U})=0$. While on the other hand we have $BH_{\mathbf{U}}=\begin{pmatrix}B_1H_{\mathbf{U}}\\B_2H_{\mathbf{U}}\end{pmatrix}$. From which, we deduce that

$$B_1H_{\mathbf{U}}=0, \tag{8}$$
$$B_2H_{\mathbf{U}}=0. \tag{9}$$

Considering the particular case $Z=\sigma^{-1}\mathbf{V}$ and $\Sigma=I_{n-1}$, for $A_1=H_{\mathbf{U}},b_1=0$ and $c_1=0$, we have $y_1=\sigma^{-2}\mathbf{V}'H_{\mathbf{U}}\mathbf{V}$ and for $A_2=0,b_2=\frac{1}{2}B_1'$ and $c_2=0$, we have $y_2=\sigma^{-1}B_1\mathbf{V}$. Then, according to the corollary, a necessary and sufficient condition for y_1 and y_2 to be independently distributed is $H_{\mathbf{U}}B_1'=0$.

As $H_{\mathbf{U}}$ is symmetric, then $H_{\mathbf{U}}B_1'=(B_1H_{\mathbf{U}})'$, and using equation Eq. (8), we obtain that $H_{\mathbf{U}}B_1'=0$ and therefore y_1 and y_2 are independently distributed.

By applying repeatedly the corollary for $A_2=0$, $b_2=\frac{1}{2}B_2'$ and $c_2=0$, from which $y_2=\sigma^{-1}B_2\mathbf{V}$, and then the condition of independence between y_1 and y_2 is $H_{\mathbf{U}}B_2'=0$ and by similar reasoning, using Eq. (9), we can show that $H_{\mathbf{U}}B_2'=0$ and therefore y_1 and y_2 are independently distributed.

Finally, we deduce that $\hat{\mathbf{a}}$ and $\hat{\sigma}^2$ are independently distributed.

3.2.3. Sufficiency and completeness

By subtracting and adding $\mathbf{U}'\hat{\mathbf{a}}$ to $\mathbf{V}-\mathbf{U}'\mathbf{a}$, the expression Eq. (5) becomes

$$\mathbb{L}_{\mathbf{V}}(\mathbf{a},\sigma^2)=\frac{1}{(2\pi\sigma^2)^{\frac{n-1}{2}}}\exp\left(-\frac{1}{2\sigma^2}[(n-1)\hat{\sigma}^2+(\hat{\mathbf{a}}-\mathbf{a})'\mathbf{U}\mathbf{U}'(\hat{\mathbf{a}}-\mathbf{a}))]\right),$$

which means that $(\hat{\mathbf{a}},\hat{\sigma}^2)$ is conjointly sufficient for $(\mathbf{a},\sigma^2)$.

The completeness follows by means of a similar reasoning to that established for the maximum likelihood estimators of the parameters of the multivariate normal distribution (see, for example, Anderson [33]).

And so the estimators $\hat{\mathbf{a}}$ and $\frac{(n-1)\hat{\sigma}^2}{(n-3)\sigma^2}$ are the UMVUE for $\mathbf{a},\sigma^2$, respectively.

3.3. Parameter confidence intervals

The γ% confidence interval for the parameter σ^2 is given, by

$$\left(\frac{(n-1)\hat{\sigma}^2}{\chi^2_{n-3,\frac{\gamma}{2}}},\frac{(n-1)\hat{\sigma}^2}{\chi^2_{n-3,1-\frac{\gamma}{2}}}\right) \tag{10}$$

and the γ% confidence interval for the parameter α is given, by

$$P\left(\alpha\in\left[\hat{\alpha}\pm\hat{\sigma}\left(\frac{n-1}{n-3}A_{11}F_{1,n-3,\gamma}\right)^{1/2}\right]\right)=1-\gamma, \tag{11}$$

where $\chi^2_{n,\gamma}$ and $F_{m,n,\gamma}$ are the upper 100γ per cent points of the χ^2 with n degrees of freedom and the F-distribution with m and n degrees of freedom, respectively, and A_{11} is the first elements of the principal diagonal of the matrix $(\mathbf{U}\mathbf{U}')^{-1}$ and where $\mathbf{U}\mathbf{U}'$ is

$$\mathbf{U}\mathbf{U}'=\begin{pmatrix}\sum_{i=2}^n\log^2(t_i/t_{i-1}) & \log(t_n/t_1)\\ \log(t_n/t_1) & n-1\end{pmatrix}.$$

According to the reparametrization $\mathbf{a}=(\alpha,-(\beta+\sigma^2/2))'$, it is not possible to obtain confidence interval for the parameter β.

3.4. Approximate confidence intervals of the estimated trend functions

On the one hand, from Zehna's theorem [34], we can obtain the estimated trend function (ETF) and the conditional trend function (ECTF) of the process, by substituting the parameters by their estimators, given in Eqs. (6) and (7), in expressions (3) and Eq. (4). Then the ECTF is given by

$$\widehat{\mathbb{E}}[x(t)\mid x(s)=x_s]=x_s\left(\frac{t}{s}\right)^{\hat{\alpha}}e^{-\hat{\beta}(t-s)}. \tag{12}$$

The ETF of the process is

$$\widehat{\mathbb{E}}[x(t)]=x_1\left(\frac{t}{t_1}\right)^{\hat{\alpha}}e^{-\hat{\beta}(t-t_1)}. \tag{13}$$

On the other hand, we still lack a confidence interval for the parameter β. It is possible to obtain approximate confidence intervals of the ETF and ECTF of the model; these consist in replacing in equations Eqs. (3 and 4) the parameter α by the extreme values of its confidence interval: the lower limit of α (α_{ll}) and the upper limit of $\alpha(\alpha_{ul})$ which are given in expression (11). For the parameter β, as we do not have a confidence interval for it, we assume the upper and lower limits of the parameter β to be equal to $\hat{\beta}$. Then, the lower limit of the ECTF (ECTF_{ll}) is given by

$$\widehat{\mathbb{E}}_{ll}[x(t)\mid x(s)=x_s]=x_s\left(\frac{t}{s}\right)^{\hat{\alpha}_{ll}}e^{-\hat{\beta}(t-s)} \tag{14}$$

and the upper limit of the ECTF (ECTF_{ul}) is

$$\widehat{\mathbb{E}}_{ul}[x(t)|x(s)=x_s] = x_s\left(\frac{t}{s}\right)^{\hat{\alpha}_{ul}} e^{-\hat{\beta}(t-s)}. \tag{15}$$

In the same way, we obtain the lower limit of the ETF (ETF_{ll}) and this latter has the following form

$$\widehat{\mathbb{E}}_{ll}[x(t)] = x_1\left(\frac{t}{t_1}\right)^{\hat{\alpha}_{ll}} e^{-\hat{\beta}(t-t_1)} \tag{16}$$

and the lower limit of the ETF (ETF_{ul}) is

$$\widehat{\mathbb{E}}_{ul}[x(t)] = x_1\left(\frac{t}{t_1}\right)^{\hat{\alpha}_{ul}} e^{-\hat{\beta}(t-t_1)}. \tag{17}$$

These functions are used in the last Section to fit and predict the future evolution of the stochastic diffusion process under consideration.

4. Application to modelling the total stock of private cars-petrol in Spain

The land transport sector is an important source of emission of greenhouse gases, and in particular of CO_2, which is why this sector is currently the object of particular attention with respect to, by example, the dynamic evolution of its size (total number of units), the energy consumption corresponding to the type of fuel utilized, and the fiscal policy corresponding to the level of greenhouse gas emissions, among other aspects. Such studies are particularly relevant taking into account the severe problems currently being addressed (especially as of the signing of the Kyoto Protocol in 1992) with respect to the environment and climate change.

Numerous studies have been made of the land transport sector, normally concerning specific countries or geopolitical areas (such as the EU, the USA or the UK). These studies are mainly of an econometric nature, seeking to explain the evolution of the sector or of given subsectors by means of economic variables (for example, GDP, prices, energy costs).

At present, Spain constitutes a flagrant example of violation of the maximum annual quantities of CO_2 assigned under the Kyoto agreement (see Gutiérrez et al. [17]), and studies are being made of the land transport sector using a modelling procedure based on stochastic diffusions (for example, Gutiérrez et al. [24]). In this line of investigation, the present study uses gamma diffusion to model the total stock of private cars-petrol in Spain. This is a subsector that makes a notable contribution to the total emission of CO_2 in Spain and to that made by the land transport sector, in particular.

This modelling is based on fitting the gamma diffusion process examined in the present study. Statistical fitting is performed by applying statistical inference (estimating the parameters and trend functions, and testing different hypotheses) as described in Section 3, and on the basis of real observations corresponding to the period 1986–2006. These data consist of the total stock of private cars-petrol in Spain, at 31 December of each year, and are published as official statistics in the National Transport Yearbook of the Spanish Ministry of Finance; they can be consulted at http:www.ine.es. These data are expressed in 10^9 cars-petrol and are included in Table 2.

The process to be modelled corresponds to the stochastic dynamic variable $X(t)$, which for each instant of time t takes the value of the total stock of private cars-petrol in Spain during the annual period that ends at this instant of time. The above-mentioned real observed values constitute a set of observations of $X(t)$ in discrete time at the end of each natural year.

The first 19 data of the above time series (i.e. up to 2004) are employed to estimate the parameters of the process, using the methods described in Section 3. The data corresponding to 2005 and 2006 are reserved for later comparison with the value predicted using the fitted SGDP.

The trend functions (ETF and ECTF) given by Eqs. (12) and (13) are fitted to their confidence intervals given by Eqs. (14)–(17). Prior to this, we calculate the estimators of the parameters, given by Eqs. (6) and (7) and their confidence intervals given by Eqs. (10) and (11). The related results are summarized in Table 1.

Table 2 summarizes the fit and the forecast obtained using the ETF with its ETF_{ll} and ETF_{ul}. Table 3 summarizes the fit and the forecast obtained using the ECTF with its $ECTF_{ll}$ and $ECTF_{ul}$.

Fig. 2 shows the pattern of the real data with respect to the fit and the forecast obtained using the ECT with ETF_{ll} and ETF_{ul}. Fig. 3 shows the fit and the forecast obtained using the ECTT of the model with respect to the real data.

All calculations were performed using MATLAB 7.0.1 mathematical software.

Remarks

In the trend analysis methodology based on fitting stochastic diffusion processes that the present authors have been developing in recent years (see References), the ECTF has been used as an instrument for detecting and analyzing possible trend changes. Indeed, a significant discrepancy between the fitted values and those predicted by the ETF and the ECTF would imply an instability in the trend. It is noteworthy that the ECTF values incorporate, step-by-step, the values observed after the fitting process. Therefore, and to the extent that, in fact, the ECTF values do not significantly improve those given by the ETF, we confirm the goodness of the fit and forecast. Such is the case in the present study.

5. Comments and conclusions

(1) Specifically, the statistically fitted model is expressed as follows:

Table 1
Estimators and their confidence intervals

	Estimator	Confidence interval
$\hat{\alpha}$	17245.122886	(17238.198158; 17252.047614)
$\hat{\sigma}^2$	0.001287	(0.000803; 0.003354)
$\hat{\beta}$	8.627309	–

Table 2
Real data vs. ETF, ETF_l and ETF_u

	Year	Real data	ETF	ETF_{low}	ETF_{upp}
	1986	0.888525	0.888525	0.888525	0.888525
	1987	0.934703	0.937685	0.934422	0.940959
	1988	0.980033	0.985251	0.978408	0.992142
	1989	1.036063	1.030723	1.020006	1.041553
	1990	1.077489	1.073604	1.058749	1.088667
	1991	1.121962	1.113409	1.094190	1.132965
	1992	1.164066	1.149678	1.125911	1.173946
	1993	1.183863	1.181980	1.153529	1.211132
	1994	1.192754	1.209925	1.176707	1.244080
	1995	1.215313	1.233171	1.195158	1.272392
	1996	1.236245	1.251428	1.208652	1.295720
	1997	1.249061	1.264472	1.217021	1.313774
	1998	1.268121	1.272139	1.220163	1.326329
	1999	1.280297	1.274335	1.218041	1.333230
	2000	1.274697	1.271037	1.210689	1.334394
	2001	1.279573	1.262295	1.198207	1.329810
	2002	1.272871	1.248224	1.180759	1.319545
	2003	1.209587	1.229011	1.158571	1.303735
	2004	1.203509	1.204904	1.131926	1.282587
Prediction	2005	1.181565	1.176208	1.101157	1.256373
Prediction	2006	1.136270	1.143280	1.066641	1.225425

Please cite this article in press as: Gutiérrez R et al., The trend of the total stock of the private car-petrol in Spain: ..., Appl Energ (2008), doi:10.1016/j.apenergy.2008.03.016

Table 3
Real data vs. ECTF $ECTF_l$ and $ECTF_u$

	Year	Real data	ECTF	$ECTF_{low}$	$ECTF_{upp}$
	1986	0.888525	0.888525	0.888525	0.888525
	1987	0.934703	0.937685	0.934422	0.940959
	1988	0.980033	0.982182	0.978702	0.985546
	1989	1.036063	1.025264	1.021700	1.028841
	1990	1.077489	1.079166	1.075416	1.082928
	1991	1.121962	1.117438	1.113558	1.121333
	1992	1.164066	1.158509	1.154488	1.162545
	1993	1.183863	1.196772	1.192620	1.200939
	1994	1.192754	1.211853	1.207651	1.216070
	1995	1.215313	1.215670	1.211457	1.219898
	1996	1.236245	1.233307	1.229035	1.237594
	1997	1.249061	1.249130	1.244805	1.253470
	1998	1.268121	1.256634	1.252285	1.260998
	1999	1.280297	1.270310	1.265916	1.274719
	2000	1.274697	1.276984	1.272569	1.281414
	2001	1.279573	1.265929	1.261555	1.270319
	2002	1.272871	1.265310	1.260940	1.269695
	2003	1.209587	1.253279	1.248952	1.257620
	2004	1.203509	1.185860	1.181769	1.189966
Prediction	2005	1.181565	1.174846	1.170795	1.178912
Prediction	2006	1.136270	1.148487	1.144528	1.152459

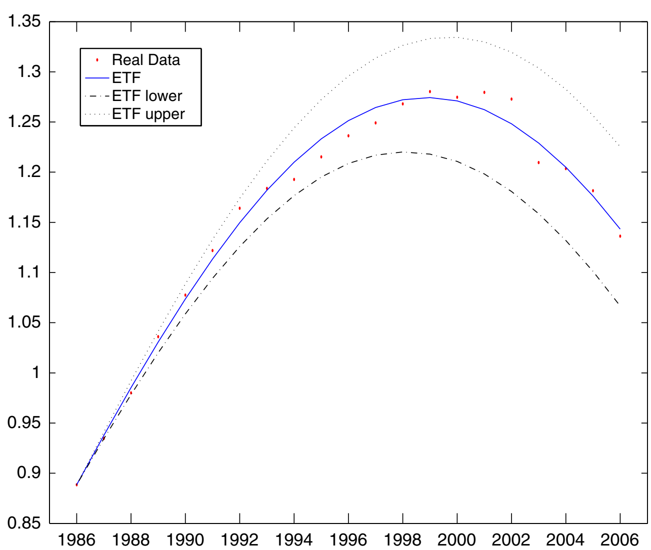

Fig. 2. Real data vs. ETF.

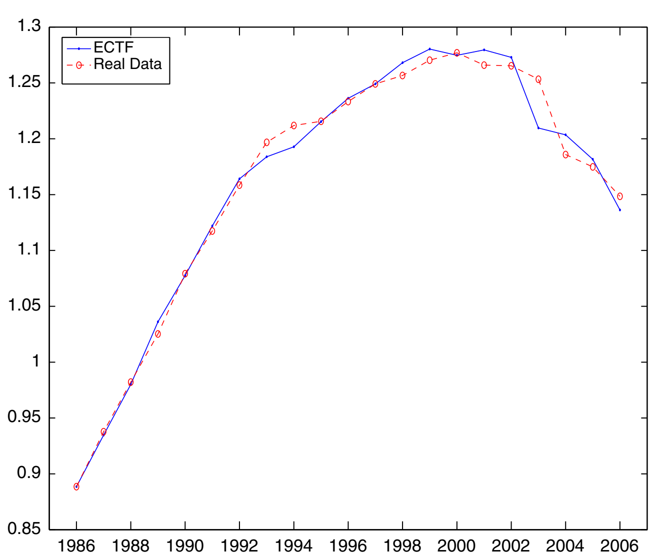

Fig. 3. Real data vs. ECTF.

$$\mathrm{d}x(t) = \left(\frac{17245.122}{t} - 8.627\right)x(t)\mathrm{d}t + 0.035x(t)\mathrm{d}w(t).$$

This is obtained from Eq. (1), substituting the estimators of the parameters (see Table 1). Note that the slowdown estimated by the fitted model for the total stock of private petrol-driven cars in Spain (2000–2006), measured in terms of the estimated value of the coefficient $\hat{\beta}$ has a value of 8.627.

(2) From a theoretical point of view, we conclude that the gamma process presented, which is of a non homogeneous nature, is such that we can explicitly establish its probability transition density function in terms of a lognormal distribution Eq. (2) together with its moment functions, and in particular its trend functions Eq. (4). We can also establish parameter estimation results using the maximum likelihood method and construct approximated confidence intervals, on the basis of discrete sampling. Therefore, the gamma process we describe is accompanied by a set of statistical results that enable it to be applied to real data.

(3) Concerning the developed application, we conclude that the gamma model proposed is adequate both for modelling the phenomenon under consideration and for forecasting purposes. In particular, the sharp variation in the trend of the total stocks of private cars-petrol in Spain that can be observed between 2000 and 2006 (See Fig. 1) is well modelled by the gamma model utilized. The variation over the period 1986–1999 is of an exponential nature, and so the stochastic model that is appropriate for this period can be based on diffusion processes of an increasing exponential trend function, such as lognormal or even Gompertz diffusion. As remarked in Section 1, the evolution of the total stocks of private cars (1986–2006) is clearly different in the subsectors "petrol" and "diesel". In the latter case, the evolution of the trend is exponential, while in the case of petrol-driven cars, the model used must reflect the change by means of a non-exponential trend diffusion. Thus, the conclusion to be drawn from the present study is that the gamma diffusion model is capable of representing the total period under consideration (1986–2006), including the above-mentioned change of trend.
With regard to the "forecasting capacity" of the proposed gamma model, we may conclude that it is valid for short and medium-term forecasts of the future evolution of the variable $x(t)$ under consideration. Furthermore, the statistical inference established in this study for the gamma process enables us to evaluate the precision of the forecasts made. Note that the real value for the year 2005 and 2006 (which were not used for the statistical fit, see Section 3) are clearly included in the interval $[1.101157; 1.256373] \times 10^9$ and $[1.066641; 1.225425] \times 10^9$ provided by the 95% confidence interval of the trend function for $t = 2005$ and $t = 2006$, calculated by Eq. (16) and (17). The prediction value for the year 2005, by using the ETF Eq. (13) is 1.176208×10^9 and for 2006 is 1.143280×10^9.

(4) The stochastic model proposed, based on the new stochastic gamma diffusion process, as described in Section 2, and which we have analyzed in probabilistic and statistical terms (Section 3) is of a non homogeneous nature. This non-homogeneity is "intrinsic" to the process, in the sense that it is not defined on the basis of a prior homogeneous version that is subsequently extended to a non homogeneous one, but, rather, it is non homogeneous from the beginning (see comments in Section 1). Therefore, an open question that remains is that of studying a version that might be termed "doubly non homogeneous" of the gamma process presented in this paper, which would be obtained

with the introduction of "exogenous factors" (of the type examined in Section 1) in the drift coefficient given by Eq. (1). This methodology has been used as a means of obtaining non homogeneous versions on the basis of prior homogeneous ones in various types of diffusions (see, for example, Gutiérrez et al. [35]). In this latter situation, when the statistical fits are performed on the models, it is necessary to possess sampling information on these "exogenous factors". It is to be hoped that when this new type of gamma process is studied it can be used to obtain improvements in the statistical fits for real cases, such as the one considered in this paper, that of the total stock of private cars-petrol in Spain. For this purpose, we will have to locate the "exogenous factors" that are assumed to influence the real pattern of the trend function, and that are capable of incorporating into the global model the "irregularities" (the discrepancies between the observed values and those predicted by the estimated trend) that can be observed when the SGDP is used, as shown in the present paper – for example, explaining the sharp change in the trend of the total stock of private cars-petrol between 2000 and 2006 in terms of the evolution of certain exogenous factors such as prices and/or fiscal measures with respect to the quantity of CO_2 emissions.

Acknowledgement

The authors are very grateful to the Editor and Referees for constructive comments and suggestions. This work was supported partially by research project MTM 2005-09209, Ministerio de Educación y Ciencia, and project P06-FQM-02271, Junta de Andalucía, Spain.

References

[1] Skiadas CH. Methods of growth functions formulation. In: Gutiérrez R, Valderrama M, editors. Selected topics on stochastic modelling. World Scientific; 1994. p. 296–310.
[2] Katsamaki A, Skiadas CH. Analytic solution and estimation of parameters on a stochastic exponential model for a technological diffusion process application. Appl Stoch Mod Data Anal 1995;11:59–75.
[3] Skiadas CH, Giovani AN. A stochastic bass innovation diffusion model for studying the growth of electricity consumption in Greece. Appl Stoch Mod Data Anal 1997;13:85–101.
[4] Giovanis AN, Skiadas CH. A stochastic logistic innovation diffusion model studying the electricity consumption in Greece and the United States. Technol Forecast Soc Change 1999;61:253–64.
[5] Bibby BM, Sorensen M. Martingale estimation functions for discretely observed diffusion processes. Bernoulli 1995;1(1/2):17–39.
[6] Ait-Sahalia Y. Maximum-likelihood estimation of discretely-sampled diffusion: a closed-form approximation approach. Econometrica 2002;70:223–62.
[7] Durham GB, Gallant AR. Numerical techniques for maximum likelihood estimation of the continuos-times diffusion processes. J Bus Econ Stat 2002;20(3):297–316.
[8] Eugene MC. Maximum likelihood estimations of a class of one-dimensional stochastic differential equation models from discrete data. J Time Ser Anal 2000;22(5):505–15.
[9] Prakasa Rao BSL. In: Arnold, editor. Statistical inference for diffusion type process. New York: London and Oxford University Press; 1999.
[10] Kutoyants YA. Statistical inference for Ergodic diffusion processes. Springer series in statistics. London: Springer-Verlag; 2004.
[11] Chan KC, Karolyi GA, Longstaff FA, Sanders AB. An empirical comparison of alternative models of the short-term interest rate. J Financ 1992;3:1209–27.
[12] Arapis M, Gao J. Empirical comparison in short-term interest rate models using nonparametric methods. J Financ Econ 2006;4(2):310–45.
[13] Elerian O, Chib S, Shephad N. Likelihood inference for discretely observed nonlinear diffusions. Econometrica 2006;69:959–93.
[14] Gutiérrez R, Román P, Torres F. Inference and first-passage-time for the lognormal diffusion process with exogenous factors: application to modelling in economics. Appl Stoch Mod Bus Ind 1999;15(4):325–32.
[15] Gutiérrez R, Román P, Torres F. Inference on some parametric functions in the univariate lognormal diffusion process with exogenous factors. Test 2001;10(2):357–73.
[16] Gutiérrez R, Gutiérrez-Sánchez R, Nafidi A. Forecasting total natural-gas consumption in Spain by using the stochastic Gompertz innovation diffusion model. Appl Energ 2005;80(2):115–24.
[17] Gutiérrez R, Gutiérrez-Sánchez R, Nafidi A. Trend analysis using nonhomogeneous stochastic diffusion processes. emission of CO_2; Kyoto protocol in Spain. Stochc Environ Res Risk Assess 2008;22:55–66.
[18] Gutiérrez R, Nafidi A, Gutiérrez-Sánchez R, Melliani S. A non-homogenous stochastic Vasicek diffusion model: application to global CO_2 emission in Marocco. In: International conference on approximation methods and numerical modelling in environment and natural research; 2007. p. 323–9
[19] Gutiérrez R, Rico N, Román P, Romero D, Serrano JJ, Torres F. Approximating the nonhomogenous lognormal diffusion process via polynomial exogenous factors. Cybernet Syst 2006;37(4):293–309.
[20] Ferrante L, Bompade S, Possati L, Leone L, Montanari MP. A stochastic formulation of the Gompertzian growth model for in vitro bactericidad kinetics: parameter estimation and extinction probability. Biometrical J 2005;47(3):306–18.
[21] Albano G, Giorno V. A stochastic model in tumor growth. J Theor Biol 2006;242(2):329–36.
[22] Gutiérrez R, Román P, Romero D, Serrano JJ, Torres F. A new gompertz-type diffusion process with application to random growth. Math Biosci 2007;208:147–65.
[23] Gutiérrez R, Angulo JM, González A, Pérez R. Inference in lognormal multidimensional diffusion processes with exogenous factors: application to modelling in economics. Appl Stoch Mod Data Anal 1991;7:295–316.
[24] Gutiérrez R, Gutiérrez-Sánchez R, Nafidi A. A bivariate stochastic Gompertz diffusion model: statistical aspects and application to the joint modelling of the gross domestic product and CO_2 emissions in Spain. Environmetrics [in press]. doi:10.1002/env.906.
[25] Gutiérrez R, Gutiérrez-Sánchez R, Nafidi A. The stochastic Rayleigh diffusion model: statistical inference and computational aspects. Applications to modelling of real cases. Appl Math Comput 2006;175:628–44.
[26] Gutiérrez R, Gutiérrez-Sánchez R, Nafidi A, Ramos E. A diffusion model with cubic drift: statistical and computational aspects and application to modelling the global CO_2 emission in Spain. Environmetrics 2007;18:55–69.
[27] Gutiérrez R, Gutiérrez-Sánchez R, Nafidi A [Emissions of the greenhouse gases attributable to the activities of the land transport: modelling and analysis using I-CIR stochastic diffusion. The case of Spain]. Environmetrics 2008;19:137–61.
[28] Ricciardi LM. Diffusion processes and related topics in biology. Lecture notes in biomathematics. Berlin: Springer-Verlag; 1977.
[29] Liptser RS, Shiryayev AN. Statistics of random processes I. New York: Springer-Verlag; 1977.
[30] Wong E, Hajek B. Stochastic processes in engineering systems. New York; 1985.
[31] Tintner G, Sengupta JK. Stochastic economics. Academic Press; 1972.
[32] Srivastava MS, Khatri CG. An introduction to multivariate analysis. New York: North-Holland Scientific Publishers; 1979.
[33] Anderson TW. An introduction to multivariate statistical analysis. 2nd ed. New York: Wiley; 1984.
[34] Zehna PW. Invariance of maximum likelihood estimators. Ann Math Stat 1966;37:744.
[35] Gutiérrez R, Gutiérrez-Sánchez R, Nafidi A. Electricity consumption in Morocco: stochastic Gompertz exogenous factors diffusion analysis. Appl Energ 2006;83:1139–51.