Monografías del Seminario Matemático García de Galdeano 34, 117–125 (2008)

A STOCHASTIC DIFFUSION MODEL BASED ON THE GAMMA DENSITY FUNCTION: STATISTICAL INFERENCE

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

Abstract.

In this paper we propose a new stochastic diffusion process based on the Gam sity function (the trend of this process is proportional to the Gamma density f Such a process can be considered as an extension of the lognormal diffusion From the corresponding Ito's stochastic differential equation and the Kolmogo tions, we obtain the analytical expression of this process, its transition probabili function and its trend function. We then study the statistical inference on t Based on discrete sampling, we obtain the likelihood estimators of the parame distribution and the confidence interval of the parameters. Finally, an applica on simulated data is considered.

Keywords: Gamma diffusion process, Trend function, Statistical inference i process, Discrete sampling, Simulation.

AMS classification: 60J60, 62M05

§1. Introduction

Diffusion processes, which play a fundamental role in stochastic modelling, ... either from the standpoint of the corresponding Ito stochastic differential equations or from that of the associated Kolmogorov (Fokker-Planck and backward) partial differential equations. This role can be seen in applications in fields such as biology, physics, demography, economics, finance and environmental sciences.

Questions of statistical inference and parameter estimation in these processes have received considerable attention in recent years, both when the process is observed continuously and when discretely. In most cases, parameter estimation is based on approximating the maximum likelihood (ML) methodology. A large body of literature addresses this question, and important studies in general or in particular cases include Bibby and Sorensen [4]; Durham and Gallant [6], Eugene [8] and the extensive review given in Prakasa-Rao [16] and Kutoyants [15], without overlooking the early works focused on other methodologies, such as the generalized method of moments by Chan et al. [5], the non parametric method of Arapis and Gao [2], and a method based on Bayesian analysis by Elerian et al. [7].

Diffusion processes, with their associated statistical methodology, have been applied, for example, in the growth of cancer cells by Ferrante et al. [9] and by Albano and Giorno [1], using homogeneous and non homogeneous Gompertz processes respectively. They have been applied in the field of energy consumption by Skiadas et al. [20] and by Giovanis et al. [10] in the case of electrical consumption in Greece and the USA, using logistic and Bass diffusion processes, respectively. Gutiérrez et al. [11] applied a Gompertz model to the modelling of natural gas consumption in Spain, and Gutiérrez et al. [12] used one to study electricity consumption in Morocco. In environmental sciences, diffusion processes have been used by Gutiérrez et al. [13] to model CO_2 emissions in Spain (using a diffusion process with cubic drift) and by Gutiérrez et al. [14] with respect to the emissions of greenhouse gases attributable to the activities of land transport (using an inverse CIR diffusion process).

In the present study, we examine a new type of stochastic gamma diffusion process (SGDP), based on the methods of growth function formulation (see Skiadas [19]). The SGDP presented is obtained by adding white noise fluctuations to the ordinary differential equation characterizing the density function of a Gamma distribution. The paper is organized as follows: in the next section, we introduce the model from the point of view of the Kolmogorov equation as the stochastic differential equation. From this initial approach, we obtain the transition probability density function (TPDF) and the moments of the process. In Section 3, we study the statistical inference on the model. Based on discrete sampling, we obtain the likelihood estimators of the parameters, their distribution and corresponding confidence interval. In Section 4, we study an application based on simulated data.

§2. The SGDP and its characteristics

2.1. The model

The SGDP can be defined as a one-dimensional diffusion process $\{x(t), t \in [t_1, T], t_1 > 0\}$ on $(0, \infty)$ with infinitesimal moments (drift parameter and diffusion coefficients) given by

$$\begin{aligned} a(x,t) &= \left(\frac{\alpha}{t} - \beta\right) x \\ b(x,t) &= \sigma^2 x^2 \end{aligned} \tag{1}$$

where α, β and σ are time-independent real parameters and to be estimated.

The TPDF of the process is denoted by $f(y,t \mid x,s)$, and then f is the unique solution to the following equations, known as the forward Fokker Planck and the backward Kolmogorov equations:

$$\begin{aligned} \frac{\partial f(y,t \mid x,s)}{\partial t} &= -\frac{\partial}{\partial x}[a(y,t) f(y,t \mid x,s)] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[b(y,t) f(y,t \mid x,s)] \\ \frac{\partial f(y,t \mid x,s)}{\partial s} &= -a(x,t)\frac{\partial f(y,t \mid x,s)}{\partial x} - \frac{1}{2} b(x,t)\frac{\partial^2 f(y,t \mid x,s)}{\partial x^2} \end{aligned}$$

with the delta type initial condition $\lim_{t \to s} f(y,t \mid x,s) = \delta(y - x)$

Remark 1. Alternatively, the above-defined process can be considered as the solution to Itô's stochastic differential equation (SDE)

$$dx(t) = \left(\frac{\alpha}{t} - \beta\right) x(t)dt + \sigma x(t)dw(t) \qquad (2)$$

with the initial condition $P[x(t_1) = x_{t_1}] = 1$, x_{t_1} is positive real, and where $w(t)$ is a one-dimensional standard Wiener process.

Note that in the absence of white noise (i.e $\sigma = 0$), the solution of the ordinary differential equation associated with the SDE Eq. (2) is $x(t) = kt^{\alpha}e^{-\beta t}$, which is proportional to the Gamma density function.

Remark 2. For $\alpha = 0$ and $\beta < 0$, we find the stochastic homogeneous lognormal diffusion process studied by Tintner et al. [22].

2.2. The TPDF and the moments

The standard solution to the above equations can be obtained using Ricciardi's theorem (see Ricciardi [17]) for the transformation of the diffusion process into the Wiener process. The infinitesimal moments Eq.(1) verify the conditions of the cited theorem ; therefore such a transform exists and has the following form:

$$\phi(t) = t$$
$$\psi(x,t) = \frac{1}{\sigma}\log(x) - \frac{1}{\sigma}\left[\alpha\log(t) - (\beta + \frac{\sigma^2}{2})t\right].$$

From the above, the TPDF for the considered process is

$$f(y,t \mid x,s) = \left[2\pi\sigma^2(t-s)\right]^{-1/2} \exp\left(-\frac{\left[\log(y) - \mu(s,t,x)\right]^2}{2\sigma^2(t-s)}\right) \qquad (3)$$

where $\mu(s,t,x)$ is given by

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

This TPDF is the density of the one-dimensional lognormal distribution: $\Lambda_1[\mu(s,t,x), \sigma^2(t-s)]$.

Taking into account that with the initial condition $P(x(t_1) = x_1) = 1$, the random variable $x(t)$ is distributed as $\Lambda_1\left[\mu(t_1,t,x_{t_1}), \sigma^2(t-t_1)\right]$ and bearing in mind the properties of this distribution, the rth moment (r is a non negative integer) of the process is expressed by

$$\mathbb{E}(x^r(t)) = \exp\left(r\mu(s,t,x_{t_1}) + \frac{r^2\sigma^2}{2}(t-t_1)\right)$$
$$= x_{t_1}^{x}\left(\frac{t}{t_1}\right)^{r\alpha} e^{-r\beta(t-t_1)} \exp\left\{\frac{r(r-1)}{2}\sigma^2(t-t_1)\right\}.$$

From which, the trend function of the process is

$$\mathbb{E}\left(x(t)\right) = \frac{x_{t_1} e^{\beta t_1}}{t_1^{\alpha}} t^{\alpha} e^{-\beta t} \tag{4}$$

and the variance function is

$$\text{Var}(x(t)) = x_1^2 \left(\frac{t}{t_1}\right)^{2\alpha} e^{-2\beta(t-s)} \left(e^{\sigma^2(t-t_1)} - 1\right).$$

Note also that the trend function Eq.(4) is proportional to the density function of the Gamma distribution.

§3. Statistical inference on the model

3.1. Parameter likelihood estimation

Let us consider a discrete sampling of the process, that is, for fixed times $t_1, t_2; \ldots, t_n$ we observe the variables $x(t_1), \ldots, x(t_n)$ whose values $x_1, \ldots, x_n$ make up the basic sample from which the inferential process is carried out. From Eq.(2), and assuming the initial condition $\mathrm{P}[x(t_1) = x_1] = 1$, the likelihood function for the sample can be expressed as:

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

Now we transform the observed values $x_1, \ldots, x_n$ as follows:

$$v_i = (t_i - t_{i-1})^{-1/2} \left(\log(x_i) - \log(x_{i-1})\right), \quad i = 2, \ldots, n.$$

Then, the likelihood function for the transformed sample is

$$\mathbb{L}(v_2, \ldots, v_n, \mathbf{a}, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n-1}{2}}} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{V} - \mathbf{U}'\mathbf{a})'(\mathbf{V} - \mathbf{U}'\mathbf{a})\right) \tag{5}$$

where $\mathbf{V} = (v_2, \ldots, v_n)'$, $\mathbf{a} = \left(\alpha, -(\beta + \sigma^2/2)\right)'$ and $\mathbf{U}$ is the $2 \times (n-1)$ matrix, whose rank is assumed to be 2, and is given by: $\mathbf{U} = (\mathbf{u}_2, \ldots, \mathbf{u}_n)$, and where, for $i = 2, \ldots, n$, $u_i = (t_i - t_{i-1})^{-1/2} \left(\log(t_i/t_{t-1}), t_i - t_{i-1}\right)'$.

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

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

After various algebraic operations (omitted here), the maximum likelihood estimators (MLEs) of the parameters are found to be:

$$\widehat{\mathbf{a}} = (\mathbf{U}\mathbf{U}')^{-1}\mathbf{U}\mathbf{V} \tag{6}$$
$$(n-1)\hat{\sigma}^2 = \mathbf{V}'\mathrm{H}_{\mathbf{U}}\mathbf{V} \tag{7}$$

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

3.2. Properties of likelihood estimators

3.2.1. Distributions of MLEs

Eq.(5) is also the density function of the random vector $\mathbf{V}$, and can be rewritten as

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

from which, we deduce that

$$\mathbf{V} \sim \mathcal{N}_{n-1}\left[\mathbf{U}'\mathbf{a}, \sigma^2 \mathbf{I}_{n-1}\right].$$

The rank of the matrix $\mathbf{U}$ is assumed to be 2. Then the matrix$(\mathbf{U}\mathbf{U}')^{-1}\mathbf{U}$ has the same rank, and therefore

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

and finally, we have

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

Using a known results in multivariate analysis (see for example [18]) we have:

$$\frac{\mathbf{V}'}{\sigma} \mathrm{H}_{\mathrm{U}} \frac{\mathbf{V}}{\sigma} = \frac{(n-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{(n-3)}$$

and $\widehat{\mathbf{a}}$ and $\hat{\sigma}^2$ are independently distributed.

3.2.2. Sufficiency and Completeness

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

$$L_{v_2,\ldots,v_n}(\mathbf{a}, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n-1}{2}}} \exp\left(-\frac{1}{2\sigma^2}\left[(n-1)\hat{\sigma}^2 + (\widehat{\mathbf{a}} - \mathbf{a})'\mathbf{U}\mathbf{U}'(\widehat{\mathbf{a}} - \mathbf{a})\right)\right]\right)$$

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

The completeness follows by means of reasoning similar to that established for the maximum likelihood estimators of the parameters of the multivariate normal distribution (see, for example, Anderson [3]).

As the estimators $\widehat{\mathbf{a}}$ and $\dfrac{(n-1)\hat{\sigma}^2}{(n-3)\sigma^2}$ are sufficient for $\mathbf{a}$ and σ^2 respectively, then we can deduce that they are the UMVUE.

3.3. Parameter confidence intervals

The $\gamma\%$ confidence interval for the parameter σ^2 is given, by

$$\left(\frac{(n-1)\hat{\sigma}^2}{{\chi^2}_{n-3,\frac{\gamma}{2}}}, \frac{(n-1)\hat{\sigma}^2}{{\chi^2}_{n-3,1-\frac{\gamma}{2}}}\right) \tag{8}$$

and the $\gamma\%$ confidence interval for the parameter α is given, by

$$\mathbf{P}\left(\alpha \in \left[\hat{\alpha} \pm \hat{\sigma}\left(\frac{n-1}{n-3}A_{11}F_{1,n-3,\gamma}\right)^{1/2}\right]\right) = 1-\gamma \tag{9}$$

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 $(\mathbf{UU}')^{-1}$ and where $\mathbf{UU}'$ is

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

§4. Simulation

The trajectory of the model can be obtained by simulating the exact solution of SDE Eq.(2). This solution can be obtained by means of Itô's formula applied to the transform $\log(x(t))$, from which we obtain the following SDE

$$d\left[\log(x(t))\right] = [\alpha - (\beta + \sigma^2/2)]dt + \sigma dw(t).$$

By integrating and simplifying, the solution of the SDE Eq.(2) is given as

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

From this explicit solution, the simulated trajectories of the process are obtained from the following discretizing time interval $[t_1, T]$: $t_i = t_1 + (i-1)h$, for $i = 1, \ldots, N$ (N is an integer and h is the discretization step), taking into account that the random variable in the latter expression $\sigma(w(t) - w(t_1))$ is distributed as a one-dimensional normal distribution $\mathcal{N}_1\left(0, \sigma^2(t-t_1)\right)$.

In this simulation, we consider M process trajectories, each of which has N observations, estimating the parameters by means of equations Eq.(6) and Eq.(7). In total M estimators are obtained for each parameter (i.e. one vector of M components), from which we compute the sample mean and the standard error (SE) of each estimator.

Let us now study the evolution of the mean and the standard error of the estimators with respect to the variation in the number N and h. The results of this study are shown in Table 1.

A MatLab program was implemented to carry out the calculation required for this study. The true parameter values considered in this simulation are $\alpha = 2$, $\beta = 0.5$, $\sigma = 0.1$ and the start point is $x_{t_1} = 2$, and $t_1 = 0.5$.

Figure (Fig.1) shows some simulated trajectories of the process and the estimated trend function (ETF) of the process obtained using the Zehna theorem, replacing the parameters

Table 1: Mean and standard error of the estimators

h	num.obs.	mean($\hat{\alpha}$)	SE($\hat{\alpha}$)	mean($\hat{\beta}$)	SE($\hat{\beta}$)	mean($\hat{\sigma}$)	SE($\hat{\sigma}$)
0.05	100	2.0071	0.2279	0.4983	0.0921	0.0991	0.0071
0.05	500	2.0059	0.1336	0.5003	0.0273	0.0998	0.0029
0.05	1000	1.9893	0.1259	0.4984	0.0176	0.0998	0.0023
0.1	100	1.9981	0.1797	0.5012	0.0543	0.0986	0.0072
0.1	500	1.9957	0.1220	0.4996	0.0171	0.0999	0.0033
0.1	1000	1.9989	0.1177	0.5005	0.0111	0.0997	0.0022
0.5	100	2.0016	0.1250	0.4995	0.0182	0.0983	0.0069
0.5	500	1.9944	0.1059	0.5001	0.0066	0.0996	0.0032
0.5	1000	1.9938	0.1097	0.5002	0.0043	0.0997	0.0023
1	100	2.0018	0.1171	0.5008	0.0118	0.0989	0.0071
1	500	2.0083	0.1010	0.5004	0.0046	0.0996	0.0032
1	1000	2.0027	0.1049	0.5002	0.0033	0.0999	0.0023

by their estimators in Eq.(4). In this figure (Fig.1 the dark line represents the ETF. In this simulation we assume $h = 0.1$, $N = 100$ and $M = 20$.

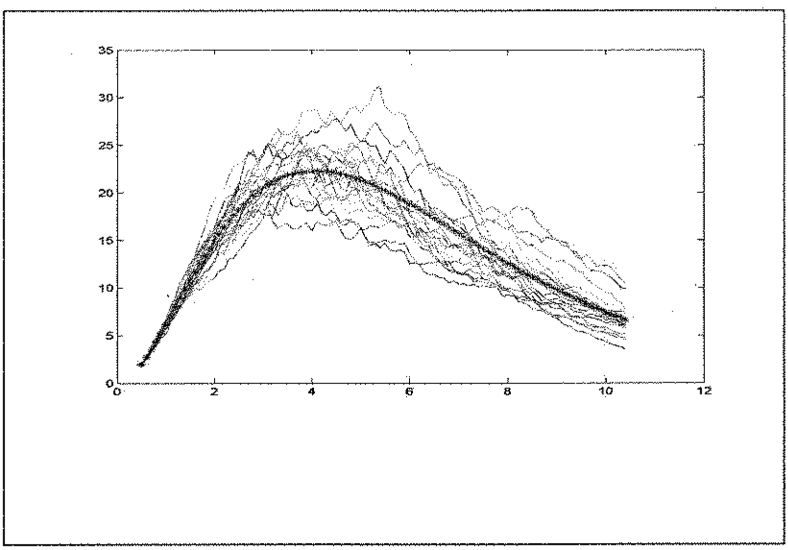

Figure 1: Simulated trajectories of the SGDP versus to ETF

Acknowledgements

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] ALBANO, G., AND GIORNO, V. A stochastic model in tumor growth. *Journal of Theoretical Biology 242*, 2 (2006), 329-336.

[2] ARAPIS, M., AND GAO, J. Empirical comparison in short-terme interest rate models using nonparametric methods. *Journal of Financial Econometrics 4*, 2 (2006), 310-345.

[3] ANDERSON, T.W. *An introduction to multivariate statistical analysis.* Second Edition. Wiley. New York, 1984.

[4] BIBBY, B.M., AND SORENSEN, M. Martingale estimation functions for discretely observed diffusion processes, *Bernoulli 1*, 1/2 (1995), 17-39.

[5] CHAN, K.C., KAROLYI, G.A., LONGSTAFF, F.A., AND SANDERS, A.B. An empirical comparison of alternative models of the short-term interest rate. *The Journal of Finance 3* (1992), 1209-1227.

[6] DURHAM, G.B., AND GALLANT, A.R. Numerical techniques for maximum likelihood estimation of the continuos- times diffusion processes. *Journal of Business and Economic Statistics 20*,3 (2002), 297-316.

[7] ELERIAN, O., CHIB, S., AND SHEPHAD, N. Likelihood inference for discretely observed non-linear diffusions. *Econometrica 69* (2006), 959-993.

[8] EUGENE, M.C. Maximum likelihood estimations of a class of one-diminsional stochastic differential equation models from discrete data. *Journal of Time Series Analysis 22*,5 (2000), 505-515.

[9] FERRANTE, L., BOMPADE, S., POSSATI, L., AND LEONE, L. 2000. Parameter estimation in a Gompertzian stochastic-model for tumor growth. *Biometrics 56* (2000), 1076-1081.

[10] GIOVANIS, A.N., AND SKIADAS, C.H. 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.

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

[12] 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.

[13] GUTIÉRREZ, R., GUTIÉRREZ-SÁNCHEZ, R., NAFIDI, A., AND RAMOS, E. A diffusion model with cubic drift: statistical and computational aspects and application to modelling the global CO_2 emission in Spain. *Environmetrics 18* (2007), 55-69.

[14] GUTIÉRREZ, R., GUTIÉRREZ-SÁNCHEZ, R., AND NAFIDI, A. Emissions of Greenhouse gases attributable to the activities of the land transport: Modelling and analysis using I-CIR stochastics diffusion. The case of Spain. *Environmetrics 2008*, DOI:10.1002/env.862

[15] KUTOYANTS, Y.A. *Statistical inference for ergodic diffusion processes*. Springer Series in Statistics. Springer-Verlag, London, 2004.

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

[17] RICCIARDI, L.M. On the transformation of difusion processes into the Wiener processes. *Journal of Mathematical Analysis and Applications 54* (1976), 185-99.

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

[19] SKIADAS, C.H. Method of Growth functions formulation . In *Selected Topics on Stochastic Modelling*, editors R. Gutiérrez, and M. Valderrama, Granada University, Granada, 1994, 296-310.

[20] SKIADAS, C.H., AND GIOVANI, A.N. A stochastic Bass innovation diffusion model for studying the growth of electricity consumption in Greece. *Applied Stochastics Model and Data Analysis 13* (1997), 85-101.

[21] SRIVASTAVA, M.S., AND KHATRI, C.G. *An introduction to multivariate analysis*. North-Holland Scientific Publisshers, New-York, 1979.

[22] TINTNER G, SENGUPTA JK. *Stochastic Economics*. Academic Press, 1972.

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

Merbouha A.
Department of Applied Mathematics and Informatics
University of Sultan Mouly Slimane, Faculté des Sciences et Techniques
B.P: 523, Béni Mellal, Morocco
merbouhak@yahoo.fr