Stoch Environ Res Risk Assess (2008) 22:57–66
DOI 10.1007/s00477-006-0097-7

ORIGINAL PAPER

Trend analysis using nonhomogeneous stochastic diffusion processes. Emission of CO_2; Kyoto protocol in Spain

R. Gutiérrez · R. Gutiérrez-Sánchez · A. Nafidi

Published online: 19 December 2006

Abstract In this study, we propose a methodology to analyse the gradual secular trends present in the time evolution of certain endogenous variables, which are of particular interest in environmental research. This methodology is based on modelling such variables by nonhomogeneous stochastic diffusion processes, the trend functions of which may be made to depend on other, exogenous, variables, which are controllable and which affect and model, in turn, the possible irregularities of such trends. The methodology is applied to analyse the evolution of the emission of CO_2 in Spain, and it is shown that the evolution of the Spanish GDP affects the trend component. These circumstances are considered in the context of Spain's non-compliance with the Kyoto protocol on controlling the emission of greenhouse gases.

Keywords Non-homogeneous lognormal and Gompertz diffusion process · Exogenous factors · Trend functions · Likelihood estimation in discrete sampling · GDP and global CO_2 emission

1 Introduction: an alternative methodology in secular trend analysis

The problem of analysing the secular trends present in time series, taken as a tendency to increase (or decrease) continuously for an extended period of time in a systematic manner, is a common problem in many phenomena of interest in environmental science (see, for example, McCuen 2003), If the secular trend is of a gradual type, i.e. if it is present over a large proportion of the time considered in the observation of the phenomenon, it is normally modelled in time series theory by using deterministic models, generally nonlinear, of logistic, Gompertz, Bass or lognormal growth curves, among other possibilities. In this deterministic, traditional analysis of trends, non-statistical fitting techniques are used and it is a complex business to analyse the possible influences of other exogenous variables on such trends, and to achieve a global model of the behaviour of the endogenous variable being studied by means of a chronological series that will, at the same time, include its own secular trend and the possible exogenous variables mentioned above. In the present paper, we propose a statistical methodology for the joint analysis of the variable and the trend, which will enable us to consider exogenous variables influencing such trends and to make forecasts using statistical criteria concerning their evolution over time. The proposed methodology is based on the examination of stochastic diffusion models with a time-continuous increasing (or decreasing) trend function. Thus, trends are modelled by means of the global modelling of the endogenous variable under examination, such that the stochastic model of the latter variable has a trend function that statistically fits the secular trend that is present in the real data observed.

In particular, in this study we use models based on stochastic diffusion processes, generally nonhomogeneous, of the lognormal and Gompertz types (NHSLDP and NHSGDP, respectively). These models have been thoroughly examined, and applied in the

R. Gutiérrez (✉) · R. Gutiérrez-Sánchez · A. Nafidi
Department of Statistics and Operations Research,
Facultad de Ciencias, University of Granada,
Campus de Fuentenueva s/n, 18071 Granada, Spain
e-mail: rgjaimez@ugr.es

homogenous case; for example, the lognormal model is described by Gutiérrez et al. (1991, 1997) and (1999). The Gompertz model has recently been discussed in Gutiérrez et al. (2005a, c, 2006a) among others, with the problem being applied to real problems such as energy consumption and the price of new housing in Spain. These studies, together with their statistical methodology are cited in Albano and Giorno (2006), and Meade and Islam (2006).

2 The problem: emission of CO_2 and Kyoto protocol in Spain

Climate change is one of the most important problems the world is currently facing; in Spain, particularly, it has severe effects, including desertification, erosion, falling water table levels, rising temperatures (to 1.5° above the annual mean value of the last 30 years) and rising sea levels. To attempt to alleviate as far as possible the severe ecological, sanitary, economic and social disruptions that such a change will provoke, the UN promoted the Kyoto Protocol (1997) at a global level, with special reference to the industrialised countries, concerning the control of greenhouse gases. The gases considered are those with a direct effect, i.e. carbon dioxide (CO_2), methane (CH_2), nitrous oxide (N_2O), hydrofluorocarbons (HFC), perfluorocarbons (PFC) and sulphur hexafluoride (SF_6), and those with indirect effects, such as carbon monoxide and nitrogen, sulphur and ammonia oxides. This protocol was ratified by Spain in February 2005; according to its provisions, the EU must reduce its global emissions of greenhouse gases by 8% over the period 2008–2012 with respect to the values corresponding to 1990, this change being apportioned among the Member States such that Spain is allocated an increase in emissions of 15% over the same period. At present, Spain's situation with respect to these provisions gives cause for concern from several standpoints, which may be summarised as follows:

- The emission of greenhouse gases in Spain is mainly comprised of CO_2 (83.67%), followed by methane and nitrous oxide.
- Over the period 1990–2006, the emission of greenhouse gases has increased in Spain by 52.88%, well above the level permitted by the EU (15%) for the period 2008–2012. Nevertheless, Spain would be satisfied (according to official information provided by the Ministry of the Environment) to achieve a final increase of 24% in the level of emissions by the end of the latter period, which would represent 9 percentage points more than the value allocated and 29 percentage points less than what is being emitted at present (in 2006).
- Spain is the industrialised country that has most increased its emission of greenhouse gases, mainly because its model of economic growth is based on sectors that emit large amounts of CO_2. In Spain, the main producers of CO_2 are the energy sector, which is responsible for 78% [electricity generation (30.5%), road transport (28%), oil refining, industrial consumption of electricity, air transport, household consumption and services], livestock farming and agriculture (11%), industrial processes (7.65%) and waste materials (2.9%).
- Another significant factor underlying the serious situation in Spain is the emission associated with the housing sector. In Spain, 10 times more new housing is constructed than in the United Kingdom, and the same amount as in Germany, France and Italy combined. In the period 1990–2000, the area of land granted planning permission for the construction of housing increased by 25%.
- Among other consequences, from the economic standpoint it is now foreseeable that Spain will have to acquire CO_2 emission rights (as regulated in the Kyoto Protocol and in subsequent agreements) from 2008, as the current level of total emissions (53% higher than in 1990) would be equivalent to some 6,000 million Euros.

Faced with this situation, Spain is seeking to improve its index of "energy intensity" (the amount of energy required to produce one point of GDP), but in this respect, too, the state of affairs is alarming. Since 1990, this index has increased by 0.5% per year, while in the EU as a whole it has decreased by 1.3% per year. Attempts are being made to increase the production of renewable energy (thus, in 2004, wind energy supplied 7.78% of the total demand in Spain, overtaking hydroelectric energy). These kinds of policies will help to reduce the emissions produced by the energy sector. Another important sector in which action must be taken is that of transport; this must be reoriented, as in 2010 it will represent 40% of all emissions of greenhouse gases.

According to official EU data, the Spanish GDP is growing at a notable, sustained rate, 25% over the last 10 years, well above the mean value for Europe. The question being asked today in Spanish society is whether this marked rise in the GDP has been achieved (and continues to be achieved) in a way that is closely related to worsening environmental conditions. In this paper, we address the analysis of CO_2 emission, which as remarked above is the most significant factor in the global emission of greenhouse gases in Spain, in relation to the behaviour pattern of the GDP.

The Gompertz and lognormal, homogeneous and nonhomogeneous, are applied in the present study to analyse the trend of the variable emission of CO_2 in Spain, based on observations of the behaviour of this variable over the period 1986–2003, taking the Spanish GDP as an exogenous variable that affects the emission of this gas.

Finally, we remark that, alternatively, the analysis of the emission of CO_2 in Spain has also been addressed by Gutiérrez et al. (2006c). In the latter study, a cubic-type diffusion process was used. This non-Gompertz model produces good results in the study of the variable in question, but does not allow us to introduce exogenous variables into the study of the trend, unlike the models presented in the present paper. Let us note, too that the nonhomogeneous lognormal models that are considered here have been extended to 2D random fields in Gutiérrez et al. (2005d). Furthermore, homogeneous 1D Gompertz diffusion processes have been extended and statistically applied to the case of the existence of a threshold parameter, a case which is of potential interest in environmental science, in Gutiérrez et al. (2006d).

In the research into climate change carried out in recent years, many studies have focused on the relation between economic growth and CO_2 emission or energy consumption, separately. Recently, in Ramanathan (2006), the linkages between CO_2 emission, the growth in GDP and energy consumption have been studied simultaneously by using Data Envelopment Analysis (DEA). These studies, aimed fundamentally at defining different indices, suffer from one drawback, in general, and this is that they are carried out on a world-wide scale, while the conclusions at this macro level are of little interest to specific regions or countries. In no case were stochastic processes used in these macroeconomic approximations to the problem.

3 The proposed models: probabilistic characteristics and statistical methodology

The models proposed in this paper are the 1D NHSGDP and NHSLDP processes with one exogenous factor; these can be considered equivalent to a $\{X_t;\ t_0 \le t \le T;\ t_0 \ge 0\}$ process with values in $(0, \infty)$, which is the solution to the following Itô's stochastic differential equation (SDE):

$$\begin{aligned} &\mathrm{d}X_t = [(\alpha_0 + \alpha_1 g(t))X_t - \beta X_t \log(X_t)]\mathrm{d}t + \sigma X_t\,\mathrm{d}W_t \\ &P(X_{t_0} = x_{t_0}) = 1; \quad t \in [t_0, T] \end{aligned} \tag{1}$$

where α_0, α_1,β and $\sigma > 0$ are real parameters (to be estimated), g is a time continuous function in $[t_0,\ T]$, $x_{t_0} > 0$ is a real fixed number and W_t is the standard Wiener process. It can be proved that the functionals

$$A(t,x) = [\alpha_0 + \alpha_1 g(t)]x - \beta x \log(x) \quad \text{and} \quad b(t,x) = \sigma x \tag{2}$$

are nonanticipatives and satisfy the Lipschitz and the growth conditions and consequently, it exists a unique strong solution of the Eq. 1 [see for example, Liptser and Shiryayev (1977), Theorem 4.6].

Furthermore, it can be easily proved that these functionals are Borel measurable and satisfy the uniform Lipschitz condition and the γ-Holder, and particular Holder of order 1, conditions (see, for example, Wong and Hajek 1985, Propositions 4.1 and 7.1]. Consequently, there exists a separable, measurable and almost surely (a.s.) sample continuous diffusion process $\{X_t;\ t\in[t_0,\ T]\}$ which is the unique (a.s.) solution of the Itô differential equation (Eq. 1) with drift parameter and diffusion coefficients given, respectively, by

$$A(t,x) = [\alpha_0 + \alpha_1 g(t)]x - \beta x \log(x) \quad \text{and} \quad B(t,x) = \sigma^2 x^2 \tag{3}$$

and with the corresponding backward and forward differential Kolmogorov equations.

This unique solution can be obtained by application of Itô's formula to the time dependent transformation $Y_t = \mathrm{e}^{\beta t}\log(X_t)$; thus we obtain the following SDE:

$$\mathrm{d}Y_t = (\alpha_0 - \sigma^2/2 + \alpha_1 g(t))\mathrm{e}^{\beta t}\mathrm{d}t + \sigma \mathrm{e}^{\beta t}\mathrm{d}W_t \tag{4}$$

the solution to which is

$$Y_t = Y_{t_0} + \int_{t_0}^{t} [(\alpha_0 - \sigma^2/2) + \alpha_1 g(\tau)]\mathrm{e}^{\beta\tau}\mathrm{d}\tau + \sigma \int_{t_0}^{t} \mathrm{e}^{\beta\tau}\mathrm{d}W_\tau \tag{5}$$

from which we obtain the explicit solution to our original SDE; the corresponding expression for the Markov type diffusion process is

$$\begin{aligned} X_t = \exp\Bigg(&\mathrm{e}^{-\beta(t-t_0)}\log(x_{t_0}) + \int_{t_0}^{t} [(\alpha_0 - \sigma^2/2) \\ &+ \alpha_1 g(\tau)]\mathrm{e}^{-\beta(t-\tau)}\mathrm{d}\tau\Bigg) \exp\left(\sigma \int_{t_0}^{t} \mathrm{e}^{-\beta(t-\tau)}\mathrm{d}W_\tau\right) \end{aligned} \tag{6}$$

The stochastic diffusion process $\{X_t;\ t_0 \leq t \leq T\}$ thus obtained is the nonhomogeneous Gompertz diffusion process with the exogenous factor $g(t)$, when $\beta \neq 0$. When $\beta = 0$, this process contains as a particular case the univariate nonhomogeneous lognormal diffusion process, with exogenous factor $g(t)$; this process has been studied in depth in Gutiérrez et al. (1997b) and in Gutiérrez et al. (2001). Note that in Eq. 1 the parameter β, which affects the term $-X_t \log(X_t)$, when this is non null, acts to slow down the drift parameter of the respective lognormal process, the drift parameter of which is of the type $(\alpha_0 + \alpha_1\ g(t))X_t$.

Let us now establish the probabilistic and statistical inference characteristics in the case of the ($\beta \neq 0$) NHSGDP model. The respective characteristics of the NHSLDP model can be obtained from the corresponding limit when the β parameter tends to zero.

3.1 Study of the NHSGDP

3.1.1 Basic probabilistic characteristics of the model

Taking into account that the random variable $\int_{t_0}^{t} e^{-\beta(t-\tau)} dW_\tau$ in Eq. 6 has a 1D normal distribution $\mathcal{N}_1(0, \int_{t_0}^{t} e^{-2\beta(t-\tau)} d\tau)$, we conclude that the distribution of the random variable $X_t/X_s = x_s$ is a 1D lognormal distribution $\Lambda(\mu(s,t,x_s),\sigma^2\lambda^2(s,t))$, where

$$\begin{aligned}\mu(s,t,x_s) = &\ e^{-\beta(t-s)}\log(x_s) + \frac{\alpha_0 - \sigma^2/2}{\beta}(1 - e^{-\beta(t-s)}) \\ &+ \alpha_1 \int_s^t g(\tau) e^{-\beta(t-\tau)} d\tau \\ \lambda^2(s,t) = &\ \frac{1}{2\beta}(1 - e^{-2\beta(t-s)})\end{aligned} \tag{7}$$

Therefore, the probability density function (p.d.f) of the process is

$$f(y,t \mid x,s) = [2\pi\sigma^2\lambda^2(s,t)]^{-\frac{1}{2}} y^{-1} \times \exp\left\{-\frac{[\log(y) - \mu(s,t,x)]^2}{2\sigma^2\lambda^2(s,t)}\right\} \tag{8}$$

By the properties of the above distribution, the rth conditional moment of the process is expressed by

$$\mathbb{E}(X_t^r|X_s) = \exp\left(r\mu(s,t,X_s) + \frac{r}{2}\lambda^2(s,t)\right) \tag{9}$$

In other words,

$$\begin{aligned}\mathbb{E}(X_t^r|X_s) = \exp\Bigg(& re^{-\beta(t-s)}\log(X_s) + \frac{r(\alpha_0 - \sigma^2/2)}{\beta} \\ &\times (1 - e^{-\beta(t-s)}) + r\alpha_1 \int_s^t g(\tau) e^{-\beta(t-\tau)} d\tau \\ &+ \frac{r^2\sigma^2}{4\beta}(1 - e^{-2\beta(t-s)})\Bigg)\end{aligned} \tag{10}$$

from which the trend and the conditional trend functions ($r = 1$) follow.

3.1.2 Parameters estimation

As the distribution of the process transitions is known, 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. 8). Let us examine the following process sample, $x_1, ..., x_n$ at the instants $t_1, t_2, ..., t_n$. Assuming that the sample is taken at time intervals with a length equal to one, i.e: $t_j - t_{j-1} = 1$, then with the additional assumption on 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:

$$\mathbb{L}(x_1, \ldots, x_n, \alpha_0, \alpha_1, \beta, \sigma^2) = \prod_{i=2}^{n} f(x_i, t_i | x_{i-1}, t_{i-1}) \tag{11}$$

Let us now implement 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:

$$v_1 = x_1, v_{i,\beta} = \lambda_\beta^{-1}(\log(x_i) - e^{-\beta}\log(x_{i-1})) \quad \text{for } i = 2, \ldots, n$$

Then, the likelihood function can be written as

$$\mathbb{L}_{v_2,\ldots,v_n}(\mathbf{a}, \beta, \sigma^2) = \left[2\pi\sigma^2\lambda_\beta^2\right]^{-(n-1)/2} \times \exp\left(-\frac{1}{2\sigma^2}(\mathbf{v}_\beta - \mathbf{U}'_\beta\mathbf{a})'(\mathbf{v}_\beta - \mathbf{U}'_\beta\mathbf{a})\right) \tag{12}$$

where $\mathbf{a} = (\alpha_0 - \frac{\sigma^2}{2}, \alpha_1)'$, $\mathbf{v}_\beta = (v_{2,\beta},...,v_{n,\beta})'$, $\lambda_\beta = \lambda(t_{i-1}, t_i)$ and $\gamma_\beta = (1 - e^{-\beta})/\beta$, $\mathbf{U}_\beta$ is the $2 \times (n - 1)$ matrix, whose rank is assumed to be 2, given by $\mathbf{U}_\beta = (\mathbf{u}_{2,\beta},...,\mathbf{u}_{n,\beta})$, con $\mathbf{u}_{i,\beta} = \lambda_\beta^{-1}(\gamma_\beta, \int_{t_{i-1}}^{t_i} g(\tau)e^{-\beta(t_i-\tau)}d\tau)'$ para $i = 2,...,n$.

By applying the maximum likelihood method (deriving with respect to the parameters and equalling to zero), the likelihood equations for the estimators of $\mathbf{a}$ and σ^2 are

$$\mathbf{U}_\beta \mathbf{v}_\beta (\mathbf{U}_\beta \mathbf{U}'_\beta \mathbf{a}) = 0 \tag{13}$$

$$(\mathbf{v}_\beta - \mathbf{U}'_\beta \mathbf{a})'(\mathbf{v}_\beta - \mathbf{U}'_\beta \mathbf{a}) = (n-1)\sigma^2 \tag{14}$$

The third likelihood equation for the estimator of β is given by

$$(n-1)\lambda_\beta^{-1}\frac{\partial \lambda_\beta}{\partial \beta} + \left(\frac{\partial \mathbf{v}_\beta}{\partial \beta} - \frac{\partial \mathbf{U}'_\beta}{\partial \beta}\mathbf{a}\right)'(\mathbf{v}_\beta - \mathbf{U}'_\beta \mathbf{a}) = 0 \tag{15}$$

From the first two equations, Eqs. 13 and 14, we obtain the expressions for the estimators $\hat{\mathbf{a}}$ and $\hat{\sigma}^2$.

$$\hat{\mathbf{a}} = (\mathbf{U}_{\hat{\beta}} \mathbf{U}'_{\hat{\beta}})^{-1}\mathbf{U}_{\hat{\beta}} \mathbf{v}_{\hat{\beta}} \tag{16}$$

$$(n-1)\hat{\sigma}^2 = \mathbf{v}'_{\hat{\beta}} \mathbf{H}_{\mathbf{U},\hat{\beta}} \mathbf{v}_{\hat{\beta}} \tag{17}$$

Making use of the fact that $\mathbf{v}_\beta = \lambda_\beta^{-1}(\mathbf{I}_x - e^{-\beta}\mathbf{J}_x)$, where $\mathbf{I}_x = (\log(x_2),..., \log(x_n))'$ and $\mathbf{J}_x = (\log(x_1),..., \log(x_{n-1}))'$, and taking into account Eqs. 16 and 17, then after various operations that need not be detailed, the estimator $\hat{\beta}$ of the parameter β satisfies the following nonlinear equation:

$$\left(\lambda_\beta^{-1} e^{-\beta}\mathbf{J}'_x - \mathbf{v}'_\beta \mathbf{U}'_\beta (\mathbf{U}'_\beta \mathbf{U}_\beta)^{-1}\frac{\partial \mathbf{U}_\beta}{\partial \beta}\right)\mathbf{H}_{\mathbf{U},\beta}\mathbf{v}_\beta = 0 \tag{18}$$

where $\mathbf{H}_{\mathbf{U},\beta}$ denotes a symmetric and idempotent matrix, the expression for which is given by $\mathbf{H}_{\mathbf{U},\beta} = \mathbf{I}_{n-1} - \mathbf{U}'_\beta(\mathbf{U}_\beta \mathbf{U}'_\beta)^{-1}\mathbf{U}_\beta$.

3.2 Conclusions derived from the NHSLDP study

As mentioned above, the characteristics of the NHSLDP are obtained from the prior results derived for NHSGDP. Thus, by determining $\beta \to 0$, for example, in Eq. 10, the conditioned moment of order r of the NHSLDP is given by

$$\mathbb{E}(X_t^r/X_s) = X_s^r \exp\left(r(\alpha_0 + \frac{(r-1)}{2}\sigma^2)(t-s) + r\alpha_1 \int_s^t g(\tau)\mathrm{d}\tau\right) \tag{19}$$

The parameter estimators are obtained in the same way, by determining $\beta \to 0$ in Eqs. 16 and 17. The resulting estimators are

$$\hat{\mathbf{a}} = (\mathbf{U}\mathbf{U}')^{-1}\mathbf{U}\mathbf{v} \tag{20}$$

$$(n-1)\hat{\sigma}^2 = \mathbf{v}'\mathbf{H}_\mathbf{U}\mathbf{v} \tag{21}$$

In this case $\mathbf{v} = (v_2, ..., v_n)'$; $v_i = \log(x_i/x_{i-1})$; $u_i = (1, \int_{t_{i-1}}^{t_i} g(\tau)\mathrm{d}\tau)$ for $i = 2, ..., n$ and $\mathbf{U} = (u_2, ..., u_n)$; $\mathbf{H}_\mathbf{U} = \mathbf{I}_{n-1} - \mathbf{U}' (\mathbf{U}\,\mathbf{U}')^{-1}\mathbf{U}$.

4 Application to real case

4.1 Practical consideration and trend analysis

The exogenous factor $g(t)$ is a time-continuous function, but in practice it is normally known only for discrete time values (often, calendar years). In consequence, it must be constructed as a continuous function by means of approximation methods, for example as polygonal functions, which is the method we apply in this study. Normal practice is not to take the real value of the exogenous variable in each observation time, but rather the increment relative to the previous instant. Then, if we denote these values by y_i; $i = 2, ..., n$ with $y_1 = 0$, and assuming these variables $g(t)$ to be polygonal function in time in the intervals $[t_{i-1}, t_i]$ for $i = 2, ..., n$ and that they have the following form:

$$g(t) = y_{i-1} + (y_i - y_{i-1})(t - t_{i-1});\quad t_{i-1} \le t \le t_i \text{ and } i = 2, \ldots, n$$

Then, we obtain

$$\int_{t_{i-1}}^{t_i} g(\tau)e^{\beta(t_i-\tau)}\mathrm{d}\tau = \gamma_\beta\left(y_{i-1} + (y_i - y_{i-1})\frac{\beta - 1 + e^{-\beta}}{\beta(1 - e^{-\beta})}\right) \tag{22}$$

Then the columns of $\mathbf{U}_\beta$ are found to be $\mathbf{u}_{i,\beta} = \lambda_\beta^{-1}\gamma_\beta(1, z_i(\beta))'$; for $i = 2,..., n$, with $z_i(\beta) = y_{i-1} + \xi_\beta (y_i - y_{i-1})$ and $\xi_\beta = \frac{\beta-1+e^{-\beta}}{\beta(1-e^{-\beta})}$.

The trend analysis in this study is based on the trend functions of the model; this requires analysis to obtain fits and forecasts to real data by means of the estimated trend function (ETF) and the estimated conditional trend function (ECTF). The latter functions are obtained for the moment of the first order and the maximum likelihood estimators by using Zehna's theorem; for example, in the case of NHSGDP, the ECTF is obtained by taking $r = 1$ in Eq. 10 and replacing the parameters in it by their estimators as given in Eqs. 16, 17 and 18.

Note that from the hypothesis, we have taken for $i = 2, ..., n$: $t_i - t_{i-1} = 1$, and thus we have $t_i = (i - 1) + t_1$. This result is also valid if we have a sample of size N that is larger than n, and so we obtain $t_i = (i - 1) + t_1$ for $i = 2, ..., N$.

The ECTF of the NHSGDP is given by the following expression (for $i = 2, ...,N$)

$$\hat{\mathbb{E}}(X_{t_i}|x_{t_{i-1}}) = \exp\left(\mathrm{e}^{-\hat{\beta}}\log(x_{t_{i-1}}) + \frac{\hat{\alpha}_0 - \sigma^2/2}{\hat{\beta}}\left(1 - \mathrm{e}^{-\hat{\beta}}\right)\right) \times \exp\left(\frac{\hat{\sigma}^2}{4\hat{\beta}}\left(1 - \mathrm{e}^{-2\hat{\beta}}\right) + \hat{\alpha}_1\gamma_{\hat{\beta}}\left[\xi_{\hat{\beta}}y_i + (1 - \xi_{\hat{\beta}})y_{i-1}\right]\right) \quad (23)$$

and by the initial condition $P[\, X(t_1) = x_1] = 1$, the ETF of the NHSGDP is

$$\hat{\mathbb{E}}(X_{t_i}) = \exp\left(\log(x_1)\mathrm{e}^{-\hat{\beta}(t_i - t_1)} + \frac{\hat{\alpha}_0 - \sigma^2/2}{\hat{\beta}}\left(1 - \mathrm{e}^{-\hat{\beta}(t_i - t_1)}\right)\right) \times \exp\left(\frac{\hat{\sigma}^2}{4\hat{\beta}}(1 - \mathrm{e}^{-2\hat{\beta}(t_i - t_1)}) + \hat{\alpha}_1\gamma_{\hat{\beta}}\left[\xi_{\hat{\beta}}(y_i - y_1) + \sum_{k=2}^{i} y_{k-1}\right]\right) \quad (24)$$

In the case of the NHSLDP, the ECTF is

$$\hat{\mathbb{E}}(X_{t_i}/x_{t_{i-1}}) = x_{t_{i-1}}\exp\left(\hat{\alpha}_0 + \frac{\hat{\alpha}_1}{2}(y_i + y_{i-1})\right) \quad (25)$$

and the ETF, by assuming that $P[X(t_1) = x_1] = 1$ is

$$\hat{\mathbb{E}}(X_{t_i}) = x_1\exp\left(\hat{\alpha}_0(t_i - t_1) + \frac{\hat{\alpha}_1}{2}\sum_{k=2}^{i}(y_k + y_{k-1})\right) \quad (26)$$

These estimated functions are subsequently used to obtain the fit and the forecast of the variable being studied, taking into account the influence of the exogenous variable under consideration.

4.2 Application to the study of the global emission of CO_2 in Spain (1986–2003)

We shall now study the evolution of the trend corresponding to the global emission of CO_2 in Spain during the period 1986–2003, making use of the NHSGDP and NHSLDP nonhomogeneous diffusion processes described above. In general, the following methodology is used: for each year ($t = 1, 2,..., n$), we observe the values of the global emission of CO_2 in Spain, together with the values of the possible exogenous variables (such as the GDP) for the same years. The annual values of the endogenous variable (emission of CO_2) are fitted statistically by means of the value provided by the conditioned and nonconditioned trend functions of the fitted process, for the year in question, ($t_i = 1, 2,..., n$). This process is fitted by making a statistical estimate of its parameters on the basis of the above-mentioned observations of the endogenous and exogenous variables. With these estimated parameters, we also obtain the fitted trends. We now develop this methodology of trend analysis, which is carried out in two stages, as described below.

4.2.1 *Joint 2D analysis of the emission of CO_2 and the GDP*

In the first stage, we perform a joint fit of the emission of CO_2 and the Spanish GDP, for the period 1986–2003. The intention of this stage is to confirm that the behaviour of the GDP affects the emission of CO_2, and so by fitting the nonhomogeneous models NHSGDP and NHSLDP, which are one dimensional but with the possibility that an exogenous variable may be included, this GDP may really be considered an exogenous variable. Specifically, we calculate the dynamic correlation, statistically estimated, for each of the years (values of $t = 1, 2, ..., n$) between the CO_2 emission data and the annual GDP values. In order to do so, we must jointly model the two variables, using a 2D homogeneous stochastic process, from which we may explicitly identify the time-correlation function. Obviously, the 2D model chosen is of the lognormal, or Gompertz type, as the aim of this study (see the second stage for further details) is to fit the nonhomogeneous Gompertz and lognormal models described to the emission of CO_2, using the GDP as the exogenous variable. For the sake of example, we shall now describe the 2D lognormal case. To do so, let us consider the theoretical model of multidimensional lognormal diffusion, with the corresponding statistical inference, as established in Gutiérrez et al. (1991), with the computational aspects as developed in Gutiérrez et al. (1997a), and let us examine, in particular, the vector of variables ($X_1(t)$ = emission of CO_2; $X_2(t)$ = GDP), fitting a 2D lognormal diffusion (with no exogenous factors) to the data observed for the period (1986–2003). For such a model, we consider the correlation function as estimated by maximum likelihood on the basis of discrete sampling in time (as is the case in question) that is expressed as

$$R(\widehat{X_1(t_i); X_2(t_i)}) = (\mathrm{e}^{\hat{a}_{12}t_i} - 1)(\mathrm{e}^{\hat{a}_{11}t_i} - 1)^{-1/2}(\mathrm{e}^{\hat{a}_{22}t_i} - 1)^{-1/2} \quad (27)$$

In the above, $\hat{a}_{11}$, $\hat{a}_{22}$ and $\hat{a}_{12}$ are the maximum likelihood estimators of the theoretical parameters a_{11}, a_{22} and a_{12} that constitute the diffusion matrix of the process under consideration. Table 1 shows the values of the correlations fitted for each year. Note that throughout this period, the correlations remain, approximately, between 0.48 and 0.55. It can reasonably be concluded that a significant degree of covariation between the two variables is thus made apparent.

4.2.2 1D analysis of the global emission of CO_2, with the GDP as the exogenous factor

In order to carry out a more detailed analysis, let us adopt a finer viewpoint than in the above 2D analysis. We shall now consider the endogenous variable to be the $X_1(t)$ = emission of CO_2, and this is modelled using a 1D (Gompertz and lognormal) diffusion process, such that the GDP can be considered an exogenous factor that affects the dynamic behaviour of the $X_1(t)$. Appropriate theoretical models for this situation are the Gompertz or the lognormal non homogeneous processes, whose infinitesimal drift and trend are affected by the mentioned exogenous variable. The choice of the GDP, among other possible exogenous factors, is justified by the sufficiently significant covariation function previously obtained with this factor, which corroborates the evolution of the two variables between 1986 and 2003.

The data concerning these variables are annual values; GDP is expressed in billions of euros, and CO_2 emission is expressed in thousand millions of metric tons of carbon. The Spanish CO_2 data are taken from Boden et al. (1995), Andres et al. (1999) and Marland et al. (2005) and may be consulted at http://www.cdiac.esd.ornl.gov/ftp/ndp030/global.1751-2002.ems. The GDP data are from the Spanish Statistical Office (INE) and may be consulted at http://www.ine.es.

The methodology described in this study was applied to the cases of the NHSGDP and NHSLDP models. The data corresponding to the years 1986–2002 (t_i, para $i = 1, ..., n$) are used to estimate the parameters of the model, by means of Eqs. 16, 17 and 18 for NHSGDP and Eqs. 20 and 21 for NHSLDP.

The forecast value of CO_2 emission in the year 2003 is calculated from the estimated trend functions. In particular, we consider the value predicted by the conditioned trend function, calculating the forecasts for 2003 on the basis of the known 2002 values of the CO_2 emission and of the exogenous factor, the GDP for the same year. These predicted values are then compared with the real ones for the same period, values that were not used in the estimation phase.

The computational aspects of this study were resolved using Mathematica software, especially as concerns the nonlinear equation Eq. 18, which was solved using the Newton–Raphson method. By means of Eq. 22, the value of the estimator of the resulting β parameter is found to be $\hat{\beta} = 0.06235$, and the corresponding values for the other parameters in the case of the NHSGDP model are $\hat{\mathbf{a}} = (0.0924; 1.6835)'$ and $\hat{\sigma}^2 = 0.001497$, while for NHSLDP they are $\hat{\mathbf{a}} = (0.0153; 1.5621)'$ and $\hat{\sigma}^2 = 0.001470$.

Table 2 describes the results obtained for the fits and forecasts based on the ECTF and ETF functions for the two models in question.

Figure 1 shows the fits obtained by lognormal and Gompertz homogeneous 1D processes (with no exogenous factor), for the emission of CO_2.

Figure 2 shows the fits and forecasts obtained using the ETF of the NHSGDP and NHSLDP process, while Fig. 3 shows those obtained when the ECTF of the NHSGDP and NHSLDP process are used.

5 Conclusions and discussion

- The main aim of the present study, from a theoretical standpoint, is to show that the trends present in the evolution of certain growth phenomena, which are frequently to be encountered in environmental science, can be analysed by means of a statistical fit of the trend functions (conditioned or non-conditioned) of models based on stochastic diffusion processes to the observed data. With these considerations in mind, we present a methodology for trend analysis as an alternative to other statistical

Table 1 Correlation between emission of CO_2 and GDP

1986	1987	1988	1989	1990	1991	1992	1993	1994
0.000	0.5509	0.5507	0.5505	0.5503	0.55	0.5498	0.546	0.5447
1995	1996	1997	1998	1999	2000	2001	2002	2003
0.4709	0.4766	0.4712	0.4851	0.4883	0.491	0.5479	0.5477	0.55

Table 2 Fit and prediction of emission of CO_2

Years	CO_2 emission	GDP	ETFG	ETFL	ECTFG	ECTFL
1986	47.61	336.643	47.61	47.61	47.61	47.61
1987	48.57	355.317	49.3282	51.1565	49.3282	51.1565
1988	49.856	373.418	53.1697	54.5767	52.3713	51.8173
1989	56.841	391.443	56.8879	57.9811	53.4202	52.9659
1990	57.814	406.245	60.1413	60.5922	60.0956	59.4007
1991	59.096	416.588	62.3595	62.1011	60.0025	59.2538
1992	61.657	420.459	63.157	62.0481	59.9275	59.0455
1993	56.504	416.122	62.1513	60.1106	60.7128	59.7317
1994	59.09	426.036	61.9038	61.4491	56.4035	57.7623
1995	63.672	437.792	63.5438	63.1914	60.7247	60.7654
1996	63.732	448.456	65.2151	64.6529	65.3517	65.1446
1997	66.961	466.513	67.5725	67.8256	66.0782	66.8595
1998	68.645	486.742	71.022	71.5017	70.4035	70.5903
1999	75.23	506.849	74.6254	75.134	72.1933	72.1321
2000	77.099	527.613	78.1763	78.9082	78.8083	79.009
2001	77.47	542.166	80.9195	81.144	79.8427	79.2836
2002	82.998	556.651	82.7542	83.3295	79.3172	79.5566
2003	84.34	570.556	84.4126	85.3402	84.6716	85.0007

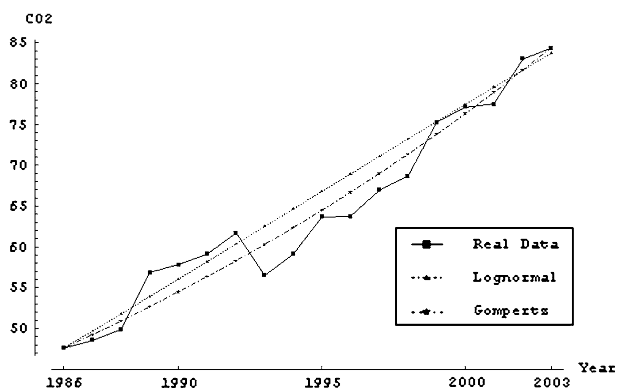

Fig. 1 Fits and forecasts using the Lognormal and Gompertz

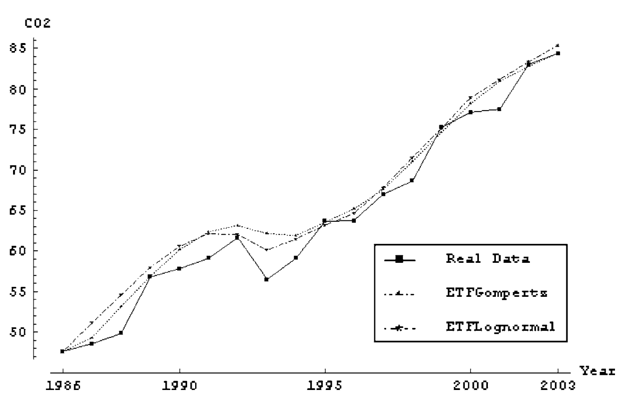

Fig. 2 Fits and forecasts using the ETF

techniques such as the analysis of secular trends of time series, dynamic regression or fitting the growth of deterministic curves.

- This study implements, in particular, two types of stochastic model, based on lognormal and

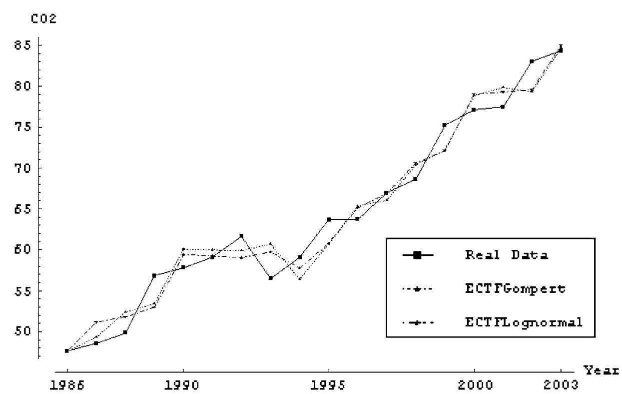

Fig. 3 Fits and forecasts using the ECTF

Gompertz stochastic diffusion processes, homogeneous or nonhomogeneous, which are found to be appropriate for fitting the dynamic variable ‘‘global emission of CO_2 in Spain‘‘. One of the possibilities presented by these models, moreover, is that they may be considered nonhomogeneous in time (the models NHSGDP and NHSLDP). The nonhomogeneity is introduced by making the drift of the diffusion depend on an exogenous variable (which is time continuous and ‘‘externally’’ controllable with respect to the endogenous variable being modelled). In consequence, the model’s trends (mean functions) will depend on this exogenous variable. The methodology proposed in this paper is implicitly utilised at a theoretical level in the nonhomogeneous Gompertz models discussed by Ferrante et al. (2000, 2005) and Albano and (Giorno 2006), in the context of modelling cellular and bacterial growth.

- As a result, we can introduce the possibility of analysing the behaviour of the trends of the model via the time evolution of the exogenous variable. In the particular case being analysed, we fit trends that depend on the GDP in Spain (the exogenous factor) to the variable "emission of CO_2".
- From a practical standpoint, and with reference to the real case being studied, we might conclude as follows: in general, nonhomogeneous lognormal and Gompertz models fitted to the observed data (see Table 2; Figs. 2, 3) are appropriate for describing the trend and for making medium-term forecasts. The two models were selected from among various based on different diffusion processes, such as lognormal, Gompertz, Brennan-Schwartz, Raleigh and logistic, which have been studied statistically and applied in Gutiérrez et al. (1999, 2005a, b, c, 2006a, b, c) and Giovanis and Skiadas (1999).
- Of all the models considered, the Gompertz type was found to be the most useful, which indicates the existence of a "slowdown" (parameter value $\beta = 0.06235$), according to the interpretation of the β of the stochastic equation Eq. 1. All the above facts, with respect to Spain, both during the period 1986–2003 and at present, confirm the hypothesis of a significant correlation between the increase in the emission of CO_2 and the growth of the GDP. In fact, comparison of the fits of the homogeneous models (Fig. 1) with the respective nonhomogeneous ones (Figs. 2, 3) confirms the marked improvement in the fits achieved by including the influence of the GDP as an exogenous factor to the emission of CO_2.
- Finally, let us remark on some extensions to this study that are currently being investigated. From the theoretical standpoint, the models used here can be extended by considering various exogenous factors. This multiple situation (which would be equivalent, in regression, to considering Multiple Linear Regression, with the presence of various regressors) has been addressed technically, for example, in Gutiérrez et al. (1997a), for the lognormal case, and by Nafidi (1997) and by Gutiérrez-Sánchez (2005) for the nonhomogeneous Gompertz case. From the practical point of view, it is to be expected that the methodology of trend analysis described in the present paper, when extended to models with exogenous factors, will lead, in turn, to an improvement of the statistical fits achieved. In particular, in the case of the emission of CO_2 in Spain, it seems clear that the addition of other exogenous factors, such as the GDP considered in this paper, would be justified and perhaps more optimal fits would be attained. An example of another, presumably important, factor would be the variable "number of new houses constructed annually in Spain", which in the above-mentioned period and also at present is the highest in the EU (see Introduction to this paper) and the effects of which on the increased emission of CO_2 are undeniable.
- Also, it would be possible to study the asymptotic normality of estimators of Eq. 12 by using for example, the delta method which would allow to obtain asymptotic confidence interval of the estimators of the parameters of the models considered.

Acknowledgments The authors are indebted to the Editors and referees for constructive comments on the preliminary version of this paper. This work was supported partially by research project MTM 2005-09209, Ministerio de Educación y Ciencia, Spain.

References

Albano G, Giorno V (2006) A stochastic model in tumor growth. J Theor Biol (in press) DOI 10.1016/j.jtbi.2006.03.001

Andres RJ, Fielding DJ, Marland G, Boden TA, Kumar N (1999) Carbon dioxide emissions from fossil-fuel use, 1751–1950. Tellus 51B:759–765

Boden TA, Marland G, Andres RJ (1995) Estimates of global, regional and national annual CO_2 emissions from fossil-fuel burning, hydraulic cement production, and gas flaring: 1950–1992. ORNL/CDIAC-90,NDP-30/R6. Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge

Ferrante L, Bompade S, Possati L, Leone L (2000) Parameter estimation in a Gompertzian stochastic-model for tumor growth. Biometrics 56:1076–1081

Ferrante L, Bompade S, Possati L, Leone L, Montanari MP (2005) A stochastic formulation of the Gompertzian growth model for in vitro bactericidad kinetics: parameter estimation and extinction probability. Biomet J 47,3:309–318

Giovanis AN, Skiadas CH (1999) A stochastic logistic innovation diffusion model for studying the electricity consumption in Greece and United States. Technol Forecast Soc Change 61:235–246

Gutiérrez-Sánchez R (2005) Stochastic diffusion Lognormal and Gompertz nonhomogeneous. Rayleigh Process. Applications. Ph.D thesis, University of Granada

Gutiérrez R, Angulo JM, González A, Pérez R (1991) Inference in lognormal multidimensional diffusion process with exogenous factors: application to modelling in economics. Appl Stoch Models Data Anal 7:295–316

Gutiérrez R, González A, Torres F (1997a) Estimation in multivariate lognormal diffusion process with exogenous factors. Appl Stat 46(1):140–146

Gutiérrez R, Ricciardi L, Román P, Torres F (1997b) First-passage time densities for time-non-homogeneous diffusion processes. J Appl Prob 34(3):623–631

Gutiérrez R, Román P, Torres F (1999) Inference and first-passage-time for the lognormal diffusion process with exogenous factors: application to modelling in economics. Appl Stochast Models Bus Indus 15(4):325–332

Gutiérrez R, Román P, Torres F. (2001) Inference on some parametric functions in the univariate lognormal diffusion process with exogenous factors. Test 10(2):357–373

Gutiérrez R, Gutiérrez-Sánchez R, Nafidi A (2005a) Forecasting total natural-gas consumption in Spain by using the stochastic Gompertz innovation diffusion model. Appl Energy 80(2):115–124

Gutiérrez R, Gutiérrez-Sánchez R, Nafidi A (2005b) The stochastic Brennan–Schwartz Diffusion Process: Statistical Inference and Applications. CIMMA: International Mediterranean Congress of Mathematics, 47–57, Almería

Gutiérrez R, Gutiérrez-Sánchez R, Nafidi A, Román P, Torres F (2005c) Inference in Gompertz type nonhomogeneous stochastic systems by means of discrete sampling. Cybern Syst 36:203–216

Gutiérrez R, Roldan C, Gutiérrez-Sánchez R, Angulo JM (2005d) Estimation and prediction of a 2D lognormal diffusion random field. Stochast Environ Res Risk Assess 19(4):258–265

Gutiérrez R, Gutiérrez-Sánchez R, Nafidi A (2006a) Electricity consumption in Morocco: stochastic Gompertz diffusion analysis with exogenous factors. Appl Energy 83:1139–1151

Gutiérrez R, Gutiérrez-Sánchez R, Nafidi A (2006b) The Stochastic Rayleigh diffusion model: Statistical inference and computational aspects. Applications to modelling of real cases. Appl Math Comput 175:628–644

Gutiérrez R, Gutiérrez-Sánchez R, Nafidi A, Ramos E (2006c) A diffusion model with cubic drift: statistical and computational aspects and application to modelling the global CO_2 emission in Spain. Environmetrics (in press) DOI 10.1002/env.801

Gutiérrez R, Gutiérrez-Sánchez R, Nafidi A, Ramos E (2006d) Threshold parameter in stochastic Gompertzian process: computational inference and application in modelling. Appl Math Comput (in press) DOI 10.1016/j.amc.2006.05.099

Liptser RS, Shiryayev AN (1977) Statistics of random processess I. Springer, Berlin Heidelberg New York

Marland G, Boden TA, Andres RJ (2005) Global, regional, and national fossil fuel CO_2 emissions. In trends: a compendium of data on global change. Carbon dioxide information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge

McCuen R (2003) Modelling hydrologic change. Statistical methods. Lewis, Boca Raton

Meade N, Islam T (2006) Modelling and forecasting the diffudion of innovation. A 25-years review. International Int J Forecast 22(3):519–545

Nafidi A (1997) Lognormal diffusion process with exogenous factors, Extensions from the Gompertz diffusion process. Ph.D thesis, University of Granada

Ramanathan R (2006) A multi-factor efficiency perspective to the relationships among world GDP, energy consumption and carbon dioxide emissions. Technol Forecast Soc Change (in press) DOI 10.1016/j.techfore. 2005.06.012

Wong E, Hajek B (1985) Stochastic processes in engineering systems. Springer, Berlin Heidelberg New York