Monografías del Seminario Matemático García de Galdeano **33**, 273–280 (2006)

A GENERALIZATION OF THE GOMPERTZ DIFFUSION MODEL: STATISTICAL INFERENCE AND APPLICATION

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

Abstract. The aim of this work is to study a non-homogenous extension of the Gompertz diffusion process (cf. [1], [8]), based on the fact that only the deceleration factor in the drift is a time-dependent function (this version can be considered as a Gompertz diffusion with exogenous factors). A particular case of this model has been considered (cf. [4]) in the study of the first passage time problem in the non-homogeneous diffusion process. The proposed extended non-homogeneous model is studied by the methodology based on Kolmogorov equations, whereas in [1, 8], the homogeneous process is studied by a methodology using Ito's equations. Firstly, we obtain the probability density function (p.d.f.) of the process and its trend functions (non conditional and conditional). Then, the statistical inference in the model is achieved, estimating the parameters by the maximum likelihood method using discrete sampling, and obtaining the distributions of the resulting estimators and the confidence intervals of the parameters. Finally, the proposed model is applied to real data for electricity consumption in Morocco.

Keywords: Gompertz diffusion process, deceleration factor, statistical inference, electricity consumption.

AMS classification: 60J60, 62M05.

§1. Introduction

In recent decades, various diffusion-type stochastic models have been developed, and these have been successfully applied to the fitting and prediction of phenomena in diverse fields, such as biology, physics, medicine, economics and finance. These models include stochastic diffusion processes such as lognormal [5] and [16]; Bass [15]; Rayleigh [9]; Gompertz [8]; and Logistic [3]. From the point of view of stochastic differential equations, the homogeneous Gompertz stochastic diffusion process (SGDP) was introduced by Ricciardi [14] in a theoretical form, and subsequently applied by Ferrante et al. [1] (growth of cancer cells) and by Gutiérrez et al. [8] (consumption of natural gas in Spain). From the perspective of the Kolmogorov equations, the model was defined by Nafidi [12] in a general form, and later applied by Gutiérrez et al. [6] in a study of the stock of motor vehicles in Spain. The non-homogeneous form of the process (with exogenous factors) has been addressed by Nafidi [12] in a very general context. Later, Gutiérrez et al. [7, 10] studied the case in which only the growth rate in the drift is affected by exogenous factors in a linear way, and applied this both to the growth in the price of new housing in Spain and to the consumption of electricity in Morocco. Finally, Ferrante et al. [2] considered a non-homogeneous version in which the growth rate is the sum of two exponential functions that are exogenous factors. The case in which the deceleration factor is affected by exogenous factors has not been addressed to date, and the study of a model with such a hypothesis is the prime concern of the present paper.

Thus, in the present article we discuss a non-homogeneous version of the Gompertz diffusion process based on the fundamental fact that the deceleration coefficient is a function of the time. The paper is structured as follows: firstly we obtain the p.d.f. of the process and its trend functions (non conditional and conditional). Then, the statistical inference in the model is addressed, estimating the parameters by the maximum likelihood method using discrete sampling, and obtaining the distributions of the resulting estimators and the confidence intervals of the parameters. Finally, a particular case of the proposed model is applied to the real case of the electricity consumption in Morocco.

§2. A generalization of the SGDP

2.1. Formulation of the model

The homogeneous SGDP model has been studied by Ricciardi [14], by Ferrante et al. [1] and by Gutiérrez et al. [8], and it is defined as the solution to Ito's SDE:

$$dx(t) = (\alpha x(t) - \beta x(t)\log(x(t)))\,dt + \sigma x(t)dw(t).$$

The α constant is the *intrinsic growth rate*; the β constant is the *deceleration factor* and the σ constant is the *diffusion coefficient*.

It is well known (see, for example [8]) that, by means of the transform $y(t) = e^{\beta t}\log(x)$, the homogeneous SGDP can be transformed into a Wiener process. On the basis of this fact, to define the required generalization of this process, let us consider the SDE in a general form:

$$dx(t) = a(t,x(t))\,dt + \sigma x(t)\,dw(t), \quad t \in [t_0, T],$$

and seek the condition that satisfies the drift $a(t,x(t))$ so that the above SDE is derived for a Wiener-type process. For this purpose, let us consider a function $g(t)$ that is derivable in $[t_0, T]$. Then, by applying Ito's formula to the transform $y(t) = g(t)\log(x)$ we obtain

$$dy(t) = g(t)\left[\frac{g'(t)}{g(t)}\log(x(t)) + \frac{a(t,x(t))}{x(t)} - \frac{\sigma^2}{2}\right]dt + \sigma g(t)\,dw(t).$$

The condition for $y(t)$ to be a Wiener process is that the right-hand term in the above SDE should be a function that depends solely on the time, such that

$$\frac{g'(t)}{g(t)}\log(x(t)) + \frac{a(t,x(t))}{x(t)} = k(t).$$

From this, we obtain a non-homogeneous version of the SGDP in the intrinsic growth rate and in the deceleration factor:

$$dx(t) = \left(k(t)x(t) - \frac{g'(t)}{g(t)}x(t)\log(x(t))\right)dt + \sigma x(t)\,dw(t).$$

This latter model in the particular case in which $\frac{g'(t)}{g(t)}$ is constant has been studied by Gutiérrez et al. [7, 10] (the case in which $k(t)$ is linear in the exogenous factors), and in another particular case by Ferrante et al. [2] (the case in which $k(t)$ is the sum of two exponential functions as exogenous factors). In these studies, estimation of the parameters requires the use of numerical approximation methods, and then it is not possible to determine, theoretically, the qualities of the estimators. In order to carry out a complete inferential study (estimation, distribution of the estimators and confidence intervals), we shall here concentrate on the case in which only the deceleration parameter is dependent on the time, that is $k(t) = a \in \mathbb{R}$. In this case, the generalization to be considered is the family of diffusion processes $\{x(t), t \in [t_0, T]\}$ with values in $(0, \infty)$ and with infinitesimal moments that are given by

$$A_1(t,x) = ax(t) - \frac{g'(t)}{g(t)} x(t) \log(x(t)), \qquad A_2(t,x) = \sigma^2 x^2(t). \tag{1}$$

Remark 1.

- If $g(t)$ is constant, the process is lognormal homogeneous (Tintner et al. [16]).
- If $g(t) = e^{\beta t}$, the process is Gompertz homogeneous (Gutiérrez et al. [8]).
- If $g(t) = t$, the process is Gompertz non-homogeneous (Gutiérrez et al. [4]).

2.2. The p.d.f. of the process

Let us take $f(y,t \mid x,s)$ to denote the p.d.f. of the model being considered (1). This function complies with the Kolmogorov equations, and so, for example, the forwards equation is:

$$\frac{\partial f}{\partial t} = -\frac{\partial [a(t,y) f]}{\partial y} + \frac{1}{2} \frac{\partial^2 [b(t,y) f]}{\partial y^2}.$$

The infinitesimal moments of the process (1) fulfil the necessary and sufficient condition of the theorem of Ricciardi [13] on the transformation of a diffusion process into a Wiener process. The relevant transform in this case is as follows:

$$\Phi(t) = \int^t g^2(\theta)\, d\theta, \quad \Psi(t,x) = \frac{g(t)}{\sigma} \log(x) - \frac{a - \sigma^2/2}{\sigma} \int^t g(\theta)\, d\theta.$$

Therefore, the p.d.f. of the resulting process is

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

which is the density of a lognormal distribution $\Lambda_1\left(\mu(s,t,x), \sigma^2 \nu^2(s,t)\right)$, with

$$\mu(s,t,x) = \frac{g(s)}{g(t)} \log(x) + \frac{a - \sigma^2/2}{g(t)} \int_s^t g(\theta) d\theta \quad \text{and} \quad \nu^2(s,t) = \frac{1}{g^2(t)} \int_s^t g^2(\theta) d\theta.$$

2.3. Trends of the process

Taking into account that the random variable $x(t) \mid x(s) = x_s \sim \Lambda_1\left(\mu(s,t,x_s), \sigma^2\nu^2(s,t)\right)$ and bearing in mind the properties of this distribution, the conditional trend function of the process is

$$\mathbf{E}(x(t) \mid x(s)) = \exp\left(\frac{g(s)}{g(t)}\log(x(s) + \frac{a-\sigma^2/2}{g(t)}\int_s^t g(\theta)\,d\theta + \frac{\sigma^2}{2g^2(t)}\int_s^t g^2(\theta)\,d\theta\right). \quad (3)$$

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

$$\mathbf{E}(x(t)) = \exp\left(\frac{g(t_1)}{g(t)}\log(x_{t_1}) + \frac{a-\sigma^2/2}{g(t)}\int_{t_1}^t g(\theta)d\theta + \frac{\sigma^2}{2g^2(t)}\int_{t_1}^t g^2(\theta)d\theta\right). \quad (4)$$

2.4. Statistical inference in the model

2.4.1. Parameter estimation

Let us consider a discrete sample of the process $(x_1,\ldots,x_n)$ at the instants of time $(t_1,\ldots,t_n)$, under the initial condition $\mathrm{P}(x(t_1) = x_1) = 1$, and let us assume, moreover, that $g(t)$ depends solely on the time; the likelihood function associated with the process is then

$$\mathbf{L}(x_1,\ldots,x_n,\alpha,\sigma^2) = \prod_{j=2}^{n} f(x_j,t_j \mid x_{j-1},t_{j-1}).$$

From (2), the above expression can be rewritten as

$$\mathbf{L}(x_1,\ldots,x_n,\alpha,\gamma) = \prod_{j=2}^{n}\left[2\pi\sigma^2\nu^2(t_{j-1},t_j)\right]^{1/2} \times \frac{1}{x_j}\exp\left(-\frac{\left[\log(x_j) - \mu(t_{j-1},t_j,x_{j-1})\right]^2}{2\sigma^2\nu^2(t_{j-1},t_j)}\right).$$

In order to work with a known likelihood function and to calculate the estimators in the simplest possible way, the discrete sampling is transformed as follows: for $j = 2,\ldots,n$,

$$\mathbf{u}_j = \frac{\nu_j^{-1}}{g(t_j)}\int_{t_{j-1}}^{t_j} g(\theta)\,d\theta, \quad \mathbf{v}_j = \nu_j^{-1}\left(\log(x_j) - \frac{g(t_{j-1})}{g(t_j)}\log(x_{j-1})\right).$$

and thus, the likelihood function can be written as

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

where $\mathbf{a} = a - \sigma^2/2$, $\mathbf{v} = (v_2,\ldots,v_n)'$, $\nu_j = \nu(t_{j-1},t_j)$ and $\mathbf{U} = (\mathbf{u}_2,\ldots,\mathbf{u}_n)$. By differentiating the log-likelihood function with respect to $\mathbf{a}$ and σ^2 and after some algebraic rearrangement, the likelihood estimators yield

$$\hat{\mathbf{a}} = (\mathbf{U}\mathbf{U}')^{-1}\mathbf{U}\mathbf{V}, \quad (5)$$

$$(n-1)\hat{\sigma}^2 = \mathbf{V}'(\mathbf{I}_{n-1} - \mathbf{U}'(\mathbf{U}\mathbf{U}')^{-1}\mathbf{U})\mathbf{V}. \quad (6)$$

2.4.2. Confidence intervals of the parameters

The distribution of the above estimators is

$$\hat{\mathbf{a}} \sim \mathcal{N}_1\left(\mathbf{a}, \sigma^2(\mathbf{U}\mathbf{U}')^{-1}\right) \quad \text{and} \quad (n-1)\hat{\sigma}^2/\sigma^2 \sim \chi^2_{n-2}.$$

It can thus be shown that $(\hat{\mathbf{a}}, \hat{\sigma}^2)$ are conjointly sufficient and complete for $(\mathbf{a}, \sigma^2)$, and so the estimators $\hat{\mathbf{a}}$ and $\frac{(n-1)}{n-2}\hat{\sigma}^2$ are the UMVUE for $\mathbf{a}$ and σ^2, respectively.

A (1-α)% confidence interval for $\mathbf{a}$ and σ^2 are given respectively by

$$\left[\hat{\mathbf{a}} - \hat{\sigma}.t_{\alpha/2,n-1}/\sqrt{n-1}, \hat{\mathbf{a}} + \hat{\sigma}.t_{\alpha/2,n-1}/\sqrt{n-1}\right], \tag{7}$$

$$\left[(n-1)\hat{\sigma}^2/\chi^2_{\alpha/2,n-1}, (n-1)\hat{\sigma}^2/\chi^2_{1-\alpha/2,n-1}\right], \tag{8}$$

where $\chi^2_{\alpha,n}$ and $t_{\alpha,n}$ are the upper 100α per cent points of the chi squared distribution and the Student distribution, respectively, with n degrees of freedom.

§3. Application to electricity consumption in Morocco

The model examined in this study was applied to real data for the total consumption of electricity in Morocco (expressed in 10^9 Kwh) during the period 1980 to 2002 (including distribution and transport losses). These data correspond to sales by ONE, the Moroccan authority, and can be consulted at [11]. The methodology can be summarised in the following two phases:

- Step1: Use the first 20 data in the series of observations considered to estimate the parameters of the model, using expressions (5) and (6). Then, determine the corresponding confidence intervals using equations (7) and (8).
- Step2: For the years 2000, 2001 and 2002, predict the corresponding values for electricity consumption in Morocco using the estimated trend function (ETF) and the estimated conditional trend function (ECTF), obtained by replacing the parameters with their estimators in expressions (3) and (4), and compare the results with the corresponding observed data for the same years.

A Matlab program was implemented to carry out the calculations required for this study. Considering, for example, the function $g(t) = t^{-4}$, the values of the corresponding estimators and the confidence intervals are: $\hat{\mathbf{a}} = 0.061645$ and $\hat{\sigma}^2 = 1.793344.10^{-4}$ with confidence intervals $(0.055215; 0.068076)$ and $(1.037172; 3.825685).10^{-4}$. Table 1 summarises the prediction results, i.e. the observed data, the values predicted by ETF and ECTF and the lower and upper limits of these functions, denoted by LL-ETF, UL-ETF and LL-ECTF and UL-ECTF, respectively.

Figure 1 shows the fits and the predictions made using the ETF and the ECTF.

Acknowledgements

This work was supported in part by the Ministerio de Ciencia y Tecnología, Spain, under Grant BFM 2002-03633.

Times	Data	ETF	LL-ETF	UL-ETF	ECTF	LL-ECTF	UL-ECTF
2000	12.8380	12.9886	11.4420	14.7624	12.9599	12.8764	13.0448
2001	13.4520	13.7442	12.0327	15.7193	13.5852	13.4977	13.6741
2002	14.0850	14.5421	12.6526	16.7362	14.2336	14.1420	14.3268

Table 1: Forecasting based on ETF and ECTF

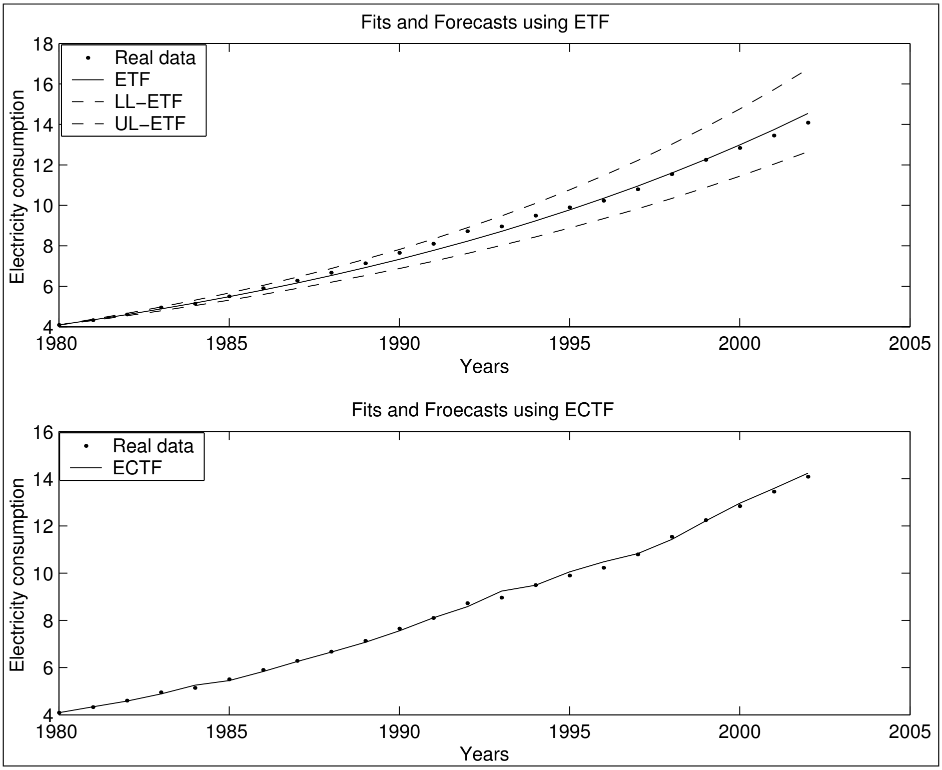

Figure 1: Fits and predictions made using the ETF (above) and the ECTF (below)

References

[1] FERRANTE, L., BOMPADE, S., POSSATI, L., AND LEONE, L. Parameter estimation in a Gompertzian stochastic-model for tumor growth. *Biometrics 56* (2000), 1076–81.

[2] FERRANTE, L., BOMPADE, S., POSSATI, L., LEONE, L., AND MONTANARI, M.P. A stochastic formulation of the Gompertzian growth model for in vitro bactericidad kinetics: Parameter estimation and extinction probability. *Biometrical Journal 47*, 3 (2005), 309–318.

[3] GIOVANIS, A. N., AND SKIADAS, C. H. A stochastic logistic innovation diffusion-model studying the electricity consumption in Greece and the United States. *Technol. Forecast. & Soc. Change 61* (1999), 253–64.

[4] GUTIÉRREZ, R., ROMÁN, P., AND TORRES, F. A note on the Volterra integral equation for the first-passage-time density. *Journal of Applied Probability 53*, 3 (1995), 635–648.

[5] GUTIÉRREZ, R., ROMÁN, P., AND TORRES, F. Inference and first-passage-time for the lognormal diffusion process with exogenous factors: Application to modelling in economics. *Appl. Stochastic Models in Business and Industry 15*, 4 (1999), 325–332.

[6] GUTIÉRREZ, R., GUTIÉRREZ-SÁNCHEZ, R., NAFIDI, A., AND RAMOS, E. Studying the vehicle park in Spain by using the lognormal and Gompertz diffusion processes. In *Proceedings of SEIO'04 Spain*. Cadiz, 2004, pp. 171–172.

[7] GUTIÉRREZ, R., NAFIDI, A., GUTIÉRREZ-SÁCHEZ, R., ROMÁN, P., AND TORRES, F. Inference in Gompertz type non homogeneous stochastic systems by means of discrete sampling. *Cybernetics and Systems 36* (2005), 203–216.

[8] GUTIÉRREZ, R., NAFIDI, A., AND GUTIÉRREZ-SÁCHEZ, R.. Forecasting total natural gas consumption in Spain by using the stochastic Gompertz innovation diffusion model. *A*ppl Energy 80, 2 (2005), 115–124.

[9] GUTIÉRREZ, R., GUTIÉRREZ-SÁCHEZ, R., AND NAFIDI, A. The Stochastic Rayleigh diffusion model: Statistical inference and computational aspects. Applications to modelling of real cases. *A*ppl. math. and comp., 175(2006), 628-644.

[10] GUTIÉRREZ, R., GUTIÉRREZ-SÁCHEZ, R., AND NAFIDI, A. Electricity consumption in Morocco: Stochastic Gompertz diffusion analysis with exogenous factors. *Appl. Energy*, (in press).

[11] *Le Maroc en chiffres*. Direction de la Statistique, Ministère de la Prévision Economique et du Plan, Morocco, 1980-2002.

[12] NAFIDI, A. *Lognormal Diffusion with Exogenous Factors, Extensions of the Gompertz Diffusion Process*. Ph.D. Thesis, Granada University, 1997 [in Spanish].

[13] RICCIARDI, L. M. On the transformation of diffusion processes into the Wiener processes. *J. of Math. Anal. and Appl. 54* (1976), 185–99.

[14] RICCIARDI, L. M. Diffusion processes and related topics in Biology. *Lecture Notes in Biomathematics 14*, (1977).

[15] SKIADAS, C. H., AND GIOVANIS, A. N. A stochastic Bass innovation diffusion model for studying the growth of electricity consumption in Greece. *Appl. Stoch. Mod. & Data Anal. 13* (1997), 85–101.

[16] TINTNER, G., AND SENGUPTA, J. K. *Stochastic Economics*. Academic Press, 1972.

R. Gutiérrez, A. Nafidi and R. Gutiérrez Sánchez
Departamento de Estadística e Investigacíon Operativa
Universidad de Granada
Avenida Fuentenueva s/n
18071 Granada
`rgjaimez@ugr.es`, `ramongs@ugr.es` and `nafidiah@ugr.es`