Use of lognormal diffusion process with exogenous factors without external information

Ramón Gutiérrez, Nuria Rico, Patricia Román, Desirée Romero and Francisco Torres.
Department of Statistics and Operational Research
University of Granada, Spain
Avda. Fuentenueva s/n. 18071 Facultad de Ciencias, Granada, Spain.
email: rgjaimez@ugr.es, nrico@ugr.es, proman@ugr.es, deromero@ugr.es, fdeasis@ugr.es

Abstract

In this paper we propose the use of polynomial exogenous factors in the lognormal diffusion process as an approximation in the case that no external information about the process is available or the functional form of the exogenous factors is unknown. We illustrate this situation with an example in environmental sciences.

1 Introduction

The use of diffusion processes with exogenous factors in their trend is common in many application fields. The reason of its application is the usual presence of deviations of the observed data with respect to the trend of some known homogenous diffusion process, in some time intervals. These factors are time dependent functions that allow, on one hand, a best fit to the data and, on the other hand, an external control to the process behaviour. The factors must be totally or partially known, that is, their functional form or some aspects about their time evolution must be available.

One of the mentioned kind of processes is the lognormal diffusion process with exogenous factors. This process is defined as a diffusion process $\{X(t) : t_0 \leq t \leq T\}$ with infinitesimal moments $A_1(x,t) = xh(t)$ and $A_2(x,t) = \sigma^2 x^2$, where $\sigma > 0$ and h is a continuous function in $[t_0, T]$ containing the external information to the process. Frequently, there are more of one external information sources, so it is usual to take h as a linear combination of continuous functions F_i, $i = 1, \ldots, q$ (from now, factors).

Some topics, as the inference and first-passage-time through varying boundaries, are already studied. With respect to the inference, some of the treated aspects are the maximum likelihood (ML) estimation of the weights of the factors and the constant σ^2 [Gutiérrez *et al.*, 1997; 1999] as well as the estimation of certain interesting parametric functions that include, as particular cases, the trend, mode and quantiles functions (and their conditional versions). For these parametric functions, their ML and uniformly unbiased minimum variance (UMVU) estimators are been obtained [Gutiérrez *et al.*, 2001a; 2001b; 2003a]. It is important to note that for the inferential process is not necessary the functional form of the exogenous factors but the value of their integrals between two time values included in the interval $[t_0, T]$ (analogous to some interpolation problems) whereas the probability density function of the first-passage-time of the process through a boundary includes the factors explicitly [Gutiérrez *et al.*, 1999].

However, it is usual in practice that the functional form of the exogenous factors is not available or, indeed, the external influences are unknown. This paper shows an approximation in such situations by using polynomial exogenous factors, that is, in this case $h(t) = \sum_{j=0}^{k} \beta_j^{k)} P_j(t)$, where P_j is a j-degree polynomial, $j = 1, \ldots, k$ and $P_0(t) = 1$.

2 An illustrative example

Stern and Kaufmann [1998] published a study on global man-made emissions of methane, from 1860 to 1994. In this study the authors provided the estimation, in the mentioned period, for total anthropogenic emissions as well for seven component categories. The global emissions is the amount of the individual components whereas the partial emissions were estimated from other variables such as population or coal production. The final goal of this process was to establish a first approximation to actual emissions and, as the authors noted, the estimates for the 1980s of total anthropogenic methane emissions, and other related to fossil fuels, were consistent with estimates from the Intergovernmental Panel on Climate Change.

Figure 1 illustrates the obtained chronological series, showing an exponential trend. For this reason we can try to fit a diffusion model to the observed data, being the lognormal diffusion process a valid, in advance, candidate.

However, after using the homogeneous lognormal diffusion process to fit the observed data (figure 2), we can note that the estimated trend shows deviations from the observed values. For this, it is reasonable to think that they are some external influences to the process that the homogeneous model does not consider. These influences must be time-dependent variables that affect the trend of the process. But, what influences?, and, what can we do if these external variables are not known?

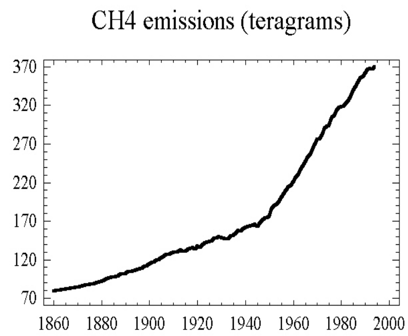

Figure 1: Global historical CH_4 emissions in teragrams (1 Tg= 10^{12}gr.)

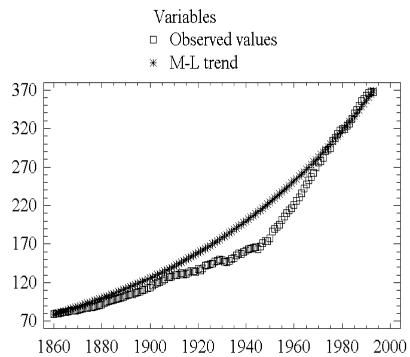

Figure 2: Estimated trend from a homogeneous lognormal diffusion process.

An approximation to solve this situation is showed later in section 6 by using polynomials to approach the unknown exogenous factors. Previously, in the next sections, we give the methodology and a brief summary of the inferential procedure on the process that will be used later.

3 Methodology

In Gutiérrez *et al.* [1999] it was proposed a methodology for building a theoretical model of lognormal diffusion process with exogenous factors that fits the data, i. e., a method for searching the h function. The development realized in that paper provided for the case that external information to the process is available. Such information is introduced in the model by means of external variables to the system that have been observed at the same time interval as the endogenous variable.

In the actual situation, we suppose that they are not additional information but only we have values $x_1, \ldots, x_n$ of the endogenous variable in times $t_1, \ldots, t_n$. If we suppose $\mathrm{P}[X(t_1) = x_1] = 1$, it is known that $\ln\left(\frac{\mathrm{E}[X(t)]}{x_1}\right) = \int_{t_1}^{t} h(s)ds = H(t)$ and we can consider the values

$$f_i = \ln\left(\frac{x_i}{x_1}\right), \quad i = 1, \ldots, n \qquad (1)$$

as an approximation to $H(t_i)$. So, with these values we fit a $(k+1)-$degree polynomial $P(t) = \sum_{i=1}^{k+1} a_i Q_i(t)$ and we can approach the lognormal diffusion process $\{X(t) : t_0 \leq t \leq T\}$ by the lognormal diffusion process with polynomial exogenous factors $\{X^{k)}(t) : t_0 \leq t \leq T\}$ whose infinitesimal moments are

$$\begin{aligned} A_1^{k)}(x,t) &= x\left[\sum_{j=0}^{k} \beta_j^{k)} P_j^{k)}(t)\right] \\ A_2^{k)}(x,t) &= \sigma_k^2 x^2 \end{aligned} \qquad (2)$$

taking, in this case, $P_j^{k)}(t) = a_{j+1}Q'_{j+1}(t)$, $j = 1, \ldots, k$, $P_0^{k)}(t) = 1$. We note that we do not take $\sum_{i=1}^{k} P_i^{k)}(t)$ as only one factor because it is possible that, after a posterior study, some of the $P_i^{k)}$ functions were not relevant.

4 The model

Let $\{X^{k)}(t);\ t_0 \leq t \leq T\}$ be the lognormal diffusion process with polynomial exogenous factors whose infinitesimal moments are given by (2) where $P_j^{k)}$ is a j-degree polynomial, $j = 1, \ldots, k$, $P_0^{k)}(t) = 1$ and $\sigma_k > 0$.

Some interesting characteristics of the process are the mean, mode and percentile functions (and their conditional versions). The point estimation of the mean and mode functions will allow to make point estimations for the expected value and most probably value of the endogenous variable, respectively. On the other hand, the obtention of the confidence bands for the mean and mode functions will lead to make confidence interval for the expected value and most probably value of the endogenous variable, respectively. Finally, the point estimation of the percentile functions for suitable α values, provides, for each t (non conditional versions) and for each t given s (conditional versions), the estimation of an interval which the endogenous variable belongs to with a predetermined probability.

These characteristics can be expressed in terms of a kind of parametric functions

$$\theta_k(C, \mathbf{A}(t,s), B(t,s), l) = C\mathrm{e}^{\mathbf{A}'(t,s)\mathbf{a}_k + B(t,s)\sigma_k^l}$$

with $C > 0$, $l \in \mathbb{N}$, $\mathbf{A}(t,s) \in \mathbb{R}^{k+1}$ and $B(t,s) \in \mathbb{R}$. Concretely, if $\mathrm{P}[X^{k)}(t_0) = x_0] = 1$ and, for the conditional versions, we assume as known the values x_s taken for $X^{k)}(s)$, $s < t$ (as it is usual in forecasting), then

$$
\begin{aligned}
m_k(t) =& \ \mathrm{E}[X^{k)}(t)] \\
=& \ \theta_k\left(x_0, \bar{\mathbf{u}}_k(t), \tfrac{1}{2}(t-t_0), 2\right) \\
m_k(t|s) =& \ \mathrm{E}[X^{k)}(t)|X^{k)}(s)] \\
=& \ \theta_k\left(x_s, \bar{\mathbf{u}}_k(t,s), \tfrac{1}{2}(t-s), 2\right) \\
Mo_k(t) =& \ \mathrm{Mode}[X^{k)}(t)] \\
=& \ \theta_k\left(x_0, \bar{\mathbf{u}}_k(t), -(t-t_0), 2\right) \\
Mo_k(t|s) =& \ \mathrm{Mode}[X^{k)}(t)|X^{k)}(s)] \\
=& \ \theta_k\left(x_s, \bar{\mathbf{u}}_k(t,s), -(t-s), 2\right) \\
P_\alpha^k(t) =& \ \alpha\text{th-percentile}[X^{k)}(t)] \\
=& \ \theta_k\left(x_0, \bar{\mathbf{u}}_k(t), z_\alpha\sqrt{t-t_0}, 1\right) \\
P_\alpha^k(t|s) =& \ \alpha\text{th-percentile}[X^{k)}(t)|X^{k)}(s)] \\
=& \ \theta_k\left(x_s, \bar{\mathbf{u}}_k(t,s), z_\alpha\sqrt{t-s}, 1\right)
\end{aligned}
$$

where $t > s$ in the conditional functions and

- $\mathbf{a}_k = (a_0^{k)}, \beta_1^{k)}, \ldots, \beta_k^{k)})'$ with $a_0^{k)} = \beta_0^{k)} - \frac{1}{2}\sigma_k^2$.
- $\bar{\mathbf{u}}_k(t,s) = \left(t-s, \int_s^t P_1^{k)}(\tau)d\tau, \ldots, \int_s^t P_k^{k)}(\tau)d\tau\right)'$, $\bar{\mathbf{u}}_k(t) = \bar{\mathbf{u}}_k(t, t_0)$
- z_α is the αth-percentile of a standard normal distribution.

The point estimation of the θ_k functions (maximum-likelihood and minimum variance unbiased estimation) was developed in Gutiérrez *et al.* [2001a] for $l = 2$ and for a generic l (with the particular case of $l = 1$) in Gutiérrez *et al.* [2003a].

On the other hand, the mean and mode functions (and their conditional versions) can be written in the form $\exp\left(\mu_k(t,s) + \lambda\sigma_k^2(t,s)\right)$ with

	Mean	Cond. Mean
$\mu_k(t,s)$	$\ln x_0 + \bar{\mathbf{u}}_k'(t)\mathbf{a}_k$	$\ln x_s + \bar{\mathbf{u}}_k'(t,s)\mathbf{a}_k$
λ	$1/2$	$1/2$
$\sigma_k^2(t,s)$	$(t-t_0)\sigma_k^2$	$(t-s)\sigma_k^2$

	Mode	Cond. Mode
$\mu_k(t,s)$	$\ln x_0 + \bar{\mathbf{u}}_k'(t)\mathbf{a}_k$	$\ln x_s + \bar{\mathbf{u}}_k'(t,s)\mathbf{a}_k$
λ	-1	-1
$\sigma_k^2(t,s)$	$(t-t_0)\sigma_k^2$	$(t-s)\sigma_k^2$

being the problem of building confidence bands for them solved in Gutiérrez *et al.* [2003a].

5 Inference on the model

Now we give a brief summary of the inferential procedure. Let us consider $x_1, \ldots, x_n$ the observed values obtained by discrete sampling of the process in times $t_1, \ldots, t_n$, $(n > k+2)$, and suppose $\mathrm{P}\left[X^{k)}(t_1) = x_1\right] = 1$. After transforming these values by means of $v_1 = y_1$ and $v_i = (t_i - t_{i-1})^{-1/2}\ln(y_i/y_{i-1})$, $i = 2, \ldots, n$, the ML estimators of the parameters $\mathbf{a}_k$ and σ_k^2 (see Gutiérrez *et al.* [1997] for the multivariate version of the process) are $\widehat{\mathbf{a}}_k = \mathbf{V}_k\mathbf{v}$ and $\widehat{\sigma}_k^2 = \frac{1}{n-1}\mathbf{v}'\mathbf{H}_k\mathbf{v}$.

In these last expressions we have denoted $\mathbf{v} = (v_2, \ldots, v_n)'$, $\mathbf{V}_k = (\mathbf{U}_k\mathbf{U}_k')^{-1}\mathbf{U}_k$ and $\mathbf{H}_k = \mathbf{I}_{n-1} - \mathbf{U}_k'\mathbf{V}_k$, being $\mathbf{U}_k$ the matrix, whose rank is assumed to be $k+1$, given by $\mathbf{U}_k = (\mathbf{u}_2^{k)}, \ldots, \mathbf{u}_n^{k)})$ with $\mathbf{u}_i^{k)} = (t_i - t_{i-1})^{-1/2}\bar{\mathbf{u}}^{k)}(t_i, t_{i-1})$.

From these estimators one can obtain the corresponding ML and UMVU estimators of the θ functions (see Gutiérrez *et al.* [2003a]). In addition, confidence bands for the functions $\exp\left(\mu_k(t,s) + \lambda\sigma_k^2(t,s)\right)$ for each t and s, can be calculated. For this, the procedure described in Gutiérrez *et al.* [2003a] can be followed. With this procedure we have, for each t and s, a confidence interval for $\mu_k(t,s) + \lambda\sigma_k^2(t,s)$. By repeating this process for each t in the time interval, for fixed s, with the same confidence level, and taking exponentials we obtain the desired confidence band for $\exp\left(\mu_k(t,s) + \lambda\sigma_k^2(t,s)\right)$.

Now, we focus our attention to the particular case treated in this paper, that is, the polynomial exogenous factors. One question that can be formulated in this model is what the optimal degree of the polynomial is, or equivalently, how many exogenous factors have to be taken. The answer is not immediate because it will depend on the data. But one can consider a similar strategy to that used in polynomial regression. For example, a forward procedure can be followed by introducing successively polynomials in the model. In such situation, it is specially interesting to get recursive expressions in order to employ, in the inferential process after including a $k+1$-degree polynomial $P_{k+1}^{k+1)}$, the calculations derived for the model containing a k-degree polynomial $P_k^{k)}$.

After introducing $P_{k+1}^{k+1)}$, the information about the exogenous factors is contained in the matrix $\mathbf{U}_{k+1} = (\mathbf{U}_k'|\mathbf{d}_{k+1})'$, where the $(i-1)$-th component of the d_{k+1} vector is given by $(t_i - t_{i-1})^{-\frac{1}{2}}\int_{t_{i-1}}^{t_i} P_{k+1}^{k+1)}(\tau)d\tau$, $i = 2, \ldots, n$. Taking into account the expression of the inverse of a partitioned matrix, and denoting by $e_{k+1} = \mathbf{d}_{k+1}'\mathbf{H}_k\mathbf{d}_{k+1}$, one obtains

Recursive expressions for the ML estimators of the process

$$
\widehat{\mathbf{a}}_{k+1} = \frac{1}{e_{k+1}}\left[\begin{pmatrix} e_{k+1}\mathbf{I}_{k+1} + \mathbf{V}_k\mathbf{d}_{k+1}\mathbf{d}_{k+1}'\mathbf{U}_k' \\ -\mathbf{d}_{k+1}'\mathbf{U}_k' \end{pmatrix}\widehat{\mathbf{a}}_k + \begin{pmatrix} -\mathbf{V}_k\mathbf{d}_{k+1} \\ 1 \end{pmatrix}\mathbf{d}_{k+1}'\mathbf{v}\right]
$$

$$
\widehat{\sigma}_{k+1}^2 = \widehat{\sigma}_k^2 - \frac{\mathbf{v}'\mathbf{H}_k\mathbf{d}_{k+1}\mathbf{d}_{k+1}'\mathbf{H}_k\mathbf{v}}{e_{k+1}}
$$

from we can obtain recursive expressions for the estimators of the parametric θ functions and the confidence bands for $\exp\left(\mu_k(t,s) + \lambda\sigma_k^2(t,s)\right)$ [Gutiérrez *et al.*, 2003b].

In this development it is necessary to note that the $\mathbf{U}_k$ matrix contains powers of the t_i, so the explained procedure is more reliable for low degrees because of the conditioning of the matrix $\mathbf{U}_k\mathbf{U}_k'$. In this sense, the above recursive expressions can ameliorate the accuracy of the calculations. Nevertheless, for some instances, orthogonal polynomials or more smoothed functions (like splines) will be preferred.

6 Application to the emissions of methane data

Returning to the example in section 2, the procedure proposed in this paper can be used with notable results under the assumption of no additional external information available about the process.

From the above theoretical development, the raised question is to fit a polynomial function to the transformed values given by (1). The main problem in this situation is to choose the degree of the polynomial. The selection criterion must not consider only the goodness of fit to the data because this property can not remain when one uses the model to make forecasts outside the rank of the observations. Therefore we need to find a balanced solution that provides for both questions: the fit of the model to the observed data and its possibilities for forecasting purposes. In this sense, we have fitted the model by using the data from 1860 to 1993 and then we have made forecast for 1994. This value is actually known and will serve to check the fitted model.

In our example, let us consider $\mathrm{P}[X(t_1) = x_1] = 1$ with $t_1 = 1860$ and x_1 the corresponding observed value. Now we consider $f_i = \ln(x_i/x_1)$ $(i = 1, \ldots, 134)$ as the values that we desire to fit. At this point it is important to note that as $f_1 = H(t_1) = 0$, we have taken $Q_i(t) = (t - 1860)^j$, $j = 1, 2, \ldots$ as the generators of the polynomials. Also, we can propose, in advance, a possible selection of the degree after fitting a regression line to the f_i data and computing the number of overcrossing with the observed trend. In this line, we can think in 4 or 5 as the possible degree, that is, the maximum degree of the exogenous factors will be 3 or 4 (after the posterior discussion about the choice of the degree, we will appreciate that this intuition is near to reality).

With these ideas in mind, we have realized an iterative procedure by fitting polynomials of degrees 2 to 6, and so we have obtained, from their derivatives, the exogenous factors related to the models $X^{k)}$, $k = 1, \ldots, 5$. Then, for each model, from the maximum-likelihood estimators of the parameters, we have obtained the ML and UMVU estimation of the mean, mode, 2.5th percentile and 97.5th percentile functions together with their conditional versions and confidence bands for the mean and mode functions. From those we have built the tables 3 and 4 that contain the values of the point and interval forecasts for the year 1994, respectively.

	$X^{1)}(t)$	$X^{2)}(t)$	$X^{3)}(t)$
ML Mean	375.033	374.852	373.142
ML Cond. Mean	373.114	372.935	371.296
ML Mode	369.262	369.088	367.591
ML Cond. Mode	373.071	372.892	371.255
UMVUE Mean	373.084	372.905	371.266
UMVUE Cond. Mean	373.114	372.935	371.295
UMVUE Mode	367.256	367.039	365.572
UMVUE Cond. Mode	373.070	372.890	371.252

	$X^{4)}(t)$	$X^{5)}(t)$
ML Mean	369.413	368.912
ML Cond. Mean	367.795	367.300
ML Mode	364.545	364.061
ML Cond. Mode	367.759	367.264
UMVUE Mean	367.768	367.271
UMVUE Cond. Mean	367.793	367.297
UMVUE Mode	362.733	362.214
UMVUE Cond. Mode	367.755	367.259

Table 3: Point forecasts for the year 1994 from the models $X^{k)}(t)$, $k = 1, \ldots, 5$.

	$X^{1)}(t)$	$X^{2)}(t)$	$X^{3)}(t)$
ML Percentiles	305.689 455.375	305.560 455.131	305.222 451.640
ML Cond. Percentiles	366.732 379.578	366.558 379.394	365.051 377.619
UMVUE Percentiles	303.474 453.803	303.099 453.893	302.584 450.799
UMVUE Cond. Percentiles	366.671 379.639	366.472 379.481	364.942 377.729
Mean Band	306.052 459.737	305.683 459.887	305.074 456.641
Cond. Mean Band	371.863 374.176	371.037 374.429	368.777 373.281
Mode Band	301.140 452.368	300.741 452.462	300.265 449.452
Cond. Mode Band	371.932 374.210	371.164 374.625	368.962 373.558

	$X^{4)}(t)$	$X^{5)}(t)$
ML Percentiles	305.875 442.220	305.520 441.544
ML Cond. Percentiles	361.973 373.686	361.492 373.177
UMVUE Percentiles	303.243 441.882	302.657 441.518
UMVUE Cond. Percentiles	361.848 373.810	361.342 373.323
Mean Band	305.410 447.072	304.811 443.811
Cond. Mean Band	364.800 370.158	363.625 370.204
Mode Band	301.112 440.788	300.495 440.453
Cond. Mode Band	365.032 370.502	363.918 370.635

Table 4: Interval forecasts for the year 1994 from the models $X^{k)}(t)$, $k = 1, \ldots, 5$.

The final question is to decide what the most appropriated model is. Since that the value of the global emissions of methane for the year 1994 is 371 Tg, we can conclude that the conditional estimations provide the better approximation and, in particular, those related to 3 degree model. As illustration, figure 5 shows, for the year 1994, the ML estimation of the conditional mean, ML estimation of the conditional percentiles (for $\alpha = 0.025$ and $\alpha = 0.975$) and 95% confidence interval for the conditional mean. Similar conclusions can be made by considering another estimations.

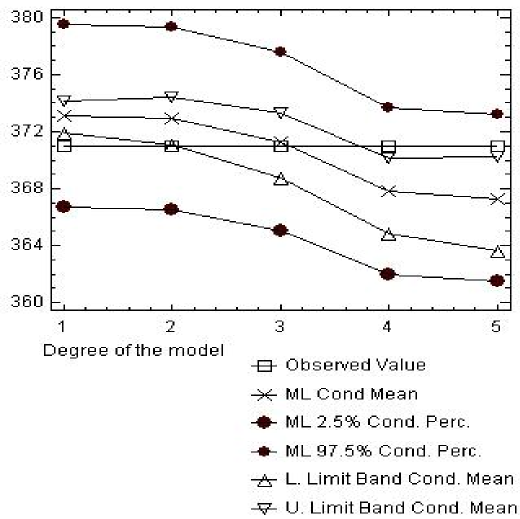

Figure 5: Estimated values of methane emissions for 1994

Furthermore, to confirm the above conclusion, we have realized, for each estimated model, 100 simulated paths following the recursive schema given by Rao *et al.* [1974] and 100 simulated values of the distribution of $X(1994)$ conditioned to $X(1993) = 367.2$. Finally, the median square error (MSE) has been calculated for the simulated distributions. The obtained results have been summarized in tables 6 and 7.

Model	Mean	5% trimmed mean
$X^{1)}(t)$	374.252	373.010
$X^{2)}(t)$	380.240	379.413
$X^{3)}(t)$	373.758	372.449
$X^{4)}(t)$	366.368	365.702
$X^{5)}(t)$	363.933	362.987

Table 6: Mean values for the year 1994 from the simulated paths

Model	Mean	5% trimmed mean	MSE
$X^{1)}(t)$	373.175	373.121	16.780
$X^{2)}(t)$	372.974	373.002	13.754
$X^{3)}(t)$	371.221	371.237	9.958
$X^{4)}(t)$	367.644	367.601	21.138
$X^{5)}(t)$	367.229	367.255	21.074

Table 7: Mean and MSE of the simulated values of the distributions of $X(1994)|X(1993) = 367.2$

One can observe that the model that offers the closest forecast to the observed value and which minimices the MSE is $X^{3)(t)}$, that is, a lognormal diffusion process with infinitesimal moments

$$A_1(x,t) = [0.0109222 - 0.000292911(t-1860)+ \\ + 6.98254 \times 10^{-6}(t-1860)^2- \\ -3.57962 \times 10^{-8}(t-1860)^3]\, x$$
$$A_2(x,t) = 7.457051282638727 \times 10^{-5} x^2$$

Figure 8 shows the observed values and the ML conditioned trend for the selected model $X^{3)}(t)$.

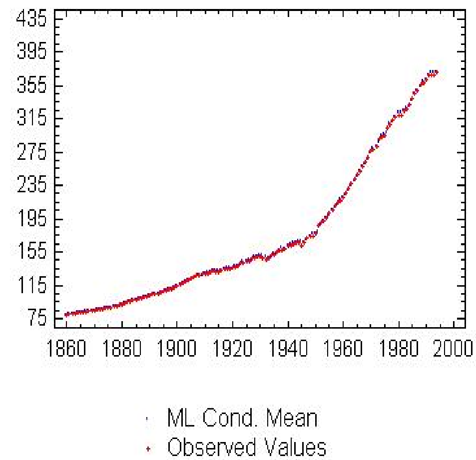

Figure 8: ML conditional trend estimated and observed values

This is the model that we propose in order to get the best forecast in 1994.

Acknowledgments

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

References

[Gutiérrez *et al.*, 1997] R. Gutiérrez, A. González and F. Torres. Estimation in multivariate lognormal diffusion process with exogenous factors. *Applied Statistics*, 46(1):140–146, 1997.

[Gutiérrez *et al.*, 1999] R. Gutiérrez, P. Román and F. Torres. 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*, 15:325–332, 1999.

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

[Gutiérrez *et al.*, 2001b] R. Gutiérrez, P. Román, D. Romero and F. Torres. Umvu estimation of the trend and covariance function for the univariate lognormal diffusion process with exogenous factors. In J. Jansen and Ch. Skiadas, editors, *Proc. of the 10th International Symposium on A.S.M.D.A.*, 1:516–521, Compiègne, June 2001.

[Gutiérrez *et al.*, 2003a] R. Gutiérrez, P. Román, D. Romero and F. Torres. Application of the univariate lognormal diffusion process with exogenous factors in forecasting. *Cybernetics & Systems*, 34(8):709–724.

[Gutiérrez *et al.*, 2003b] R. Gutiérrez, N. Rico, P. Román, D. Romero and F. Torres. Application of the lognormal diffusion process with polynomial exogenous factors in forecasting. (*submitted*).

[Rao *et al.*, 1974] N.J. Rao, J.D. Borwankar and D. Ramkrishna. Numerical solution of Itô integral equations. *SIAM J. Control*, 12:24–139, 1974.

[Stern and Kaufman, 1998] D.I. Stern and R.K. Kaufmann. Annual estimates of global anthropogenic methane emissions: 1860-1994. In J. Carbon dioxide information analysis center, Oak Ridge National Laboratory, U.S. Department of Energy, *Trends Online: a compendium of data on global change* Oak Ridge, Tenn., U.S.A. 1998.