Monografías Matemáticas *García de Galdeano* **35**, 263–270 (2010)

STATISTICAL INFERENCE IN THE STOCHASTIC GAMMA DIFFUSION PROCESS WITH EXTERNAL INFORMATION

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

Abstract. In this work, we consider a new extension of the one-dimensional stochastic gamma diffusion process (cf. [11]) by introducing external time functions as exogenous factors, in the same way as exogenous factors have been introduced into lognormal process (cf. [14]), the Gompertz process (cf. [10]) and the Vasicek process (cf. [12]), among others. Firstly, we determine the probabilistic characteristics of the process as its analytical expression, the transition probability density function and the trend functions. Secondly, we study the statistical inference in this process: the parameters present in the model are studied by using the maximum likelihood estimation method on the basis of the discrete sampling, thus obtaining the expression of the likelihood estimators and their properties (statistical distribution, sufficiency and completeness), together with the confidence intervals of the parameters.

Keywords: Stochastic gamma diffusion process, exogenous factors, statistical inference in diffusion process.

AMS classification: 60J60, 62M05.

§1. Introduction

Stochastic processes are used in fields as diverse as physics, biology, economics and finance to model and analyze dynamic systems. One particular class of stochastic processes which has attracted considerable attention is that of diffusion processes. And one of the questions that has provoked greatest theoretical and practical interest concerning diffusions, and which has been the object of many studies in recent years, is the problem of establishing the corresponding statistical inference, a question that may be approached by the use of either continuous or discrete sampling. This inference, and in particular the estimation of parameters, has been studied in the general case by various authors, such as Bibby and Sorensen [3], Prakasa Rao [17], Ait-Sahalia [1] and Egorov et al. [4], among many others. And in the case of particular diffusions, it has been considered, for example, by Giovanis et al. [7] in the logistic case, Gutiérrez et al. [8] in the Gompertz case, Gutiérrez et al. [9] in the Rayleigh case and Forman et al. [6] in the case of Pearson diffusions, among other important diffusions.

Due to the need to use stochastic diffusions to accurately model real phenomena that are becoming more and more complex, various extensions of these processes have been considered, such as non-homogeneous extensions and, in particular non-homogeneous extensions with exogenous factors, which have been defined, studied and applied, for example, in the case of the lognormal process by Gutiérrez et al. [14], in the case of the Gompertz process by Gutiérrez et al. [10] and Ferrante [5], in the case of the Vasicek process by Gutiérrezet al. [12], and by Picchini et al. [16] in the case of the Brennan-Schwartz diffusion process, among others.

In the present study, based on the methodology established for the consideration of exogenous factors affecting drift , described in [14], [10] and [12], we define a new diffusion process with external information, modelled by time deterministic functions (exogenous variables) that affect the drift of the Gamma diffusion process, as studied in [13] and [11]. We go on to examine a new Gamma diffusion process with exogenous factors, investigating its main probabilistic characteristics and the corresponding statistical inference.

The remainder of the paper is organised as follows. In the next section, we first define the model and consider its probabilistic characterisations, such as the explicit expression, the probability transition density function (ptdf) and the moments (in particular the trend functions). In the third section, we study the statistical inference in the proposed process using discrete sampling, obtaining the likelihood estimators, their statistical properties and the confidence parameter intervals.

§2. The model and its basic probabilistic characteristics

2.1. The proposed model and their analytical expression

The model considered is the one dimensional process $\{x(t),\ t \in [t_1, T],\ t_1 > 0\}$ with values in $(0, \infty)$ and governed by the following Ito's stochastic differential equation (SDE)

$$dx(t) = a(t, x(t))dt + b^{1/2}(t, x(t))dw(t), \quad \mathrm{P}[x(t_1) = x_{t_1}] = 1, \tag{1}$$

where $a(t, x)$ and $b(t, x)$ are given by

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

In the first coefficient $a(t, x)$, the function h is considered as a linear combination of the exogenous factors, and is given by

$$h(t) = \beta_0 + \sum_{i=1}^{q} \beta_i g_i(t)$$

where g_i (for $i = 1, \ldots, q$) are called exogenous factors (external information) and are a time-continuous function in $[t_1, T]$, α, β_i (for $i = 0, \ldots, q$) and $\sigma > 0$ are time-independent real parameters (to be estimated).

It can be proved that the functionals $a(t, x)$ and $b(t, x)$ 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 [15], 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 [19], Propositions 4.1 and 7.1]. Consequently, there exists a separable, measurable and almost surely (a.s.) sample continuous diffusion process $\{x(t)\,;\, t \in [t_1, T]\}$ which is the unique (a.s.) solution to Ito's SDE Eq.(1) with infinitesimal moments (drift and diffusion coefficients) given, respectively, by $a(t, x)$ and $b(t, x)$.

2.2. The ptdf and moments of the model

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

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

By integrating and substituting, we deduce that the analytical expression of the solution to the SDE Eq.(1) is

$$x(t) = x_{t_1}\left(\frac{t}{t_1}\right)^{\alpha} \exp\!\left(-\int_{t_1}^{t}\left(h(\tau) - \frac{\sigma^2}{2}\right)d\tau + \sigma(w(t) - w(t_1))\right),$$

then, $x(t)$ 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

$$\mu(s, t, x) = \log(x) + \alpha\log(t/s) - (\beta_0 + \sigma^2/2)(t - s) - \sum_{i=1}^{q}\beta_i\int_s^t g_i(\tau)d\tau,$$

and therefore, the tpdf of the process has the following form

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

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

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

Then, the conditional trend function ($r = 1$) of the process is

$$\mathbb{E}\left[x(t) \mid x(s) = x_s\right] = x_s\left(\frac{t}{s}\right)^{\alpha} e^{-\beta_0(t-s)-\sum_{i=1}^{q}\beta_i\int_s^t g_i(\tau)d\tau}.$$

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

$$\mathbb{E}\left[x(t)\right] = x_{t_1}\left(\frac{t}{t_1}\right)^{\alpha} e^{-\beta_0(t-t_1)-\sum_{i=1}^{q}\beta_i\int_{t_1}^t g_i(\tau)d\tau}.$$

And the variance of the process is given by

$$Var\left[x(t)\right] = x_{t_1}^2\left(\frac{t}{1}\right)^{2\alpha} e^{-2\beta_0(t-t_1)-2\sum_{i=1}^{q}\beta_i\int_{t_1}^t g_i(\tau)d\tau}\left(e^{\sigma^2(t-t_1)} - 1\right).$$

§3. Statistical inference on the model

3.1. Likelihood parameter estimation

In the present study, with discrete sampling, we estimate the parameters α, σ^2 and β_i (for $i = 1, \ldots, q$) of the model by applying maximum likelihood estimation (MLE) methodology. Let us consider a discrete sampling of the process $x_1, \ldots, x_n$ for times $t_1, t_2, \ldots, t_n$ and assume an initial distribution $\mathrm{P}[x(t_1) = x_1] = 1$. Then the associated likelihood function can be obtained from Eq.(2) by the following expression

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

An implementation based on the change of variable can be used 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 following reparametrization $\Gamma = \left(\alpha, -(\beta_0 + \sigma^2/2), -\beta_1, \ldots, -\beta_q\right)'$ and

$$u_i = (t_i - t_{i-1})^{-1/2}\left(\log(t_i/t_{t-1}), t_i - t_{i-1}, \int_{t_i}^{t_{i-1}} g_1(\tau)\, d\tau, \ldots, \int_{t_i}^{t_{i-1}} g_q(\tau)\, d\tau\right)'.$$

Then, the likelihood function for the transformed sample is

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

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

$$\mathbb{L}_{\mathbf{V}}(\Gamma, \sigma^2) = \left[2\pi\sigma^2\right]^{-(n-1)/2} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{V} - \mathbf{U}'\Gamma)'(\mathbf{V} - \mathbf{U}'\Gamma)\right). \tag{3}$$

After calculating the derivatives of the log-likelihood function with respect to the parameter matrix Γ and the coefficient σ^2, the likelihood equations are

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

From which, the likelihood estimators of the parameters are

$$\widehat{\Gamma} = (\mathbf{U}\mathbf{U}')^{-1}\mathbf{U}\mathbf{V},$$
$$(n-1)\hat{\sigma}^2 = \mathbf{V}'\mathrm{H}_{\mathbf{U}}\mathbf{V},$$

where $\mathrm{H}_{\mathbf{U}} = \mathrm{I}_{n-1} - \mathbf{U}'(\mathbf{U}\mathbf{U}')^{-1}\mathbf{U}$ is an idempotent symmetric matrix.

Remark 1. In the absence of exogenous factors (i.e: $g_i = 0$, for $i = 1, \ldots, q$), we obtain the stochastic Gamma diffusion process studied in Gutiérrezet al. [11, 13], and it can be shown that all the results established in the present study generalize those obtained in the two papers cited.

3.2. Properties of likelihood estimators

3.2.1. Distribution and independence of MLEs

Using Eq.(3), it can be deduced that $\mathbf{V} \sim \mathcal{N}_{n-1}\left[\mathbf{U}'\Gamma, \sigma^2\mathbf{I}_{n-1}\right]$.

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

$$\widehat{\Gamma} \sim \mathcal{N}_{q+2}\left[\Gamma, \sigma^2(\mathbf{UU}')^{-1}\right].$$

On the one hand, we have $\sigma^{-1}(\mathbf{V} - \mathbf{U}'\Gamma) \sim \mathcal{N}_{n-1}(0, \mathbf{I}_{n-1})$, as $\mathrm{H_U}$ is idempotent, then by a known multivariate analysis result (see for example, [18, Theorem 2, p. 57]), we have

$$\sigma^{-2}(\mathbf{V} - \mathbf{U}'\Gamma)'\mathrm{H_U}(\mathbf{V} - \mathbf{U}'\Gamma) \sim \chi^2_{\mathrm{rank}(\mathrm{H_U})}.$$

And by taking into account that $\mathrm{H_U}$ is symmetric and idempotent, we have $\mathrm{rank}(\mathrm{H_U}) = \mathrm{tr}(\mathrm{H_U}) = n - q - 3$, and therefore

$$\sigma^{-2}(\mathbf{V} - \mathbf{U}'\Gamma)'\mathrm{H_U}\sigma^{-1}(\mathbf{V} - \mathbf{U}'\Gamma) = \sigma^{-2}\mathbf{V}'\mathrm{H_U}\mathbf{V} \sim \chi^2_{n-q-3}.$$

From which, we deduce that

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

On the other hand, as $(\mathbf{UU}')^{-1}\mathbf{U}\mathrm{H_U} = 0$, then by Theorem 3 in [18, p. 59], we have $(\mathbf{UU}')^{-1}\mathbf{UV}$ and $\mathbf{V}'\mathrm{H_U}\mathbf{V}$ are independently distributed, which means that $\widehat{\Gamma}$ and $\hat{\sigma}^2$ are independently distributed.

3.2.2. Sufficiency and Completeness of MLEs

By substracting and adding $\mathbf{U}'\widehat{\Gamma}$ to $\mathbf{V} - \mathbf{U}'\Gamma$, expression Eq.(3) becomes

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

This shows that $\left(\widehat{\Gamma}, \hat{\sigma}^2\right)$ is conjointly sufficient for $\left(\Gamma, \sigma^2\right)$.

The completeness follows by means of similar reasoning to that established for the maximum likelihood estimators of the parameters of the multivariate normal distribution (see, for example, Anderson [2]).

Finally it can be deduced that the estimators $\widehat{\Gamma}$ and $\frac{(n-1)\hat{\sigma}^2}{(n-q-3)\sigma^2}$ are the UMVUE for the parameters Γ and σ^2 respectively.

3.3. Parameter confidence intervals

On the basis of the above results, it can be deduced that the $(1-\gamma)\%$ confidence interval for the parameter σ^2 is given, by

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

And the $(1-\gamma)\%$ concentration ellipsoid for $\Gamma_* = \left(-\beta_1, \ldots, -\beta_q\right)'$ is given by

$$\left(\Gamma_* - \hat{\Gamma}_*\right)\left[A_{(22)}\right]^{-1}\left(\Gamma_* - \hat{\Gamma}_*\right)' \leq \frac{(n-1)q}{n-q-3}\hat{\sigma}^2 F_{q,n-q-3,\gamma},$$

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, $A_{(22)}$ is $q \times q$-matrix and given in

$$(\mathbf{UU}')^{-1} = \begin{pmatrix} A_{(11)} & A_{(12)} \\ A_{(21)} & A_{(22)} \end{pmatrix}.$$

§4. Conclusions

The Gamma process, from the outset, is a non-homogenenous diffusion process, as its drift depends explicitly on the time t. In the present paper, we have introduced a new type of Gamma diffusion, including a second source of non-homogeneneity, which is derived from making the function $h(t)$, which forms part of the drift of the initial diffusion, depend on q exogenous factors, $g_1(t)$, $i = 1, \ldots, q$. These factors are external (or exogenous) to the process $x(t)$ itself, and act as "regressors" and thus the drift of the diffusion varies, as do its trend functions. In consequence, through an appropriate choice of such exogenous factors, it is possible to fit the Gamma diffusion introduced, and in particular its trend functions, to a real phenomenon, in a way that is more suitable in statistical terms than if this were done with the initial Gamma diffusion (without exogenous factors). This is so because, thanks to these factors, we can model the influence of certain exogenous factors on the dynamic behaviour of the endogenous variable $x(t)$.

This fit can be applied, in practice, to the Gamma diffusion examined in the present study because it has been possible to develop the basic results of statistical inference (the estimation and testing of hypotheses) for the model defined in Eq(1).Thus, we have a method for adjusting, and for analyzing the goodness of fit, that is suitable for practical implementation.

Acknowledgements

This work was supported partially by research project MTM 2008-05785, Ministerio de Educación y Ciencia, and project P06-FQM-02271, Junta de Andalucía, Spain.

References

[1] Ait-Sahalia, Y. Maximum likelihood estimation of discritely sampled diffusions: a closed- froma approximation approach. *Econometrica* (2002), 223–262.

[2] Anderson, T. *An introduction to multivariate statistical analysis, Second Edition.* New York, 1984.

[3] Bibby, B., and Sorensen, M. Martingale estimation functions for discretely observed diffusion processes. *Bernoulli 1 (1/2)* (1995), 17–39.

[4] EGOROV, A., LI, H., AND XU, Y. Maximum likelihood estimation of time-homogeneous diffusions. *Journal of Econometrics 114* (2003), 107–139.

[5] FERRANTE, L., BOMPADE, S., POSSATI, L., LEONE, L., AND MONTANARI, M. A stochastic formulation of the Gompertzian growth model for in vitro bactericidad kinetics: Parameter estimation and extinction probability. *Biometrical Journal 47(3)* (2005), 306–318.

[6] FORMAN, J., AND SORENSEN, M. The pearson diffusions: A class of statistically tractable diffusion processes. *Scandinnavian Journal of Statistics 35* (2005), 438–465.

[7] GIOVANIS, A., AND SKIADAS, C. A stochastic logistic innovation diffusion-model studying the electricity consumption in greece and the united states. *Technological Forecasting and Social Change 61* (1999), 253–264.

[8] GUTIÉRREZ, R., GUTIÉRREZ-SÁNCHEZ, R., AND NAFIDI, A. Electricity consumption in morocco: Stochastic gompertz diffusion analysis with exogenous factors. *Applied Energy 83* (2006), 1139–1151.

[9] GUTIÉRREZ, R., GUTIÉRREZ-SÁNCHEZ, R., AND NAFIDI, A. Trend analysis and computational statistical estimation in a stochastic rayleigh model: Simulation and application. *Mathematics and Computers in Simulation 77* (2008), 209–217.

[10] GUTIÉRREZ, R., GUTIÉRREZ-SÁNCHEZ, R., AND NAFIDI, A. Trend analysis using nonhomogeneous stochastic diffusion processes. emission of CO_2; kyoto protocol in spain. *Stochastic Environmental Research and Risk Assessment 22* (2008), 57–66.

[11] GUTIÉRREZ, R., GUTIÉRREZ-SÁNCHEZ, R., AND NAFIDI, A. The trend of the total stock of the private car-petrol in spain: Stochastic modelling using a new gamma diffusion process. *Applied Energy 86* (2009), 18–24.

[12] GUTIÉRREZ, R., GUTIÉRREZ-SÁNCHEZ, R., NAFIDI, A., AND EL MELLIANI, S. A non homogenous stochastic vasicek diffusion model: Application to global CO_2 emission in morocco. In *MAMERN-07* (2007), B. Amaziane, D. Barrera, and D. Sbibih, Eds., pp. 323–329.

[13] GUTIÉRREZ, R., GUTIÉRREZ-SÁNCHEZ, R., NAFIDI, A., AND MERBOUHA, A. A stochastic diffusion model based on the gamma density: Statistical inference. *Monografías del Seminario Matemático García de Galdeano 34* (2008), 117–125.

[14] 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. *Applied Stochastic Models in Business and Industry 15* (1999), 325–332.

[15] LIPTER, R., AND SHIRYAYEV, A. *Statistics of Random Processes II. Applications*. Springer-Verlag, New York, 1978.

[16] PICCHINI, U., DITLEVSEN, S., AND DE GAETANO, A. Maximum likelihood estimation of a time-inhomogeneous stochastic differential model of glucose dynamics. *Mathematical Medicine and Biology 25* (2008), 141–155.

[17] PRAKASA-RAO, B. *Statistical inference for diffusion type processes*. Arnold, London and Oxford University Press, New York, 1999.

[18] SEARLE, S. *Linear Models*. John Wiley & Sons, 1971.

[19] WONG, E., AND HAJEK, B. *Stochastic processes in engineering systems*. Springer-Verlag, New York, 1985.

A. Nafidi
Université Hassan 1er
Ecole Supérieure de Technologie- Berrechid
B.P. 218, Berrechid, Maroc
nafidiah@ugr.es

R. Gutiérrez R. and R. Gutiérrez-Sánchez
Department of Statistics
and Operations Research
University of Granada, Facultad de Ciencias
Campus de Fuentenueva s/n,
18071 Granada, Spain
rgjaimez@ugr.es and ramongs@ugr.es