On the structure of the stochastic process of mortgages in Spain

- Paula R. Bouzas, Ana M. Aguilera, Mariano J. Valderrama, Nuria Ruiz-Fuentes
- On the structure of the stochastic process of mortgages in Spain
- *Computational Statistics 21, 73–89 (2006)*
- DOI: https://doi.org/10.1007/s00180-006-0252-0

Computational Statistics (2006) 21:73-89
DOI 10.1007/s00180-006-0252-0

On the structure of the stochastic process of mortgages in Spain

P.R. Bouzas[2], A.M. Aguilera[2], M.J. Valderrama[2] and N. Ruiz-Fuentes[3]

[2] Department of Statistics and Operational Research, University of Granada, 18071 Spain
[3] Department of Statistics and Operational Research, University of Jaén, 23071 Spain

Summary

The number of mortgages in Spain is a counting process that can be modelled as a doubly stochastic Poisson process (DSPP). A modelling method for the intensity of a DSPP is proposed. A first step consists on estimating discrete sample paths of it from observed ones of the DSPP, then a continuous modelling is derived by means of Functional Principal Component Analysis. The method is validated by a simulation. Finally, it is applied to the real process of the mortgages in Spain discussing the interpretation of the principal components and factors.

Keywords: Doubly Stochastic Poisson Process, Point Estimation, Functional Principal Component, Mortgages.

1 Introduction

The number of mortgages in a country is an important index to evaluate the economic development of that country. This paper is focused in modelling the point process of the number of mortgages in Spain through a previous modelling of its intensity. The number of mortgages is considered as a doubly stochastic Poisson process (DSPP), that is a Poisson process conditioned to an external process; so its intensity is also a stochastic process (see Snyder 1991 and Daley and Vere-Jones 1988, among others). We have chosen the DSPP to model this real counting process in order to consider a more complex structure of its intensity than in homogeneous and non-homogeneous Poisson processes.

We have available large data of the DSPP , it is of the number of mortgages in Spain; the monthly data of the stochastic process from 1991 to 2000 of the 52 Spanish provinces. These data can be seen as 52 sample paths of the DSPP.

When the first two moments of the intensity process are known, its estimation has been deeply studied in literature. Even though, when there is not statistical knowledge about the intensity process, Snyder and Miller (1991) propose to estimate it as a moving average of an observed trajectory of the DSPP and to make a regression so that the intensity can be modelled as a function of time. This estimation looses the stochastic nature of the intensity process of the DSPP which is the fundamental characteristic of this generalization of the Poisson process, so the DSPP is simplified to be a non-homogeneous one. The novelty of this paper is to present an estimating method of the intensity process by modelling its stochastic structure without any statistical assumptions on it.

In Section 2, a point estimator for the intensity process of a DSPP is defined when a long enough trajectory of the process is observed but no other statistical knowledge is available. Then, we propose a method of estimating a discrete sample path of the intensity process from each observed one of the DSPP.

Section 3 is devoted to model the stochastic structure of the intensity process from the estimated trajectories calculated in the previous section by means of Functional Principal Components Analysis (FPCA). Therefore, the intensity is expressed in terms of an orthogonal expansion of the principal components (p.c.'s). See for example Ramsay and Silverman, 1997 and Valderrama et al. 2000, for deeper study of FPCA and examples.

In Section 4, the inference on the intensity process proposed is applied to a specific case of DSPP, a compound DSPP. This process is a random combination of four DSPP's. The simulation is used not only to illustrate the calculation of the point estimator and the FPCA, but also to show the stability of the mentioned estimator and to validate the modelling method.

Finally, Section 5 applies the validated methodology to the real observed case of the number of mortgages in Spain mentioned above (see Bouzas et al., 2002 for an introductory work). Even more, the first three p.c's and principal factors are displayed and it is also treated their interpretation as well as a discussion on the estimation error.

2 Estimated trajectories of the intensity

The estimation problem of the intensity process is described as follows. Let $\{N(t) : t \geq t_0\}$ be a DSPP with intensity process $\{\lambda(t, x(t)) : t \geq t_0\}$ where $\{x(t) : t \geq t_0\}$ is the information process. Our purpose is to estimate the stochastic process $\lambda(t, x(t))$, having no previous knowledge about its structure, from a trajectory of $N(t)$ observed on the interval $[t_0, t_0 + rT)$.

From now on, we will write $\lambda(t)$ instead of $\lambda(t, x(t))$ to simplify the notation. We will considered that the data available are k sample paths of the DSPP that will be denoted by N_ω, $\omega = 1, \ldots, k$. Let us now explain a way to estimate shorter k sample paths of $\lambda(t)$ from the observed data.

Due to the independence of the increments of a DSPP, each trajectory, N_ω, can be cut into r subtrajectories by dividing the time interval in r subintervals so that they start at the same time; this rescaling can be expressed as follows

$$N_{\omega,i}(t) = N_\omega\left(t + (i-1)T\right) - N_\omega\left(t_0 + (i-1)T\right), \qquad t \in (t_0, t_0 + T]$$

with $i = 1, \ldots, r$ and $\omega = 1, \ldots, k$. The subintervals have equal length to use all the information in the estimator that will be proposed below.

We will define a point estimator of the intensity process in several instants of time of $(t_0, t_0 + T]$ for each of the k sample paths, using the r new subtrajectories of each one. Let us choose a partition in $(t_0, t_0 + T]$, defined by the knots

$$t_0 < t_{\omega,1} < \ldots < t_{\omega,j} < \ldots < t_{\omega,p_\omega} = t_0 + T$$

with $\omega = 1, \ldots, k$. We denote the last knot by t_{ω,p_ω} because the number of knots may be different for each k. Then, the point estimator is defined by

$$\hat{\lambda}(t_{\omega,j}) = \frac{1}{r} \sum_{i=1}^{r} \frac{N_{w,i}(t_{\omega,j}) - N_{w,i}(t_{\omega,j-1})}{t_{\omega,j} - t_{\omega,j-1}} \tag{1}$$

with $j = 1, \ldots, p_\omega$ and $\omega = 1, \ldots, k$. Therefore, we have p_ω estimations of $\lambda(t)$ in $(t_0, t_0 + T]$, for each $\omega = 1, \ldots, k$. They can be considered as k estimated discrete trajectories of the intensity in $(t_0, t_0 + T]$.

Estimating the discrete sample paths of the intensity process from observed data of the DSPP has various steps explained above. They are outlined in Figure 1.

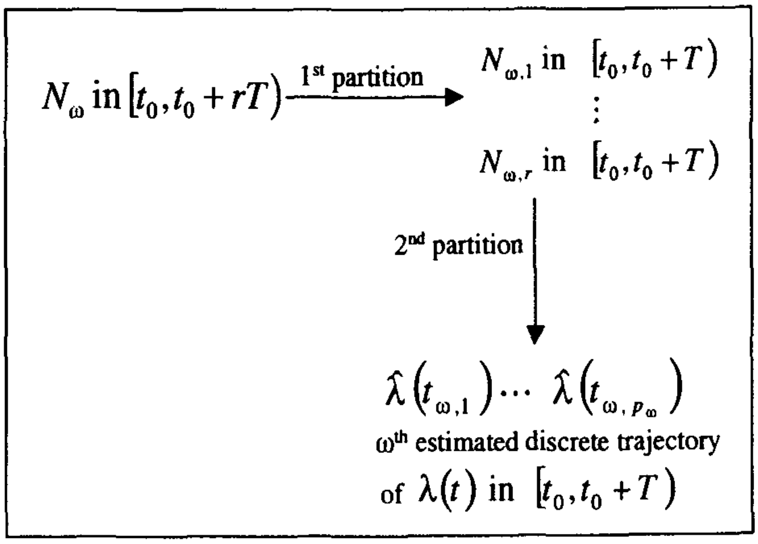

Figure 1: Diagram of the steps of the estimation of a trajectory of $\lambda(t)$.

2.1 Properties and confidence intervals

Next, we study the main properties of this estimator. Using the fact that $N_{\omega,i}(t_{\omega,j-1}, t_{\omega,j})$ are DSPP's with mean $\int_{t_{\omega,j-1}}^{t_{\omega,j}} \lambda(\sigma)\, d\sigma$, for every $i = 1, \dots, r$, let us notice that

$$E[\hat{\lambda}(t_{\omega,j})] = \frac{1}{r} \sum_{i=1}^{r} \frac{E[N_{\omega,i}(t_{\omega,j-1}, t_{\omega,j})]}{t_{\omega,j} - t_{\omega,j-1}} = \frac{\int_{t_{\omega,j-1}}^{t_{\omega,j}} \lambda(\sigma)\, d\sigma}{t_{\omega,j} - t_{\omega,j-1}}$$

and if $\max_{j=1,\dots,p}(t_{\omega,j} - t_{\omega,j-1}) \to 0$, when $p \to \infty$, then the estimator is unbiased:

$$E[\hat{\lambda}(t_{\omega,j})] \to \lambda(t_{\omega,j}).$$

It is also known that the variance of a DSPP coincides with its mean, so we can prove that

$$Var[\hat{\lambda}(t_{\omega,j})] = \frac{1}{r^2} \frac{\sum_{i=1}^{r} Var[N_{\omega,i}(t_{\omega,j-1}, t_{\omega,j})]}{(t_{\omega,j} - t_{\omega,j-1})^2} = \frac{1}{r} \frac{\int_{t_{\omega,j-1}}^{t_{\omega,j}} \lambda(\sigma)\, d\sigma}{(t_{\omega,j} - t_{\omega,j-1})^2} \overset{r \to \infty}{\longrightarrow} 0 \tag{2}$$

so that the estimator is consistent.

Let us extend the point estimation of the intensity process to estimation in a whole interval. First of all, we need an statistic with known distribution in which the parameter is involved. In order to find it, let us take into account the distribution of various statistics.

Due to the independence of the subtrajectories, $N_{\omega,i}(t)$, remembering that $N_{\omega,i}(t_{\omega,j-1}, t_{\omega,j})$ are DSPP's with mean $\int_{t_{\omega,j-1}}^{t_{\omega,j}} \lambda(\sigma)\, d\sigma$, for every $i = 1, \ldots, r$ and using the Central Limit Theorem,

$$\frac{\sum_{i=1}^{r} N_{\omega,i}(t_{\omega,j-1}, t_{\omega,j})}{r} \to N\left(\int_{t_{\omega,j-1}}^{t_{\omega,j}} \lambda(\sigma)\, d\sigma, \frac{\int_{t_{\omega,j-1}}^{t_{\omega,j}} \lambda(\sigma)\, d\sigma}{r}\right)$$

so then,

$$\hat{\lambda}(t_{\omega,j}) = \frac{\sum_{i=1}^{r} N_{\omega,i}(t_{\omega,j-1}, t_{\omega,j})}{r\, (t_{\omega,j} - t_{\omega,j-1})} \to N\left(\frac{\int_{t_{\omega,j-1}}^{t_{\omega,j}} \lambda(\sigma)\, d\sigma}{t_{\omega,j} - t_{\omega,j-1}}, \frac{\int_{t_{\omega,j-1}}^{t_{\omega,j}} \lambda(\sigma)\, d\sigma}{r\, (t_{\omega,j} - t_{\omega,j-1})^2}\right).$$

Taking into account this last distribution and choosing a partition verifying $\max_{j=1,\ldots,p}(t_{\omega,j} - t_{\omega,j-1}) \to 0$, when $p \to \infty$, we obtain that

$$\hat{\lambda}(t_{\omega,j}) \to N\left(\lambda(t_{\omega,j}), \frac{\lambda(t_{\omega,j})}{r\, (t_{\omega,j} - t_{\omega,j-1})}\right).$$

Finally, we have found out that the limit distribution of our point estimator if the amplitude of the partition tends to zero. In practice, this can be translated to be the distribution of $\hat{\lambda}(t_j)$ for r large enough and a fine partition. Therefore,

$$\frac{\hat{\lambda}(t_{\omega,j}) - \lambda(t_{\omega,j})}{\sqrt{\frac{\lambda(t_{\omega,j})}{r\, (t_{\omega,j} - t_{\omega,j-1})}}} \to N(0,1).$$

Due to the consistency of $\hat{\lambda}(t_{\omega,j})$ (see equation (2)), we can also assure that replacing $\lambda(t_{\omega,j})$ in the standard deviation of the Normal distribution by $\hat{\lambda}(t_{\omega,j})$ the distribution is still the same. Using the usual method to derive a confidence interval, we obtain

$$\left[\hat{\lambda}(t_{\omega,j}) - z_{1-\alpha/2}\sqrt{\frac{\hat{\lambda}(t_{\omega,j})}{r\, (t_{\omega,j} - t_{\omega,j-1})}}, \hat{\lambda}(t_{\omega,j}) + z_{1-\alpha/2}\sqrt{\frac{\hat{\lambda}(t_{\omega,j})}{r\, (t_{\omega,j} - t_{\omega,j-1})}}\right] \tag{3}$$

where $z_{1-\alpha/2}$ is the value of a standard Normal distribution whose distribution function is $1 - \alpha/2$, therefore the confidence level of the interval is $(1-\alpha)\, 100\,\%$.

Calculating the confidence interval for each knot of the partition, the set of all the inferior extremes and the set of all the superior extremes form a confidence band for the intensity process of a DSPP in the whole interval $[t_0, t_0 + T)$.

All the calculations needed for deriving the estimations and the confidence bands, have been implemented with Matlab 6.0. Therefore, our program can give discrete sample paths of the intensity process of a real DSPP from its observed sample paths.

3 Continuous modelling of the intensity

In the previous section we have developed point estimations of the intensity process in some instants of time. In this section, we will model the intensity in a continuous way by applying FPCA. Therefore, it will be expressed in terms of an orthogonal expansion of its principal components which is a continuous estimation of its stochastic structure.

In Section 2, the point estimator of the intensity process given by equation (1) provided k discrete trajectories of it. Under the assumption that the intensity is a process continuous in quadratic mean and with squared integrable sample paths, we propose now to apply a FPCA to these trajectories in order to model the structure of their process. The number r is chosen such that these trajectories have the same length not to loose any information applying the FPCA.

We will obtain an orthogonal expansion of the intensity process in terms of its first q principal components

$$\lambda_\omega^q(t) \simeq \hat{\mu}_\lambda(t) + \sum_{s=1}^{q} \hat{\xi}_{\omega,s} \hat{f}_s(t)$$

where $\hat{\mu}_\lambda(t)$ is the estimation of the expectation of $\lambda(t)$ from the k estimated sample paths:

$$\hat{\mu}_\lambda(t) = \frac{1}{k} \sum_{\omega=1}^{k} \hat{\lambda}_\omega(t)$$

and $\hat{f}_s(t)$ are the eigenfunctions, also called principal factors, associated to the eigenvalues of the second order integral equation

$$\int_{t_0}^{t_0+T} \hat{C}_\lambda(t,u) \hat{f}_s(u)\, du = \hat{\lambda}_s \hat{f}_s(t), \qquad t \in [t_0, t_0 + T] \tag{4}$$

where $\hat{C}_\lambda$ is the estimator of the covariance function of $\lambda(t)$:

$$\hat{C}_\lambda(t,u) = \frac{1}{k-1}\sum_{\omega=1}^{k}\left[\hat{\lambda}_\omega(t) - \hat{\mu}_\lambda(t)\right]\left[\hat{\lambda}_\omega(u) - \hat{\mu}_\omega(u)\right]$$

and $\hat{\xi}_{\omega,s}$ is the s-*th* principal component:

$$\hat{\xi}_{\omega,s} = \int_{t_0}^{t_0+T}\left[\hat{\lambda}_\omega(t) - \hat{\mu}_\lambda(t)\right]\hat{f}_s(t)\,dt.$$

Usually, trajectories are not completely known but observed in discrete instants of time (in the real case of mortgages in Spain that we will study in Section 5, not observed but estimated). On the other hand, equation (4) must be solved by numerical methods and the FPCA estimated from discrete observations. In any case, it is usual to assume that the trajectories belong to a finite dimension space, generated by a basis denoted by $\Phi_1, \ldots, \Phi_n$, so that

$$\hat{\lambda}_\omega(t) = \sum_{l=1}^{n}\alpha_{\omega,l}\Phi_l(t).$$

where $\alpha_{\omega,l}$ are the corresponding coefficients. If the basis is orthonormal, the FPCA is reduced to multivariate PCA of the matrix whose rows are the coordinates of each estimated sample path, $\hat{\lambda}_\omega(t)$, with respect to the basis. The computation of the coefficients can be solved by different numerical methods like different types of interpolation, projection, etc. In this paper, we will use natural cubic spline interpolation on the discrete estimated sample paths of the intensity (see Aguilera 1996).

A specific Turbo Pascal program called SMCP[2] (see Aguilera 1999) was developed in order to calculate the estimation of the FPCA. The program is composed by an executable program and twenty five libraries that allow us to use different methods of estimation of the FPCA.

4 Application to simulations

The estimator and confidence interval proposed in Section 2 have been applied to a simulated example in order to illustrate the methodology. Using that the real intensity of the simulation is known, we validate the inference by observing the behavior of the estimations of the intensity and its comparison to the real values.

If we simulate k trajectories of the DSPP according to a chosen intensity. Then, we can obtain their corresponding k shorter estimated sample paths of the intensity process. We propose the following measurement of the error committed in the estimation of each intensity trajectory

$$\varepsilon_\omega = \frac{1}{r}\sum_{i=1}^{r} \varepsilon_{\omega,i}\,, \quad \omega = 1,\ldots,k \tag{5}$$

where

$$\varepsilon_{\omega,i} = \frac{1}{p_\omega}\sum_{j=1}^{p_\omega}[\lambda_{\omega,i}(t_{\omega,j}) - \hat{\lambda}_{\omega,i}(t_{\omega,j})]^2, \quad i = 1,\ldots,r.$$

The simulation example we are going to develop is a compound DSPP. It is formed from four DSPP's in which each random intensity is uniformly distributed in $[0,1]$ for every $t \in [t_0, t_0 + T)$, a four-dimensional boolean vector regulates the random occurrence of each of the four DSPP's at the same time. An illustration of how to derive this compound DSPP can be seen in Figure 2.

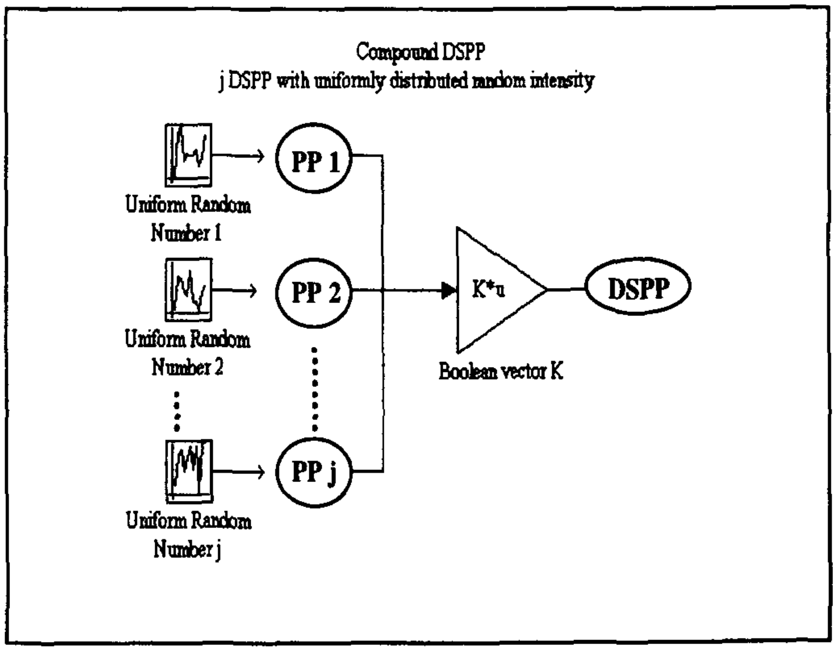

Figure 2: Diagram of a compound doubly stochastic Poisson process.

There have been simulated $K = 150$ trajectories of the compound DSPP in $[t_0, t_0 + rT) \equiv [0, 100)$. We have taken $r = 10$ and $p - \omega = 20$, $\omega = 1,\ldots,150$. Calculating the point estimator, there has been obtained other 150 shorter trajectories of the intensity in $[0, 10)$ estimated in 20 equally spaced knots of the interval. As an illustration, we can see two simulated subtrajectories of the intensity (dots) and their estimated ones of the compound DSPP (continuous line) in Figures 3A and 3B, as well as the confidence band (dashed line).

The 150 simulated sample paths were also used to calculate the errors in estimation. We have got that the error ε_ω is very small for every $\omega = 1, \ldots, 150$. All of them can be seen displayed in Figure 3C. Moreover, we can conclude that the estimation error is small and keeps stable as the error mean among estimated trajectories of the intensity process has been 0.5711 and the standard deviation 0.0775.

We have performed the FPCA in this simulation from natural cubic spline interpolation as mentioned above; the variances of the p.c.'s and their accumulated variability appear in Table 1.

p.c.	Variance	Accum. Variance
ξ_1	0.19	9.81%
ξ_2	0.16	18.44%
ξ_3	0.15	26.44%
ξ_4	0.15	34.08%
ξ_5	0.13	40.82%
ξ_6	0.12	47.25%
ξ_7	0.12	53.50%
ξ_8	0.11	59.38%
ξ_9	0.11	64.97%
ξ_{10}	0.10	70.23%
ξ_{11}	0.09	75.13%
ξ_{12}	0.08	79.56%
Total Variance	1.9	height

Table 1: Variances and accumulated variances of the p.c.'s.

Let us observe that it is necessary to select at least 12 p.c.'s in order to accumulate nearly 80% of the total variability. As examples, two of the sample paths (continuous line) and their orthogonal representation in terms of the first 12 p.c.'s (dashed line) are shown in Figures 3D and 3E.

The mean square error committed by this p.c.'s representation is given by

$$MSE^p(t) = \frac{1}{k}\sum_{\omega=1}^{k}\left(\hat{\lambda}_\omega(t) - \hat{\lambda}^p(t)\right)^2 \tag{6}$$

where p is the number of p.c.'s taken for the approximation and it has been drawn in Figure 3F.

We have observed that in the application of our modelling methodology of the intensity process of a simulated compound DSPP, the estimation error

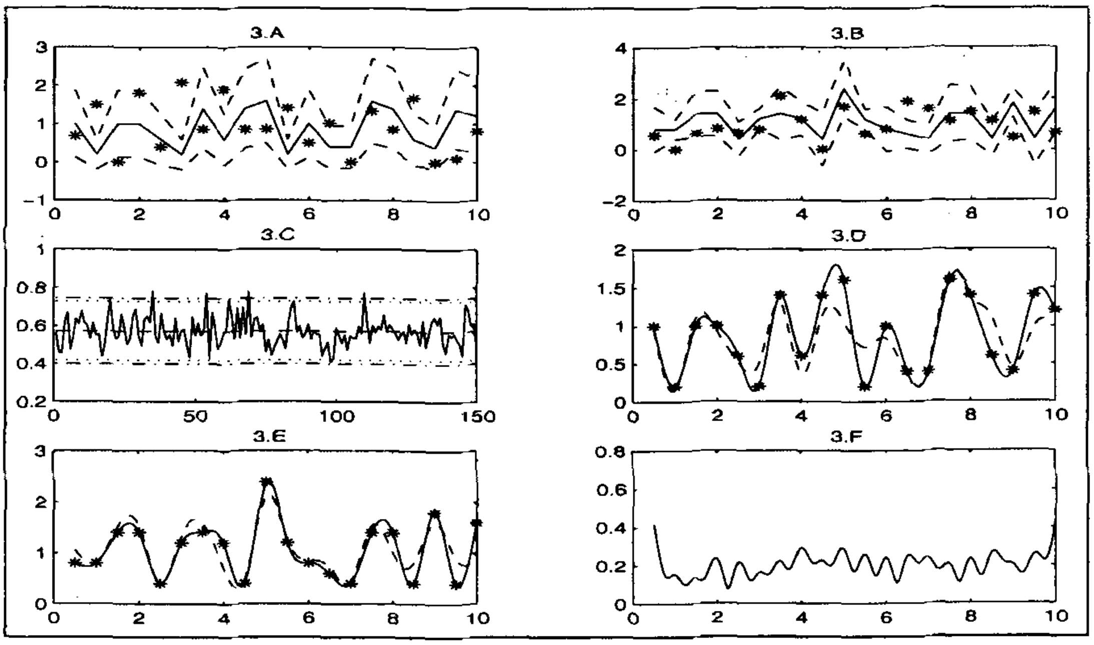

Figure 3: Plots for the compound DSPP simulation.

is small and stable having simulated 150 sample paths, so this validates the methodology and in consequence it is adequate to be applied to a real case of DSPP.

5 Mortgages in Spain

The number of mortgages in the 52 provinces of Spain are different sample paths of the process which is going to be modelled by a DSPP. As the different provinces have different populations, we consider the number of mortgages per 10000 inhabitants in order to work with comparable data.

In our case, $N(t)$ is observed from $t_0 = 0$, the beginning of 1991, to $t_0 + rT$, $(T = 12,\ r = 10)$, the end of 2000. Therefore, each subtrajectory is defined on a complete year $(t_0, t_0 + T]$. In our real process of mortgages, the initial sample paths had the same length, so if we take the same number of knots of the first partition, r, the estimated sample paths of the intensity will also have the same length and it will be useful for us to apply FPCA in order to use all the information from the observed data.

The second partition has been chosen so that the knots are the months of the year. It is, $t_{\omega,j}$, $j = 1, \ldots, p_\omega$ are equally spaced and $p_\omega = 12$, $\omega = 1, \ldots, 150$. Calculating the point estimator of equation (1) for the 12 months of the year $(j = 1, \ldots, 12)$ and the 52 provinces $(\omega = 1, \ldots, 52)$, we have got 52 estimated trajectories of the intensity process each one formed by its estimation in 12 points of the year, in the twelve months. The estimated trajectories of the intensity of two of the provinces can be seen as examples in Figure 4 as well as the confidence bands derived from the confidence intervals at every point estimation from equation (3).

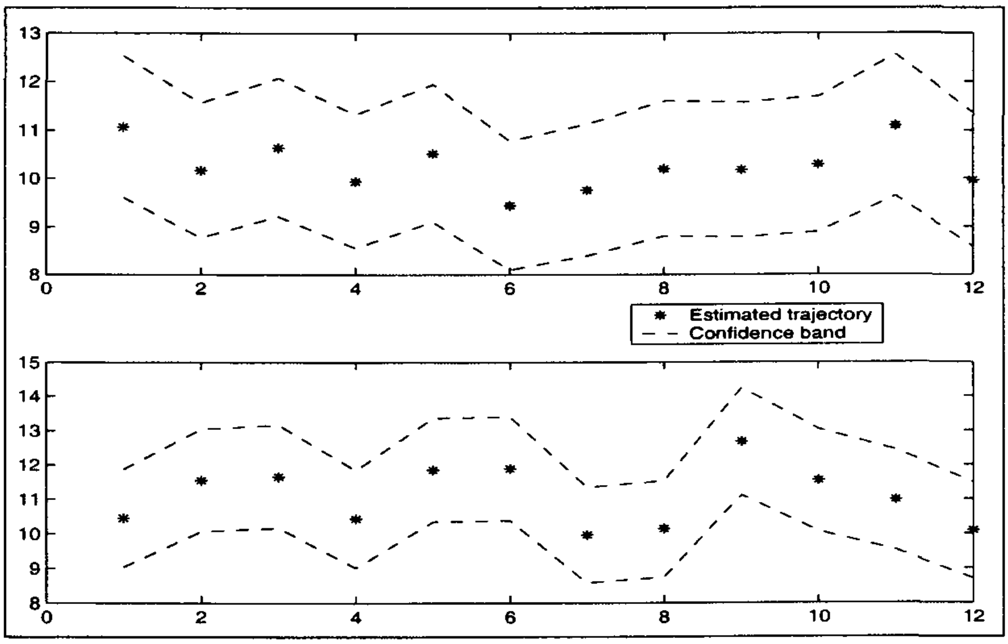

Figure 4: Two subtrajectories of $\hat{\lambda}(t)$ of the mortgages in Spain, the estimated trajectory and confidence bands.

In the real case of the number of mortgages in Spain, the 52 discrete estimated sample paths of $\lambda(t)$ in a year for the 52 provinces, have also been interpolated by natural cubic splines. The performance of the FPCA gives us the variances of the p.c.'s and their accumulated variability which appear in Table 2.

Principal component	% accumulated variance
ξ_1	96.348
ξ_2	97.227
ξ_3	98.027
ξ_4	98.486
ξ_5	98.884
ξ_6	99.202
Total variance	181.919

Table 2: Accumulated variances of the p.c.'s

This table shows that the first three p.c.'s accumulate approximately 98% of the total variability and the first six ones more than 99%. As the first one accumulates so much variability, more than 96%, we have decided to take into account the first six in order not to have a trivial representation (we will give further explanation in next subsection). As an illustration, Figure 5 shows two (among the 52) of the sample paths and its orthogonal representation in terms of the first six p.c.'s.

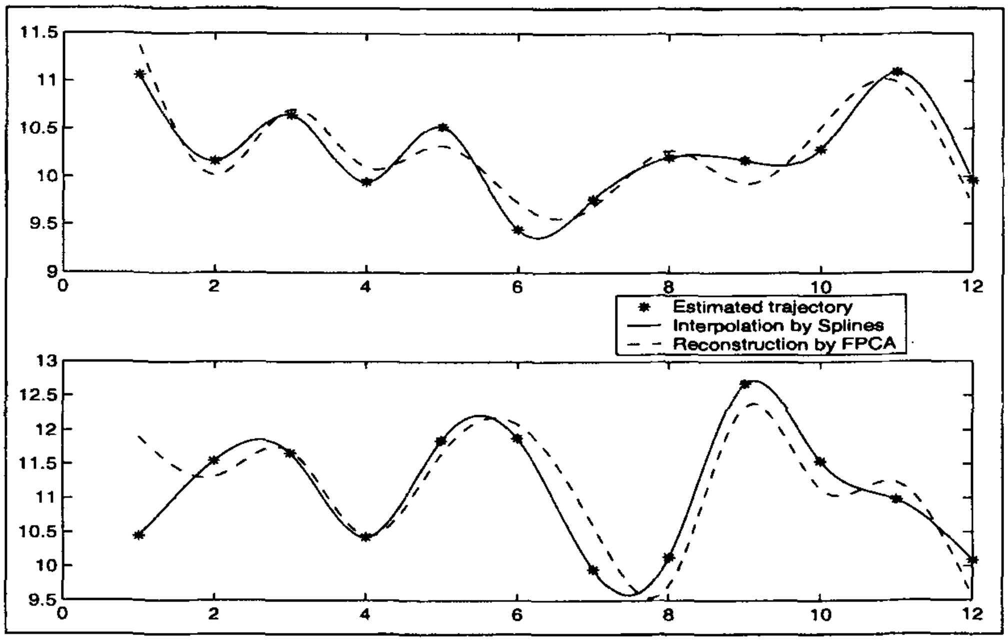

Figure 5: Two sample paths of $\hat{\lambda}(t)$ of the mortgages in Spain and their orthogonal representation.

5.1 Discuss on the principal components and factors

It would be desirable to be able to interpret the principal factors and components and maybe it could help to choose the number of p.c.'s (q) on the expansion but it is most times very difficult. See Ramsay and Silverman, 1997 for a further study on the interpretation of the p.c's and principal factors.

Let us make an attempt to study them; therefore we present the plot of the first three principal factors in Figure 6 and the first three p.c.'s are plotted one versus another one in Figures 7 - 9 where the provinces are noted by its first letters.

Let us observe Figure 6 which displays the first three principal factors of the intensity of the Spanish mortgages. It can be seen that the first principal factor is positive throughout the year and nearly the same for all the months.

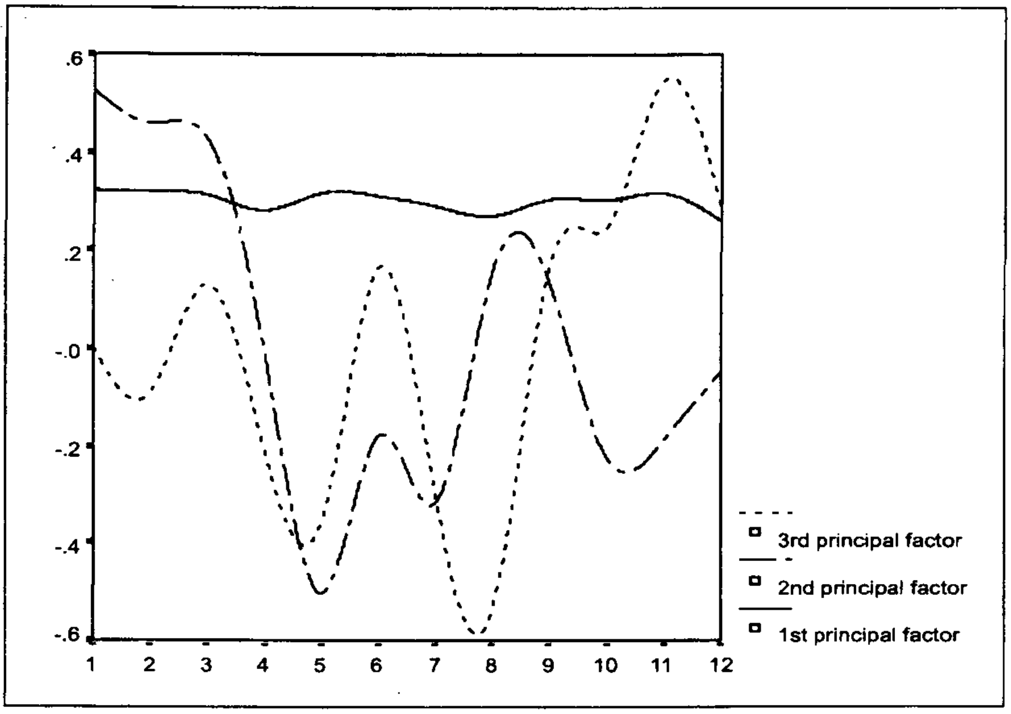

Figure 6: First three principal factors.

This means that it corresponds to a measure of the uniformity of the mortgages through the year. In fact, the variability of the monthly means of the point estimations of the intensity from the observed data of the DSPP is very small as we can see in Table 3.

Month	Mean	Month	Mean
Jan.	11.40	July	9.72
Feb.	10.93	August	9.34
March	11.06	Sep.	10.63
April	10.14	Oct.	10.90
May	10.96	Nov.	10.95
June	10.61	Dec.	9.62

Table 3: Monthly mean of the point estimations of the intensity from the observed DSPP

On the other hand, it is surprising to observe that the order of the 52 provinces on the first p.c. is the same as the order of the provinces provided by their annual means of the point estimations of the intensity. As a consequence, the first p.c. that explains 96% of the total variability of the intensity process is related with the mean of the DSPP ($E[N(t)] = \int_{t_0}^{t} \lambda(\sigma)\, d\sigma$) and so with the global amount of mortgages in a Spanish province that is

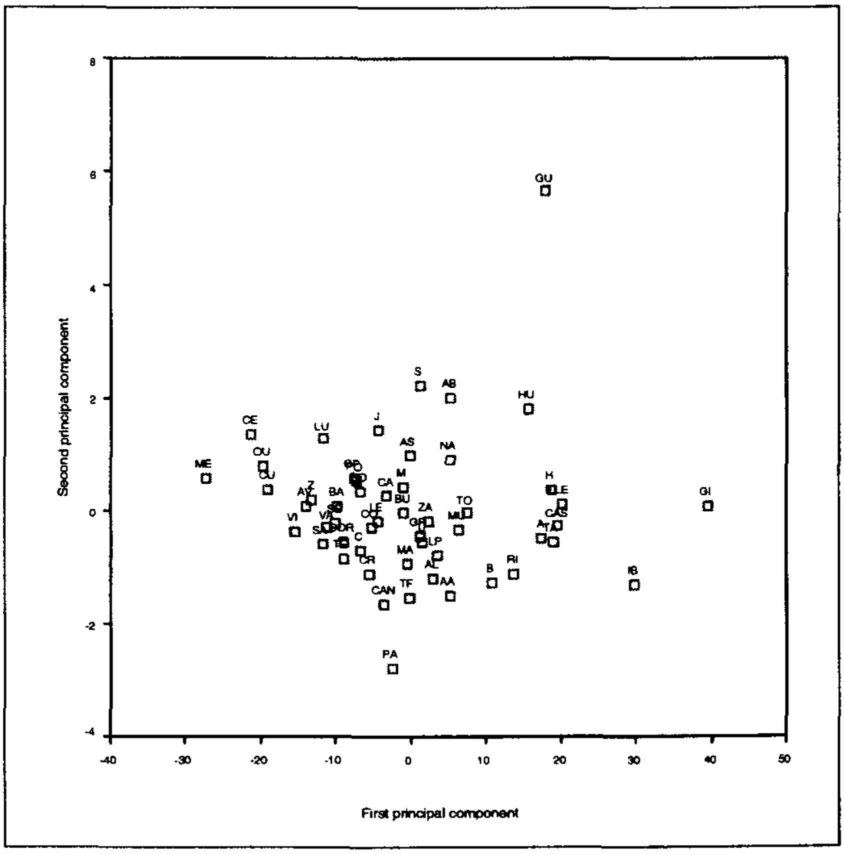

Figure 7: First principal component versus second one.

nearly invariant in time.

Taking into account that the second and third p.c.'s explain a very small percentage of variability and they are orthogonal to the first one and as well as to each other, their associated eigenfunctions represent small variation from the general behavior and so, lightly specific patterns of the intensity of some provinces with major or minor values of these two p.c.'s.

We can point out about the third principal factor that has an oscillatory form with increasing amplitude and phase until august and then the function increases and has a maximum in november. In can be seen that the variation, month to month, of the monthly mean of the data (see Table 3) follows this same structure with a small translation in time. The provinces with high values in any of the second or third p.c.'s (as it can be observed Guadalajara (GU) in Figures 7 -9) have a very similar behavior to the associated principal factor.

Let us amplify the reason of the decision of taking into account the first six p.c.'s. It is hold by the comparison of the mean square error on the

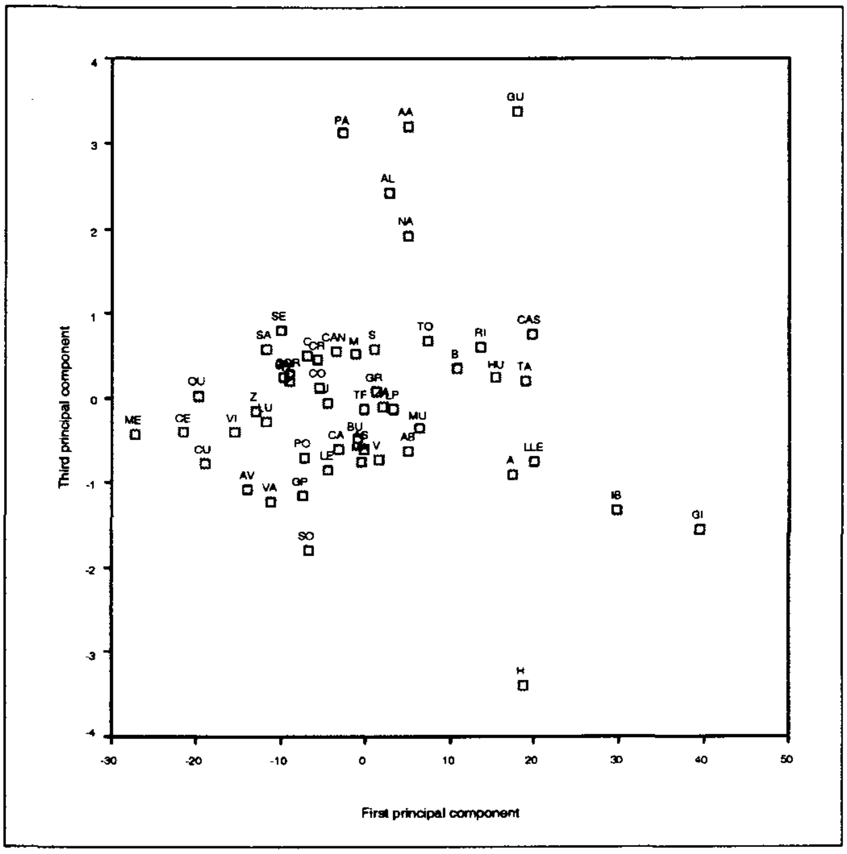

Figure 8: First principal component versus third one.

estimated intensity process (see equation (6)) committed by taking $p = 3$ or $p = 6$ in the expansion; the error in the intensity process estimation is notably smaller with $p = 6$. Figure 10 displays the comparison of the mean square errors committed by both representations.

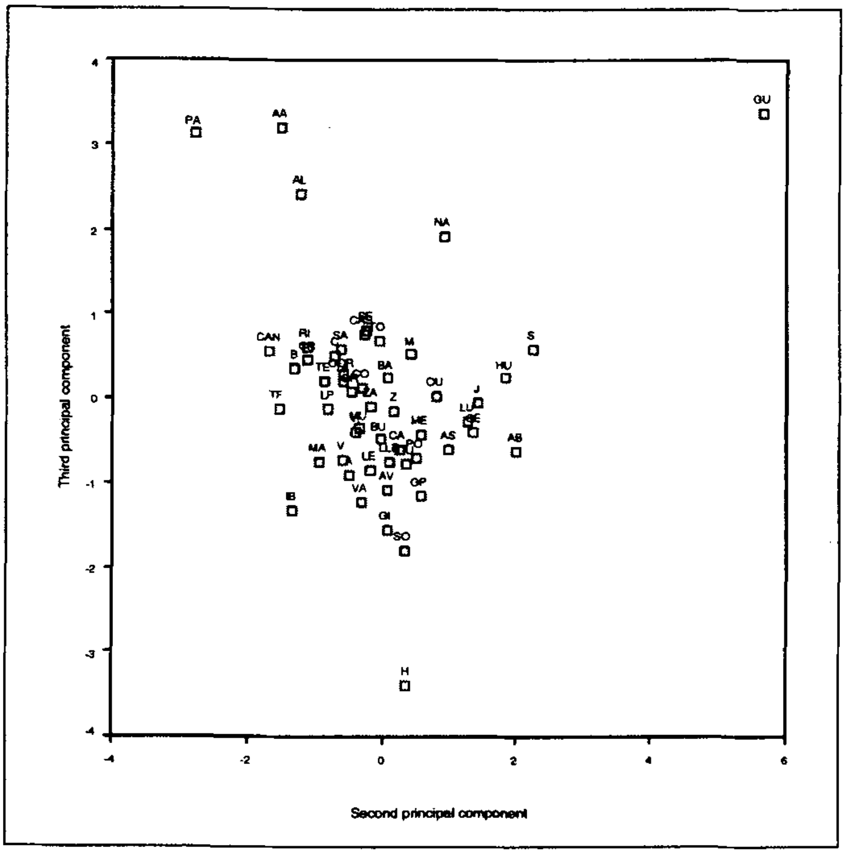

Figure 9: Second principal component versus third one.

References

Aguilera, A.M., Gutiérrez, R. and Valderrama, M.J. (1996), 'Approximation of estimators in the PCA of a Stochastic Process using B-splines', *SIAM Communications in Statistics* **25**, 671–690.

Aguilera, A.M., Valderrama,M.J., Ocaña, F.A. and Ruiz-Molina, J.C. (1999), *Implementación de metodologías para la modelización de datos temporales de las estadísticas públicas de Andalucía,* Instituto de Estadística de Andalucía. Sevilla.

Bouzas, P.R., Ruiz-Fuentes, N., Aguilera, A.M. and Valderrama, M.J. (2002), *Estimation of the structure of the stochastic process of mortages in Spain,* Proc. of the Euro-Japanese Workshop on Stochastic Modelling for Finance, Insurance, Production and Reliability. Chamonix.

Scott, D. (1988), *An Introduction to the Theory of Point Processes,* Springer-Verlag. New York.

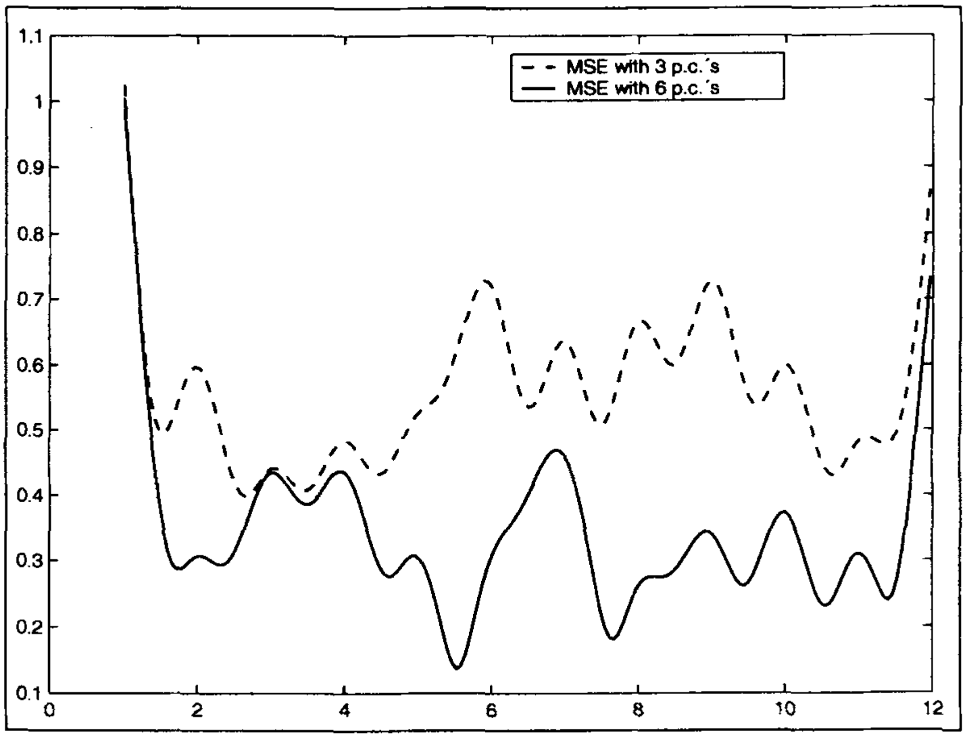

Figure 10: Comparison of the mean square error of the representation with 3 and 6 p.c.'s.

Ramsay, J.O. and Silverman, B.M. (1997), *Functional Data Analysis*, Springer-Verlag. New York.

Snyder, D.L. and Miller,M.I. (1991), *Random Point Processes in Time and Space. 2nd Edition*, Springer-Verlag. New York.

Valderrama, M.J., Aguilera, A.M. and Ocaña, F.A. (2000), *An Predicción dinámica mediante análisis de datos funcionales*, La Muralla-Hespérides, Madrid.