ISSN 1645-6726

INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

REVSTAT
Statistical Journal

Special Issue on Computational Statistics in Economics and Finance

Guest Editors:
Paulo Teles
Pedro Duarte Silva
Pilar Muñoz
Paula Brito

Volume 7, No.1
April 2009

REVSTAT

Statistical Journal

Catalogação Recomendada

REVSTAT. Lisboa, 2003-
Revstat : statistical journal / ed. Instituto Nacional de Estatística. - Vol. 1, 2003- . - Lisboa I.N.E., 2003- . - 30 cm
Semestral. - Continuação de : Revista de Estatística =
ISSN 0873-4275. - edição exclusivamente em inglês
ISSN 1645-6726

CREDITS

- **PUBLISHER**
  - *Instituto Nacional de Estatística, I.P. (INE, I.P.)*
    Av. António José de Almeida, 2
    1000-043 LISBOA
    PORTUGAL
    Tel.: + 351 21 842 61 00
    Fax: + 351 21 842 6364
    Web site: http://www.ine.pt
    Customer Support Service
    (National network) : 808 201 808
    Other networks: + 351 22 605 07 48
- **COVER DESIGN**
  - *Mário Bouçadas, designed on the stain glass window at INE by the painter Abel Manta*
- **LAYOUT AND GRAPHIC DESIGN**
  - *Carlos Perpétuo*
- **PRINTING**
  - *Instituto Nacional de Estatística, I.P.*
- **EDITION**
  - *350 copies*
- **LEGAL DEPOSIT REGISTRATION**
  - *N.º 191915/03*

PRICE

[VAT 5% included]

- Single issue …… **€ 10**
- Annual subscription (No. 1 Special Issue, No. 2 and No.3)… **€ 24**
- Annual subscription (No. 2, No. 3) …… **€ 16**

INDEX

REVSTAT – Statistical Journal
Volume 7, Number 1, April 2009, 1–22

A METHOD OF TREND EXTRACTION USING SINGULAR SPECTRUM ANALYSIS

Author: Theodore Alexandrov
– Center for Industrial Mathematics, University of Bremen, Germany
theodore@math.uni-bremen.de

Abstract:

- The paper presents a new method of trend extraction in the framework of the Singular Spectrum Analysis (SSA) approach. This method is easy to use, does not need specification of models of time series and trend, allows to extract trend in the presence of noise and oscillations and has only two parameters (besides basic SSA parameter called window length). One parameter manages scale of the extracted trend and another is a method specific threshold value. We propose procedures for the choice of the parameters. The presented method is evaluated on a simulated time series with a polynomial trend and an oscillating component with unknown period and on the seasonally adjusted monthly data of unemployment level in Alaska for the period 1976/01–2006/09.

Key-Words:

- *time series; trend extraction; Singular Spectrum Analysis.*

AMS Subject Classification:

- 37M10, 15A18, 60G35, 62M15.

1. INTRODUCTION

Trend extraction is an important task in applied time series analysis, in particular in economics and engineering. We present a new method of trend extraction in the framework of the Singular Spectrum Analysis approach.

Trend is usually defined as a smooth additive component containing information about time series global change. This definition is rather vague (which type of smoothness is used? which kind of information is contained in the trend?). It may sound strange, but there is no more precise definition of the trend accepted by the majority of researchers and practitioners. Each approach to trend extraction defines trend with respect to the mathematical tools used (e.g. using Fourier transformation or derivatives). Thus in the corresponding literature one can find various specific definitions of the trend. For further discussion on trend issues we refer to [2].

Singular Spectrum Analysis (SSA) is a general approach to time series analysis and forecast. Algorithm of SSA is similar to that of Principal Components Analysis (PCA) of multivariate data. In contrast to PCA which is applied to a matrix, SSA is applied to a time series and provides a representation of the given time series in terms of eigenvalues and eigenvectors of a matrix made of the time series. The basic idea of SSA has been proposed by [5] for dimension calculation and reconstruction of attractors of dynamical systems, see historical reviews in [10] and in [11]. In this paper we mostly follow the notations of [11].

SSA can be used for a wide range of tasks: trend or quasi-periodic component detection and extraction, denoising, forecasting, change-point detection. The present bibliography on SSA includes two monographs, several book chapters, and over a hundred papers. For more details see references at the website SSAwiki: `http://www.math.uni-bremen.de/~theodore/ssawiki`.

The method presented in this paper has been first proposed in [3] and is studied in detail in the author's unpublished Ph.D. thesis [1] available only in Russian at `http://www.pdmi.ras.ru/~theo/autossa`.

The proposed method is easy to use (has only two parameters), does not need specification of models of time series and trend, allows one to specify desired trend scale, and extracts trend in the presence of noise and oscillations.

The outline of this paper is as follows. Section 2 introduces SSA, formulates properties of trends in SSA and presents the already existing methods of trend extraction in SSA. Section 3 proposes our method of trend extraction. In Section 4 we discuss the frequency properties of additive components of a time series and present our procedure for the choice of first parameter of the method,

a low-frequency boundary. Section 5 starts with investigation of the role of the second method parameter, the low-frequency contribution, based on a simulation example. Then we propose a heuristic strategy for the choice of this parameter. In Section 6, applications of the proposed method to a simulated time series with a polynomial trend and oscillations and on the unemployment level in Alaska are considered. Finally, Section 7 offers conclusions.

2. SINGULAR SPECTRUM ANALYSIS

Let us have a time series $F = (f_0, ..., f_{N-1})$, $f_n \in \mathbb{R}$, of length N, and we are looking for some specific additive component of F (e.g. a trend). The central idea of SSA is to embed F into high-dimensional euclidean space, then find a subspace corresponding to the sought-for component and, finally, reconstruct a time series component corresponding to this subspace. The choice of the subspace is a crucial question in SSA. The basic SSA algorithm consists of decomposition of a time series and reconstruction of a desired additive component. These two steps are summarized below; for a detailed description, see page 16 of [11].

Decomposition. The decomposition takes a time series of length N and comes up with an $L \times K$ matrix. This stage starts by defining a parameter L ($1 < L < N$), called the window length, and constructing the so-called trajectory matrix $\mathbf{X} \in \mathbb{R}^{L \times K}$, $K = N - L + 1$, with stepwise taken portions of the original time series F as columns:

$$F = (f_0, ..., f_{N-1}) \;\rightarrow\; \mathbf{X} = [X_1 : ... : X_K] \,, \qquad X_j = (f_{j-1}, ..., f_{j+L-2})^{\mathrm{T}} \,. \tag{2.1}$$

Note that $\mathbf{X}$ is a Hankel matrix and (2.1) defines one-to-one correspondence between series of length N and Hankel matrices of size $L \times K$. Then Singular Value Decomposition (SVD) of $\mathbf{X}$ is applied, where j-th component of SVD is specified by j-th eigenvalue λ_j and eigenvector U_j of $\mathbf{X}\mathbf{X}^{\mathrm{T}}$:

$$\mathbf{X} = \sum_{j=1}^{d} \sqrt{\lambda_j}\, U_j V_j^{\mathrm{T}} \,, \qquad V_j = \mathbf{X}^{\mathrm{T}} U_j \Big/ \sqrt{\lambda_j} \,, \qquad d = \max\{j\colon \lambda_j > 0\} \,.$$

Since the matrix $\mathbf{X}\mathbf{X}^{\mathrm{T}}$ is positive-definite, their eigenvalues λ_j are positive. The SVD components are numbered in the decreasing order of eigenvalues λ_j. We define j-th Empirical Orthogonal Function (EOF) as the sequence of elements of the j-th eigenvector U_j. The triple $(\sqrt{\lambda_j}, U_j, V_j)$ is called j-th eigentriple, $\sqrt{\lambda_j}$ is called the j-th singular value, U_j is the j-th left singular vector and V_j is the j-th right singular vector.

Reconstruction. Reconstruction goes from an $L\times K$ matrix into a time series of length N. This stage combines (i) selection of a subgroup $\mathcal{J}\subset\{1,...,L\}$ of SVD components; (ii) hankelization (averaging along entries with indices $i+j=\text{const.}$) of the $L\times K$ matrix from the selected $\mathcal{J}$ components of the SVD; (iii) reconstruction of a time series component of length N from the Hankel matrix by the mentioned one-to-one correspondence (like in (2.1) but in the reverse direction, see below the exact formulae). The result of the reconstruction stage is a time series additive component:

$$\mathbf{X}_{\mathcal{J}}=\sum_{j\in\mathcal{J}}\sqrt{\lambda_j}\,U_jV_j^{\mathrm{T}}\ \rightarrow\ G=(g_0,...,g_{N-1})\ .$$

For the sake of brevity, let us describe the hankelization of the matrix $\mathbf{X}_{\mathcal{J}}$ and the subsequent reconstruction of a time series component G as being applied to a matrix $\mathbf{Y}=\left\{y_{ij}\right\}_{i,j=1}^{i=L,j=K}$ as it is introduced in [11]. First we introduce $L^*=\min\{L,K\}$, $K^*=\max\{L,K\}$ and define an $L^*\times K^*$ matrix $\mathbf{Y}^*$ as given by $\mathbf{Y}^*=\mathbf{Y}$ if $L\leqslant K$ and $\mathbf{Y}^*=\mathbf{Y}^{\mathrm{T}}$ if $L>K$. Then the elements of the time series $G=(g_0,...,g_{N-1})$ formed from the matrix $\mathbf{Y}$ are calculated by averaging along cross-diagonals of matrix $\mathbf{Y}^*$ as

$$g_n=\begin{cases}\dfrac{1}{n+1}\displaystyle\sum_{m=1}^{n+1}y^*_{m,n-m+2}\,, & 0\leqslant n<L^*-1\ ,\\[2ex] \dfrac{1}{L^*}\displaystyle\sum_{m=1}^{L^*}y^*_{m,n-m+2}\,, & L^*-1\leqslant n<K^*\,,\\[2ex] \dfrac{1}{N-n}\displaystyle\sum_{m=n-K^*+2}^{N-K^*+1}y^*_{m,n-m+2}\,, & K^*\leqslant n<N\ .\end{cases}\tag{2.2}$$

Changing the window length parameter and, what is more important, the subgroup $\mathcal{J}$ of SVD components used for reconstruction, one can change the output time series G. In the problem of trend extraction, we are looking for G approximating a trend of a time series. Thus, the trend extraction problem in SSA is reduced to (i) the choice of a window length L used for decomposition and (ii) the selection of a subgroup $\mathcal{J}$ of SVD components used for reconstruction. The first problem is thoroughly discussed in section 1.6 of [11]. In this paper, we propose a solution for the second problem.

Note that for the reconstruction of a time series component, SSA considers the whole time series, as its algorithm uses SVD of the trajectory matrix built from all parts of the time series. Therefore, SSA is not a local method in contrast to a linear filtering or wavelet methods. On the other hand, this property makes SSA robust to outliers, see [11] for more details.

An essential disadvantage of SSA is its computational complexity for the calculation of SVD. This shortcoming can be reduced by using modern [9] and parallel algorithms for SVD. Moreover, for trend revision in case of receiving new data points, a computationally attractive algorithm of [12] for updating SVD can be used.

It is worth to mention here that the similar ideas of using SVD of the trajectory matrix have been proposed in other areas, e.g. in signal extraction in oceanology [8] and estimation of parameters of damped complex exponential signals [13].

2.1. Trend in SSA

SSA is a nonparametric approach which does not need a priori specification of models of time series and trend, neither deterministic nor stochastic ones. The classes of trends and residuals which can be successfully separated by SSA are characterized as follows.

First, since we extract any trend by selecting a subgroup of all d SVD components, this trend should generate less than d SVD components. For an infinite time series, a class of such trends coincides with the class of time series governed by finite difference equations [11]. This class can be described explicitly as linear combinations of products of polynomials, exponentials and sines [6]. An element of this class suits well for representation of a smooth and slow varying trend.

Second, a residual should belong to a class of time series which can be separated from a trend. The separability theory due to [14] helps us determine this class. In [14] it was proved that (i) any deterministic function can be asymptotically separated from any ergodic stochastic noise as the time series length and window length tend to infinity; (ii) under some conditions any trend can be separated from any quasi-periodic component, see also [11]. These properties of SSA make this approach feasible for trend extraction in the presence of noise and quasi-periodic oscillating components.

Finally, as trend is a smooth and slow varying time series component, it generates SVD components with smooth and slow varying EOFs. Eigenvectors represent an orthonormal basis of a trajectory vector space spanned on the columns of trajectory matrix. Thus each EOF is a linear combination of portions of the corresponding time series and inherits its global smoothness properties. This idea is considered in detail in [11] for the cases of polynomial and exponential trends.

2.2. Existing methods of trend extraction in SSA

A naive approach to trend extraction in SSA is to reconstruct a trend from several first SVD components. Despite its simplicity, this approach works in many real-life cases for the following reason. An eigenvalue represents a contribution of the corresponding SVD component into the form of the time series, see section 1.6 of [11]. Since a trend usually characterizes the shape of a time series, its eigenvalues are larger than the other ones, that implies small order numbers of the trend SVD components. However, the selection procedure fails when the values of a trend are small enough as compared with a residual, or when a trend has a complicated structure (e.g. a high-order polynomial) and is characterized by many (not only by the first ones) SVD components.

A smarter way of selecting trend SVD components is to choose the components with smooth and slow varying EOFs (we have explained this fact above). At present, there exist only one parametric method of [15] which follows this approach. In [15] it was proposed using the Kendall correlation coefficient for testing for monotonic growth of an EOF. Unfortunately, this method is far from perfect since it is not possible to establish which kinds of trend can be extracted by its means. This method seems to be aimed at extraction of monotonic trends because their EOFs are usually monotonic. However, even a monotonic trend can produce non-monotonic EOF, especially in case of noisy observations. An example could be a linear trend which generates a linear and a constant EOFs. If there is a noise or another time series component added, then this component is often mixed with trend components corrupting its EOFs. Then, even in case of very small corruption, the constant EOF can be highly non monotonic. Naturally, the method using the Kendall correlation coefficient does not suit for non monotonic trends producing non monotonic EOFs. For example, a polynomial of low order which is often used for trend modelling usually produces non monotonic EOFs, for details see e.g. [11].

3. PROPOSED METHOD FOR TREND EXTRACTION

In this section, we present our method of trend extraction. First, following [11], we introduce the periodogram of a time series.

Let us consider the Fourier representation of the elements of a time series X of length N, $X = (x_0, ..., x_{N-1})$, see e.g. section 7.3 of [7]:

$$x_n \;=\; c_0 \;+\; \sum_{1 \leqslant k \leqslant \frac{N-1}{2}} \Big(c_k \cos(2\pi n k/N) + s_k \sin(2\pi n k/N) \Big) \;+\; (-1)^n c_{N/2} \;,$$

where $k \in \mathbb{N}$, $0 \leqslant n \leqslant N-1$, and $c_{N/2} = 0$ if N is an odd number. Then the periodogram of X at the frequencies $\omega \in \{k/N\}_{k=0}^{\lfloor N/2 \rfloor}$ is defined as

$$I_X^N(k/N) \;=\; \frac{N}{2} \begin{cases} 2c_0^2\,, & k = 0\ , \\ c_k^2 + s_k^2\,, & 0 < k < N/2\ , \\ 2c_{N/2}^2\,, & \text{if } N \text{ an even number and } k = N/2\ . \end{cases} \tag{3.1}$$

Note that this periodogram is different from the periodogram usually used in spectral analysis, see e.g. [4] or [7]. To show this difference, let us denote the k-th element of the discrete Fourier transform of X as

$$\mathcal{F}_k(X) \;=\; \sum_{n=0}^{N-1} e^{-i2\pi nk/N} x_n\ ,$$

then the periodogram $I_X^N(\omega)$ at the frequencies $\omega \in \{k/N\}_{k=0}^{\lfloor N/2 \rfloor}$ is calculated as

$$I_X^N(k/N) \;=\; \frac{1}{N} \begin{cases} 2\left|\mathcal{F}_k(X)\right|^2, & \text{if } 0 < k < N/2\ , \\ \left|\mathcal{F}_k(X)\right|^2, & \text{if } k = 0 \text{ or } N \text{ is even and } k = N/2\ . \end{cases}$$

One can see that in addition to the normalization different from that in [4] and [7], the values for frequencies in the interval $(0, 0.5)$ are multiplied by two. This is done to ensure the following property:

$$\|X\|_2^2 \;=\; \sum_{n=0}^{N-1} x_n^2 \;=\; \sum_{k=0}^{\lfloor N/2 \rfloor} I_X^N(k/N)\ . \tag{3.2}$$

Let us introduce the cumulative contribution of the frequencies $[0, \omega]$ as $\pi_X^N(\omega) = \sum_{k:0 \leq k/N \leq \omega} I_X^N(k/N)$, $\omega \in [0, 0.5]$. Then, for a given $\omega_0 \in (0, 0.5)$, we define the contribution of low frequencies from the interval $[0, \omega_0]$ to $X \in \mathbb{R}^N$ as

$$\mathcal{C}(X, \omega_0) \;=\; \pi_X^N(\omega_0)/\pi_X^N(0.5)\ . \tag{3.3}$$

Then, given parameters $\omega_0 \in (0, 0.5)$ and $\mathcal{C}_0 \in [0, 1]$, we propose to select those SVD components whose eigenvectors satisfy the following criterion:

$$\mathcal{C}(U_j, \omega_0) \geqslant \mathcal{C}_0\ , \tag{3.4}$$

where U_j is the corresponding j-th eigenvector. One may interpret this method as selection of SVD components with EOFs mostly characterized by low-frequency fluctuations. It is worth noting here that when we apply $\mathcal{C}$, π or I (defined above for a time series) to a vector, they are simply applied to a series of elements of the vector.

Having the trend SVD components selected using (3.4), one reconstructs the trend according to Section 2. The question is how to select ω_0 and how to define the threshold $\mathcal{C}_0$. These issues are discussed in Sections 4 and 5, respectively.

4. THE LOW-FREQUENCY BOUNDARY ω_0

The low-frequency boundary ω_0 manages the scale of the extracted trend: the lower is ω_0, the slower varies the extracted trend. Selection of ω_0 can be done a priori based on additional information about the data thus prespecifying the desired scale of the trend.

For example, if we assume to have a quasi-periodic component with known period T, then we should select $\omega_0 < 1/T$ in order not to include this component in the trend. For extraction of a trend of monthly data with possible seasonal oscillations of period 12, we suggest to select $\omega_0 < 1/12$, e.g. $\omega_0 = 0.075$.

In this paper we also propose a method of selection of ω_0 considering a time series periodogram. Since a trend is a slow varying component, its periodogram has large values close to zero frequency and small values for other frequencies. The problem of selecting ω_0 is the problem of finding such a low-frequency value that the frequencies corresponding to the large trend periodogram values are inside the interval $[0, \omega_0]$. At the same time, ω_0 cannot be too large because then an oscillating component with a frequency less than ω_0 can be included in the trend produced. Considering the periodogram of a trend, we could find the proper value of ω_0 but for a given time series its trend is unknown.

What we propose is to choose ω_0 based on the periodogram of the original time series. The following proposition substantiates this approach.

Proposition 4.1. *Let us have two time series $G = (g_0, ..., g_{N-1})$ and $H = (h_0, ..., h_{N-1})$ of length N, then for each k: $0 \le k \le \lfloor N/2 \rfloor$ the following inequality holds:*

$$\left| I^N_{G+H}(k/N) - I^N_G(k/N) - I^N_H(k/N) \right| \leqslant 2\sqrt{I^N_G(k/N)\, I^N_H(k/N)}\,. \tag{4.1}$$

Proof of Proposition 4.1: Let us first consider the case when $0 < k < N/2$. We denote as $c_{k,X}$ and $s_{k,X}$ the coefficients of Fourier representation of a time series X used in the periodogram definition (3.1). Then, by this definition,

$$\begin{aligned} I^N_{G+H}(k/N) - I^N_G(k/N) - I^N_H(k/N) \;&= \\ &= \frac{N}{2}\left(c^2_{k,G+H} + s^2_{k,G+H} - c^2_{k,G} - s^2_{k,G} - c^2_{k,H} - s^2_{k,H}\right). \end{aligned}$$

Since $c_{k,G+H} = \frac{2}{N}\,\Re\mathcal{F}_k(G+H) = c_{k,G} + c_{k,H}$ (where $\Re z$ denotes a real part of a complex number z) and, analogously, $s_{k,G+H} = s_{k,G} + s_{k,H}$, we have

$$I^N_{G+H}(k/N) - I^N_G(k/N) - I^N_H(k/N) \;=\; N\big(c_{k,G}c_{k,H} + s_{k,H}s_{k,H}\big)\,. \tag{4.2}$$

Let us consider the periodograms multiplication used in the right part of (4.1):

$$I_G^N(k/N)\, I_H^N(k/N) \;=\; \frac{N^2}{4}\left(c_{k,G}^2 + s_{k,G}^2\right)\left(c_{k,H}^2 + s_{k,H}^2\right)\,. \tag{4.3}$$

Since for all real a, b, c and d it holds that $(a^2+b^2)\,(c^2+d^2) = (|a\,c| + |b\,d|)^2 + (|a\,d| - |b\,c|)^2$, then

$$\begin{aligned} I_G^N(k/N)\, I_H^N(k/N) \;&= \\ &= \frac{N^2}{4}\left(|c_{k,G}\,c_{k,H}| + |s_{k,G}\,s_{k,H}|\right)^2 + \left(|c_{k,G}\,s_{k,H}| - |c_{k,H}\,s_{k,G}|\right)^2. \end{aligned} \tag{4.4}$$

Finally, taking the square of (4.2), dividing it by four and taking into account (4.4), we have

$$\begin{aligned} \frac{1}{4}\left(I_{G+H}^N(k/N) - I_G^N(k/N) - I_H^N(k/N)\right)^2 \;&= \\ &= \frac{N^2}{4}\left(c_{k,G}\,c_{k,H} + s_{k,G}\,s_{k,H}\right)^2 \\ &\leqslant \frac{N^2}{4}\left(|c_{k,G}\,c_{k,H}| + |s_{k,G}\,s_{k,H}|\right)^2 \\ &\leqslant \frac{N^2}{4}\left(|c_{k,G}\,c_{k,H}| + |s_{k,G}\,s_{k,H}|\right)^2 + \left(|c_{k,G}\,s_{k,H}| - |c_{k,H}\,s_{k,G}|\right)^2 \\ &= I_G^N(k/N)\, I_H^N(k/N) \end{aligned}$$

and the inequality in (4.1) holds $0 < k < N/2$.

Second, we consider the case when $k = 0$ or $k = N/2$. Again, by the definition of the periodogram

$$2\sqrt{I_G^N(k/N)\, I_H^N(k/N)} \;=\; 2\sqrt{N^2\, c_{k,G}^2\, c_{k,H}^2} \;=\; 2\,N\,|c_{k,G}\,c_{k,H}|\,.$$

At the same time,

$$\left|I_{G+H}^N(k/N) - I_G^N(k/N) - I_H^N(k/N)\right| = N\left|c_{k,G+H}^2 - c_{k,G}^2 - c_{k,H}^2\right| = N\left|2\,c_{k,G}\,c_{k,H}\right|$$

which leads for $k = 0$ or $k = N/2$ to

$$\left|I_{G+H}^N(k/N) - I_G^N(k/N) - I_H^N(k/N)\right| \;=\; 2\sqrt{I_G^N(k/N)\, I_H^N(k/N)}$$

and the result in (4.1) holds with equality. □

Corollary 4.1. *Let us define for a time series F of length N the frequency support of the periodogram I_F^N as a subset of frequencies $\{k/N\}_{k=0}^{\lfloor N/2\rfloor}$ such that $I_F^N(k'/N) > 0$ for k'/N from this subset. If the frequency supports of two time series G and H are disjoint then $I_{G+H}^N(k/N) = I_G^N(k/N) + I_H^N(k/N)$.*

Let us demonstrate that when supports of periodograms of time series G and H are nearly disjoint, the periodogram of the sum $G+H$ is close to the sum of their periodograms.

The fact that the periodograms of G and H are very different at k/N can be expressed as

$$I_G^N(k/N) \,/\, I_H^N(k/N) \;=\; d \;\gg\; 1 \;,$$

since without loss of generality we can assume $I_G^N(k/N) > I_H^N(k/N)$. Then using Proposition 4.1 we have that

$$\begin{aligned}\left|I_{G+H}^N(k/N) - I_G^N(k/N) - I_H^N(k/N)\right| \;\leqslant \\ \leqslant\; 2\sqrt{I_G^N(k/N)\; I_H^N(k/N)} \;=\; \frac{2}{\sqrt{d}}\; I_G^N(k/N) \;\ll\; I_G^N(k/N) \;,\end{aligned}$$

that means that the difference $\left|I_{G+H}^N(k/N) - I_G^N(k/N) - I_H^N(k/N)\right|$ is significantly smaller than the value of the largest periodogram (of I_G^N, I_H^N) at the point k/N.

In many applications, the given time series can be modelled as made of a trend with large periodogram values at low-frequency interval $[0, \omega_0]$, oscillations with periods smaller than $1/\omega_0$, and noise whose frequency contribution spreads over all the frequencies $[0, 0.5]$ but is relatively small. In this case the periodogram supports of the trend and the residual can be considered as nearly disjoint. Therefore, from Corollary 4.1, we conclude that the periodogram of the time series is approximately equal to the sum of the periodograms of the trend, oscillations and noise.

For a time series X of length N, we propose to select the value of the parameter ω_0 according to the following rule:

$$\omega_0 \;=\; \max_{k/N,\, 0 \leqslant k \leqslant N/2} \left\{ k/N \colon\; I_X^N(0), ..., I_X^N(k/N) < M_X^N \right\}, \tag{4.5}$$

where M_X^N is the median of the values of periodogram of X. The modelling of a time series as a sum of a trend, oscillations and a noise (let us suppose to have a normal noise) motivates this rule as follows. Since the frequency supports of the trend and oscillating components do not overlap, only the values of the noise periodogram can mix with the values of the trend periodogram. First, the values of the noise periodogram for neighboring ordinates are asymptotically independent (see e.g. section 7.3.2 of [7]). Second, supposing a relatively long time series and narrow frequency supports of trend and oscillating components, the median of values of the time series periodogram gives an estimation of the median of the values of the noise periodogram. Since a trend is supposed to have large contribution to the shape of the time series (i.e. a large L_2-norm) compared to the noise and its frequency support is quite narrow compared to the whole interval $[0, 0.5]$, its periodogram values are relatively larger than the median of the noise periodogram values due to (3.2). Therefore, the condition used in (4.5)

is fulfilled only for such a frequency ω_0 that the trend periodogram values is close to zero (outside the trend frequency interval). Large noise periodogram values in this frequency region can lead to selecting larger than necessary ω_0. But remember that we compare the periodogram values with their median and the noise periodogram values are independent (asymptotically). Hence, with probability approximately equal to $1-0.5^m$ (e.g. this value is equal to 0.9375 for $m=4$) we select the m-th point (of the grid $\{k/N\}$) located to the right side of the trend frequency interval (where the trend peridogram values are larger then the noised periodogram median).

Note that the lengths N of the time series and L of eigenvector are different ($L<N$) which causes different resolution of their periodograms. Having estimated ω_0 after consideration of the periodogram of the original time series, one should select

$$\omega_0' = \lceil L\,\omega_0 \rceil / L\ . \tag{4.6}$$

Dependence of ω_0 on the time series resolution. Let us define the resolution ρ of the original time series as $\rho = (\tau_{n+1}-\tau_n)^{-1}$, where τ_n is the time of n-th measurement. If one have estimated ω_0 for the data with resolution ρ and there comes the same data but measured with higher resolution $\rho' = m\,\rho$ ($m \in \mathbb{R}$) thus increasing the data length in m times, then in order to extract the same trend, one should take the new threshold value $\omega_0' = \omega_0/m$. In a similar manner, after decimation of the data reducing the resolution in m times, the value $\omega_0' = m\,\omega_0$ should be taken.

Example 4.1 (The choice of ω_0 for a noised exponential trend)**.** Let us consider an example of selection of the threshold ω_0 for an exponential trend and a white Gaussian noise which also demonstrates Proposition 4.1. Let the time series $F = G + H$ be of length $N = 120$, where the components G and H are defined as $g_n = A e^{0.01n}$, $h_n = B\,\varepsilon_n$, $\varepsilon_n \sim \text{iid}\, N(0,1)$ and A, B are selected so that $\|G\|^2 = \|H\|^2 = \sum_{n=0}^{N-1} g_n = \sum_{n=0}^{N-1} h_n = 1$. The normalization is done to ensure that $\sum_{k=0}^{60} I_G^N(k/N) = \sum_{k=0}^{60} I_H^N(k/N) = 1$. Figure 1 shows a) the simulated time series F, b) its components, c) the periodograms of the components, d) the periodograms zoomed together with a line corresponding to the median of the noise periodogram values equal to 0.0126, e) the periodogram I_F^N of F and a kind of "confidence" interval of its estimation $I_G^N + I_H^N$ calculated according Proposition 4.1 and a line corresponding to the median M_F^{120} of the time series periodogram values (used for estimating ω_0), and f) the discrepancy, the difference between I_F^N and $I_G^N + I_H^N$ together with the values of this difference estimated in the right side of (4.1). Note tha the median of the periodogram values of F is equal to 0.0141, which is close to the median of the noise periodogram values equal to 0.0126. The value of ω_0 estimated according to the proposed rule (4.5) is equal to $6/120 = 0.05$.

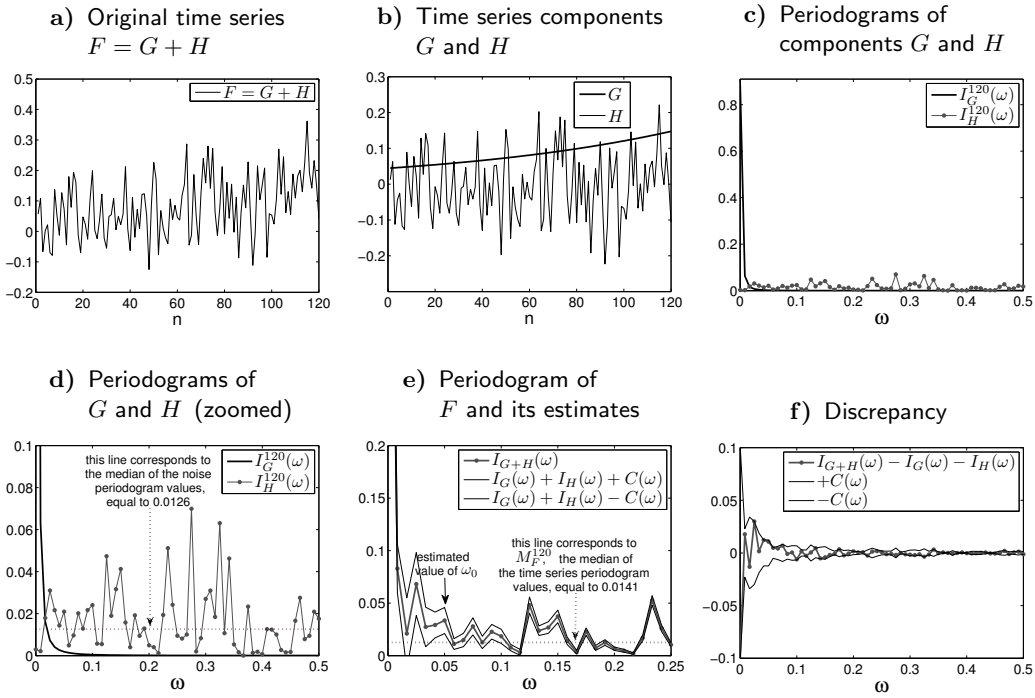

Figure 1: The choice of ω_0 for an exponential trend and Gaussian noise; The value $C(\omega)$ used in the legends is equal to $2\sqrt{I_G^N(\omega)\, I_H^N(\omega)}$.

5. THE LOW-FREQUENCY CONTRIBUTION $\mathcal{C}_0$

Before suggesting a procedure for selection of the second parameter of the proposed method, the low-frequency threshold $\mathcal{C}_0$, we investigate the effect of the choice of $\mathcal{C}_0$ on the quality of the trend extracted. For this aim, we consider a time series model with a trend that generates SVD components with known numbers. Then, for a sufficient number of simulated time series, we compare our trend extraction procedure with a SSA-based procedure which simply reconstructs the trend using the known trend SVD components.

5.1. A simulation example: an exponential trend plus a Gaussian noise

The model considered is the same as in example above. Let the time series $F = (f_0, ..., f_{N-1})$ consist of an exponential trend t_n plus a Gaussian white noise r_n:

$$f_n = t_n + r_n \ , \qquad t_n = e^{\alpha n} \ , \qquad r_n = \sigma\, e^{\alpha n}\, \varepsilon_n, \quad \varepsilon_n \sim \text{iid}\, N(0,1) \ . \tag{5.1}$$

According to [11], for such a time series with moderate noise the first SVD component corresponds to the trend. We considered only the noise levels when this is true (empirically checked). Note that the noise r_n has a multiplicative model as its standard deviation is proportional to the trend.

In the following, we consider the following properties. First, we calculate the difference between the trend $\hat{t}_n(\mathcal{C}_0)$ resulted from our method with $\mathcal{C}_0$ used and the reconstruction $\tilde{t}_n$ of the first SVD component exploiting the weighted mean square error (MSE) because this measure is more relevant for a model with a multiplicative noise than a simple MSE:

$$\mathcal{D}\big(\hat{t}_n(\mathcal{C}_0), \tilde{t}_n\big) \;=\; \frac{1}{N}\sum_{n=0}^{N-1} e^{-2\alpha n}\big(\hat{t}_n(\mathcal{C}_0) - \tilde{t}_n\big)^2 . \tag{5.2}$$

This measure compares our trend and the ideal SSA trend. Second, we calculate the weighted mean square errors between $\hat{t}_n(\mathcal{C}_0)$, $\tilde{t}_n$ and the true trend t_n separately:

$$\mathcal{D}\big(\hat{t}_n(\mathcal{C}_0)\big) = \frac{1}{N}\sum_{n=0}^{N-1} e^{-2\alpha n}\big(t_n - \hat{t}_n(\mathcal{C}_0)\big)^2, \quad \mathcal{D}(\tilde{t}_n) = \frac{1}{N}\sum_{n=0}^{N-1} e^{-2\alpha n}\big(t_n - \tilde{t}_n\big)^2 . \tag{5.3}$$

5.1.1. Scheme of estimation of the errors using simulation

The errors (5.2), (5.3) are estimated using the following scheme. We simulate S realizations of the time series F according to the model (5.1) and calculate the mean of $\mathcal{D}\big(\hat{t}_n(\mathcal{C}_0), \tilde{t}_n\big)$ for all values of $\mathcal{C}_0$ from the large grid 0:0.01:1:

$$\overline{\mathcal{D}}\big(\hat{t}_n(\mathcal{C}_0), \tilde{t}_n\big) \;=\; \frac{1}{S}\sum_{s=1}^{S} \mathcal{D}\big(\hat{t}_n^{(s)}(\mathcal{C}_0), \tilde{t}_n^{(s)}\big) , \tag{5.4}$$

where $\hat{t}_n^{(s)}(\mathcal{C}_0)$ and $\tilde{t}_n^{(s)}$ denote trends of the s-th simulated time series. The mean errors $\overline{\mathcal{D}}\big(\hat{t}_n(\mathcal{C}_0)\big)$, $\overline{\mathcal{D}}(\tilde{t}_n)$ between the true trend t_n and the extracted trends $\hat{t}_n(\mathcal{C}_0)$ and $\tilde{t}_n$, respectively, are calculated similarly. Let us also denote the minimal values of the mean errors as

$$\overline{\mathcal{D}}^{\min}(\hat{t}_n, \tilde{t}_n) = \min_{\mathcal{C}_0} \overline{\mathcal{D}}\big(\hat{t}_n(\mathcal{C}_0), \tilde{t}_n\big) , \qquad \overline{\mathcal{D}}^{\min}(\hat{t}_n) = \min_{\mathcal{C}_0} \overline{\mathcal{D}}\big(\hat{t}_n(\mathcal{C}_0)\big) \tag{5.5}$$

and the value of $\mathcal{C}_0$ providing the minimal mean error between the extracted trend and the ideal SSA trend as

$$\mathcal{C}_0^{\text{opt}} = \arg\min_{\mathcal{C}_0} \overline{\mathcal{D}}\big(\hat{t}_n(\mathcal{C}_0), \tilde{t}_n\big) ,$$

so that $\overline{\mathcal{D}}^{\min}(\hat{t}_n, \tilde{t}_n) = \overline{\mathcal{D}}\big(\hat{t}_n(\mathcal{C}_0^{\text{opt}}), \tilde{t}_n\big)$.

The simulated time series are of length $N = 47$. In order to achieve the best separability [11] we have selected the SSA window length $L = \lceil N/2 \rceil = 24$. The estimates of the mean errors are calculated on $S = 10^4$ realizations of the time series.

We consider different values of the model parameters α and σ. The values of α are 0 (corresponding to a constant trend), 0.01 and 0.02 which correspond to the increase of trend values (from t_0 to t_{N-1}) in 1, 1.6 and 2.5 times, respectively. The levels of noise are $0.2 \leqslant \sigma \leqslant 1.6$. It was empirically checked that for such levels of noise the first SVD component corresponds to the trend.

Moreover, we estimated the probability of the type I error of not selecting the first SVD component as the ratio of times when the first component is not identified as a trend component by our procedure to the number of repetitions S.

Choice of ω_0. In order to select the low-frequency threshold ω_0, we considered several simulated time series with different α and the maximal noise $\sigma = 1.6$. Two examples of their periodograms for $\alpha = 0$ and $\alpha = 0.02$ are depicted in Figure 2. The median values for the periodograms depicted in Figure 2 are 2.936 and 2.924 which leads to $\omega_0 = 0$ for $\alpha = 0$ and $\omega_0 = \lceil 1/N \cdot L \rceil / L = 1/24 \approxeq 0.042$ for $\alpha = 0.02$ estimated using (4.6). We decided to take the same $\omega_0 = 0.042$ (the largest one) for all α considered.

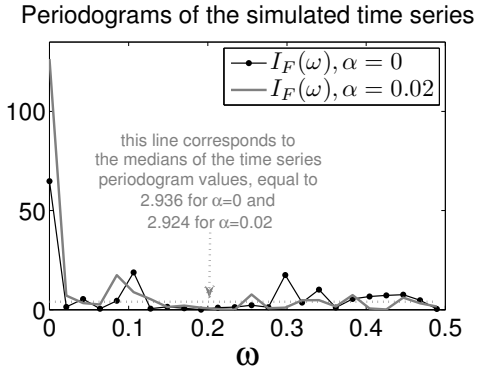

Figure 2: The periodograms of two time series of the model (5.1) with $\sigma = 1.6$ and $\alpha = 0, 0.02$.

5.1.2. Simulation results

Figure 3 shows the evolution of the square roots of the mean errors and $\mathcal{C}_0^{\text{opt}}$ as a function of σ. The values $\alpha = 0$ and $\alpha = 0.02$ are used. The square roots of the mean errors (i.e. standard deviations) are taken for better comparison with σ which is the standard deviation multiplier of the noise.

The plots of the minimal mean error $\overline{\mathcal{D}}^{\min}(\hat{t}_n, \tilde{t}_n)$ and the optimal $\mathcal{C}_0^{\text{opt}}$ for $\alpha = 0.02$ are depicted in Figure 3, where the values for $\alpha = 0$ are also shown in gray color. The estimates for $\alpha = 0.01$ are not reported here.

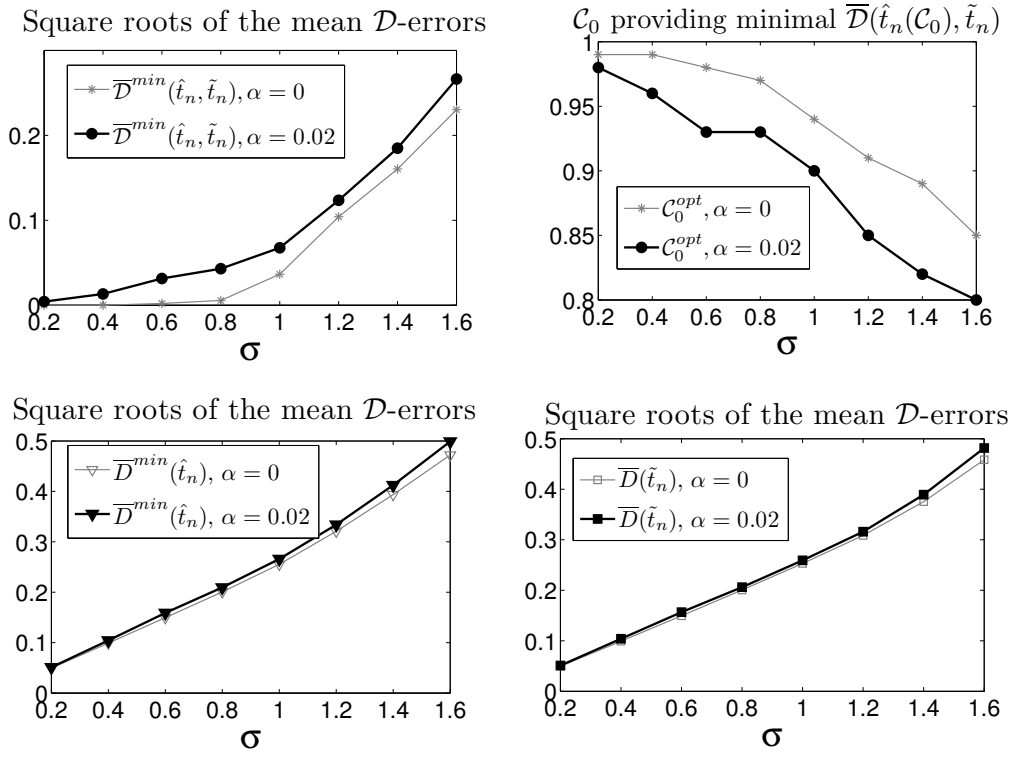

Figure 3: The square roots of the mean errors $\overline{\mathcal{D}}^{\min}(\hat{t}_n, \tilde{t}_n)$ (top left) $\overline{\mathcal{D}}^{\min}(\hat{t}_n)$ (bottom left) and $\overline{\mathcal{D}}(\tilde{t}_n)$ (bottom right) as well as the optimal $\mathcal{C}_0$ value providing a minimal mean error $\overline{\mathcal{D}}^{\min}(\hat{t}_n, \tilde{t}_n)$ between the extracted trend and the ideal SSA trend (top right); all for $\alpha = 0$ and $\alpha = 0.02$.

The interpretation of the produced results is as follows. First, the trend extracted with the optimal $\mathcal{C}_0$ is very similar to the ideal SSA trend, reconstructed by the first SVD component since $\overline{\mathcal{D}}^{\min}(\hat{t}_n, \tilde{t}_n) \ll \overline{\mathcal{D}}^{\min}(\hat{t}_n)$ (the error between our trend and the ideal trend is much smaller than the error of the ideal trend itself), especially when $\sigma \leqslant 0.8$. Moreover, the estimated probability of the type I error (i.e. the probability of not selecting the first SVD component) is less than 0.05 for $\sigma \leqslant 1.4$. All this allows us to conclude that in case of an exponential trend and a white Gaussian noise the proposed method of trend extraction with an optimal $\mathcal{C}_0$ with high probability selects the required first SVD component corresponding to the trend.

The trend $\hat{t}_n(\mathcal{C}_0^{\text{opt}})$ extracted with an optimal $\mathcal{C}_0$ estimates the true trend quite good when comparing the deviation $\sqrt{\overline{\mathcal{D}}^{\min}(\hat{t}_n)}$ with the noise standard

deviation σ. For example, for $\sigma = 1.6$ the value of $\sqrt{\overline{\mathcal{D}}^{\min}(\hat{t}_n)}$ is approximately equal to 0.5.

Note that for different α the mean errors $\overline{\mathcal{D}}^{\min}(\hat{t}_n)$ are very similar though the used optimal values of $\mathcal{C}_0$ are quite different (Figure 3). This shows that the method adapts to the change of the model parameter α.

Let us consider the dependence of inaccuracy of the proposed trend extraction method on the value of $\mathcal{C}_0$. As above, the inaccuracy is measured with the minimal mean error $\overline{\mathcal{D}}^{\min}(\hat{t}_n, \tilde{t}_n)$ between the extracted trend and the ideal SSA trend. Figure 4 shows the graphs of this error as a function of $\mathcal{C}_0$ for different exponentials α and noise levels σ.

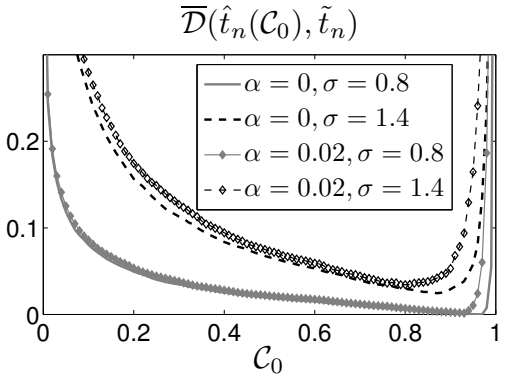

Figure 4: The error $\overline{\mathcal{D}}\big(\hat{t}_n(\mathcal{C}_0), \tilde{t}_n\big)$ as a function of $\mathcal{C}_0$ for different combinations of $\alpha = 0, 0.02$ and $\sigma = 0.8, 1.4$.

One can see that it is crucial not to select too large $\mathcal{C}_0$ since in this case the trend component can be not included in the reconstruction (that is also confirmed by the estimated probability of the type I error which is not reported here). At the same time without significant loss of accuracy one can choose $\mathcal{C}_0$ smaller than $\mathcal{C}_0^{\text{opt}}$ (corresponding to the best accuracy). This is true due to the small contribution of each of noise components which can be erroneously included for $\mathcal{C}_0 < \mathcal{C}_0^{\text{opt}}$.

5.2. Heuristic procedure for the choice of $\mathcal{C}_0$

Based on the observations of Section 5.1, we propose the following heuristic procedure for choosing the value of the method low-frequency threshold $\mathcal{C}_0$.

As discussed, trend EOFs vary slow. First we show that this property is inherited by the trend elementary reconstructed components, the time series components each reconstructed from one trend SVD component.

Proposition 5.1. *Let $(\sqrt{\lambda}, U, V)$ be an eigentriple of SSA decomposition of a time series F, $U = (u_1, ..., u_L)^T$, $V = (v_1, ..., v_L)^T$, and G be a time series reconstructed by this eigentriple. If it is true that*

$$\exists\, \delta_1, \delta_2 \in \mathbb{R}: \quad \forall\, k,\ 1 \leqslant k \leqslant L-1: \qquad |u_{k+1} - u_k| < \delta_1\,, \quad |v_{k+1} - v_k| < \delta_2\ ,$$

then for the elements of $G = (g_0, ..., g_{N-1})$ the following holds:

$$\exists\, \epsilon(\delta_1, \delta_2): \quad \forall\, n,\ L^* - 1 \leqslant n < K^*: \qquad |g_{n+1} - g_n| < \epsilon(\delta_1, \delta_2)\ ,$$

where $L^* = \min\{L, K\}$, $K^* = \max\{L, K\}$.

Proof of Proposition 5.1: One can easily prove this proposition taking into account how the elementary reconstructed component G is constructed from its eigentriple $(\sqrt{\lambda}, U, V)$, see Section 2. First, the matrix $\mathbf{Y} = \sqrt{\lambda}\, U V^{\mathrm{T}}$ is constructed. Second, the hankelization of $\mathbf{Y}$ is performed.

Let us show how to calculate ϵ using (2.2) for δ_1, δ_2 when $L \leqslant K$. For other cases $\epsilon(\delta_1, \delta_2)$ is calculated similarly.

$$\begin{aligned} |g_{n+1} - g_n| \;&=\; \frac{\sqrt{\lambda}}{L} \left| \sum_{m=1}^{L} \big(u_m\, v_{n-m+3} - u_m\, v_{n-m+2}\big) \right| \\ &<\; \frac{\sqrt{\lambda}}{L} \sum_{m=1}^{L} |u_m|\, |v_{n-m+3} - v_{n-m+2}| \\ &<\; \frac{\sqrt{\lambda}}{L}\, \delta_2 \sum_{m=1}^{L} |u_m| \;<\; \delta_2\, \frac{\sqrt{\lambda}}{L} \big(u_1 + (L-1)\, \delta_1\big)\,. \end{aligned} \qquad \square$$

Let us have a time series F and denote its trend extracted with the method with parameters $\omega_0, \mathcal{C}_0$ as $T(\omega_0, \mathcal{C}_0)$. In order to propose the procedure selecting $\mathcal{C}_0$, we first define the normalized contribution of low-frequency oscillations in the residual $F - T(\omega_0, \mathcal{C}_0)$ as:

$$\mathcal{R}_{F,\omega_0}(\mathcal{C}_0) \;=\; \mathcal{C}\big(F - T(\omega_0, \mathcal{C}_0), \omega_0\big)\, \mathcal{C}(F, \omega_0)^{-1}\,,$$

where $\mathcal{C}$ is defined in equation (3.3).

Based on Proposition 5.1, we expect that the elementary reconstructed components corresponding to a trend have large contribution of low frequencies. Thus, the maximal values of $\mathcal{C}_0$ which lead to selection of trend-corresponding SVD components should generate jumps of $\mathcal{R}_{F,\omega_0}(\mathcal{C}_0)$.

Exploiting this idea, we propose the following way of choosing $\mathcal{C}_0$:

$$\mathcal{C}_0^{\mathcal{R}} \;=\; \min\Big\{\mathcal{C}_0 \in [0,1]:\ \mathcal{R}_{F,\omega_0}(\mathcal{C}_0 + \Delta\mathcal{C}) - \mathcal{R}_{F,\omega_0}(\mathcal{C}_0) \geq \Delta\mathcal{R}\Big\}\,, \tag{5.6}$$

where $\Delta\mathcal{C}$ is a search step and $\Delta\mathcal{R}$ is the given threshold. On one hand, this strategy is heuristic and requires selection of $\Delta\mathcal{R}$, but on the other hand, the simulation results and application to different time series showed its ability to choose reasonable $\mathcal{C}_0$ in many cases. Based on this empirical experience, we suggest using $0.05 \le \Delta\mathcal{R} \le 0.1$. The step $\Delta\mathcal{C}$ is to be chosen as small as possible to discriminate identifications occurring at different values of $\mathcal{C}_0$. To reduce computational time, we commonly take $\Delta\mathcal{C} \ge 0.01$ and suggest a default value of $\Delta\mathcal{C} = 0.01$.

6. EXAMPLES

Simulated example with polynomial trend. The first example illustrates the choice of parameters ω_0 and $\mathcal{C}_0$. We simulated a time series of length $N = 300$, shown in Figure 5, containing a polynomial trend, an exponentially-modulated sine wave, and a white Gaussian noise, whose n-th element is expressed as $f_n = 10^{-11}(n-10)(n-70)(n-160)^2(n-290)^2 + \exp(0.01n)\sin(2\pi n/12) + \varepsilon_n$, ε_n is iid $N(0, 5^2)$. The period of the sine wave is assumed to be unknown.

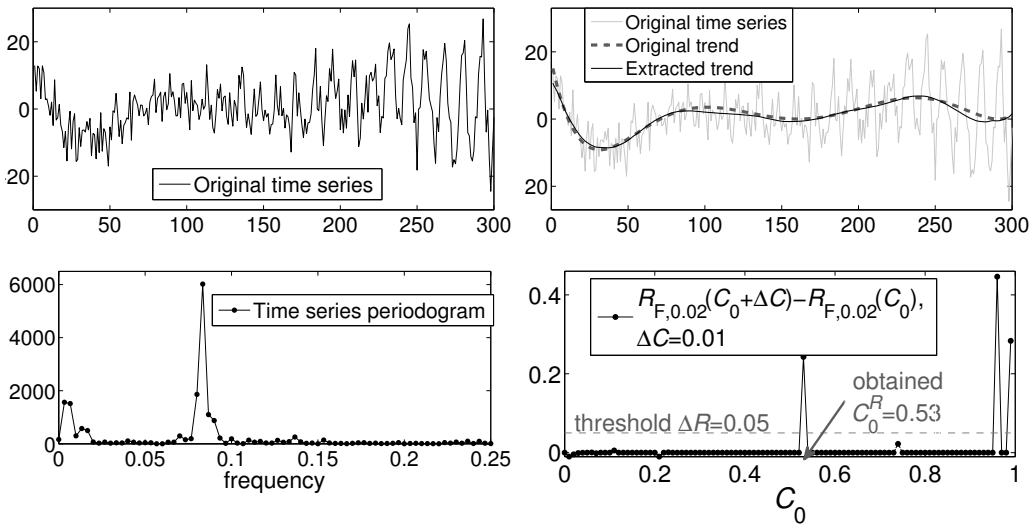

Figure 5: Simulated example with a polynomial trend: original time series (top left); the original trend and an extracted one with $L = 180$, $\Delta\mathcal{C} = 0.01$, and $\Delta\mathcal{R} = 0.05$ (top right); zoomed time series periodogram inside $\omega \in [0, 0.25]$ (bottom left); the values of $\mathcal{R}_{F,\omega_0}(\mathcal{C}_0 + \Delta\mathcal{C}) - \mathcal{R}_{F,\omega_0}(\mathcal{C}_0)$ used for the choice of $\mathcal{C}_0$ resulted in a value $\mathcal{C}_0^{\mathcal{R}} = 0.53$ (bottom right).

We have chosen the window length $L = N/2 = 150$ for achieving better separability of trend and residual. The value $\omega_0 = 6/N = 0.02$ was selected

using (4.5), where the calculated median value is $M_X^N \approxeq 37.06$. The search for $\mathcal{C}_0$ using (5.6) has been done with step $\Delta\mathcal{C} = 0.01$ and $\Delta\mathcal{R} = 0.05$. As shown in Figure 5, despite of the strong noise and oscillations, the extracted trend approximates the original one very well. The achieved mean square error is 0.79. For example, the ideal low pass filter with the cutoff frequency 0.02 produced the error of 3.14. This superiority is achieved mostly due to better approximation at the first and last 50 points of the time series. All the calculations were performed using our Matlab-based software AutoSSA available at `http://www.pdmi.ras.ru/~theo/autossa`.

Trends of the unemployment level. Let us demonstrate extraction of trends of different scale. We took the data of the unemployment level (unemployed persons) in Alaska for the period 1976/01–2006/09 (monthly data, seasonally adjusted), provided by the Bureau of Labor Statistics at `http://www.bls.gov` under the identifier LASST02000004 (Figure 6). This time series is typical for economical applications, where data contain relatively little noise and are subject to abrupt changes. Economists are often interested in the "short" term trend which includes cyclical fluctuations and is referred to as trend-cycle.

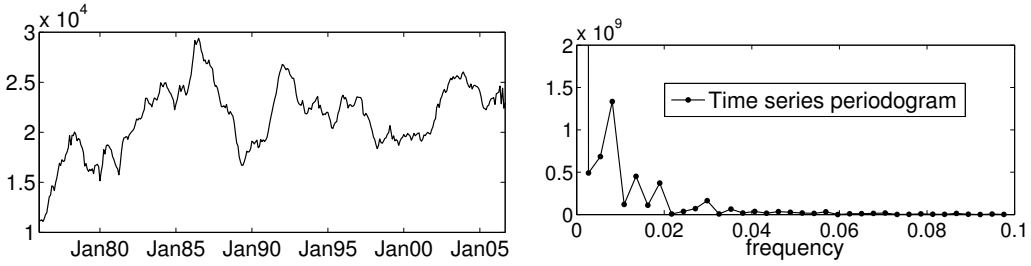

Figure 6: Unemployment level in Alaska: original data (left-hand side panel), zoomed periodogram (right-hand side panel).

The length of the data is $N = 369$. For achieving better separability of trend and residual we selected L close to $N/2$ but divisible by the period $T = 12$ of probable seasonal oscillations: $L = 12\lfloor N/24 \rfloor = 180$.

We extracted trends of different scales using the following values of ω_0: 0.01, 0.02, 0.05, 0.075 and 0.095, see Figure 7 for the results. The value $0.095 \approxeq \lceil 33/369 \cdot 180 \rceil / 180$ was selected according to (4.6), where $M_X^N \approxeq 5.19 \cdot 10^5$. The value 0.075 is the default value for monthly data (Section 4). Other values (0.01, 0.02 and 0.05) were considered for better illustration of how the value of ω_0 influences the scale of the extracted trend. The search for $\mathcal{C}_0$ was performed as described in Section 5 in the interval $[0.5, 1]$ with the step $\Delta\mathcal{C} = 0.01$ and $\Delta\mathcal{R} = 0.05$.

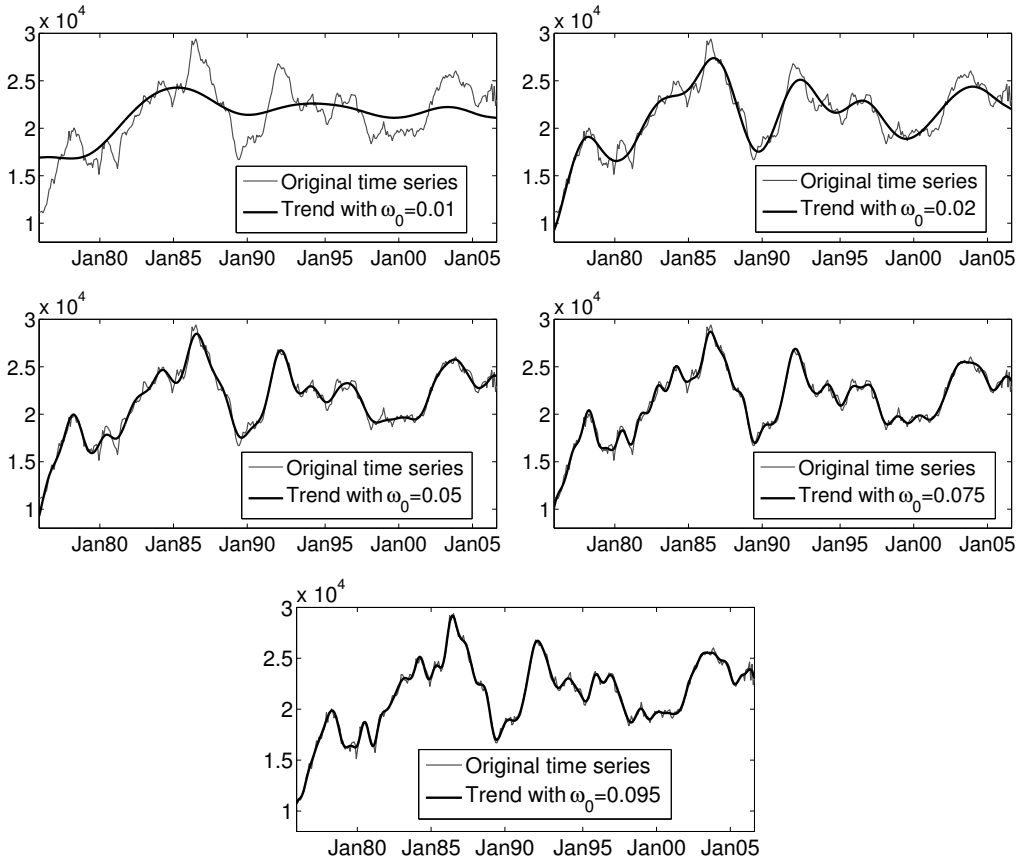

Figure 7: Unemployment level in Alaska: extracted trends of different scales with $\omega_0 = 0.01, 0.02, 0.05, 0.075$ and 0.095 ($L = 180$, $\Delta\mathcal{C} = 0.01$ and $\Delta\mathcal{R} = 0.05$).

7. CONCLUSIONS

SSA is an attractive approach to trend extraction because it: (i) requires no model specification of time series and trend, (ii) extracts trend of noisy time series containing oscillations of unknown period. In this paper, we presented a method which inherits these properties and is easy to use since it requires selection of only two parameters.

ACKNOWLEDGMENTS

The author warmly thanks his Ph.D. thesis advisor Nina Golyandina for her supervision and guidance that helped him to produce the results presented in this paper. The author greatly appreciates an anonymous reviewer for his valuable and constructive comments.

REFERENCES

[1] Alexandrov, T. (2006). *Software package for automatic extraction and forecast of additive components of time series in the framework of the Caterpillar-SSA approach*, PhD thesis, St. Petersburg State University. In Russian, available at `http://www.pdmi.ras.ru/~theo/autossa`.

[2] Alexandrov, T.; Bianconcini, S.; Dagum, E.B.; Maass, P. and McElroy, T.S. (2008). A review of some modern approaches to the problem of trend extraction, *US Census Bureau TechReport* RRS2008/03.

[3] Alexandrov, T. and Golyandina, N. (2005). Automatic extraction and forecast of time series cyclic components within the framework of SSA. In "Proc. of the 5th St. Petersburg Workshop on Simulation", 45–50.

[4] Brillinger, D.R. (2001). *Time Series: Data Analysis and Theory*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.

[5] Broomhead, D.S. and King, G.P. (1986). Extracting qualitative dynamics from experimental data, *Physica D*, **20**, 217–236.

[6] Buchstaber, V.M. (1995). Time series analysis and grassmannians, *Amer. Math. Soc. Trans.*, **162**, 1–17.

[7] Chatfield, C. (2003). *The Analysis of Time Series: An Introduction*, 6th ed., Chapman & Hall/CRC.

[8] Colebrook, J.M. (1978). Continuous plankton records — zooplankton and environment, northeast Atlantic and North Sea, 1948–1975, *Oceanol. Acta.*, **1**, 9–23.

[9] Drmač, Z. and Veselić, K. (2005). *New fast and accurate Jacobi SVD algorithm: I, II*, Tech. Rep. LAPACK Working Note 169, Dep. of Mathematics, University of Zagreb, Croatia.

[10] Ghil, M.; Allen, R.M.; Dettinger, M.D.; Ide, K.; Kondrashov, D.; Mann, M.E.; Robertson, A.; Saunders, A.; Tian, Y.; Varadi, F. and Yiou, P. (2002). Advanced spectral methods for climatic time series, *Rev. Geophys.*, **40**(1), 1–41.

[11] Golyandina, N.E.; Nekrutkin, V.V. and Zhigljavsky, A.A. (2001). *Analysis of Time Series Structure: SSA and Related Techniques*, Boca Raton, FL: Chapman&Hall/CRC.

[12] Gu, M. and Eisenstat, S.C. (1993). *A stable and fast algorithm for updating the singular value decomposition*, Tech. Rep. YALEU/DCS/RR-966, Dep. of Computer Science, Yale University.

[13] Kumaresan, R. and Tufts, D.W. (1980). *Data-adaptive principal component signal processing.* In "Proc. of IEEE Conference On Decision and Control", Albuquerque, 949–954.

[14] Nekrutkin, V. (1996). *Theoretical properties of the "Caterpillar" method of time series analysis.* In "Proc. of the 8th IEEE Signal Processing Workshop on Statistical Signal and Array Processing", IEEE Computer Society, 395–397.

[15] Vautard, R.; Yiou, P. and Ghil, M. (1992). Singular-spectrum analysis: A toolkit for short, noisy chaotic signals, *Physica D*, **58**, 95–126.

REVSTAT – Statistical Journal
Volume 7, Number 1, April 2009, 23–36

THE SVM APPROACH FOR BOX–JENKINS MODELS

Authors: Saeid Amiri
– Dep. of Energy and Technology, Swedish Univ. of Agriculture Sciences, P.O. Box 7032, SE 750 07 Uppsala, Sweden
saeid.amiri@et.slu.se

Dietrich von Rosen
– Dep. of Energy and Technology, Swedish Univ. of Agriculture Sciences, P.O. Box 7032, SE 750 07 Uppsala, Sweden
Dietrich.von.Rosen@et.slu.se

Silvelyn Zwanzig
– Department of Mathematics, Uppsala University, Box 480, SE 751 06 Uppsala, Sweden
zwanzig@math.uu.se

Abstract:

- Support Vector Machine (SVM) is known in classification and regression modeling. It has been receiving attention in the application of nonlinear functions. The aim is to motivate the use of the SVM approach to analyze the time series models. This is an effort to assess the performance of SVM in comparison with ARMA model. The applicability of this approach for a unit root situation is also considered.

Key-Words:

- *Support Vector Machine; time series analysis; unit root.*

AMS Subject Classification:

- 49A05, 78B26.

1. INTRODUCTION

Time series analysis is the study of observations made sequentially in time. It is a complicated field in statistics because of direct and indirect effects of time on the variables in the model. The essential difference between the modeling via time series and ordinary method is that data points taken over time may have an internal relation that should be accounted for. It can be a correlation structure, a trend, seasonality and so on.

Time series can be studied in the time domain and in the time frequency domain. The time domain is more known among researchers in sciences whereas the frequency domain has many applications in engineering. Time domain is modeled by two main approaches. The traditional approach has been given in Box and Jenkins (1970) in their influential book, includes a systematic class of models called autoregressive integrated moving average (ARIMA) (see, for example, Shumway and Stoffer (2000) and Pourahmadi (2001)). A defining feature of these models is that they are multiplicative models, meaning that observed data are assumed to result from the products of factors involving differential or difference equation operators responding to a white noise input.

Other approaches use additive models or structural models. In this approach, it is assumed that the observations include sum of components, each of which deals with a specified time series structure. None of them have inferential tools such as the Box–Jenkins model, for example model selection, parameter estimation and model validation. ARIMA model can therefore be considered as a benchmark model in evaluating the performance of new method. Support Vector Machine is one of the new methods in modeling that has good performance in classification and regression analysis. A few papers have tried to use it for time series, see Müller (1997) and Murkharejee (1997). They have considered dynamic models e.g., the Mackey class equation was used to show the efficiency of SVM.

We are motivated to use SVM because of its ability in dealing with stationary as well as non-stationary series. Moreover, contrary to the traditional methods of time series analysis (autoregressive or structural models that assume normality and stationarity of the series), SVM makes no prior assumptions about the data.

The paper contains five sections and is organized as follows. In Section 2, the necessary theoretical background is provided and the SVM modeling is concisely described. In Section 3, it is shown that the approach of time series modeling can be written as a SVM model. Section 4 includes the discussion of the data and also present the results. Finally some conclusions are given in Section 5.

2. SUPPORT VECTOR MACHINE

During the last decades many researchers have been working on SVM in a variety of fields and it has in fact been a very active field. SVM has impacted on improving the statistical learning method and has been used to solve problems in classification. The SVM approach has improved the modeling, especially for nonlinear models. The review of Burges (1998), Cristianini and Shaw-Taylor (2000) and Bishop (2006) help to understand the concept of SVM. For more details see Vapnik (1995) and Vapnik (1998). Let us briefly consider the SVM regression approach.

In statistics, the aim of modeling is often to find a function $f(x)$ which predicts y in a model $y = f(x) + error$. It is not easy to find $f(x)$. It can be interpolated by using mathematical methods and approximated by using statistical methods. Via some statistical criteria like sum of squares or maximum likelihood, ML, the model can be exploited. To evaluate the procedure, one needs a criterion or loss function. It is defined as "ignoring observation which error is less than ϵ",

$$L(x, y, f) \;=\; \big|y - f(x)\big|_\epsilon = \max\big(0, \big|y - f(x)\big| - \epsilon\big)\,.$$

It is called "ϵ-insensitive error function". Another loss function is Huber's loss function which is the squared distance between the observations and the function, see Cristianini and Shaw-Taylor (2000) and Hasti *et al.* (2001). In Figure 1, the points outside the tube around the function are called slack variables which is shown by ξ_{1i} and ξ_{2j} for above and below the tube, respectively. The value of the points inside the tube is zero and outside is nonzero. To find ξ_{1i} and ξ_{2j}, one should estimate parameters by the error function as below,

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^{N}(\xi_{1i} + \xi_{2i}) + \frac{\lambda}{2}\,\|W\|^2\,, \\ \text{subject to} \quad & y_i \le f + \epsilon + \xi_{1i}\,, \\ & y_i \ge f - \epsilon - \xi_{2i}\,, \\ & \xi_{1i}, \xi_{2j} \ge 0\,. \end{aligned}$$

By using the Lagrange multiplier to find parameters and optimize by the Karush–Kuhn–Tucker condition, $f(x)$ can be shown to equal

$$f(x) = \sum_{i=1}^{N} \alpha_i\, k(x, x_i)\;, \tag{2.1}$$

where α_i are support vectors, i.e. those points that contribute in the prediction. All points within the tube have $\alpha_i = 0$ and a few of α_i are nonzero. In (2.1), $k(x, x_i)$ is the kernel function, which is an inner product of variables, i.e.,

$$k(x, x_i) = \big\langle \phi(x), \phi(x_i) \big\rangle\;. \tag{2.2}$$

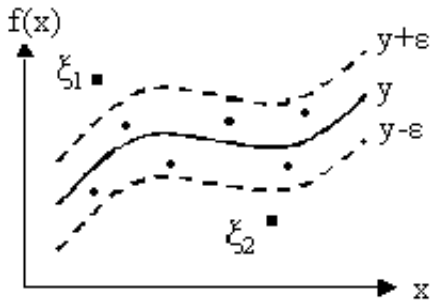

Figure 1: SVM regression with insensitive tube, slack variables ξ_1, ξ_2 and observations.

The following are some kernels:

Linear kernel	$k(x, x') = \langle x, x' \rangle$,
Polynomial kernel	$k(x, x') = \big(a\langle x, x' \rangle + k\big)^d$,
Radial Basis Function kernel (RBF)	$k(x, x') = \exp\big(-\sigma \|x - x'\|^2\big)$,
Laplacian kernel	$k(x, x') = \langle x, x' \rangle \exp\big(-\sigma \|x - x'\|\big)$.

Other kernels are the hyberbolic tangent kernel, the spline kernel, the Bessel and the ANOVA RBF kernel. The number of kernels is unlimited and new kernels can be found by combining existing ones (for more information see Burges (1998), Shaw-Taylor (2000) and Karatzoglou *et al.* (2007)). There are several advantages and disadvantages; SVM is based on the kernel, hence the suitable kernel selection is most important step. However, in practice one needs to study only a few kernel functions (Burges (1998)). The key in SVM is the transformation of a nonlinear problem to a higher dimensional linear space using the kernel function. SVM is not based on any assumptions about the distribution.

3. TIME SERIES ANALYSIS

The Box–Jenkins approach involves identifying an appropriate ARMA process by a mathematical model for forecasting. This model is a combination of AR and MA models. AR(p) is defined as bellow,

$$x_{t+1} = \sum_{j=1}^{p} \phi_j \, x_{t+1-j} + \epsilon_{t+1} \ . \tag{3.1}$$

If one considers the series to be deterministic as linear dynamic systems, a method based on the linear measure such as ARMA model can be used for analysis of the series. However, observed real data are rarely normally distributed and

tend to have marginal distributions with heavier tails. It has been shown that most of the financial time series are nonlinear (see, for example, Soofi and Cao (2002)). Based on the second scenario, we should use the method which has the capability to capture both the linearities and the nonlinearities of the series (see, for example, Hassani *et al.* (2009a) and Hassani *et al.* (2009b)). Here the nonlinear model can be written as

$$x_{t+1} = \sum_{j=1}^{p} \phi_j \, h_j(x_{t+1-j}) + \epsilon_{t+1} \ , \qquad \epsilon_{t+1} \sim N(0, \sigma^2) \ , \tag{3.2}$$

$$x_{t+1} = \big(h_1(x_t), ..., h_p(x_{t+1-p})\big) \begin{pmatrix} \phi_1 \\ \vdots \\ \phi_p \end{pmatrix} , \tag{3.3}$$

$$x = H\,\phi \ , \tag{3.4}$$

where $H = \big(h_1(\cdot), ..., h_p(\cdot)\big)$ and $\phi = (\phi_1, ..., \phi_p)^{\mathrm{T}}$. If H is known, the parameters can be estimated. To simplify assume $\underline{x}_t = (x_t, x_{t-1}, ..., x_{t+1-p})$, $p < t$. The parameters of the model can be estimated by the conditional ML:

$$\begin{aligned} L(\phi, \sigma | \underline{x}_p) &= f(x_{p+1}|\underline{x}_p)\, f(x_{p+2}|\underline{x}_{p+1}) \cdots f(x_t|\underline{x}_{t-1}) \\ &= \prod_{i=p}^{t-1} f(x_{i+1}|\underline{x}_i) \\ &= \prod_{i=p}^{t-1} \frac{1}{\sqrt{2\pi}\,\sigma} \, \exp - \frac{\big(x_{i+1} - \sum_{j=1}^{p} \phi_j \, h(x_{i+1-j})\big)^2}{2\,\sigma^2} \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{(t-p)/2} \exp - \sum_{i=p}^{t-1} \frac{\big(x_{i+1} - \sum_{j=1}^{p} \phi_j \, h(x_{i+1-j})\big)^2}{2\,\sigma^2} \ . \end{aligned} \tag{3.5}$$

Thus, one needs to minimize,

$$SS = \sum_{i=p}^{t-1} \left(x_{i+1} - \sum_{j=1}^{p} \phi_j \, h_j(x_{i+1-j}) \right)^2 = \sum_{i=p}^{t-1} (x_{i+1} - H_i\,\phi)^2 \ . \tag{3.6}$$

To improve the accuracy of the estimation procedure, one can use a penalty function,

$$SS2 = \sum_{i=p}^{t-1} (x_{i+1} - H_i\,\phi)^2 + \lambda \|\phi\| \ = \ (x - H\phi)^{\mathrm{T}} (x - H\phi) + \lambda \|\phi\| \ , \tag{3.7}$$

$$\frac{\partial SS2}{\partial \phi} = 0 \implies -H^{\mathrm{T}}(x - H\phi) + \lambda\,\phi \ = \ 0 \ ,$$

which implies that

$$H\phi \ = \ (HH^{\mathrm{T}} + \lambda\, I)^{-1} H H^{\mathrm{T}} x \ , \tag{3.8}$$

where HH^{T} is a matrix of inner product of the observations. It is quite straightforward to show that (3.8) can be written as an inner product. Therefore, the nonlinear equation can be written as a kernel function,

$$x_{t+1} = f(\underline{x}_t) + e_{t+1} = \sum_{i=1}^{p} \phi_i \, h_i(x_{t+1-i}) + e_{t+1} = \sum_{i=1}^{t} \alpha_i \, k(\underline{x}_t, \underline{x}_i) + e_{t+1} \, . \tag{3.9}$$

Another formula that can be considered is the use of time index, as independent, in the model. This is a reasonable variable as the time series data are collected during time,

$$x_t = \sum_{i=1}^{t} \alpha_i \, k(\underline{x}_t, i) \, . \tag{3.10}$$

Let us now consider the moving average model of order q, MA(q),

$$x_t = \sum_{j=0}^{q} \theta_j \, w_{t-j} \, , \qquad w_t \sim N(0, \sigma^2) \, . \tag{3.11}$$

The previous procedure follows by using a nonlinear function,

$$x_t = \sum_{j=0}^{q} \theta_j \, h(w_{t-j}) \, .$$

It is difficult to decide about the distribution of $h(\cdot)$ beforehand. With the assumption $h(w_{t-j}) \sim N(\mu_n, \sigma_n^2)$, there is no improvement for modeling. However, if the model is invertible, we can write MA as AR and follow the previous model. Hence, there are two problems: the distribution of $h(\cdot)$ and the invertibility of the model which make the behavior of MA a bit unclear for using kernel. The similar problem exists for ARMA(p, q). There are two viewpoints: first, ignorance of MA in the model and considering ARMA(p, q) as AR, and second, if ARMA(p, q) is invertible, then ARMA can be written as AR directly. At any rate, the procedure of AR process can be used.

Let us now consider a unit root process:

$$x_t = \mu + x_{t-1} + w_t = \mu + \mu + x_{t-2} + w_{t-1} + w_t = \cdots = t\mu + x_0 + \sum_{i=0}^{t} w_i \, . \tag{3.12}$$

This is a problem for the Box–Jenkins approach as it violates the stationarity condition, and therefore one can not formulate the Box–Jenkins model (see, for example, Brockwell and Davis (1991)). The modeling of the unit root has been discussed extensively in the literature. There exist some statistical tests for diagnosis and also modeling in the special conditions. Equation (3.12) tells us that the unit root has a regression form of time but because of dependency between

observations, the common regression can not be used for it. In this case, one can use SVM, using the previous discussion and rewriting it as kernel formula. It is not based on the distribution and hence the dependency does not affect on it. It should be noted that, if $\mu = 0$ then this model has major drawback and behaves randomly.

4. APPLICATIONS

In this section, the applicability of SVM for time series analysis is considered. In order to performs the comparison, two different criteria are used: sum of squared residuals (SSR) and Akaike Information Criterion (AIC). AIC is calculated based on $\ln \widehat{\sigma}_k^2 + \frac{2k}{n}$, where $\widehat{\sigma}_k^2 = \frac{SSR}{n}$, k and n are the number of parameters and observations, respectively. In the following, the SVM approach is used in the modeling of AR(2), MA(1) and ARMA$(2,1)$ process.

4.1. AR

Here we use the series that has been used in Brockwell and Davis (1991), Example 9.2.1. The series includes 200 observations. Table 1 shows SSR and AIC of AR(2) and SVM with different kernels. SVM has been calculated using equation (3.9). In the table, the results of a few kernels are presented as SSR of other kernels were larger than AR(2). The results show the efficiency of the Laplacian kernel in comparison with the Box–Jenkins modeling. It should be noted that RBF with $\sigma = 50$ fitted fairly well.

Table 1: SSR and AIC of AR(2) and SVM with different kernels.

Model	SSR	AIC
AR(2)	176.99	−0.102
RBF[1]	171.73	−0.136
RBF[2]	144.33	−0.368
Bessel[1]	161.16	−0.176
Bessel[2]	194.46	0.009
Laplacian[1]	100.83	−0.664
Laplacian[2]	202.68	0.330
linear	177.75	−0.102
poly[3]	176.43	−0.085

[1] Fitted by $\sigma = 10$.
[2] Fitted by $\sigma = 50$.
[3] With 2 degrees.

The calculations in Table 2 are based on equation (3.10). This model uses the time as an independent variable. The table shows how much fitting has been improved. The Laplacian kernel and Bessel kernel have smaller SSR than AR, but other kernels have greater SSR than AR. These values show the Bessel kernel has been fitted well, but its variation is very large. The variation of Laplacian kernel is small in comparison with the Bessel kernel, and hence it seems to be more reliable to use. The Laplacian kernel, for this model, is better than the previous models.

Table 2: Modeling directly based on time for AR(2) with different kernels.

Model	SSR	AIC
Laplacian[1]	56.60	−1.252
Laplacian[2]	21.55	−2.217
Bessel[1]	29.50	−1.830
Bessel[2]	980.17	1.619

[1] Fitted by $\sigma = 10$.
[2] Fitted by $\sigma = 50$.

Moreover, consider AR(2) with $x_t = x_{t-1} - 0.9\,x_{t-2} + \omega_t$. This model is stationary and hence the Box–Jenkins model fits very well. To compare the Box–Jenkins model with SVM, the simulation of this model is performed 1000 times with 100 observations. The results for the Box–Jenkins model and different kernels are shown in Table 3. The first two columns include the results of using (3.9)

Table 3: Percent and order of model in simulation of AR.

Model	model based on x_t		model based on t	
	percent	order	percent	order
AR(2)	0.020	6.93	0.006	2.93
RBF[1]	0.283	3.67	0.00	9.18
RBF[2]	0.000	4.43	0.00	6.00
Bessel[1]	0.023	3.77	0.00	7.90
Bessel[2]	0.000	5.85	0.994	1.00
tangent[1]	0.000	12.51	0.000	12.63
tangent[2]	0.000	12.49	0.000	12.53
splinedot	0.000	14.51	0.000	14.42
spline1	0.000	14.48	0.000	14.36
Laplacian[1]	0.540	2.17	0.000	3.92
Laplacian[2]	0.003	6.27	0.000	2.14
linear	0.020	6.36	0.000	10.21
poly[3]	0.110	5.52	0.000	10.22
ANOVA[1]	0.000	10.98	0.000	7.52
ANOVA[2]	0.000	10.01	0.000	4.99

[1] Fitted by $\sigma = 10$.
[2] Fitted by $\sigma = 50$.
[3] With 2 degrees.

and the second two columns include the results of using (3.10). The order column is the mean of orders of models in all of the simulations and the percent shows how many times the model has the smallest SSR in the simulations. As it appears from Table 3, the Laplacian kernel in 54% time has minimum SSR using x_t, but Bessel kernel has minimum SSR using time as explanatory variable. The results of Table 3 is similar to those obtained in Table 1. Therefore, the Bessel and Laplacian kernel are suitable for AR. Table 2 also shows that the fitted model based on the time index as an explanatory variable has better performance than a model based on x_t.

4.2. MA

The Example 10.4.2 of Brockwell and Davis (1991) is a MA(1) process with 160 observations. Here we use the same series to examine the performance of the SVM modeling. The results are presented in Table 4.

Table 4: SSR and AIC of MA(1) and SVM with different kernel.

Model	SSR	AIC
MA(1)	147	−0.072
Bessel[1]	227.373	0.388
Bessel[2]	198.415	0.252
Laplacian[1]	178.720	0.123
Laplacian[2]	79.282	−0.689

[1] Fitted by $\sigma = 10$.
[2] Fitted by $\sigma = 50$.

The results show that the Laplacian kernel with large σ has been fitted very well to MA(1) and also SSR of using Bessel kernel is close to MA(1), but other kernels have not good performance. As it is mentioned above, SVM has a better performance for a AR(p) model than a MA model. For a AR model, the Laplacian kernel with small σ has smallest SSR, but for MA, the Laplacian kernel with larger σ has smallest SSR. For more clarification, see Table 5 which shows the result of the simulation $y_t = \omega_t + 0.5\,\omega_{t-1}$ with 100 observations. This includes the order and the percent of different models in comparison with the Box–Jenkins model. The results confirm the previous results that indicate the Laplacian kernel with large σ has fitted better, almost 88%, than other methods.

Table 5: Percent and order of model in simulation of MA.

Model	percent	order
MA(1)	0.000	8.08
RBF[1]	0.000	8.54
RBF[2]	0.000	5.00
Bessel[1]	0.000	6.512
Bessel[2]	0.112	2.90
tangent[1]	0.000	12.59
tangent[2]	0.000	12.44
Spline[1]	0.000	14.50
Spline[2]	0.000	14.46
Laplacian[1]	0.000	3.00
Laplacian[2]	0.888	1.11
linear	0.000	10.68
poly[3]	0.000	13.68
ANOVA[1]	0.000	6.89
ANOVA[2]	0.000	4.00

[1] Fitted by $\sigma = 10$.
[2] Fitted by $\sigma = 50$.
[3] With 2 degrees.

4.3. ARMA

Next we consider ARMA$(2, 1)$ with 200 observations from Brockwell and Davis (1991), Example 9.2.3. Table 6 shows SSR and AIC of ARMA$(2, 1)$ and different kernels. The first two columns include the results of using (3.9) and the second two columns include the results of using (3.10). It admits the efficiency of Laplacian kernel for the ARMA model. As it appears from the results, the Laplacian kernel has the smallest SSR in both cases.

Table 6: SSR and AIC of ARMA and SVM with different kernels.

Model	model based on x_t		model based on t	
	SSR	AIC	SSR	AIC
ARMA$(2, 1)$	197.16	0.0157		
RBF[1]	244.16	0.209	1536.55	2.048
RBF[2]	176.26	−0.008	1216.10	1.815
Bessel[1]	201.50	0.037	1460.53	2.018
Bessel[2]	195.39	0.006	56.82	−1.228
Laplacian[1]	116.96	−0.526	350.00	0.569
Laplacian[2]	200.14	0.010	46.76	−1.443

[1] Fitted by $\sigma = 10$.
[2] Fitted by $\sigma = 50$.
[3] With 2 degrees.

To simulate ARMA$(2,1)$, consider $x_t = 0.4\,x_{t-1} + 0.5\,x_{t-2} + \omega_t + 0.2\,\omega_{t-1}$. The simulation results are based on 1000 replications of 100 observations. The results of ARMA$(2,1)$ using the Box–Jenkins and SVM, using different kernels, were presented in Table 7. The results are similar to those obtained in Table 6, which is based on a time series data. As it appears from the table, in both models, equation (3.9) and (3.10), the Laplacian kernel has better performance than others. The Laplacian kernel, using x_t and time as explanatory variables, with $\sigma = 10$ has the smallest SSR in 92.3% and 66% of the simulations, respectively.

Table 7: Percent and order of model in simulation of ARMA$(2,1)$.

Model	model based on x_t		model based on t	
	percent	order	percent	order
ARMA$(2,1)$	0.000	8.80	0.000	9.00
RBF[1]	0.020	5.14	0.000	7.99
RBF[2]	0.000	3.33	0.000	4.93
Bessel[1]	0.000	4.23	0.000	6.47
Bessel[2]	0.002	3.86	0.044	2.33
tangent[1]	0.000	12.56	0.000	12.60
tangent[2]	0.000	12.46	0.000	12.39
Spline[1]	0.000	14.56	0.000	14.47
Spline[2]	0.000	14.40	0.000	14.52
Laplacian[1]	0.923	1.14	0.660	1.48
Laplacian[2]	0.045	3.81	0.296	2.39
Linear	0.000	9.51	0.000	10.87
Poly[3]	0.000	8.67	0.000	10.12
ANOVA[1]	0.000	9.64	0.000	6.48
ANOVA[2]	0.000	7.83	0.000	3.90

[1] Fitted by $\sigma = 10$.
[2] Fitted by $\sigma = 50$.
[3] With 2 degrees.

4.4. Unit root

Let us now consider the application of SVM for a unit root process. The model $x_t = x_{t-1} + \omega_t$ with 100 observations is simulated 1000 to study the SVM performance. The results of SVM modeling for the simulated series are presented in Table 8. For a better understanding of the SVM performance in modeling, the order of the model is presented in comparison with the other competitive methods and also the percent. In this case, modeling by ARMA model is impossible because of the non stationarity property of the series. Nonstationarity can often

be associated with different trends in the signal or heterogeneous segments with different local statistical properties. Table 8 indicates that the Laplacian kernel has been fitted very well to the series.

Table 8: Percent and order of the model in simulation of a unit root process.

Model	percent	order
RBF[1]	0.00	7.68
RBF[2]	0.019	4.08
Bessel[1]	0.000	6.23
Bessel[2]	0.036	2.77
tangent[1]	0.000	11.63
tangent[2]	0.000	11.36
spline[1]	0.000	13.57
spline[2]	0.000	13.42
Laplacian[1]	0.915	1.14
Laplacian[2]	0.002	5.68
linear	0.000	9.87
poly[3]	0.000	9.08
ANOVA[1]	0.002	5.75
ANOVA[2]	0.024	2.85

[1] Fitted by $\sigma = 10$.
[2] Fitted by $\sigma = 50$.
[3] With 2 degrees.

5. CONCLUSION

Although the Box–Jenkins model is still one of the most applied model in time series analysis, there are several major drawbacks; the Box–Jenkins models are based on the stationarity, but this is often not sufficient, for example modeling unit root process using ARMA approach is impossible.

The results of this study show that the ARMA models can be expressed as SVM. The performance of the SVM modeling is studied in comparison with the Box–Jenkins modeling. Particularly, the Laplacian kernel is superior to others. It is therefore concluded that the use of SVM for the ARMA model is of great interest and should be considered (see Section 3). Moreover, the use of time index, as explanatory variable, in modeling will improve the accuracy of the results (see Tables 3, 6 and 7). To clarify the performance of the SVM for time series analysis, several examples and simulated series are used. The empirical results confirm our theoretical results. Our findings also show that the SVM based on the Laplacian kernel works very well for the unit root process.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Dr. Mats Gustafsson and the referees for the valuable suggestions that led to the improvement of this paper.

REFERENCES

[1] Bishop, C.M. (2006). *Pattern Recognition and Machine Learning*, Springer, New York.

[2] Box, G. and Jenkins, G. (1970). *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco.

[3] Brockwell, P.J. and Davis, R.A. (1991). *Time Series: Theory and Methods*, 2nd ed., Springer, New York.

[4] Burges, C.J.C. (1998). A tutorial on support vector machines for pattern recognition, *Data Mining and Knowledge Discovery*, **2**(2), 121–167.

[5] Cristianini, N. and Shaw-Taylor, J. (2000). *An introduction to Support Vector Machine*, Cambridge University Press, New York.

[6] Hassani, H.; Heravi, H. and Zhigljavsky, A. (2009). Forecasting European industrial production with singular spectrum analysis, *International Journal of Forecasting*, doi:10.1016/j.ijforecast.2008.09.007.

[7] Hassani, H.; Dionisio, A. and Ghodsi, M. (2009). The effect of noise reduction in measuring the linear and nonlinear dependency of financial markets, *Nonlinear Analysis: Real World Applications*, doi:10.1016/j.nonrwa.2009.01.004.

[8] Hastie, T.; Tibshirani, R. and Friedman, J. (2000). *Elements of Statistical learning: Data Mining, Inference and Prediction*, Springer, New York.

[9] Karatzoglou, A. *et al.* (2007). *kernel lab package*, `http://cran.r-project.org/src/contrib/Descriptions/kernlab.html`

[10] Müller, K.R. *et al.* (1997). *Predicting time series with support vector machines*, ICANN'97, Berlin, 999–1004.

[11] Murkharejee, S. *et al.* (1997). *Nonlinear Prediction of Chaotic Time Series using Support Vector Machines*, IEEE workshop on Neural Network for Signal Processing.

[12] Pourahmadi, M. (2001). *Foundations of Time Series Analysis and Prediction Theory*, Wiley, New York.

[13] Shumway, R.H. and Stoffer, D.S. (2000). *Time Series Analysis and Its Applications*, Springer, New York.

[14] Soofi, A. and Cao, L. (Eds.) (2002). *Modelling and Forecasting Financial Data: Techniques of Nonlinear Dynamics*, Kluwer Academic Publishers, Boston.

[15] Vapnik, V.N. (1995). *The Nature of Statistical Learning Theory*, Springer, New York.

[16] Vapnik, V.N. (1998). *Statistical Learning Theory*, Wiley, New York.

REVSTAT – Statistical Journal
Volume 7, Number 1, April 2009, 37–54

PREWHITENING-BASED ESTIMATION IN PARTIAL LINEAR REGRESSION MODELS: A COMPARATIVE STUDY

Authors: Germán Aneiros-Pérez
– Departamento de Matemáticas, Universidade da Coruña, Spain
ganeiros@udc.es

Juan Manuel Vilar-Fernández
– Departamento de Matemáticas, Universidade da Coruña, Spain
eijvilar@udc.es

Abstract:

- The problem of semiparametric modelling in time series is considered. For this, partial linear regression models are used, that is, regression models where the regression function is the sum of a linear and a nonparametric component. Two estimators for the nonparametric component are shown: one estimator takes into account the dependence structure in the errors of the regression function and the another estimator not. Both estimators are compared in practice for several real time series concerning economics and finance. In addition to the partial linear regression model, other regression models are included in the comparison.

Key-Words:

- *nonparametric regression; time series.*

AMS Subject Classification:

- 62G08, 62G20, 62M10, 91B84, 62P20.

1. INTRODUCTION

Linear regression modelling is a nice form for linking variables because in general the parameters have some kind of meaning or interpretation. Nevertheless, it is known that the main drawback of the linear regression models is their lack of flexibility. In practice, this fact causes that some interesting relationships can not be modelled by means of this class of models.

A way to avoid that drawback is to add to the linear regression function a nonparametric component. The resulting model, known as a partial linear regression (PLR) model, was introduced by Engle *et al.* (1986) to study the effect of weather on electricity demand. Another interesting feature of the PLR models is that they also avoid the "curse of dimensionality" (assuming low dimension for the explanatory variable that enters in a nonparametric way). From a theoretical point of view that dimension can be high, but usually is 1. Thus, we can say that the PLR models are flexible models that, in practice, can handle multiple variables.

Since the pioneer work of Engle *et al.* (1986), several papers have been published on this class of models in the setting of i.i.d. data (see, e.g., Speckman, 1988, Robinson, 1988, or Linton, 1995) as well as for dependent data (see, e.g., Gao, 1995, or Aneiros-Pérez *et al.*, 2004). In these papers, one can find asymptotic results (consistency, asymptotic normality) on estimators for each component in the PLR model, as well as on bandwidth selectors for those estimators, and even on testing of hypotheses. In addition, PLR models have demonstrated their usefulness in many field of applied sciences, such as economics, environmental studies, medicine... (see Härdle *et al.*, 2000, for a monograph and applications of the PLR model). A common feature in all these publications is that they study the same type of estimator, regardless of the data are independent or not.

In a recent paper, Aneiros-Pérez and Vilar-Fernández (2008) proposed a new estimator for the nonparametric component in a PLR model under random design and dependence conditions. To construct their estimator, these authors took into account the dependence structure in the errors of the model. Specifically, this dependence structure was used to transform the PLR model with dependent errors into a PLR model with uncorrelated errors (say, into a "whitened model"). Then, the estimator of the nonparametric component was based on this whitened model. Aneiros-Pérez and Vilar-Fernández (2008) obtained the asymptotic normality of both the estimator based on the original PLR model and the estimator based on the whitened model. As a consequence, they noted that the second estimator is asymptotically more efficient than the first estimator.

The aim of this paper is to illustrate, in practice, both the competitiveness of the PLR model and the usefulness of the prewhitening transformation.

Our paper is organized as follows. The PLR model is presented in Section 2, and estimators of both linear and nonparametric components of the model are motivated and defined. Then, a comparative study on the behavior of those estimators is deeply carried out in Section 3. For this, three real datasets in the context of economics and finance were analyzed. Concluding remarks are given in Section 4.

2. MOTIVATION AND CONSTRUCTION OF THE ESTIMATORS

2.1. The partial linear regression model

The class of the PLR models assumes that the regression function is the sum of a linear and a nonparametric component. This can be mathematically expressed through the regression model

$$Y_i = \mathbf{X}_i^{\mathrm{T}}\beta + m(\mathbf{T}_i) + \varepsilon_i \quad (i = 1, ..., n) \ , \tag{2.1}$$

where $\mathbf{X}_i = (X_{i1}, ..., X_{id_0})^{\mathrm{T}}$ and $\mathbf{T}_i = (T_{i1}, ..., T_{id_1})^{\mathrm{T}}$ ($d_0 \geq 1$, $d_1 \geq 1$) are vectors of explanatory variables, $\beta = (\beta_1, ..., \beta_{d_0})^{\mathrm{T}}$ is a vector of unknown real parameters, m is an unknown smooth real function and $\{\varepsilon_i\}$ are the random errors satisfying

$$E(\varepsilon_i \,|\, \mathbf{X}_i, \mathbf{T}_i) = 0 \quad (i = 1, ..., n) \ . \tag{2.2}$$

In this paper, we focus on estimation of m in (2.1) when both $\mathbf{X}_i$ and $\mathbf{T}_i$ are random (random design) and, in addition, the errors ε_i are dependent. Specifically, we assume that these errors follow the invertible linear process

$$\varepsilon_i = \sum_{j=0}^{\infty} c_j \, e_{i-j} \ , \quad \text{where} \ \ c_0 = 1 \ \text{and} \ e_i \ \text{are i.i.d. with} \ E(e_i) = 0 \ . \tag{2.3}$$

In general, estimators proposed in the statistical literature to estimate m in (2.1) do not have into account the dependence structure in the errors. However, it seems natural to think that incorporate such information in the construction of the estimator can be helpful. Aneiros-Pérez and Vilar-Fernández (2008) proposed to use that dependence structure in the following way. Let us denote $c(L) = \sum_{j=0}^{\infty} c_j L^j$, where L is the lag operator, and

$$a(L) = c(L)^{-1} = a_0 - \sum_{j=1}^{\infty} a_j L^j \qquad \text{with} \ \ a_0 = 1 \ . \tag{2.4}$$

Applying $a(L)$ to the PLR model (2.1) and rewriting the corresponding equation, we obtain the new PLR model

$$\underline{Y}_i = \mathbf{X}_i^{\mathrm{T}}\beta + m(\mathbf{T}_i) + e_i \quad (i = 1, ..., n) \ , \tag{2.5}$$

where $\underline{Y}_i = Y_i - \sum_{j=1}^{\infty} a_j\big(Y_{i-j} - \mathbf{X}_{i-j}^{\mathrm{T}}\,\beta - m(\mathbf{T}_{i-j})\big) = Y_i - \sum_{j=1}^{\infty} a_j\,\varepsilon_{i-j}$. As we see, the regression function in the PLR models (2.1) and (2.5) is the same, but in (2.5) the errors are i.i.d. Now, it should be noted that an estimator for m based on the PLR model (2.5) takes into account the dependence structure of the errors ε_i (through $\underline{Y}_i$). From now on, we will name "original PLR model" and "whitened PLR model" to the models (2.1) and (2.5), respectively.

2.2. The estimators

Aneiros-Pérez and Vilar-Fernández (2008) studied and compared (from an asymptotic point of view) two estimators for $m(\mathbf{t})$, one (say $\widehat{m}(\mathbf{t})$) based on the original PLR model (2.1) and the other (say $\underline{\widehat{m}}(\mathbf{t})$) based on the whitened PLR model (2.5). Specifically, these authors proved that, under suitable conditions, the asymptotic distribution of these estimators (properly normalized) is Gaussian. In summary, from that result one can observe that both estimators asymptotically have the same bias but different variances, the variance of $\widehat{m}(\mathbf{t})$ relative to the variance $\underline{\widehat{m}}(\mathbf{t})$ being $\sigma_\varepsilon^2/\sigma_e^2 = \sum_{j=0}^{\infty} c_j^2 \geq 1$ (the equality holding if and only if $\{\varepsilon_i\}$ is i.i.d.). Thus, we have that the estimator based on the whitened PLR model is asymptotically more efficient than the estimator based on the original PLR model.

Now, we motivate and construct both estimators $\widehat{m}(\mathbf{t})$ and $\underline{\widehat{m}}(\mathbf{t})$. We begin with $\widehat{m}(\mathbf{t})$. Let us assume that we have a root-n consistent estimator for β (say $\widehat{\beta}_{h_0}$). Then, from the original PLR model (2.1) we can write

$$Y_i - \mathbf{X}_i^{\mathrm{T}}\widehat{\beta}_{h_0} \approx m(\mathbf{T}_i) + \varepsilon_i \quad (i = 1, ..., n) \ ,$$

and, assuming that m is a smooth function, we proceed to estimate $m(\mathbf{t})$ by means of the nonparametric estimate

$$\widehat{m}_{h_0,h_1}(\mathbf{t}) = \sum_{j=1}^{n} w_{h_1,j}(\mathbf{t})\,\big(Y_j - \mathbf{X}_j^{\mathrm{T}}\widehat{\beta}_{h_0}\big)\,, \tag{2.6}$$

where $w_{h_1,j}(\mathbf{t})$ are weight functions depending on $\mathbf{t}$ and the design points $\mathbf{T}_i$, and both h_0 and h_1 are smoothing parameters or bandwidths that typically appear in any setting of nonparametric or semiparametric estimation. The weight functions considered in Aneiros-Pérez and Vilar-Fernández (2008) were local p-order polynomial type weights (for local polynomial estimation see, e.g., Fan and Gijbels, 1996, or Francisco-Fernández and Vilar-Fernández, 2001).

As we see, the estimator (2.6) is based on the original PLR model (2.1), but similar steps could be used to construct an estimator for $m(\mathbf{t})$ based on the whitened PLR model (2.5). Because in practice the response variable $\underline{Y}_i$ in (2.5) is unknown (depends on a_i, β and m), the first step in constructing such an estimator should be to propose a "reasonable" approximation for $\underline{Y}_i$. In this way, Aneiros-Pérez and Vilar-Fernández (2008) proposed to use the residuals $\widehat{\varepsilon}_i = Y_i - \mathbf{X}_i^{\mathrm{T}}\widehat{\beta}_{h_0} - \widehat{m}_{h_0,h_0}(\mathbf{T}_i)$ of the original PLR model to construct an estimate of $A_{\mathcal{T}} = (a_1, ..., a_{\mathcal{T}})^{\mathrm{T}}$, $\mathcal{T}$ being a truncation parameter large enough to avoid problems with the bias. Specifically, this estimator for $A_{\mathcal{T}}$ is constructed by means of the ordinary least squares (OLS) method applied to the model

$$\widehat{\varepsilon}_i \;=\; a_1\widehat{\varepsilon}_{i-1} + \cdots + a_{\mathcal{T}}\,\widehat{\varepsilon}_{i-\mathcal{T}} + \mathit{residual}_i \quad (i = \mathcal{T}+1, ..., n)\ . \tag{2.7}$$

In this way, the estimator

$$\widehat{A}_{\mathcal{T}} = \left(\widehat{\varepsilon}_{\mathcal{T}}^{\mathrm{T}}\,\widehat{\varepsilon}_{\mathcal{T}}\right)^{-1}\widehat{\varepsilon}_{\mathcal{T}}^{\mathrm{T}}\,\widehat{\varepsilon} \tag{2.8}$$

is obtained, where $\widehat{\varepsilon} = (\widehat{\varepsilon}_{\mathcal{T}+1}, ..., \widehat{\varepsilon}_n)^{\mathrm{T}}$ and $\widehat{\varepsilon}_{\mathcal{T}} = (\widehat{\varepsilon}_{i,j})_{\substack{1\le i\le n-\mathcal{T}\\ 1\le j\le \mathcal{T}}}$ with $\widehat{\varepsilon}_{i,j} = \widehat{\varepsilon}_{i-j+\mathcal{T}}$. Now, using $\widehat{A}_{\mathcal{T}}$ together with $\widehat{\beta}_{h_0}$ and $\widehat{m}_{h_0,h_0}$, we define

$$\underline{\widehat{Y}}_{\mathcal{T},i} \;=\; Y_i - \sum_{j=1}^{\mathcal{T}} \widehat{a}_j\Big(Y_{i-j} - \mathbf{X}_{i-j}^{\mathrm{T}}\,\widehat{\beta}_{h_0} - \widehat{m}_{h_0,h_0}(\mathbf{T}_{i-j})\Big) \quad (i = \mathcal{T}+1, ..., n)\ . \tag{2.9}$$

Finally, from (2.5) and (2.9), we can write

$$\underline{\widehat{Y}}_{\mathcal{T},i} - \mathbf{X}_i^{\mathrm{T}}\widehat{\beta}_{h_0} \approx\; m(\mathbf{T}_i) + \varepsilon_i \quad (i = \mathcal{T}+1, ..., n)\ ,$$

and, as in (2.6), we construct the estimator

$$\underline{\widehat{m}}_{\mathcal{T},h_0,h_1}(\mathbf{t}) \;=\; \sum_{i=\mathcal{T}+1}^{n} w_{h_1,i}(\mathbf{t})\left(\underline{\widehat{Y}}_{\mathcal{T},i} - \mathbf{X}_i^{T}\widehat{\beta}_{h_0}\right). \tag{2.10}$$

In summary, the steps taken to construct the estimator (2.10) are:

Step 1: Construct a root-n consistent estimator $\widehat{\beta}_{h_0}$ for β.

Step 2: Construct the residuals $\widehat{\varepsilon}_i$.

Step 3: Use $\widehat{\varepsilon}_i$ $(i = \mathcal{T}+1, ..., n)$ to construct an estimator $\widehat{a}_j$ for a_j $(j = 1, ..., \mathcal{T})$.

Step 4: Use $\widehat{a}_j$ $(j = 1, ..., \mathcal{T})$, $\widehat{\beta}_{h_0}$ and $\widehat{m}_{h_0,h_0}$ to construct an approximation $\underline{\widehat{Y}}_{\mathcal{T},i}$ for $\underline{Y}_i$ $(i = \mathcal{T}+1, ..., n)$.

Step 5: Use $\widehat{\beta}_{h_0}$ and $\underline{\widehat{Y}}_{\mathcal{T},i}$ to construct the estimator (2.10).

Finally, we motivate the root-n consistent estimator $\widehat{\beta}_{h_0}$ for β used in Aneiros-Pérez and Vilar-Fernández (2008). If we subtract $E(Y_i \,|\, \mathbf{T}_i)$ on both sides of equality (2.1), we get the linear regression model

$$Y_i - E(Y_i \,|\, \mathbf{T}_i) \,=\, \big(\mathbf{X}_i - E(\mathbf{X}_i \,|\, \mathbf{T}_i)\big)^{\mathrm{T}} \beta + \varepsilon_i \quad (i = 1, ..., n) \,. \tag{2.11}$$

Then, replacing the response and explanatory variables in (2.11) (that is, $Y_i - E(Y_i \,|\, \mathbf{T}_i)$ and $\mathbf{X}_i - E(\mathbf{X}_i \,|\, \mathbf{T}_i)$, respectively) by the corresponding residuals obtained when Y and $\mathbf{X}$ are nonparametrically adjusted for $\mathbf{T}$, we can write (in matricial form)

$$\widetilde{\mathbf{Y}}_{h_0} \approx \, \widetilde{\mathbf{X}}_{h_0}^{T} \, \beta + \varepsilon_i \quad (i = 1, ..., n) \,, \tag{2.12}$$

where, for both the n-dimensional vector $\mathbf{A} = \mathbf{Y}$ and the $(n \times d_0)$-matrix $\mathbf{A} = \mathbf{X}$, and for any real number $h_0 > 0$, we have denoted $\widetilde{\mathbf{A}}_{h_0} = (\mathbf{I} - \mathbf{W}_{h_0})\, \mathbf{A}$ with $\mathbf{W}_{h_0} = \big(w_{h_0,j}(\mathbf{T}_i)\big)_{i,j}$. Now, using OLS in (2.12), one obtains

$$\widehat{\beta}_{h_0} = \big(\widetilde{\mathbf{X}}_{h_0}^{\mathrm{T}} \, \widetilde{\mathbf{X}}_{h_0}\big)^{-1} \, \widetilde{\mathbf{X}}_{h_0}^{\mathrm{T}} \widetilde{\mathbf{Y}}_{h_0} \,.$$

At this time, four facts should be clear. First, we dispose of two estimators for m in (2.1): $\widehat{m}_{h_0,h_1}(\mathbf{t})$ and $\underline{\widehat{m}}_{\mathcal{T},h_0,h_1}(\mathbf{t})$. Second, $\widehat{m}_{h_0,h_1}(\mathbf{t})$ does not take into account the dependence structure in the errors of the PLR model (2.1). Third, $\underline{\widehat{m}}_{\mathcal{T},h_0,h_1}(\mathbf{t})$ takes into account the dependence structure in the errors of the PLR model (2.1). Fourth, Aneiros-Pérez and Vilar-Fernández (2008) proved that $\underline{\widehat{m}}_{\mathcal{T},h_0,h_1}(\mathbf{t})$ is asymptotically more efficient than $\widehat{m}_{h_0,h_1}(\mathbf{t})$.

3. APPLICATIONS TO REAL DATA

The main goal of this section is to compare the behavior of the estimators $\widehat{m}_{h_0,h_1}(\mathbf{t})$ and $\underline{\widehat{m}}_{\mathcal{T},h_0,h_1}(\mathbf{t})$ when they are applied to real data. In addition, in order to make more general the study and not only confined to the PLR model, in a first attempt we will consider a set of regression models together with their conventional estimators. Then, the accuracy of these models/estimators will be compared with that of the PLR model/estimators $\widehat{m}_{h_0,h_1}(\mathbf{t})$ and $\underline{\widehat{m}}_{\mathcal{T},h_0,h_1}(\mathbf{t})$. In a second attempt, we will take into account the fact that the prewhitening transformation (2.4) can also be applied to that set of regression models. Thus, we will include in the study the estimators based on the corresponding whitened regression models. Then, when the study is completed, we will have shown both the competitiveness of the PLR model and the usefulness of the prewhitening transformation.

Three real datasets will be analyzed, all related to the field of economics and finance. Specifically, the first example deals with market shares and prices of two dentifrices, while the second dataset is related to the building industry. Finally, in the third study we consider relationships between stock indices.

3.1. Models

In the three datasets we will study, we have a response variable (say Y) and two explanatory variables (say X and T). Thus, we will consider four classes of regression models linking Y with X and/or T. Now, we quickly present each one of these classes and give a short motivation of them:

Linear models. Maybe, the first class that comes to mind is that of the classical linear regression models. These models allow easy interpretation of the effect of each explanatory variable on the response variable. Nevertheless, it is known than its main handicap is the lack of flexibility.

Nonparametric models. In order to avoid the handicap named in the previous kind of models, the second class to be considered is that of nonparametric models. A problem of this class is the known as "curse of dimensionality", which is based on the fact that, when the number d of real explanatory variables is greater than 1, large sample sizes are required to obtain good estimates (these sample sizes increase exponentially with d). In view of this problem, we will restrict to the class of nonparametric models with only a real explanatory variable.

Partial linear models. This class was presented in a general setting in Section 2. In this practical study, we will consider PLR models that include one real variable in a linear form and another one in a nonparametric form (that is, $d_0 = d_1 = 1$ in (2.1)). Note that this class overcomes the curse of dimensionality.

Additive models. The last class is composed by the additive models with two explanatory variables, that is, nonparametric models whose regression function is the sum of two nonparametric components. It is interesting to observe that, as the previous class, this class avoids the curse of dimensionality (in fact, this is achieved in both classes by means of the same method: to express the regression function as the sum of two components).

The wide range of models named above can be seen in Table 1.

Table 1: Regression models.

Models	**Notation**
Linear models	
$Y = \alpha + X\beta + \varepsilon$	L1
$Y = \alpha + T\beta + \varepsilon$	L2
$Y = \alpha + X\beta_1 + T\beta_2 + \varepsilon$	L3
Nonparametric models	
$Y = m(X) + \varepsilon$	NP1
$Y = m(T) + \varepsilon$	NP2
Partial linear models	
$Y = X\beta + m(T) + \varepsilon$	PL1
$Y = T\beta + m(X) + \varepsilon$	PL2
Additive model	
$Y = \mu + m_1(X) + m_2(T) + \varepsilon$	ADD

3.2. Estimators

Now we indicate, for each regression model in Table 1, the kind of estimator considered. OLS estimators were used to estimate the parameters corresponding to the linear models, while the local linear polynomial estimator was considered to estimate the regression function in the nonparametric models. Both the parametric and the nonparametric component in the PLR models were estimated by means of the estimators presented in Section 2. Finally, for the additive model, a backfitting algorithm was considered (see Hastie and Tibshirani, 1990). In the last two classes of models, the weight functions were local linear polynomial type weights (as for the nonparametric case).

3.3. Choosing the parameters of the estimates

In practice, as usual in both nonparametric and semiparametric settings, it is necessary to choose some parameters related to the estimates. Specifically, we refer to the kernel and the bandwidths. In addition, for the cases where the prewhitening transformation is considered, we must give a value for the truncation parameter $\mathcal{T}$.

On the one hand, the Epanechnikov Kernel $K(u) = 0.75(1-u)^2 I_{[-1,1]}(u)$ and the truncation parameter $\mathcal{T} = 2$ were used in the estimates. On the other hand, the bandwidths were chosen by means of the cross-validation procedure.

In short, this bandwidth selector proposes to choose the value $\widehat{h}_{CV}$ that minimizes to the function

$$CV(h) \,=\, n^{-1} \sum_{i=1}^{n} \left(Y_i - \widehat{r}_h^i(X_i, T_i)\right)^2 \omega(T_i) \;,$$

where $\widehat{r}_h^i(\cdot,\cdot)$ denotes the estimator of the regression function (of each model considered) constructed without using the i-th observation, and $\omega(\cdot)$ is a weight function included to avoid boundary effects. Note that for both the PLR and the additive models h is a vector (say $h = (h_0, h_1)^{\mathrm{T}}$ and $h = (h_1, h_2)^{\mathrm{T}}$, respectively). The good (asymptotic) properties of the cross-validation selector are based on the fact that $CV(h)$ is (asymptotically) equivalent (except for a constant) to the average squared error (see Aneiros-Pérez and Quintela-del-Río 2001 for some results on this selector in the setting of PLR models).

3.4. Measure of accuracy

To compare the accuracy of the different models/estimators, we considered the Relative Cross-Validation (RCV)

$$RCV = \frac{CV\left(\widehat{h}_{CV}\right)}{\mathrm{Var}\{Y_i\}} \;,$$

where $\mathrm{Var}\{Y_i\}$ denotes the variance corresponding to the sample of responses. Observe that RCV gives a global measure of the accuracy of each model/estimator.

3.5. The dentifrice data

The first dataset analyzed consists of weekly market shares of Crest and Colgate dentifrice, together with the price of Crest dentifrice, during the period January 1, 1958 to April 30, 1963 (276 data). This dataset was used in Wichern and Jones (1977) to asses the impact of market disturbances, and can be found on the website `http://www.alianzaeditorial.es` (Book title: Análisis de Series Temporales; Author: Peña, D., Section: Ejercicios prácticos). The graphics of these time series are shown in Figure 1.

From Figure 1, we clearly observe the presence of trend in the three series. These trends were eliminated by differentiating. Then, using the transformed data, we seek in Table 1 an adequate regression model to link the data corresponding to the market share of Crest dentifrice (Y) with those of market share of Colgate dentifrice (X) and price of Crest dentifrice (T).

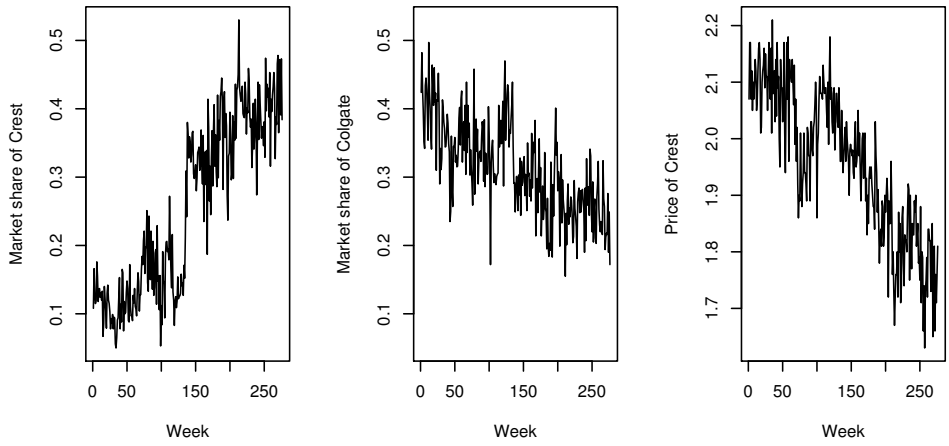

Figure 1: The market shares of Crest (left) and Colgate (middle) dentifrice, and the price of Crest dentifrice (right).

The RCV's of the models/estimates considered, as well as comparisons between them, are displayed in Table 2. From the third column in this table, we can say that the prewhitening transformation has proved useful in all the models. In addition, the fourth column indicates that the best model is the PLR model that includes the effect of market share of Colgate dentifrice and price of Crest dentifrice in a linear and a nonparametric way, respectively (that is, the PLR model PL1).

Table 2: Values of the criterion error and ratios (^{a}original model, ^{b}whitened model).

Model	**RCV**	$\mathbf{RCV}^b/\mathbf{RCV}^a$	$\mathbf{RCV}/\min \mathbf{RCV}$
L1	^{a}0.8347 ^{b}0.8313	0.9959	1.0370 1.0328
L2	0.9793 0.9748	0.9953	1.2167 1.2110
L3	0.8161 0.8108	0.9935	1.0139 1.0073
NP1	0.8359 0.8308	0.9939	1.0385 1.0321
NP2	0.9802 0.9751	0.9948	1.2177 1.2114
PL1	0.8090 0.8049	0.9950	1.0050 1.0000
PL2	0.8118 0.8077	0.9949	1.0086 1.0034
ADD	0.8219 0.8147	0.9912	1.0211 1.0121

Finally, we give some information on the estimates of the parameter β and the function m in the best model (PL1). The estimates of β using the conventional and the prewhitened-based estimators were -0.4146 and -0.4220, respectively. The corresponding estimates of m are shown in Figure 2 (from now on, in the graphics corresponding to the estimates of m, "Estimator 1" and "Estimator 2" are referred to as $\widehat{m}_{h_0,h_1}$ and $\underline{\widehat{m}}_{\mathcal{T},h_0,h_1}$, respectively).

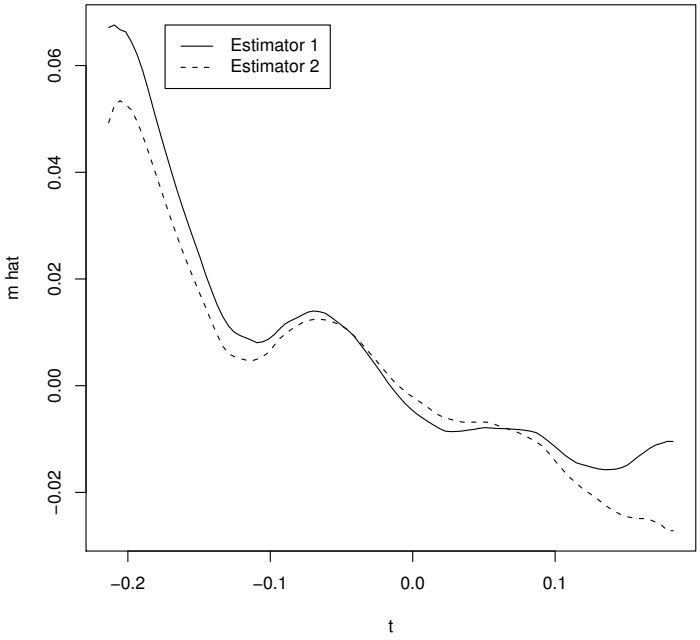

Figure 2: Estimates of the nonparametric component m in the PLR model PL1.

3.6. The building industry data

The building industry data is the second example we analyze. We have monthly observations corresponding to the number of buildings started, quantity of cement produced and number of buildings completed in Galicia (an autonomous community located in northwestern Spain) during the period January 1987 to December 2000 (168 data). These time series are available on the website `http://www.ige.eu`. Figure 3 displays these data.

Our goal is to get a model to analyze the effect of both the quantity of cement produced (X) and the number of buildings completed (T) on the number of buildings started (Y). Because, as in the previous example, our time series contain trend (see Figure 3) and therefore they are not stationary, we have worked with the differenced data.

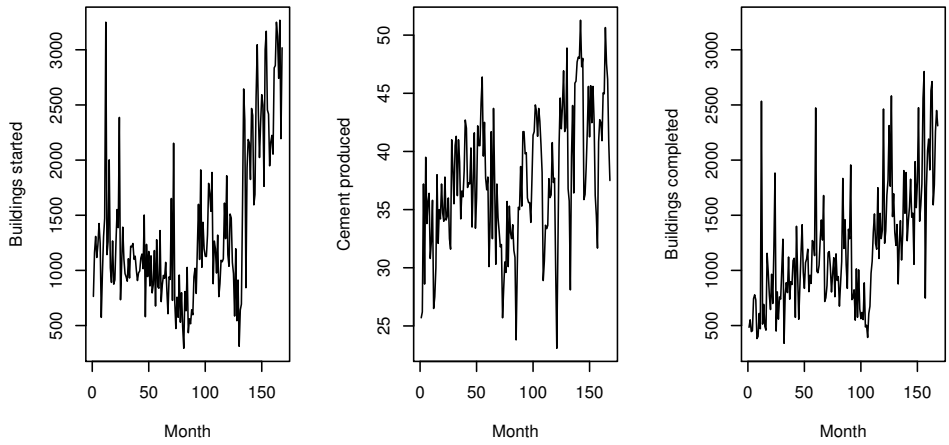

Figure 3: Number of buildings started (left), quantity of cement produced (middle) and number of buildings completed (right) in Galicia.

Table 3 shows interesting information on the accuracy of the models considered, as well as on the behavior of the different estimators. From this table, we note that the prewhitening transformation does not always improve the original model (see column 3), but there are improvements for the best two models (PL1 and ADD). Finally, the prewhitened-based estimator applied on the PLR model PL1 gives the best accuracy (see column 4).

Table 3: Values of the criterion error and ratios (^{a}original model, ^{b}whitened model).

Model	**RCV**	$\mathbf{RCV}^b/\mathbf{RCV}^a$	$\mathbf{RCV}/\min\mathbf{RCV}$
L1	^{a}0.9613 ^{b}0.9616	1.0003	1.0704 1.0707
L2	0.9872 0.9879	1.0007	1.0992 1.0999
L3	0.9488 0.9458	0.9969	1.0564 1.0531
NP1	0.9409 0.9472	1.0067	1.0476 1.0547
NP2	0.9610 0.9492	0.9877	1.0700 1.0569
PL1	0.9071 0.8981	0.9901	1.0100 1.0000
PL2	0.9170 0.9176	1.0006	1.0210 1.0216
ADD	0.9055 0.9001	0.9940	1.0082 1.0022

Focusing on the PLR model PL1, we have that the estimates of β using the conventional and the prewhitened-based estimators were 21.91 and 21.16, respectively. The corresponding estimates of m are shown in Figure 4.

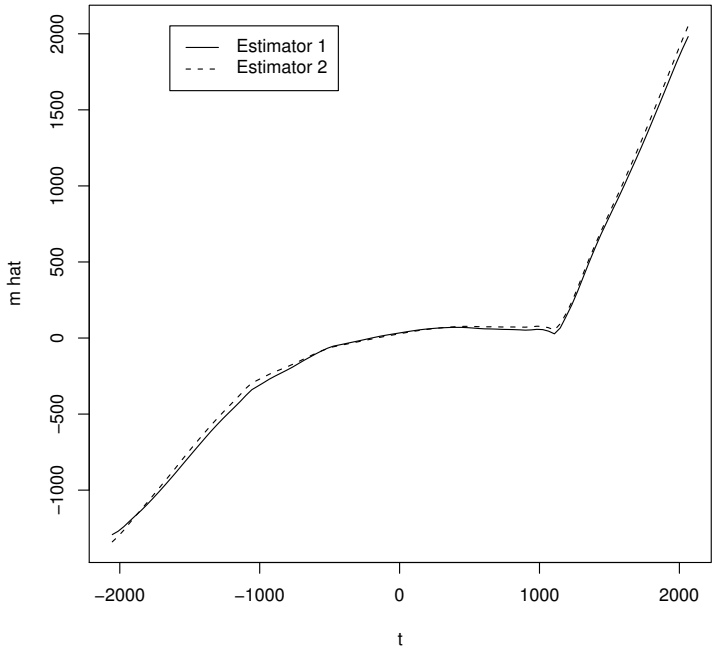

Figure 4: Estimates of the nonparametric component m in the PLR model PL1.

3.7. The stock data

Finally, we present an analysis on stock data. Specifically, our time series collect Banca Commerciale Index (Milan), FT-SE 100 Index (London) and General Index (Madrid) for each month during the period January 1988 to December 2000 (156 data). These data, which can be obtained on the website `http://www.ec.europa.eu/eurostat`, are shown in Figure 5.

From a first analysis of the data, we found the presence of both heteroscedasticity and trend. Thus, the data have been transformed using logarithms and then differentiated to achieve stationarity. Now, using the transformed data, we are interested in the construction of an adequate regression model to link the Banca Commerciale Index (Y) with the FT-SE 100 Index (X) and the General Index (T).

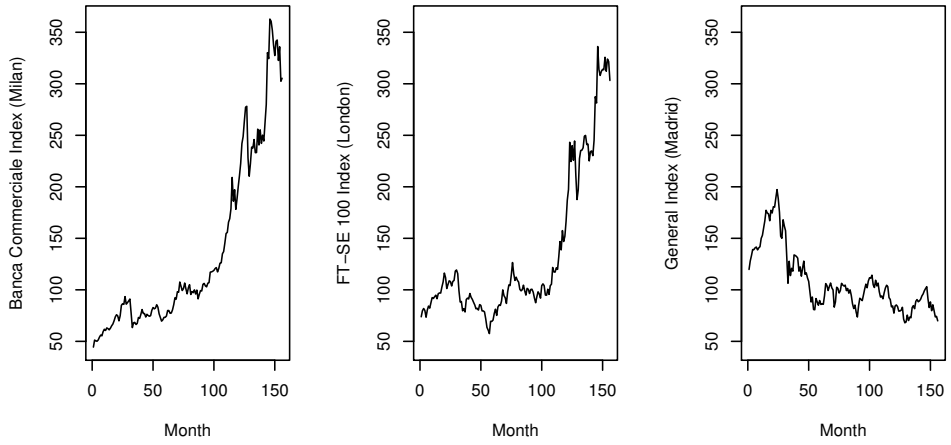

Figure 5: Banca Commerciale Index (left), FT-SE 100 Index (middle) and General Index (right).

The RCV's obtained for the different models considered, as well as comparisons between them, are given in Table 4. Some conclusions can be obtained from this table. First, only in the models L1 and NP2 the prewhitening transformation does not improve the original model (see column 3). Second, the best model is the PLR model that includes the effect of FT-SE 100 Index and General Index in a linear and a nonparametric way, respectively (that is, the PLR model PL1). Third, the prewhitened-based estimator applied on this PLR model gives the best accuracy.

Table 4: Values of the criterion error and ratios (^{a}original model, ^{b}whitened model).

Model	**RCV**	$\mathbf{RCV}^b/\mathbf{RCV}^a$	$\mathbf{RCV}/\min\mathbf{RCV}$
L1	^{a}0.7027 ^{b}0.7041	1.0019	1.0226 1.0246
L2	0.9859 0.9843	0.9984	1.4347 1.4324
L3	0.7129 0.7119	0.9986	1.0374 1.0360
NP1	0.7058 0.7044	0.9980	1.0271 1.0251
NP2	0.9768 0.9797	1.0029	1.4215 1.4257
PL1	0.6911 0.6872	0.9944	1.0057 1.0000
PL2	0.7004 0.6994	0.9986	1.0192 1.0178
ADD	0.7040 0.6978	0.9912	1.0245 1.0155

We complete the analysis showing the estimates of the parameter β and the function m in the best model (PL1). The estimates of β using the conventional and the prewhitened-based estimators were 0.4713 and 0.4666, respectively. The corresponding estimates of m are displayed in Figure 6.

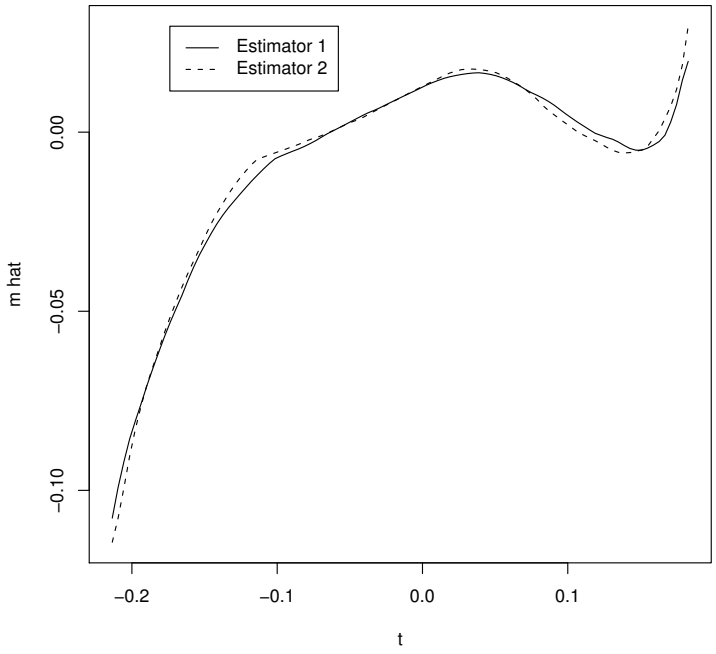

Figure 6: Estimates of the nonparametric component m in the PLR model PL1.

4. CONCLUDING REMARKS

In this paper, partial linear regression modelling in time series was dealt from a practical point of view. For this, we divided the paper into two parts. In the first part, some theory was shown. Specifically, we motivated and presented the PLR model. Then, we carefully constructed the estimator proposed in Aneiros-Pérez and Vilar-Fernández (2008), which is based on a whitened version of the original PLR model. By this motive, the estimator takes into account the dependence structure in the random errors (this fact is crucial for its good asymptotic behavior). The second part contains the main contribution of our work. It analyzes several real time series concerning economics and finance. Specifically, these time series were modelled by means of a wide range of regression models, including PLR models. Then, the corresponding regression functions were estimated. For this, both conventional and whitened-model based estimators were used. Finally, the performance of the corresponding estimators was measured.

In all the time series studied, the PLR model (estimated using the estimator proposed in Aneiros-Pérez and Vilar-Fernández, 2008) gave the best results.

We are aware that the improvement on the point-estimates is small. In fact, from the theoretical results, it is expected that a greater improvement is obtained on comparing confidence intervals (as noted in Subsection 2.2, the asymptotic difference between the conventional and the whitened-model based estimators is in their variances). Nevertheless, it should be noted that to construct confidence intervals one needs to estimate those variances, and the variability of the corresponding estimators could mask the theoretical result. For this reason, we have preferred to compare the point-estimates.

ACKNOWLEDGMENTS

This work has been supported by the grants numbers PGIDIT07PXIB105259PR and 07SIN012105PR from Xunta de Galicia (Spain), and by the grant number MTM2008-00166 from Ministerio de Ciencia e Innovación (Spain). We also acknowledge the valuable suggestions from the editor and the referees.

REFERENCES

[1] Aneiros-Pérez, G.; González-Manteiga, W. and Vieu, P. (2004). Estimation and testing in a partial linear regression model under long-memory dependence, *Bernoulli*, **10**, 49–78.

[2] Aneiros-Pérez, G. and Quintela-del-Río, A. (2001). Modified cross-validation in semiparametric regression models with dependent errors, *Communications in Statistics: Theory and Methods*, **30**, 289–307.

[3] Aneiros-Pérez, G. and Vilar-Fernández, J.M. (2008). Local polynomial estimation in partial linear regression models under dependence, *Computational Statistics and Data Analysis*, **52**, 2757–2777.

[4] Engle, R.; Granger, C.; Rice, J. and Weiss, A. (1986). Nonparametric estimates of the relation between weather and electricity sales, *Journal of the American Statistical Association*, **81**, 310–320.

[5] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*, Chapman and Hall.

[6] Francisco-Fernández, M. and Vilar-Fernández, J.M. (2001). Local polynomial regression estimation with correlated errors, *Communications in Statistics: Theory and Methods*, **30**, 1271–1293.

[7] Gao, J.T. (1995). Asymptotic theory for partly linear models, *Communications in Statistics: Theory and Methods*, **24**, 1985–2009.

[8] Härdle, W.; Liang, H. and Gao, J.T. (2000). *Partially Linear Models*, Physica Verlag.

[9] Hastie, T.J. and Tibshirani, R.J. (1990). *Generalized Additive Models*, Chapman and Hall, New York.

[10] Linton, O. (1995). Second order approximation in the partially linear regression model, *Econometrica*, **63**, 1079–1112.

[11] Robinson, P. (1988). Root-n-consistent semiparametric regression, *Econometrica*, **56**, 931–954.

[12] Speckman, P. (1988). Kernel smoothing in partial linear models, *Journal of the Royal Statistical Society: Series B*, **50**, 413–436.

[13] Wichern, D.W. and Jones, R.H. (1977). Assessing the impact of market disturbances using intervention analysis, *Management Science*, **24**, 329–337.

REVSTAT – Statistical Journal
Volume 7, Number 1, April 2009, 55–66

AN INSURANCE TYPE MODEL FOR THE HEALTH COST OF COLD HOUSING: AN APPLICATION OF GAMLSS

Authors: Robert Gilchrist
– STORM, London Metropolitan University, Holloway Road, London, N7 8DB, U.K.
r.gilchrist@londonmet.ac.uk

Alim Kamara
– STORM, London Metropolitan University, Holloway Road, London, N7 8DB, U.K.
alimskay@yahoo.co.uk

Janet Rudge
– HRSJ, London Metropolitan University, Holloway Road, London, N7 8DB, U.K.
janetrudge111@googlemail.com

Abstract:

- This paper introduces a substantive problem, namely the link between fuel poverty and excess winter morbidity amongst older people, and shows how the GAMLSS suite of programs (www.gamlss.com) can be used to provide a very flexible method of modelling both the number of hospital admissions and the corresponding lengths of stay in hospital. The approach is closely related to the models that have been used to model the number of insurance claims, and their cost (see Heller *et al.* (2007)). We fit the Beta Binomial distribution and a variety of continuous distributions.

Key-Words:

- *Beta Binomial; GAMLSS; insurance; morbidity; costs.*

AMS Subject Classification:

- 62J12, 62P12.

1. INTRODUCTION

Fuel poverty is defined as "the inability to afford adequate warmth in the home" and is related to poor energy efficiency of homes as well as householders' incomes. In the U.K. a household is defined as suffering from fuel poverty if more than 10% of their income is spent on fuel. In 2008, 4.5 millions people were defined as "fuel poor". It is projected that 5.8 million will be "fuel poor" in 2009. Older households are the group most vulnerable to fuel poverty, and are also particularly susceptible to cold-related health effects. The significant numbers recognised as fuel poor have as yet unrecognised implications for costs to public services.

Conventionally, research has referred to effects of cold homes in terms of excess winter deaths (e.g. Wilkinson *et al.* (2001)). These deaths are known to be associated with outdoor winter temperatures, but direct evidence of links to low indoor temperatures is limited. Mortality statistics disguise the full extent of potentially long-term chronic conditions exacerbated by cold. Hence we have concentrated on measuring excess winter morbidity (illness) in relation to fuel poverty, rather than mortality, because of the consequent implications for winter pressures on health services.

We have previously demonstrated links between fuel poverty risk and excess winter hospital episodes among older people in Newham, using this excess as a measure of associated morbidity (e.g. Rudge and Gilchrist (2007)). In this paper we refer to that work and describe the means by which this measure could be developed as a costing element for a health impact assessment tool. The results could contribute to the debate regarding the case for increased energy efficiency investment on public health grounds, in addition to the accepted environmental grounds. Our methodology is closely related to the models that have been used for insurance claims. The two stage aspect of our modelling, firstly estimating the probability of being an emergency respiratory admission, followed by estimating the probability of dying after admission, gives a model for the effect of fuel poverty on this form of mortality.

2. SUBSTANTIVE BACKGROUND TO FUEL POVERTY AND POOR HEALTH

The U.K. Department of Health (2001) recognises that fuel poverty affects health inequalities, particularly among older people. The potential benefits of energy efficiency investment for older fuel poor households involve improvements in comfort, health and well being. Identifying cost savings associated with such benefits is complicated by the many confounding factors involved in showing direct causal links between housing characteristics and health.

There are no current precise methods of calculating the cost to the health services of cold-related disease arising from poor housing. The newly prevailing emphasis on dealing with climate change and carbon emissions may deflect attention from the needs of the fuel poor, who cannot afford to use energy extravagantly. Energy-saving targets tend to skew energy-efficiency investment in favour of fuel-rich households. However, public health implications demand that such investment should also be health-driven.

3. DATA AND STATISTICAL METHODOLOGY

The main source of data here considered is our existing database for Newham hospital admissions over 1993-96. These data are anonymised with respect to individuals, having been provided at enumeration district (ED) level.

Our previous work examined the excess morbidity for different ages and genders in terms of a range of explanatory variables. We now propose extending this work by analysing daily episodes by length of stay and investigating the associated costs for such episodes. Our proposed methodology is based upon the modelling of the propensity for an individual to be an emergency respiratory hospital admission, together with the duration of stay in hospital for such admissions. This approach is similar to that used for insurance claims (see e.g. Heller, *et al.* (2007)) in which the probability of a claim and the size of a claim are both modelled. Having modelled the probability of being a hospital respiratory admission and the length of the consequential stays in hospital, we could use data on the average cost of such hospital admissions, adjusted for the duration of stay, to give a model for the cost of the Newham admissions. The effect of FPR will be determined by considering the excess cost in winter over that in summer.

Our methodology utilises the R-based GAMLSS package (see Rigby and Stasinopoulos (2005) and `www.gamlss.com`). GAMLSS is a suite of programs written in R (see `www.r-project.org`). We consider the probability of being admitted as following a Beta Binomial distribution, this being a more flexible extension of the more traditional Binomial distribution. Our "default" approach is to use logistic regression. We also noted that some of the elderly people die after admission to hospital. We model this probability of death also using the Beta Binomial. The corresponding length of episode is modelled from a selection of continuous distributions.

The GAMLSS package allow us easily to find the the maximum likelihood estimates of the several parameters of a wide range of distributions and to incorporate random effects and smoothing terms. We can make use of the many facilities of R, such as automatic model selection, and we can easily access the wide range of diagnostics available in R. Up to 4-distributional parameters can be modelled in terms of the risk factors. The potential risk factors are shown in the

accompanying Table 1. We utilise nominal factors ED, gender and age to allow differing parameters to be fitted for the differing numbers of "at risk" males and females, of differing ages, in each enumeration district. The definition of the fuel poverty index FPR is discussed further below. Potential confounding factors are considered, using 1991 Census data, including pensioners with limiting long term illness and ethnic composition. Daily weather data were obtained for 1993–1997.

Table 1: Explanatory variables and factors. ** SAP35 is energy rating, or measure of energy efficiency, on a scale of 0–100, where 0 is poorest. # denotes component of FPR.

Variable	Description
hh1#	% households with one or more pensioner(s)
hh2	% small households (one or two persons households)
undoc#	% households under-occupied (1 person with >4 rooms; 2 persons with >5 rooms)
lowsap#	% dwellings with poor energy efficiency (below SAP35**)
ctb#	% households in receipt of Council Tax Benefit (indicator of low income)
tow	Townsend deprivation score
ch	% households with no central heating
pens	% lone pensioner households with no central heating
pre	% dwellings built before 1945
pensm	Total male pensioners as % of total population
pensf	Total female pensioners as % of total population
penswh	% of white pensioners in the ED
FPR	Fuel Poverty Risk Index = $(\text{hh1} * \text{undoc} * \text{lowsap} * \text{ctb}) * 10^{-3}$
pwh	White pensioners (% total pensioners)
mmeant	Monthly mean air temperature, °C
mmaxt	Monthly maximum air temperature, °C
mmint	Monthly minimum air temperature, °C
mrain	Monthly rainfall totals, mm
msun	Monthly sunshine hours
mmwd	Monthly mean wind speed
msol	Monthly solar radiation, W hr m-2
mtdif	Difference from previous month mean temperatures, °C
dwigs	Total number of dwellings
house	Total number of households
pop	% population 65 years old or more
age	(1) 65–74, (2) 75–84, (3) 85+
nage	Age with 2 levels only: (1) 65–84, (2) 85+
sex	(1) Male, (2) Female
q	Season factor with 3 levels: (1) "Summer", (2) November, January, February, (3) December
nq	Season factor with 2 levels: (1) Not December, (2) December
z	Factor with 48 levels denoting month
E	Factor with 450 levels specifying enumeration district (ED)

The lagged influence of weather is considered, together with maximum, minimum and mean monthly temperature and average monthly rainfall, wind speed, hours of sunshine, and solar radiation levels.

The chance of a repeat admission of an individual appears to be low, although this is not easy to determine precisely as the original data predated inclusion of patient identifier codes. Hence, although an assumption of independence of the observed admissions and of the observed lengths of episode is not too unreasonable, some correlation between occurrences might be expected, and perhaps some correlation between lengths of episode. GAMLSS makes it possible to incorporate a possible random effect in our linear predictor to allow for over-dispersion caused by the unknown correlation. Moreover, the Beta Binomial distribution is itself a form of over dispersed Binomial distribution; i.e. a Binomial distribution with a Beta distributed random effect.

4. DEFINITION OF THE FUEL POVERTY RISK INDEX

Our population-based study of the London Borough of Newham involved creating a Fuel Poverty Risk Index (FPR), derived from known risk factors, to compare with a cold-related health indicator, based on excess winter emergency respiratory hospital admissions (see Rudge and Gilchrist (2005)). Our data level was limited to small areas, rather than individuals, for patient anonymity reasons.

Datasets were collated for enumeration districts (EDs), which contain, on average, about 220 households: we collected data on household age, size and tenure from the 1991 Census; Council Tax Benefit (CTB) for 1998; estimated energy ratings for dwellings, based on classification by tenure (census data), size and type (from a drive round survey and census) and building age; numbers of emergency episodes for all respiratory diagnoses for patients aged above 64 years for August '93 to July '97 from Hospital Episode Statistics (HES). (Emergency admissions are more likely to reflect seasonal effects than elective admissions.)

The FPR was calculated for EDs as a product of the following (unweighted) factors, all as percentages of total households or total dwellings:

- households with one or more pensioners;
- households in receipt of CTB (indicating low income);
- dwellings with poor energy efficiency (i.e. below the 1991 national average energy rating);
- under occupancy (small households occupying relatively large homes).

5. A STATISTICAL MODEL FOR THE EXPECTED TOTAL DURATION OF EMERGENCY RESPIRATORY HOSPITAL ADMISSIONS

We here develop a model to explain the observed illness counts in each ED, in each month, in terms of the potential explanatory variables, and notably FPR. We model the counts for males and females, and for the three age categories. We consider data for each of 48 months. Our particular interest is in the difference between the counts observed in summer and winter, and whether we can explain this difference in terms of the explanatory variables. To examine this we model the probability p_{ijkl} of an individual of gender i, in age group j, in ED k, being ill in month l, $i=1,2,;\ j=1,..,3;\ k=1,..,450,\ l=1,..,48$.

Our count data consists of the number of people who are ill in a given month, as a proportion of the total number at risk. Perhaps the most natural model for such data is the Binomial distribution, with the observed counts restricted by a "Binomial Denominator". We here use a logistic Beta-Binomial assumption which can allow for potential "over-dispersion" in our counts.

Thus we assume we have observed numbers Y_{ijkl} of emergency respiratory admissions of gender i, age j, in ED k, in month l, $i=1,2;\ j=1,2,3;\ k=1,..,450;\ l=1,..,48$. The number of people at risk in each "cell" is n_{ijkl}. We assume that Y_{ijkl} follows a Beta Binomial distribution, $\mathbf{BB}(n_{ijkl}, p_{ijkl}, \sigma_{ijkl})$. Our basic assumption is that we have a logit link, i.e. $p_{ijlk} = 1/(1+\exp(-\eta_{ijkl}))$, where η_{ijkl} is a linear predictor based upon the explanatory variables in Table 1.

5.1. Distributional assumption

In defining the probability function, we drop the suffices i, j, k, l for clarity of exposition. The probability function of a random variable, Y which follows the Beta Binomial distribution denoted here as $\mathbf{BB}(n, p, \sigma)$, is given by

$$p_Y(y|p,\sigma) \;=\; \frac{\Gamma(n+1)}{\Gamma(y+1)\,\Gamma(n-y+1)}\;\frac{\Gamma\left(\frac{1}{\sigma}\right)\,\Gamma\left(y+\frac{p}{\sigma}\right)\,\Gamma\left[n+\frac{(1-p)}{\sigma}-y\right]}{\Gamma\left(n+\frac{1}{\sigma}\right)\,\Gamma\left(\frac{p}{\sigma}\right)\,\Gamma\left(\frac{1-p}{\sigma}\right)}$$

for $y=0,1,2,...,n$, where $0<p<1$ and $\sigma>0$ (and n is a known positive integer). Note that $E(Y)=np=\mu$, say, and $\mathrm{Var}(Y)=np\,(1-p)\left[1+\frac{\sigma}{1+\sigma}\,(n-1)\right]$. Here σ may be viewed as a random effect parameter.

For our modelling we have a r.v. Y_{ijkl}, where we model p_{ijlk} and σ_{ijkl} in terms of our explanatory variables and factors. We assume that the duration d_{ijkl} of observed stays of patients for cell i, j, k, l are such that $d_{ijkl} \sim D(\psi_{ijkl}, \alpha_{ijkl}, \beta_{ijkl}, \gamma_{ijkl})$ where D is one of the many 4-parameter distributions available in GAMLSS. (We here restrict ourselves to distributions with a closed form for the mean and variance, as this is more convenient for derivation of the expectation and variance of the cost to the NHS of fuel poverty). Our default approach is to assume a log link, i.e. $E(d_{ijkl}) = \psi_{ijkl} = \exp(\zeta_{ijkl})$, where ζ_{ijkl} is a linear predictor based on the explanatory variates in Table 1.

5.2. A model for the probability of dying after admission

A proportion of the emergency respiratory admission die whilst in hospital. We model the probability of dying using the Beta Binomial with logit link, in a similar way to the modelling of the probability of being admitted. The full range of possible covariates and factors was considered.

5.3. Modelling the length of stay in hospital

The length of stay in hospital is different for those who survived and those who died. We model each distribution using a range of continuous densities available in GAMLSS, such as the Gamma, Generalised (3 parameter) Gamma, Inverse Gaussian and a Generalised (3 parameter) Inverse Gaussian.

5.4. Model selection strategy

We illustrate our selection strategy for the 2-parameter Beta Binomial. (We extended this naturally for distributions with more parameters). We initially used the step Akaike criterion to select a model for $\mu_{ijkl} = n_{ijkl} * p_{ijkl}$, keeping σ_{ijkl} constant. We then used a step Akaike approach to fit σ_{ijkl}, for the current "best" linear predictor for μ_{ijkl} (with any remaining parameters constant for the more general case). Using the current "best" linear predictors for σ_{ijkl}, the model for μ_{ijkl} was refitted, and so on. We finally removed the terms whose removal was not significant on a χ^2 scale. We combined levels of factors where this did not result in significant deterioration in scaled deviance.

6. RESULTS

6.1. The probability of being an emergency respiratory admission

From the 1991 Census, there were about 25,000 people in Newham over 64 years old. The total count of emergency respiratory episodes amongst this age group was 3378 (over 4 years), 16% of which ended in death. Respiratory episodes far outnumber those for other possible cold-related diagnoses.

We fitted Beta Binomial models to explain morbidity counts in terms of the wide range of explanatory variables, removing any that were not statistically significant. We attempted to avoid a so-called ecological fallacy by using a wide range of explanatory variables. Investigation of the monthly data for 450 EDs determined that "winter" was better defined as November–February, rather than the traditional UK use of December–March.

The accompanying Table 2 shows our "best" model, using a logit link, for the probability of being an emergency respiratory hospital admission, and a log link for the σ coefficient. The linear predictor for p_{ijkl} has an interaction between "season" and FPR, showing that morbidity counts rise with increasing fuel poverty risk index in "winter", with a notably large effect in December. This is over and above the underlying effect of winter itself, irrespective of FPR. Effects are evident for age, with higher counts for older people, and sex, with lower counts for women.

Table 2: Best fitting BB model. Logit link for p, using log link for σ. age2 represents age level 2, etc., mmaxtt[z] denotes mmaxtt indexed over months z, hh1[E] denotes hh1 indexed over enumeration districts E, etc.

Variable	Estimate	Std. Error	t value	$\Pr(>\|t\|)$
(Intercept)	−6.802 e+00	1.724 e−01	−39.462	0.000 e+00
age2	6.971 e−01	3.966 e−02	17.574	4.958 e−69
age3	1.803 e+00	5.004 e−02	36.022	2.701 e−282
sex2	−7.115 e−01	3.582 e−02	−19.861	1.311 e−87
q2	1.878 e−01	5.790 e−02	3.244	1.179 e−03
q3	4.647 e−01	8.970 e−02	5.180	2.223 e−07
mmaxt[z]	−1.245 e−02	5.162 e−03	−2.413	1.584 e−02
hh2[E]	1.019 e−02	1.976 e−03	5.157	2.521 e−07
lowsap[E]	−1.996 e−03	9.576 e−04	−2.085	3.711 e−02
ctb[E]	8.214 e−03	1.163 e−03	7.062	1.655 e−12
ch[E]	2.602 e−02	2.883 e−03	9.026	1.807 e−19
pens[E]	−4.041 e−02	4.919 e−03	−8.215	2.148 e−16
pre[E]	−2.538 e−03	8.277 e−04	−3.067	2.164 e−03
fpr06[E]	−5.392 e−05	4.175 e−05	−1.291	1.966 e−01
fpr06[E]:nq2	1.692 e−04	7.009 e−05	2.414	1.578 e−02

There was a strong month effect. To understand this further, we considered monthly weather-related factors. Of all these, maximum temperature was most significant, with a higher maximum leading to lower morbidity counts. Having allowed for the maximum temperature effect, other weather related variables were not significant. The log link proved most convenient and as before, we considered all possible covariates and factors.

The σ coefficient (a random effect) depends only upon the age of the people and their gender; see Table 3. (The linear predictor of the σ coefficient is always negative; it is larger for the over 84 year olds than for the over 64 year olds, and is larger for men than women. As a log link is used, the actual value of σ is calculated from the exponential of the linear predictor).

Table 3: The linear predictor for the sigma parameter in the best fitting BB model for the probability of admission (log link). nage2 represents the 85+ years old.

Variable	Estimate	Std. Error	t value	$\Pr(>\|t\|)$
(Intercept)	−4.567	1.362 e−01	−33.517	6.690 e−245
nage2	3.315	1.777 e−01	18.654	1.600 e−77
sex2	−1.175	1.777 e−01	−6.612	3.800 e−11

6.2. Modelling the probability of a patient dying after admission

We fitted the Beta Binomial with logit link and found that only age was significant in explaining the probability π of death of admitted emergency respiratory patients. Older patients were more likely to die; the probabilities of dying were 0.12, 0.16 and 0.24, respectively, for ages 65–74, 75–84 and 85+. The random effect parameter, which we will call λ, did not depend upon the covariates (its value being 1.76).

Table 4: Best fitting Beta Binomial model for probability π of an individual's death, using a logit link for π and log link for λ. The best fit λ is constant. age2 represents age level 2, etc.

Variable	Estimate	Std. Error	t value	$\Pr(>\|t\|)$
(Intercept)	−1.944	9.710 e−02	−20.02	2.42 e−84
age2	0.263	1.154 e−01	2.278	2.28 e−02
age3	0.780	1.227 e−01	6.355	2.36 e−10

6.3. Modelling the length of stay (episode)

We found that, for both those who died and those who survived, the Inverse Gaussian distribution with log link gave the best fit (smallest AIC) amongst the distributions we considered. Our best model for the survivors had a linear predictor for the mean depending only upon age (older people staying longer), with the dispersion decreasing with age. For those who died, the length of stay depended upon gender, with women staying longer. The dispersion parameter was a constant.

Table 5: Linear predictor for length of stay of survivors (Inverse Gaussian, log link). age2 represents age level 2, etc.

Variable	Estimate	Std. Error	t value	$\Pr(>\|t\|)$
(Intercept)	1.965	4.30 e−02	45.790	0.000 e+00
age2	0.233	6.10 e−02	3.830	1.32 e−04
age3	0.530	8.52 e−02	6.230	5.49 e−10

Table 6: Linear predictor for length of stay of those who died in hospital (Inverse Gaussian, log link). S2 represents female.

Variable	Estimate	Std. Error	t value	$\Pr(>\|t\|)$
(Intercept)	2.547	1.31 e−01	19.46	0.000 e+00
S2	0.672	2.36 e−01	3.830	1.32 e−04

7. CONCLUSION

We model both the propensity to be ill and the probability of survival after hospital emergency respiratory admission by the Beta Binomial distribution. We model the subsequent probability of dying in hospital and the length of stay in hospital, thereby providing a potential model for the cost of excess winter morbidity attributable to fuel poverty. Our approach is similar to that used in modelling the probability and cost of insurance claims. The GAMLSS software enables us not only to use the Beta Binomial in modelling the probabilities of admission and survival, but also to use a wide range of continuous distributions to model the length of time that a patient stays in hospital. It may be noted that

mortality due to fuel poverty is a topical issue in the UK. Our emphasis has been on morbidity. With 16% dying, the determining of a direct relationship between mortality and "fuel poverty" would require substantially more data than our 2835 admissions. However, we could give an estimate of the mortality (after emergency respiratory admission)attributable to FPR by combining the probability of being admitted and the subsequent probability of dying (the latter only depending on age).

REFERENCES

[1] Department of Health (2001). *National Service Framework for Older People.* At: `http://www.doh.gov.uk/nsf/olderpeople/pdfs/nsfolderpeople.pdf` [Accessed 12/3/03].

[2] Heller, G.; Stasinopoulos, D.M.; Rigby, R.A. and de Jong, P. (2007). Mean and dispersion modelling for policy claim costs, *Scandinavian Actuarial Journal*, 1–12.

[3] Rigby, R. and Stasinopoulos, D.M. (2005). Generalized additive models for location, scale and shape (with Discussion), *Applied Statistics*, **54**, 507–554.

[4] Rudge, J. and Gilchrist, R. (2005). Excess winter morbidity among older people at risk of cold homes, *Journal of Public Health*, **27**(4), 353–358.

[5] Rudge, J. and Gilchrist, R. (2007). Measuring the health impact of temperatures in dwellings: investigating excess winter morbidity and cold homes in the London Borough of Newham, *Energy and Buildings*, **39**, 847–858.

[6] Wilkinson, P.; Landon, M.; Armstrong, B. *et al.* (2001). *Cold Comfort: the social and environmental determinants of excess winter death in England, 1986–96*, The Policy Press, Bristol.

REVSTAT – Statistical Journal
Volume 7, Number 1, April 2009, 67–85

DETECTING SOCIAL INTERACTIONS IN BIVARIATE PROBIT MODELS WITH AN ENDOGENOUS DUMMY VARIABLE: SOME SIMULATION RESULTS

Author: Johannes Jaenicke
– Faculty of Economics, Law and Social Sciences,
University of Erfurt, Germany
Johannes.Jaenicke@uni-erfurt.de

Abstract:

- This paper analyzes the possibility of detecting observable and non-observable social interactions in a bivariate probit model with an endogenous dummy regressor via Monte Carlo simulation. In small samples, we find severe size distortions of the Wald test and only a low probability of detecting observable and non-observable social interactions. In large samples, however, we find this test to be very powerful.

Key-Words:

- *parameter tests; bivariate probit model; Monte Carlo study.*

AMS Subject Classification:

- 62F03, 62P20.

1. INTRODUCTION

For the researcher, interactions between two persons, e.g., spouses or brothers and sisters may only partly be observable, due to measurement problems. Respondents may not wish to reveal the influence of the other person or may not be aware of them. In order to detect the neglected or non-observable interactions between the respective decision processes a bivariate probit model is recommendable (Jaenicke, 2004).

The identification of discrete choice models with social interactions is studied by Brock and Durlauf (2001, 2007). They show that it is possible to overcome Manski's (1993) famous reflection problem discussed in the context of reference group behaviour by using discrete choice models. In order to account for the neglected or non-observable interactions between discrete decisions in the respective decision processes, a bivariate probit model is a useful empirical description of this process. In this model, an endogenous dummy variable represents the binary decision of the peer that may have an influence on the second person. The bivariate probit model with an endogenous dummy variable is introduced by Maddala and Lee (1976) and belongs to a general class of simultaneous equation models discussed by Heckmann (1978), Maddala (1983), Wilde (2000, 2004), and Greene (2008).

In some applications of the bivariate probit model, e.g. Dean (1995), and Greene (1998) only small samples with 76 or 132 observations are available. However, even in data sets with 500, 1,000 or 2,000 observations, parameter tests may be crucial, as shown by Monfardini and Radici (2008).

Our intention is to find out for different sample sizes whether, in the presence of social interactions, it is possible to detect these interactions in a bivariate probit model with an endogenous dummy regressor. Hence we analyze the distribution of the estimated parameters and size and power of the usual z-coefficient test concerning the parameters of the observable and non-observable interactions, i.e. the endogenous dummy variable and the residual covariance between both equations of this bivariate probit model.

2. A BIVARIATE PROBIT MODEL OF SOCIAL INTERACTIONS

The maximum likelihood estimation of a bivariate probit model involves the numerical problem of the evaluation of double integrals over the normal distribution. This estimation procedure is implemented in several statistic software packages and widely used in practice. We use a two equation binary choice model with an endogenous dummy regressor, first proposed by Maddala and Lee (1976).

The regression equations of the individual I and the peer P are

$$\begin{aligned} Y_I^* &= X_I\beta_1 + Y_P\beta_2 + u_I\,, & Y_I &= 1 \quad \text{if } Y_I^* > 0, \quad 0 \text{ otherwise}\,, \\ Y_P^* &= X_P\gamma_1 + u_P\,, & Y_P &= 1 \quad \text{if } Y_P^* > 0, \quad 0 \text{ otherwise}\,, \\ [u_I, u_P] &\sim \Phi_2(0,0,1,1,\rho)\,, \end{aligned}$$

with the observable discrete choice behavior Y, latent variables Y_I^*, and exogenous variables X. The residual vector $[u_I, u_P]$ is bivariate normal distributed with $E(u_i) = 0$, $\text{var}(u_i) = 1$, $i = I, P$, and $\text{cov}(u_I, u_P) = \rho$. As a condition of identification, we only need exclusion restrictions if there is no variation of the exogenous regressors (Wilde, 2000).

In our model, the observable part of the social interactions, the influence of the decision of the peer P on the behavior of the individual I is tested by the hypothesis $H_0 \colon \beta_2 = 0$. The non-observable part of the social interactions may be revealed through the residual covariance structure. A residual covariance $\text{cov}(u_I, u_P)$, i.e. ρ, significantly different from zero, may serve as an indicator of unobserved social interactions between the two decisions or as an indicator of simultaneously neglected third-party effects. Restricting residual correlation of the bivariate probit model to zero may result in biased and inconsistent estimations (Murphy, 1995). Fitting separate probit models for the first- and the second decision equation can involve significant endogeneity biases in the estimation (Lollivier, 2001). The joint estimation of the two equations provides substantial efficiency gains compared to separate estimation based on two-stage technique.

3. MONTE CARLO RESULTS FOR THE BIVARIATE PROBIT MODEL

In a small Monte Carlo study, we analyze the size and the power of the usual z-coefficient tests concerning the parameters of the observable and non-observable social interactions, β_2 and ρ. The test statistics are $z(\beta_2) = \hat{\beta}_2 / se(\hat{\beta}_2)$ and $z(\rho) = \hat{\rho} / se(\hat{\rho})$ and their squares result in the standard Wald test (Greene 2008, p. 820).

The non-observable influences stem from missing variables. These may be uncorrelated, weakly or strongly correlated or identically for the two persons. To create the non-observable interactions, we use an omitted variable vector $[v_I, v_P] \sim \Phi_2(0,0,1,1,r)$ with $r \in [0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 1.0]$ in one set of experiments. In this case, the residuals u_i, $i = I, P$, are the sum $u_i = v_i + \varepsilon_i$ with $[\varepsilon_I, \varepsilon_P] \sim \Phi_2(0,0,1,1,0)$, therefore $[u_I, u_P] \sim \Phi_2(0,0,2,2,\frac{r}{2})$. Because the assumption of the unit residual variance $\text{var}(u_i)$ is not met, we expect some problems resulting from the misspecification of the model.

In the experiments with the extended model, we include v_i as additional explanatory variables. In this case, the residuals are $u_i = \varepsilon_i$ and are independent normal distributed. Because of the independence of the residuals, $\rho = 0$, the model is overparametrized. Two single equation models would be more efficient. Anyway, since we do not know the true parameter set in the empirical research situation, in the simulation experiments we continue with bivariate probit models.

The variables X_i, $i = I, P$, are standard normal distributed, $X_i \sim N(0, 1)$, $i = I, P$. All parameters β and γ in the omitted variable model and in the extended model are set equal to one. We use the econometric software package Limdep 7.0. It performs well in nonlinear estimation benchmark tests (McCullough, 1999). We estimate the bivariate models with the default settings of the procedure (algorithm: BFGS; maximum iterations: 100). The Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm is rather time consuming, but it shows a convergence rate of between 99.5 percent (in small data sets with 100 observations) and 100 percent (in data sets with 10,000 observations) in our Monte Carlo study. The number of replications in the Monte Carlo experiment is $N = 1{,}000$. The number of observations varies systematically from $T = 100$ to $T = 2{,}000$. In some experiments, we estimate data sets up to $T = 10{,}000$ and in one up to $T = 40{,}000$ observations. Due to the nonlinear estimation problem, experiments with these huge data sets are computational intensive, e.g., some experiments need around 84 hours on a Pentium(R), 3.2 GHz to estimate two bivariate probit models 1,000 times.

The estimated parameters $\widehat{\rho}$ show no severe bias. Figure 1 presents the density estimation with the Epanechnikov kernel function in the case that the true parameter $\rho = 0$. With increasing sample size from $T = 200$ to $T = 1{,}000$,

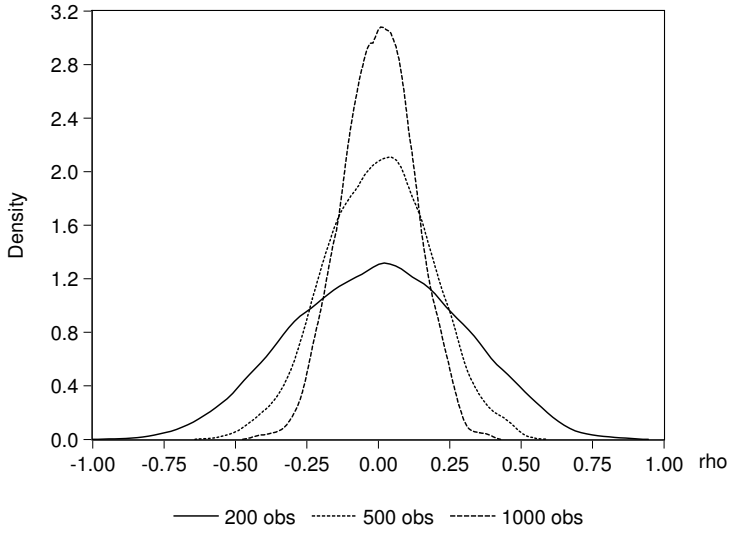

Figure 1: Kernel estimation, $\rho = 0$.

the dispersion of the estimated parameters becomes smaller. The parameters are distributed more or less symmetrically (with skewness S_T between -0.083 and -0.040) and means (with $\overline{\widehat{\rho_T}}$ between -0.001 and 0.003) very close to the theoretical value zero.

The picture changes if we assume with $\rho = 0.5$ strong residual correlation between both decision equations in small data sets. In the case of $T = 200$, we find with $\overline{\widehat{\rho_{200}}} = 0.485$ some deviations from the theoretical parameter value. In all three cases, the distribution of the estimated parameters $\widehat{\rho_{200}}$ is left skewed with a skewness S_T between -0.568 and -0.403. In the $T = 200$ case, some $\widehat{\rho_{200}}$ are very close the theoretical limit of 1. In Figure 2, we present a Kernel density estimation for $\rho = 0.5$ and $T = 200$, $T = 500$ and $T = 1{,}000$.

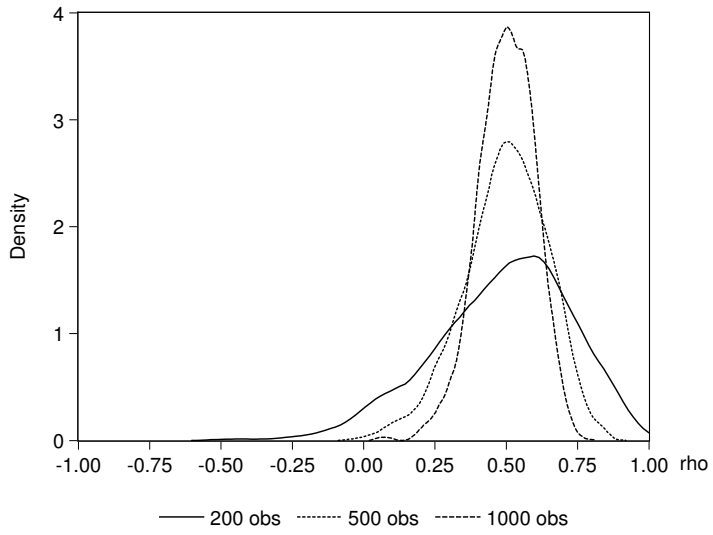

Figure 2: Kernel estimation, $\rho = 0.5$.

Looking at the z-statistics in the case $\rho = 0$, we find some indication that the z-statistics may not be normally distributed in small samples. To analyze this problem graphically, we compare the quantiles of these statistics with the standard normal distribution in Figure 3 for $T = 200, 500, 1000$ observations. Especially in the case of $\rho = 0$, $T = 200$ observations, we find some deviation from normality.

We start our analysis studying possible size distortions of the z_ρ-parameter test in the bivariate probit model with an endogenous dummy variable. The parameter describing the influence of the endogenous dummy variable is set equal to one, $\beta_2 = 1$. In Table 1 (see Appendix), we present percentiles of the simulated z_ρ-statistic, some descriptive statistics like mean, standard deviation, skewness, kurtosis, and the size of the test statistic with the nominal significance level equal to 1, 5 and 10 percent.

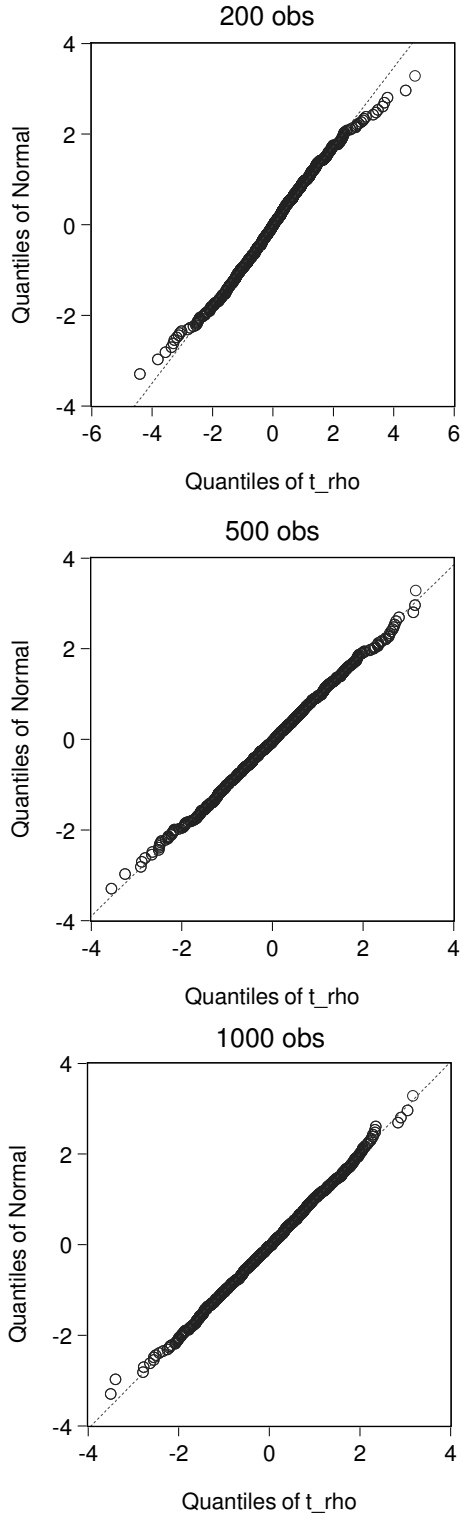

Figure 3: QQ-Plots for t_ρ, with $\rho = 0$ and $T = 200, 500, 1000$.

From Table 1, we see severe size distortions for the z_ρ-parameter test in small data sets. In the case of $T = 100, 150$ observations, the test statistic is excessively liberal. E.g., in the case of a nominal 5-percent level, the empirical size is more than twice as high. The deviations from the nominal level become stronger with a higher significance level. E.g., on the 1 percent level, the empirical size is 7 percent in the extreme case of $T = 100$ and 5 percent in the case of $T = 150$ observations. In these two cases the z_ρ-parameter test shows strong deviations from normality. The distribution is negatively skewed and reveals strong excess kurtosis. As expected, deviations from normality are not pronounced in the case of medium size or huge data sets.

In Table 2, we present a set of experiments using our bivariate probit model with an endogenous dummy variable but restricting the data generating mechanism to $\beta_2 = 0$, i.e., there is no endogenous dummy variable in the true model. The results are generally in line with the ones of Table 1, but with stronger deviations from normality in the small sample cases.

Using different correlation structures in Table 3 ($r = 0$) and Table 4 ($r = 1$), we analyse the small sample behaviour of the test concerning the observable interactions, the z_{β_2}-parameter test. We find again some size distortions for data sets with $T \leq 200$ observations but no systematic influence of the correlation r.

How will the size of the test statistic be affected if it is possible to include the omitted variable vector $[v_I, v_P]$ as additional regressors? Does the correlation r have an influence on the outcome of the test? Comparing Tables 2 and 5, we see that the inclusion of additional explanatory variables makes the size distortions decrease only slightly. The comparison of Tables 5 and 6 makes clear that correlation between explanatory variables negatively affects the size of the z_ρ-test. In the case of $r = 1$, the empirical p-value of 10 percent is shifted to 23 percent in the $T = 100$ observation case. In this strong correlation case, we need more than $T = 2{,}000$ observations to obtain good size properties of the z_ρ-test. Summarizing Table 1 to 6, we find that in all small sample cases the test statistic is too liberal.

In the next simulation experiments, we have a look at the power properties of the test statistic. To answer the question whether it is possible to detect non-observable social interactions, we run 57 different Monte Carlo experiments. In the case of low residual correlation, the probability of finding significant interactions lies below one half in small or medium sized data sets. E.g., in the case of $r = 0.2$ and $T = 2{,}500$, only 35 percent of the z_ρ-tests are significant at the 10 percent level and in small data sets, e.g. $r = 0.2$ and $T = 100$, only 17 percent of the z_ρ-tests are significant at the 10 percent level. Taking into account the size distortions of the test statistic, the results will become even more unsatisfactionary. As expected, with more observations, and stronger correlation, the power increases. A power of at least 90 percent (at the nominal 10 percent level) can be found in the cases of $r = 0.2$ and $T = 15{,}000$ or $r = 1$ and $T = 2{,}500$.

The power properties of the z_{β_2}-test, presented in Table 8, are (at the 10 percent level) around one third in the small sample case and nearly 100 percent in $T \geq 1{,}000$ data sets. The changing of the correlation r has more or less no influence on the Monte Carlo results.

The power can be dramatically increased if it is possible to include the neglected variable vector $[v_I, v_P]$. The results of these Monte Carlo experiments are presented in Table 9. The inclusion of these additional explanatory variables will shift the power properties significantly towards more than 50 percent even in the $T = 100$ case. With 250 observations, the power is close to the 90 percent level or higher. This is true although the extended model is overparametrized, because the true correlation-coefficient is zero in this model. In line with the results from Table 8, the results are nearly robust regarding the correlation structure r. Nevertheless, we have to take into account the too liberal behavior of our test in small samples.

4. CONCLUSIONS

In our paper, we find that the power of z-parameter tests concerning the residual correlation between the two decision equations in the bivariate probit model is very low in small samples. This is is especially true for weak correlations. The power of the parameter test concerning the endogenous dummy variable is around one third in small samples. If it is possible to find omitted variables, the power of this test can be increased notedly.

Additionally, our simulation results reveal severe over-rejection rates in small samples. Using a likelihood ratio test may result in better size behavior at least in medium sized samples (Monfardini and Radici, 2008).

From an empirical point of view, we may often fail to find significant social interactions in the data sets although they exist. An extensive search for omitted variables may therefore be essential to prove social interactions in empirical models.

ACKNOWLEDGMENTS

The author wishes to thank the anonymous referee of this paper and the participants of the discussion at COMPSTAT 2008, Annual Meeting of the Austrian Economic Association 2008, Pfingsttagung der Deutschen Statistischen Gesellschaft 2008, and Uwe Hassler, Goethe University, Frankfurt a. M. for valuable comments. I am especially grateful to Margot Petersen-Jaenicke for her valuable help.

REFERENCES

[1] Brock, W.A. and Durlauf, S.N. (2001). Discrete choice with social interactions, *Review of Economic Studies*, **68**, 235–260.

[2] Brock, W.A. and Durlauf, S.N. (2007). Identification of binary choice models with social interactions, *Journal of Econometrics*, **140**, 52–75.

[3] Dean, J.M. (1995). Market disruption and the incidence of VERs under the MFA, *Review of Economics and Statistics*, **77**, 383–388.

[4] Greene, W. (1998). Gender economics courses in liberal art colleges: Further results, *Journal of Economic Education*, **29**, 291–300.

[5] Greene, W.H. (2008). *Econometric Analysis*, 6th ed., Pearson, Prentice Hall.

[6] Heckman, J. (1978). Dummy endogenous variables in a simultaneous equation system, *Econometrica*, **9**, 255–268.

[7] Jaenicke, J. (2004). *Observable and Non-Observable Social Interactions in Labor Supply*, Discussion paper No. 2003/05, Rev. version, May 2004, University of Osnabrück.

[8] Lollivier, S. (2001). Endogénéité d'une variable explicative dichotomique dans le cadre d'un modèle probit bivarié, *Annales d'Économie et de Statistique*, **62**, 251–269.

[9] Maddala, G.S. (1983). *Limited Dependent and Quantitative Variables in Econometrics*, Cambridge University Press, Cambridge.

[10] Maddala, G.S. and Lee, L.-F. (1976). Recursive models with qualitative endogenous variables, *Annals of Economic and Social Measurement*, **5**, 525–545.

[11] Manski, C.F. (1993). Identification of endogenous social effects: the reflection problem, *Review of Economic Studies*, **60**, 531–542.

[12] McCullough, B.D. (1999). Econometric software reliability: EViews, LIMDEP, SHAZAM and TSP, *Journal of Applied Econometrics*, **14**, 191–202.

[13] Monfardini, C. and Radice, R. (2008). Testing exogeneity in the bivariate probit model: A Monte Carlo study, *Oxford Bulletin of Economics and Statistics*, **70**, 271–282.

[14] Murphy, A. (1995). Female labour force participation and unemployment in Northern Ireland: Religion and family effects, *Economic and Social Review*, **27**, 67–84.

[15] Wilde, J. (2000). Identification of multiple equation probit models with endogenous dummy regressors, *Economics Letters*, **69**, 309–312.

[16] Wilde, J. (2004). Estimating multiple equation hybrid models with endogenous dummy regressors, *Statistica Neerlandica*, **58**, 296–312.

Appendix

Table 1: Finite sample behavior of the z_ρ - parameter test concerning the non-observable interactions in the case of r=0 and β2=1

		percentiles																p-value		
r	T	0.01	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99	mean	std.	skew.	kurt.	0.01	0.05	0.1
0.0	100	-4.07	-2.09	-1.50	-0.91	-0.53	-0.27	-0.02	0.21	0.50	0.80	1.51	2.11	4.85	-0.04	1.60	-1.47	23.14	0.07	0.12	0.17
0.0	150	-4.00	-1.97	-1.46	-0.82	-0.51	-0.26	0.00	0.21	0.47	0.81	1.35	1.87	3.85	-0.06	1.42	-2.04	22.25	0.05	0.10	0.15
0.0	200	-3.04	-1.78	-1.36	-0.87	-0.53	-0.24	0.01	0.27	0.53	0.91	1.41	1.90	3.01	0.02	1.14	0.09	3.94	0.03	0.08	0.14
0.0	250	-2.36	-1.60	-1.24	-0.81	-0.49	-0.25	0.00	0.24	0.55	0.86	1.44	1.99	2.91	0.05	1.06	0.36	3.57	0.02	0.07	0.12
0.0	500	-2.50	-1.61	-1.24	-0.82	-0.50	-0.25	0.03	0.29	0.56	0.86	1.34	1.73	2.59	0.03	1.03	0.00	3.18	0.02	0.06	0.10
0.0	750	-2.54	-1.81	-1.29	-0.84	-0.56	-0.23	0.02	0.27	0.54	0.90	1.33	1.71	2.54	0.01	1.04	-0.01	3.07	0.02	0.06	0.12
0.0	1000	-2.35	-1.58	-1.27	-0.82	-0.50	-0.23	0.04	0.29	0.54	0.82	1.29	1.67	2.24	0.02	0.99	-0.04	3.07	0.01	0.05	0.09
0.0	1500	-2.58	-1.66	-1.24	-0.83	-0.52	-0.20	0.04	0.29	0.56	0.87	1.34	1.63	2.38	0.03	1.02	-0.11	3.25	0.02	0.04	0.10
0.0	2000	-2.22	-1.63	-1.21	-0.79	-0.51	-0.22	0.07	0.32	0.58	0.87	1.32	1.63	2.25	0.04	0.98	-0.08	2.89	0.01	0.04	0.09
0.0	2500	-2.21	-1.59	-1.23	-0.79	-0.51	-0.25	0.03	0.28	0.54	0.83	1.30	1.61	2.42	0.03	0.98	0.09	3.04	0.01	0.05	0.09
0.0	5000	-2.20	-1.61	-1.17	-0.76	-0.44	-0.22	0.02	0.27	0.53	0.90	1.26	1.62	2.35	0.04	0.97	0.02	2.97	0.01	0.04	0.09
0.0	7500	-2.15	-1.57	-1.25	-0.78	-0.45	-0.20	0.07	0.31	0.55	0.85	1.24	1.50	2.44	0.04	0.96	-0.03	2.91	0.01	0.04	0.09
0.0	10000	-2.33	-1.60	-1.20	-0.80	-0.44	-0.20	0.07	0.29	0.55	0.86	1.28	1.63	2.30	0.04	0.98	-0.01	3.09	0.01	0.05	0.09

Table 2: Finite sample behavior of the z_ρ - parameter test concerning the non-observable interactions in the case of r=0 and β2=0

		percentiles																p-value		
r	T	0.01	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99	mean	std.	skew.	kurt.	0.01	0.05	0.1
0.0	100	-4.22	-1.97	-1.48	-0.90	-0.57	-0.30	-0.05	0.18	0.48	0.83	1.38	2.02	4.09	-0.08	1.56	-2.82	31.83	0.06	0.10	0.16
0.0	150	-3.06	-2.00	-1.47	-0.88	-0.54	-0.25	0.01	0.23	0.49	0.87	1.46	1.94	2.63	-0.07	1.90	-15.57	387.95	0.03	0.10	0.15
0.0	200	-2.81	-1.76	-1.37	-0.85	-0.52	-0.25	0.02	0.26	0.50	0.83	1.38	1.90	2.97	0.01	1.11	0.05	3.73	0.03	0.08	0.14
0.0	250	-2.52	-1.55	-1.21	-0.78	-0.48	-0.25	0.02	0.29	0.59	0.90	1.34	1.76	3.14	0.06	1.05	0.22	4.31	0.02	0.06	0.10
0.0	500	-2.59	-1.65	-1.24	-0.78	-0.48	-0.24	0.02	0.27	0.56	0.92	1.31	1.67	2.50	0.03	1.02	-0.13	3.39	0.02	0.05	0.10
0.0	750	-2.50	-1.63	-1.25	-0.85	-0.55	-0.27	0.01	0.24	0.54	0.86	1.35	1.77	2.51	0.01	1.05	0.10	3.36	0.02	0.07	0.11
0.0	1000	-2.41	-1.54	-1.22	-0.82	-0.52	-0.24	0.02	0.28	0.55	0.86	1.35	1.65	2.33	0.03	0.99	-0.01	3.02	0.01	0.04	0.09
0.0	1500	-2.42	-1.60	-1.28	-0.86	-0.54	-0.26	0.04	0.29	0.59	0.89	1.39	1.70	2.29	0.03	1.03	-0.03	3.10	0.01	0.04	0.10
0.0	2000	-2.48	-1.69	-1.28	-0.84	-0.49	-0.18	0.09	0.36	0.59	0.94	1.35	1.68	2.45	0.05	1.03	-0.10	3.05	0.01	0.05	0.11
0.0	2500	-2.20	-1.62	-1.25	-0.80	-0.48	-0.26	0.01	0.29	0.56	0.90	1.32	1.69	2.43	0.03	1.00	0.07	2.94	0.01	0.05	0.10
0.0	5000	-2.32	-1.63	-1.25	-0.80	-0.47	-0.20	0.06	0.31	0.58	0.91	1.27	1.57	2.28	0.04	0.98	-0.09	2.90	0.01	0.04	0.09
0.0	7500	-2.25	-1.56	-1.20	-0.79	-0.49	-0.24	0.06	0.33	0.55	0.81	1.30	1.61	2.28	0.04	0.97	-0.01	2.95	0.01	0.04	0.08
0.0	10000	-2.31	-1.51	-1.23	-0.82	-0.53	-0.24	0.07	0.32	0.56	0.87	1.31	1.69	2.34	0.03	0.99	-0.06	3.05	0.01	0.05	0.09

Table 3: **Finite sample behavior of the z_β2 - parameter test concerning the observable interactions in the case of r=0 and β2=0**

		percentiles																	p-value		
r	**T**	**0.01**	**0.05**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**0.95**	**0.99**	**mean**	**std.**	**skew.**	**kurt.**	**0.01**	**0.05**	**0.1**
0.0	100	-4.76	-1.95	-1.37	-0.84	-0.52	-0.28	-0.02	0.26	0.54	0.88	1.38	1.86	3.15	-0.05	1.33	-1.32	11.06	0.05	0.09	0.14
0.0	150	-2.71	-1.76	-1.34	-0.87	-0.56	-0.27	-0.01	0.25	0.55	0.91	1.42	1.89	2.65	0.00	1.19	-0.83	12.08	0.03	0.08	0.13
0.0	200	-2.75	-1.81	-1.46	-0.92	-0.53	-0.25	0.01	0.27	0.59	0.88	1.35	1.73	2.57	-0.02	1.09	-0.11	3.20	0.02	0.07	0.13
0.0	250	-2.49	-1.77	-1.38	-0.89	-0.56	-0.26	-0.05	0.23	0.50	0.77	1.17	1.54	2.42	-0.06	1.01	-0.13	3.25	0.01	0.06	0.10
0.0	500	-2.33	-1.69	-1.34	-0.93	-0.61	-0.29	0.00	0.22	0.46	0.78	1.24	1.60	2.47	-0.04	1.02	0.14	3.24	0.01	0.06	0.10
0.0	750	-2.76	-1.61	-1.37	-0.90	-0.59	-0.27	0.07	0.30	0.57	0.87	1.35	1.79	2.54	0.00	1.06	-0.02	3.07	0.02	0.06	0.10
0.0	1000	-2.28	-1.63	-1.31	-0.90	-0.59	-0.31	0.03	0.27	0.53	0.81	1.22	1.61	2.37	-0.02	1.00	0.06	2.97	0.01	0.05	0.09
0.0	1500	-2.31	-1.73	-1.37	-0.90	-0.59	-0.25	-0.03	0.25	0.54	0.86	1.28	1.62	2.22	-0.03	1.03	0.04	3.00	0.01	0.06	0.11
0.0	2000	-2.27	-1.66	-1.33	-0.89	-0.60	-0.29	-0.03	0.21	0.48	0.80	1.28	1.68	2.24	-0.03	1.01	0.11	3.23	0.01	0.04	0.11
0.0	2500	-2.33	-1.69	-1.32	-0.82	-0.49	-0.25	0.00	0.22	0.49	0.84	1.25	1.56	2.30	-0.01	0.99	-0.06	3.12	0.01	0.05	0.09
0.0	5000	-2.28	-1.59	-1.26	-0.83	-0.58	-0.30	-0.06	0.23	0.52	0.83	1.29	1.66	2.42	-0.01	0.99	0.12	2.86	0.01	0.05	0.10
0.0	7500	-2.22	-1.55	-1.21	-0.82	-0.53	-0.30	-0.03	0.20	0.47	0.77	1.24	1.62	2.19	-0.02	0.96	0.05	3.03	0.01	0.05	0.09
0.0	10000	-2.26	-1.63	-1.20	-0.83	-0.54	-0.27	-0.04	0.21	0.50	0.77	1.21	1.60	2.55	-0.02	0.98	0.14	3.29	0.01	0.05	0.10

Table 4: **Finite sample behavior of the z_β2 - parameter test concerning the observable interactions in the case of r=1 and β2=0**

		percentiles																	p-value		
r	**T**	**0.01**	**0.05**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**0.95**	**0.99**	**mean**	**std.**	**skew.**	**kurt.**	**0.01**	**0.05**	**0.10**
1.0	100	-4.96	-2.50	-1.64	-0.93	-0.59	-0.29	-0.03	0.21	0.50	0.79	1.18	1.58	2.13	-0.17	1.28	-1.23	6.76	0.11	0.09	0.14
1.0	150	-4.32	-2.22	-1.60	-0.96	-0.59	-0.29	-0.01	0.23	0.51	0.82	1.23	1.62	2.21	-0.14	1.25	-1.34	7.78	0.08	0.09	0.14
1.0	200	-4.12	-2.04	-1.58	-1.01	-0.60	-0.25	0.03	0.25	0.53	0.81	1.21	1.48	2.21	-0.13	1.23	-1.61	10.72	0.03	0.08	0.13
1.0	250	-3.15	-1.92	-1.38	-0.93	-0.60	-0.27	-0.04	0.19	0.47	0.77	1.10	1.41	2.08	-0.12	1.05	-0.51	3.89	0.03	0.06	0.10
1.0	500	-2.71	-1.79	-1.37	-0.94	-0.60	-0.28	-0.02	0.19	0.43	0.72	1.18	1.52	2.36	-0.09	1.01	-0.14	3.25	0.02	0.06	0.10
1.0	750	-2.90	-1.72	-1.33	-0.90	-0.54	-0.27	-0.01	0.25	0.51	0.79	1.28	1.65	2.31	-0.04	1.05	-0.30	3.55	0.02	0.06	0.11
1.0	1000	-2.52	-1.72	-1.33	-0.94	-0.60	-0.30	0.01	0.25	0.48	0.77	1.17	1.52	2.12	-0.06	1.00	-0.12	3.01	0.01	0.06	0.09
1.0	1500	-2.53	-1.77	-1.39	-0.93	-0.62	-0.30	-0.04	0.26	0.54	0.81	1.21	1.53	2.25	-0.05	1.02	-0.07	3.05	0.01	0.05	0.10
1.0	2000	-2.42	-1.70	-1.32	-0.92	-0.62	-0.31	-0.06	0.22	0.48	0.81	1.27	1.65	2.18	-0.05	1.01	-0.01	3.10	0.01	0.05	0.11
1.0	2500	-2.49	-1.70	-1.31	-0.85	-0.51	-0.24	-0.02	0.26	0.50	0.85	1.23	1.61	2.14	-0.03	0.99	-0.14	3.12	0.01	0.05	0.10
1.0	5000	-2.22	-1.65	-1.20	-0.84	-0.54	-0.29	-0.03	0.23	0.50	0.81	1.28	1.65	2.43	-0.01	0.98	0.06	2.89	0.01	0.05	0.10
1.0	7500	-2.36	-1.60	-1.22	-0.79	-0.53	-0.27	-0.02	0.22	0.49	0.79	1.26	1.58	2.18	-0.02	0.96	-0.02	2.99	0.01	0.05	0.09
1.0	10000	-2.25	-1.66	-1.24	-0.77	-0.55	-0.26	-0.03	0.22	0.48	0.79	1.22	1.61	2.46	-0.01	0.98	0.08	3.14	0.01	0.04	0.10

Table 5: Finite sample behavior of the z_ρ - parameter test in the case of r=0, β2=0 and additional explanatory variables

		percentiles																	p-value		
r	T	0.01	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99	mean	std.	skew.	kurt.	0.01	0.05	0.1
0.0	100	-4.14	-2.27	-1.67	-0.95	-0.58	-0.31	-0.06	0.21	0.50	0.86	1.42	1.99	3.65	-0.08	1.39	-0.66	7.78	0.06	0.12	0.17
0.0	150	-3.34	-2.00	-1.46	-0.90	-0.55	-0.27	-0.04	0.22	0.49	0.83	1.31	1.73	2.93	-0.06	1.43	-1.37	40.10	0.04	0.09	0.13
0.0	200	-3.16	-1.94	-1.35	-0.84	-0.52	-0.23	0.01	0.24	0.50	0.88	1.38	1.76	2.57	-0.02	1.14	-0.20	4.29	0.03	0.08	0.13
0.0	250	-2.91	-1.73	-1.30	-0.82	-0.51	-0.23	0.01	0.27	0.54	0.85	1.36	1.74	2.49	0.00	1.11	-0.53	7.04	0.02	0.07	0.12
0.0	500	-2.23	-1.68	-1.22	-0.80	-0.48	-0.23	0.00	0.27	0.54	0.88	1.33	1.73	2.35	0.03	1.01	0.01	3.03	0.01	0.05	0.11
0.0	750	-2.33	-1.66	-1.24	-0.81	-0.52	-0.24	0.02	0.25	0.47	0.77	1.23	1.60	2.29	-0.01	0.98	-0.08	3.43	0.01	0.05	0.10
0.0	1000	-2.43	-1.69	-1.29	-0.84	-0.47	-0.26	0.02	0.29	0.53	0.86	1.32	1.68	2.46	0.02	1.02	-0.04	3.01	0.01	0.06	0.11
0.0	1500	-2.27	-1.76	-1.34	-0.82	-0.50	-0.25	0.04	0.30	0.54	0.86	1.31	1.67	2.45	0.01	1.02	0.04	2.94	0.01	0.06	0.11
0.0	2000	-2.35	-1.62	-1.27	-0.81	-0.54	-0.25	0.03	0.27	0.54	0.86	1.33	1.73	2.65	0.03	1.02	0.11	3.10	0.02	0.05	0.11
0.0	2500	-2.30	-1.61	-1.24	-0.85	-0.52	-0.28	0.00	0.29	0.55	0.88	1.35	1.69	2.30	0.02	1.01	0.12	2.97	0.01	0.04	0.11
0.0	5000	-2.57	-1.57	-1.30	-0.82	-0.48	-0.21	0.03	0.30	0.55	0.84	1.26	1.78	2.50	0.03	1.01	0.01	3.13	0.02	0.06	0.10
0.0	7500	-2.21	-1.55	-1.25	-0.79	-0.52	-0.23	0.04	0.27	0.53	0.85	1.36	1.71	2.40	0.03	0.99	0.06	3.01	0.01	0.05	0.11
0.0	10000	-2.38	-1.63	-1.26	-0.81	-0.48	-0.19	0.03	0.30	0.56	0.91	1.36	1.71	2.55	0.05	1.03	-0.02	3.10	0.01	0.06	0.11

Table 6: Finite sample behavior of the z_ρ - parameter test in the case of r=1, β2=0 and additional explanatory variables

		percentiles																	p-value		
r	T	0.01	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99	mean	std.	skew.	kurt.	0.01	0.05	0.1
1.0	100	-8.40	-3.45	-1.81	-1.01	-0.60	-0.29	-0.03	0.24	0.49	0.93	1.73	2.46	4.83	-0.15	1.98	-1.56	12.84	0.11	0.17	0.23
1.0	150	-5.33	-2.35	-1.59	-0.93	-0.57	-0.24	0.01	0.27	0.54	0.93	1.58	2.15	4.22	-0.05	1.62	-1.42	14.00	0.08	0.13	0.19
1.0	200	-6.64	-2.28	-1.59	-0.98	-0.56	-0.28	-0.01	0.22	0.50	0.92	1.52	2.23	3.43	-0.10	1.71	-2.92	28.76	0.07	0.13	0.18
1.0	250	-4.44	-1.96	-1.39	-0.84	-0.54	-0.26	-0.02	0.28	0.57	0.92	1.52	2.01	2.84	-0.05	1.60	-4.66	50.76	0.04	0.10	0.15
1.0	500	-2.98	-1.78	-1.39	-0.84	-0.53	-0.22	0.01	0.27	0.52	0.84	1.33	1.77	2.66	0.00	1.09	-0.11	3.75	0.03	0.08	0.12
1.0	750	-2.65	-1.89	-1.41	-0.89	-0.50	-0.25	0.01	0.25	0.50	0.88	1.27	1.77	2.74	-0.02	1.10	-0.09	3.63	0.02	0.07	0.13
1.0	1000	-2.62	-1.73	-1.38	-0.87	-0.51	-0.22	0.04	0.31	0.58	0.82	1.30	1.73	2.41	0.01	1.05	-0.16	3.46	0.02	0.07	0.11
1.0	1500	-2.88	-1.70	-1.31	-0.85	-0.47	-0.22	0.01	0.29	0.55	0.85	1.30	1.72	2.62	0.01	1.07	-0.14	3.77	0.02	0.07	0.12
1.0	2000	-2.54	-1.72	-1.34	-0.77	-0.46	-0.21	0.03	0.29	0.55	0.89	1.34	1.80	2.69	0.03	1.07	0.03	3.72	0.02	0.07	0.12
1.0	2500	-2.49	-1.55	-1.24	-0.81	-0.47	-0.24	0.01	0.24	0.52	0.84	1.30	1.74	2.58	0.03	1.02	0.15	3.54	0.02	0.06	0.10
1.0	5000	-2.34	-1.51	-1.18	-0.82	-0.50	-0.24	-0.01	0.27	0.53	0.90	1.38	1.76	2.56	0.04	1.01	0.07	3.13	0.02	0.05	0.10
1.0	7500	-2.30	-1.56	-1.21	-0.77	-0.49	-0.22	0.01	0.27	0.55	0.89	1.31	1.66	2.33	0.04	0.98	0.02	3.00	0.01	0.04	0.09
1.0	10000	-2.45	-1.59	-1.27	-0.78	-0.45	-0.19	0.07	0.31	0.56	0.89	1.37	1.72	2.48	0.05	1.01	-0.06	3.16	0.01	0.06	0.11

Table 7: Finite sample behavior of the z_ρ - parameter test concerning the non-observable interactions in the case of r>0 and $\beta 2=1$

		percentiles																	p-value		
r	**T**	**0.01**	**0.05**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**0.95**	**0.99**	**mean**	**std.**	**skew.**	**kurt.**	**0.01**	**0.05**	**0.1**
0.1	100	-4.07	-2.09	-1.50	-0.91	-0.53	-0.27	-0.02	0.21	0.50	0.80	1.51	2.11	4.85	-0.04	1.60	-1.47	23.14	0.07	0.12	0.17
0.1	150	-4.00	-1.97	-1.46	-0.82	-0.51	-0.26	0.00	0.21	0.47	0.81	1.35	1.87	3.85	-0.06	1.42	-2.04	22.25	0.05	0.10	0.15
0.1	200	-2.33	-1.56	-1.15	-0.73	-0.36	-0.09	0.15	0.42	0.71	1.12	1.59	2.05	3.40	0.21	1.14	0.43	4.12	0.03	0.08	0.14
0.1	250	-2.07	-1.33	-1.02	-0.63	-0.33	-0.07	0.19	0.43	0.76	1.06	1.63	2.15	3.21	0.26	1.09	0.50	3.90	0.03	0.08	0.12
0.1	500	-2.10	-1.31	-1.02	-0.59	-0.24	0.02	0.25	0.58	0.86	1.20	1.65	2.01	2.98	0.31	1.06	0.14	3.48	0.03	0.07	0.13
0.1	750	-2.14	-1.41	-0.98	-0.53	-0.22	0.07	0.33	0.63	0.87	1.25	1.70	2.02	2.86	0.35	1.06	0.05	3.16	0.02	0.07	0.14
0.1	1000	-1.91	-1.21	-0.85	-0.43	-0.14	0.16	0.41	0.70	0.94	1.26	1.72	2.13	2.83	0.42	1.00	0.06	2.99	0.01	0.07	0.13
0.1	1500	-2.03	-1.21	-0.78	-0.35	-0.04	0.25	0.51	0.78	1.05	1.37	1.81	2.18	2.93	0.51	1.03	-0.04	3.18	0.02	0.09	0.16
0.1	2000	-1.51	-1.04	-0.68	-0.29	0.08	0.35	0.60	0.87	1.15	1.46	1.89	2.22	2.91	0.60	0.99	-0.03	2.85	0.02	0.09	0.16
0.2	100	-3.97	-1.73	-1.14	-0.63	-0.28	-0.06	0.19	0.44	0.72	1.12	1.78	2.61	5.30	0.24	1.68	-2.54	35.35	0.07	0.12	0.17
0.2	150	-2.75	-1.57	-1.05	-0.56	-0.22	0.02	0.24	0.54	0.79	1.12	1.77	2.30	4.72	0.31	1.36	1.02	14.79	0.05	0.10	0.16
0.2	200	-2.13	-1.36	-1.02	-0.54	-0.18	0.10	0.31	0.54	0.86	1.27	1.78	2.33	3.86	0.38	1.19	0.14	8.12	0.04	0.10	0.15
0.2	250	-1.84	-1.16	-0.82	-0.42	-0.14	0.08	0.34	0.62	0.95	1.25	1.86	2.35	3.48	0.46	1.11	0.68	4.47	0.04	0.10	0.14
0.2	500	-1.90	-1.07	-0.73	-0.29	0.04	0.27	0.54	0.82	1.14	1.50	1.95	2.34	3.22	0.59	1.08	0.23	3.75	0.04	0.11	0.18
0.2	750	-1.77	-0.98	-0.64	-0.18	0.09	0.42	0.67	0.94	1.23	1.62	2.08	2.43	3.21	0.69	1.06	0.13	3.14	0.04	0.12	0.21
0.2	1000	-1.58	-0.81	-0.48	-0.02	0.26	0.55	0.82	1.05	1.34	1.67	2.13	2.55	3.22	0.82	1.01	0.07	2.97	0.05	0.13	0.21
0.2	1500	-1.54	-0.70	-0.29	0.13	0.44	0.71	1.04	1.26	1.52	1.88	2.34	2.65	3.37	1.00	1.04	-0.05	3.13	0.06	0.19	0.27
0.2	2000	-1.03	-0.46	-0.16	0.28	0.62	0.90	1.16	1.45	1.70	2.03	2.50	2.76	3.53	1.17	1.01	-0.01	2.83	0.08	0.23	0.32
0.2	2500	-0.85	-0.29	0.00	0.44	0.75	0.99	1.29	1.52	1.79	2.12	2.57	2.92	3.79	1.28	1.00	0.15	3.03	0.10	0.25	0.35
0.2	5000	-0.52	0.20	0.54	1.01	1.25	1.53	1.75	2.01	2.33	2.63	3.07	3.49	3.98	1.80	0.99	0.05	2.95	0.22	0.42	0.55
0.2	7500	0.09	0.56	0.92	1.37	1.65	1.98	2.21	2.45	2.70	3.02	3.35	3.73	4.59	2.19	0.96	0.02	2.97	0.35	0.61	0.70
0.2	10000	0.06	0.89	1.25	1.67	1.99	2.26	2.54	2.78	3.05	3.35	3.81	4.07	4.86	2.52	1.00	0.02	3.13	0.49	0.71	0.81
0.2	15000	0.57	1.37	1.67	2.20	2.55	2.81	3.09	3.31	3.59	3.89	4.39	4.74	5.40	3.05	1.02	-0.05	2.88	0.69	0.85	0.90
0.2	20000	1.07	1.83	2.23	2.66	2.99	3.25	3.48	3.71	4.01	4.38	4.88	5.25	5.83	3.51	1.02	0.03	2.89	0.83	0.93	0.97
0.2	30000	1.27	1.87	2.26	2.68	2.99	3.28	3.51	3.76	4.06	4.39	4.80	5.02	5.58	3.51	0.97	-0.07	2.74	0.83	0.94	0.97
0.2	40000	1.07	1.83	2.23	2.66	3.00	3.25	3.48	3.71	4.01	4.38	4.88	5.25	5.83	3.51	1.02	0.03	2.89	0.83	0.93	0.97
0.3	100	-2.82	-1.53	-0.98	-0.51	-0.16	0.08	0.30	0.56	0.85	1.32	1.97	2.82	5.70	0.44	1.59	-0.57	18.65	0.08	0.13	0.19
0.3	150	-2.55	-1.45	-0.91	-0.44	-0.12	0.18	0.42	0.70	0.98	1.36	2.00	2.66	5.91	0.51	1.46	0.43	15.46	0.06	0.13	0.18
0.3	200	-1.87	-1.12	-0.77	-0.34	-0.05	0.24	0.47	0.74	1.09	1.46	2.08	2.74	4.05	0.62	1.34	3.17	37.39	0.06	0.12	0.19
0.3	250	-1.53	-0.96	-0.59	-0.22	0.05	0.28	0.53	0.82	1.15	1.52	2.19	2.73	4.44	0.69	1.21	1.49	11.01	0.06	0.13	0.18
0.3	500	-1.50	-0.81	-0.43	-0.06	0.28	0.54	0.85	1.14	1.46	1.80	2.33	2.80	3.82	0.90	1.12	0.36	3.92	0.07	0.16	0.25
0.3	750	-1.28	-0.65	-0.32	0.15	0.46	0.71	1.01	1.30	1.60	1.99	2.47	3.00	3.80	1.06	1.10	0.27	3.30	0.08	0.21	0.29
0.3	1000	-1.19	-0.47	-0.07	0.36	0.66	0.94	1.21	1.48	1.77	2.10	2.56	2.96	3.75	1.23	1.04	0.11	2.99	0.10	0.24	0.35
0.3	1500	-1.04	-0.19	0.19	0.60	0.93	1.24	1.50	1.76	2.03	2.43	2.83	3.29	4.14	1.51	1.07	0.02	3.16	0.16	0.33	0.45
0.3	2000	-0.59	0.04	0.42	0.85	1.17	1.47	1.75	2.04	2.31	2.59	3.10	3.44	4.17	1.75	1.03	0.02	2.93	0.21	0.43	0.54

Table 7 (cont.): Finite sample behavior of the z_ρ - parameter test concerning the non-observable interactions in the case of r>0 and $\beta 2=1$

		percentiles																p-value		
r	T	0.01	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99	mean	std.	skew.	kurt.	0.01	0.05	0.1
0.4	100	-2.81	-1.31	-0.81	-0.37	-0.07	0.18	0.45	0.70	1.01	1.51	2.26	3.26	6.50	0.61	1.65	-0.26	17.31	0.09	0.16	0.20
0.4	150	-2.01	-1.23	-0.75	-0.27	0.03	0.32	0.56	0.86	1.16	1.51	2.21	3.12	6.33	0.73	1.77	2.89	68.17	0.07	0.14	0.20
0.4	200	-1.80	-1.00	-0.62	-0.19	0.13	0.41	0.67	0.96	1.28	1.65	2.36	3.15	5.05	0.83	1.45	3.56	40.80	0.08	0.16	0.22
0.4	250	-1.40	-0.81	-0.44	-0.03	0.24	0.48	0.77	1.04	1.38	1.79	2.41	3.06	4.63	0.92	1.31	2.63	27.92	0.09	0.16	0.23
0.4	500	-1.13	-0.56	-0.22	0.21	0.57	0.85	1.15	1.47	1.77	2.14	2.69	3.23	4.32	1.20	1.15	0.37	3.70	0.11	0.25	0.35
0.4	750	-0.85	-0.34	0.00	0.51	0.79	1.07	1.41	1.68	1.98	2.39	2.91	3.40	4.20	1.44	1.14	0.29	3.22	0.16	0.31	0.41
0.4	1000	-0.67	-0.08	0.32	0.71	1.10	1.36	1.63	1.92	2.23	2.55	3.01	3.47	4.31	1.66	1.07	0.15	3.03	0.19	0.38	0.50
0.4	1500	-0.69	0.29	0.60	1.09	1.44	1.75	2.01	2.26	2.60	2.95	3.43	3.86	4.77	2.02	1.10	0.04	3.09	0.31	0.53	0.62
0.4	2000	-0.04	0.62	0.98	1.44	1.77	2.04	2.31	2.64	2.90	3.20	3.70	4.10	4.88	2.33	1.06	0.07	3.03	0.42	0.63	0.74
0.5	100	-2.48	-1.06	-0.69	-0.23	0.07	0.28	0.55	0.85	1.16	1.62	2.48	3.32	5.40	0.76	1.50	0.99	8.17	0.11	0.18	0.22
0.5	150	-1.74	-1.02	-0.63	-0.14	0.22	0.46	0.72	1.04	1.35	1.74	2.54	3.45	7.56	0.98	2.23	11.31	229.04	0.10	0.17	0.23
0.5	200	-1.60	-0.73	-0.45	0.00	0.29	0.57	0.82	1.17	1.51	1.93	2.65	3.42	5.49	1.03	1.41	1.93	14.00	0.11	0.19	0.27
0.5	250	-1.19	-0.59	-0.22	0.16	0.41	0.69	0.94	1.23	1.59	1.99	2.70	3.36	5.35	1.13	1.29	1.35	8.11	0.11	0.21	0.29
0.5	500	-0.90	-0.25	0.08	0.49	0.87	1.14	1.43	1.78	2.10	2.46	3.04	3.58	4.80	1.53	1.20	0.64	4.71	0.17	0.34	0.44
0.5	750	-0.63	0.03	0.39	0.83	1.13	1.47	1.78	2.04	2.38	2.81	3.38	3.90	4.89	1.83	1.19	0.47	3.72	0.25	0.44	0.55
0.5	1000	-0.22	0.34	0.76	1.17	1.50	1.77	2.06	2.37	2.71	3.01	3.52	4.08	4.90	2.11	1.11	0.25	3.16	0.34	0.53	0.63
0.5	1500	-0.25	0.78	1.11	1.59	1.90	2.24	2.54	2.82	3.16	3.56	4.04	4.54	5.53	2.57	1.15	0.16	3.11	0.48	0.69	0.79
0.5	2000	0.46	1.17	1.55	2.05	2.34	2.61	2.98	3.26	3.54	3.90	4.38	4.83	5.77	2.97	1.11	0.16	3.11	0.62	0.82	0.89
0.5	2500	0.99	1.57	1.88	2.33	2.69	2.94	3.28	3.54	3.83	4.22	4.74	5.13	6.19	3.30	1.11	0.28	3.18	0.74	0.89	0.94
0.5	5000	2.20	2.88	3.27	3.71	4.08	4.34	4.59	4.91	5.21	5.59	6.11	6.53	7.31	4.65	1.10	0.12	2.91	0.98	1.00	1.00
0.5	7500	3.26	3.85	4.29	4.71	5.15	5.45	5.72	5.96	6.25	6.55	7.01	7.34	8.30	5.68	1.07	0.00	2.92	1.00	1.00	1.00
0.5	10000	4.01	4.74	5.12	5.58	5.97	6.29	6.54	6.83	7.11	7.45	7.92	8.34	9.33	6.54	1.11	0.12	3.20	1.00	1.00	1.00

Table 8: Finite sample behavior of the z_β2 - parameter test concerning the non-observable interactions in the case of r≥0 and β2=1

		percentiles																p-value		
r	**T**	**0.01**	**0.05**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**0.95**	**0.99**	**mean**	**std.**	**skew.**	**kurt.**	**0.01**	**0.05**	**0.1**
0.0	100	-2.08	-0.67	-0.15	0.28	0.57	0.84	1.13	1.41	1.70	2.16	3.06	3.66	5.69	1.27	1.47	0.30	8.07	0.15	0.25	0.33
0.0	150	-1.28	-0.21	0.16	0.60	0.89	1.17	1.40	1.67	2.04	2.48	3.15	3.93	5.80	1.59	1.39	1.59	10.10	0.19	0.32	0.41
0.0	200	-0.73	-0.04	0.31	0.76	1.05	1.30	1.61	1.90	2.21	2.63	3.36	3.99	5.21	1.73	1.24	0.78	4.84	0.21	0.37	0.48
0.0	250	-0.48	0.03	0.50	0.99	1.33	1.58	1.82	2.06	2.31	2.75	3.24	3.73	4.89	1.86	1.09	0.32	3.48	0.24	0.45	0.58
0.0	500	0.07	0.88	1.23	1.73	2.05	2.34	2.57	2.83	3.15	3.49	3.99	4.56	5.35	2.62	1.09	0.30	3.58	0.50	0.72	0.83
0.0	750	0.94	1.57	1.85	2.32	2.60	2.89	3.19	3.50	3.79	4.20	4.69	5.25	6.13	3.25	1.13	0.37	3.24	0.71	0.88	0.94
0.0	1000	1.40	2.07	2.44	2.85	3.15	3.39	3.65	3.94	4.27	4.63	5.13	5.63	6.39	3.73	1.06	0.23	2.93	0.86	0.96	0.98
0.0	1500	2.18	2.88	3.15	3.61	3.89	4.26	4.51	4.75	5.12	5.44	6.00	6.45	7.44	4.55	1.11	0.34	3.28	0.97	1.00	1.00
0.0	2000	2.95	3.55	3.94	4.38	4.70	4.93	5.18	5.48	5.77	6.11	6.60	7.05	7.80	5.24	1.04	0.18	3.12	0.99	1.00	1.00
0.0	2500	3.39	4.15	4.61	5.00	5.31	5.59	5.85	6.16	6.43	6.79	7.23	7.72	8.40	5.89	1.06	0.06	3.04	1.00	1.00	1.00
0.0	5000	6.02	6.64	6.97	7.37	7.71	8.02	8.30	8.58	8.83	9.19	9.67	10.07	10.73	8.30	1.05	0.09	2.75	1.00	1.00	1.00
0.0	7500	7.86	8.52	8.89	9.32	9.63	9.86	10.12	10.41	10.73	11.07	11.51	11.89	12.50	10.17	1.02	0.05	2.79	1.00	1.00	1.00
0.0	10000	9.20	10.09	10.43	10.85	11.16	11.45	11.72	11.99	12.26	12.61	13.06	13.48	14.49	11.74	1.05	0.08	3.16	1.00	1.00	1.00
0.1	100	-2.08	-0.67	-0.15	0.28	0.57	0.84	1.13	1.41	1.70	2.16	3.06	3.66	5.69	1.27	1.47	0.30	8.07	0.15	0.25	0.33
0.1	150	-1.28	-0.21	0.16	0.60	0.89	1.17	1.40	1.67	2.04	2.48	3.15	3.93	5.80	1.59	1.39	1.59	10.10	0.19	0.32	0.41
0.1	200	-0.78	-0.08	0.31	0.78	1.03	1.30	1.59	1.89	2.19	2.61	3.28	3.85	4.86	1.70	1.18	0.53	3.77	0.21	0.38	0.48
0.1	250	-0.42	0.07	0.46	0.98	1.29	1.58	1.80	2.03	2.30	2.71	3.19	3.68	4.89	1.84	1.08	0.30	3.57	0.24	0.43	0.57
0.1	500	0.12	0.95	1.23	1.71	2.05	2.30	2.54	2.76	3.14	3.50	3.97	4.46	5.21	2.60	1.08	0.31	3.67	0.48	0.72	0.82
0.1	750	0.88	1.52	1.84	2.29	2.57	2.86	3.16	3.46	3.75	4.16	4.62	5.16	6.20	3.22	1.12	0.39	3.41	0.69	0.87	0.94
0.1	1000	1.38	2.09	2.41	2.83	3.12	3.35	3.62	3.91	4.20	4.59	5.11	5.51	6.20	3.70	1.04	0.24	2.93	0.86	0.96	0.98
0.1	1500	2.19	2.86	3.11	3.55	3.92	4.23	4.45	4.72	5.05	5.37	5.95	6.30	7.30	4.51	1.08	0.31	3.23	0.97	1.00	1.00
0.1	2000	2.93	3.57	3.91	4.32	4.65	4.90	5.15	5.42	5.73	6.10	6.51	6.94	7.61	5.19	1.03	0.17	3.14	1.00	1.00	1.00
0.2	100	-1.93	-0.62	-0.14	0.28	0.59	0.85	1.09	1.37	1.63	2.07	2.78	3.45	5.08	1.21	1.28	0.32	4.98	0.13	0.24	0.32
0.2	150	-1.25	-0.21	0.16	0.59	0.89	1.18	1.39	1.65	1.95	2.47	3.01	3.70	5.52	1.53	1.30	0.87	9.44	0.18	0.30	0.41
0.2	200	-0.80	-0.06	0.31	0.79	1.06	1.30	1.60	1.87	2.14	2.59	3.23	3.77	4.96	1.70	1.23	1.83	19.71	0.20	0.37	0.49
0.2	250	-0.47	0.08	0.50	1.01	1.34	1.59	1.81	2.03	2.27	2.66	3.16	3.60	4.58	1.83	1.06	0.25	3.54	0.22	0.43	0.57
0.2	500	0.10	0.95	1.23	1.73	2.02	2.31	2.56	2.75	3.14	3.46	3.93	4.33	5.13	2.58	1.06	0.29	3.78	0.49	0.72	0.82
0.2	750	0.85	1.51	1.84	2.30	2.56	2.86	3.14	3.45	3.73	4.12	4.55	5.02	6.13	3.19	1.09	0.35	3.32	0.70	0.87	0.93
0.2	1000	1.42	2.08	2.42	2.79	3.10	3.35	3.60	3.86	4.16	4.53	5.01	5.45	6.16	3.67	1.02	0.25	2.91	0.86	0.97	0.98
0.2	1500	2.14	2.82	3.13	3.52	3.90	4.19	4.43	4.68	4.99	5.32	5.84	6.23	7.12	4.46	1.05	0.29	3.20	0.97	1.00	1.00
0.2	2000	2.86	3.51	3.89	4.26	4.65	4.87	5.09	5.34	5.66	6.00	6.49	6.82	7.42	5.15	1.00	0.15	3.07	1.00	1.00	1.00
0.2	2500	3.40	4.15	4.48	4.94	5.25	5.48	5.76	6.03	6.30	6.60	7.06	7.48	8.21	5.78	1.01	0.06	3.06	1.00	1.00	1.00
0.2	5000	5.96	6.54	6.84	7.30	7.61	7.88	8.14	8.42	8.68	9.02	9.48	9.79	10.58	8.16	1.00	0.07	2.77	1.00	1.00	1.00
0.2	7500	7.74	8.39	8.75	9.19	9.48	9.73	9.94	10.22	10.51	10.83	11.29	11.62	12.27	9.99	0.97	0.04	2.82	1.00	1.00	1.00
0.2	10000	9.17	9.99	10.30	10.69	11.00	11.27	11.54	11.77	12.05	12.35	12.80	13.22	14.10	11.53	1.00	0.09	3.20	1.00	1.00	1.00
0.2	15000	11.79	12.47	12.80	13.25	13.57	13.87	14.12	14.39	14.67	15.02	15.56	15.93	16.46	14.15	1.04	0.07	2.73	1.00	1.00	1.00
0.2	20000	14.05	14.71	14.98	15.45	15.79	16.06	16.34	16.63	16.87	17.18	17.71	18.05	18.80	16.34	1.02	0.07	2.78	1.00	1.00	1.00
0.2	30000	14.19	14.79	15.09	15.50	15.80	16.05	16.27	16.57	16.90	17.21	17.64	18.00	18.68	16.35	0.99	0.16	3.02	1.00	1.00	1.00
0.2	40000	14.05	14.71	14.98	15.45	15.79	16.06	16.34	16.63	16.87	17.18	17.71	18.06	18.80	16.35	1.05	0.48	5.75	1.00	1.00	1.00

Table 8 (cont.): Finite sample behavior of the z_β2 - parameter test concerning the non-observable interactions in the case of r≥0 and β2=1

		percentiles																p-value		
r	**T**	**0.01**	**0.05**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**0.95**	**0.99**	**mean**	**std.**	**skew.**	**kurt.**	**0.01**	**0.05**	**0.1**
0.3	100	-2.39	-0.64	-0.13	0.28	0.60	0.83	1.11	1.37	1.67	2.01	2.69	3.42	4.69	1.18	1.27	-0.06	6.11	0.12	0.22	0.32
0.3	150	-1.24	-0.20	0.11	0.59	0.87	1.14	1.38	1.64	1.92	2.44	2.94	3.69	4.95	1.49	1.26	0.64	10.01	0.17	0.29	0.40
0.3	200	-0.70	-0.05	0.30	0.78	1.06	1.28	1.56	1.86	2.14	2.54	3.13	3.71	4.66	1.66	1.14	0.31	4.16	0.19	0.36	0.48
0.3	250	-0.51	0.07	0.52	1.00	1.30	1.59	1.79	2.01	2.24	2.64	3.10	3.57	4.60	1.81	1.04	0.15	3.83	0.21	0.43	0.57
0.3	500	0.11	0.90	1.22	1.70	1.99	2.31	2.51	2.77	3.08	3.40	3.88	4.27	5.14	2.56	1.05	0.27	3.84	0.47	0.71	0.82
0.3	750	0.93	1.56	1.83	2.27	2.56	2.83	3.10	3.40	3.68	4.04	4.54	5.02	5.86	3.16	1.07	0.32	3.32	0.70	0.88	0.94
0.3	1000	1.39	2.12	2.42	2.78	3.09	3.31	3.56	3.81	4.13	4.47	4.98	5.34	6.13	3.64	1.00	0.25	2.92	0.87	0.96	0.98
0.3	1500	2.10	2.81	3.13	3.51	3.88	4.16	4.38	4.66	4.97	5.27	5.81	6.09	7.09	4.43	1.03	0.25	3.22	0.97	1.00	1.00
0.3	2000	2.91	3.49	3.88	4.26	4.61	4.86	5.05	5.32	5.61	5.96	6.40	6.74	7.36	5.11	0.98	0.14	3.03	1.00	1.00	1.00
0.4	100	-1.93	-0.59	-0.14	0.27	0.60	0.82	1.10	1.34	1.61	1.96	2.58	3.32	4.68	1.16	1.26	-0.04	6.31	0.11	0.21	0.30
0.4	150	-1.17	-0.19	0.15	0.57	0.86	1.13	1.37	1.62	1.90	2.33	2.92	3.49	4.56	1.45	1.17	-0.08	5.84	0.17	0.28	0.39
0.4	200	-0.79	-0.07	0.30	0.79	1.04	1.30	1.58	1.81	2.10	2.53	3.11	3.51	4.64	1.64	1.13	0.05	5.61	0.19	0.35	0.47
0.4	250	-0.41	0.08	0.48	1.00	1.29	1.56	1.78	1.99	2.23	2.57	3.08	3.55	4.65	1.79	1.03	0.08	4.27	0.20	0.42	0.57
0.4	500	0.09	0.94	1.21	1.70	1.99	2.27	2.51	2.74	3.04	3.36	3.88	4.28	5.10	2.54	1.03	0.28	3.78	0.47	0.71	0.81
0.4	750	0.97	1.56	1.84	2.25	2.55	2.83	3.09	3.40	3.65	4.01	4.50	4.95	5.65	3.14	1.05	0.29	3.21	0.69	0.87	0.93
0.4	1000	1.37	2.13	2.37	2.79	3.06	3.29	3.54	3.81	4.10	4.44	4.95	5.32	6.05	3.62	0.98	0.24	2.97	0.86	0.97	0.98
0.4	1500	2.15	2.79	3.16	3.51	3.86	4.13	4.40	4.64	4.93	5.20	5.74	6.10	6.91	4.41	1.02	0.23	3.21	0.97	0.99	1.00
0.4	2000	2.89	3.50	3.89	4.30	4.58	4.82	5.04	5.30	5.57	5.89	6.32	6.66	7.35	5.09	0.96	0.14	3.13	1.00	1.00	1.00
0.5	100	-3.00	-0.62	-0.08	0.26	0.60	0.82	1.08	1.33	1.58	1.90	2.51	3.07	4.10	1.10	1.21	-0.72	7.76	0.11	0.20	0.29
0.5	150	-1.19	-0.15	0.16	0.60	0.87	1.14	1.34	1.59	1.89	2.31	2.85	3.45	4.51	1.44	1.18	0.32	9.53	0.16	0.28	0.38
0.5	200	-0.90	-0.07	0.33	0.77	1.05	1.30	1.59	1.82	2.09	2.47	3.03	3.48	4.53	1.61	1.10	-0.10	5.92	0.19	0.35	0.48
0.5	250	-0.45	0.12	0.53	1.01	1.31	1.56	1.79	1.99	2.21	2.56	3.00	3.41	4.19	1.78	1.00	-0.05	4.47	0.20	0.42	0.55
0.5	500	0.13	0.98	1.26	1.70	2.03	2.27	2.52	2.69	2.97	3.31	3.82	4.14	4.98	2.52	1.00	0.19	3.75	0.46	0.73	0.81
0.5	750	0.88	1.55	1.87	2.28	2.52	2.81	3.07	3.37	3.66	3.96	4.39	4.85	5.61	3.12	1.02	0.26	3.28	0.69	0.88	0.94
0.5	1000	1.40	2.13	2.42	2.80	3.03	3.29	3.52	3.78	4.08	4.42	4.86	5.24	5.98	3.59	0.96	0.23	3.06	0.86	0.97	0.98
0.5	1500	2.05	2.77	3.16	3.50	3.86	4.11	4.40	4.63	4.88	5.15	5.71	6.08	6.94	4.39	1.00	0.19	3.13	0.97	0.99	1.00
0.5	2000	2.78	3.55	3.87	4.28	4.56	4.77	5.04	5.29	5.56	5.87	6.29	6.63	7.32	5.06	0.95	0.12	3.15	1.00	1.00	1.00
0.5	2500	3.45	4.11	4.43	4.89	5.22	5.43	5.66	5.92	6.17	6.46	7.00	7.35	8.02	5.69	0.97	0.04	3.06	1.00	1.00	1.00
0.5	5000	6.03	6.46	6.76	7.18	7.49	7.77	8.02	8.29	8.51	8.87	9.23	9.60	10.26	8.02	0.96	0.06	2.77	1.00	1.00	1.00
0.5	7500	7.74	8.33	8.67	9.05	9.31	9.58	9.80	10.04	10.31	10.61	11.09	11.36	11.98	9.83	0.92	0.02	2.86	1.00	1.00	1.00
0.5	10000	9.13	9.82	10.15	10.55	10.84	11.09	11.35	11.58	11.81	12.15	12.60	12.96	13.70	11.35	0.95	0.10	3.19	1.00	1.00	1.00

Table 9: **Finite sample behavior of the z_β2 - parameter test concerning the non-observable interactions in the case of r≥0 and β2=1 and additional explanatory variables**

		percentiles																	p-value		
r	**T**	**0.01**	**0.05**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**0.95**	**0.99**	**mean**	**std.**	**skew.**	**kurt.**	**0.01**	**0.05**	**0.1**
0.0	100	-0.32	0.20	0.54	1.00	1.33	1.54	1.79	2.04	2.37	2.78	3.34	4.12	6.06	1.94	1.21	0.91	4.89	0.25	0.43	0.56
0.0	150	0.28	0.77	1.11	1.47	1.79	2.03	2.32	2.54	2.86	3.26	3.91	4.68	6.46	2.44	1.19	1.08	5.69	0.39	0.63	0.76
0.0	200	0.61	1.15	1.51	1.86	2.16	2.43	2.70	2.97	3.29	3.70	4.22	4.86	6.36	2.81	1.17	0.78	4.67	0.54	0.77	0.87
0.0	250	0.90	1.57	1.92	2.23	2.51	2.79	3.05	3.29	3.57	4.00	4.50	5.04	6.17	3.13	1.08	0.65	4.45	0.68	0.89	0.94
0.0	500	2.25	2.73	3.15	3.59	3.94	4.18	4.46	4.72	4.97	5.30	5.81	6.23	7.03	4.47	1.04	0.16	3.23	0.96	0.99	1.00
0.0	750	3.28	3.88	4.18	4.64	5.00	5.24	5.49	5.74	6.05	6.38	6.90	7.29	8.18	5.53	1.04	0.22	3.14	1.00	1.00	1.00
0.0	1000	4.04	4.72	5.10	5.60	5.90	6.13	6.38	6.63	6.92	7.22	7.68	8.18	8.91	6.40	1.02	0.09	3.22	1.00	1.00	1.00
0.0	1500	5.40	6.01	6.50	6.96	7.26	7.55	7.76	8.06	8.34	8.65	9.17	9.59	10.28	7.81	1.05	0.11	3.12	1.00	1.00	1.00
0.0	2000	6.63	7.43	7.77	8.22	8.57	8.82	9.04	9.26	9.55	9.95	10.40	10.78	11.35	9.07	1.02	0.05	3.15	1.00	1.00	1.00
0.0	2500	7.72	8.38	8.75	9.23	9.57	9.87	10.13	10.39	10.71	11.10	11.53	11.86	12.61	10.14	1.06	0.06	2.81	1.00	1.00	1.00
0.0	5000	11.98	12.70	13.03	13.43	13.74	14.05	14.33	14.59	14.91	15.23	15.84	16.15	16.69	14.35	1.07	0.05	2.97	1.00	1.00	1.00
0.0	7500	15.32	15.98	16.27	16.70	17.02	17.30	17.51	17.77	18.09	18.43	18.91	19.26	20.33	17.57	1.03	0.24	3.29	1.00	1.00	1.00
0.0	10000	17.85	18.54	18.91	19.42	19.71	19.96	20.23	20.53	20.80	21.24	21.70	22.11	22.86	20.28	1.08	0.08	2.81	1.00	1.00	1.00
0.1	100	-0.32	0.20	0.54	1.00	1.33	1.54	1.79	2.04	2.37	2.78	3.34	4.12	6.06	1.94	1.21	0.91	4.89	0.25	0.43	0.56
0.1	150	0.28	0.77	1.11	1.47	1.79	2.03	2.32	2.54	2.86	3.26	3.91	4.68	6.46	2.44	1.19	1.08	5.69	0.39	0.63	0.76
0.1	200	0.55	1.13	1.44	1.84	2.16	2.42	2.68	2.96	3.27	3.69	4.22	4.92	6.25	2.80	1.17	0.75	4.59	0.53	0.76	0.86
0.1	250	0.77	1.53	1.86	2.22	2.48	2.78	3.00	3.30	3.58	3.98	4.51	4.97	6.18	3.12	1.10	0.76	5.09	0.67	0.89	0.94
0.1	500	2.11	2.71	3.11	3.60	3.91	4.14	4.43	4.67	4.96	5.28	5.83	6.28	7.18	4.45	1.06	0.21	3.34	0.97	0.99	1.00
0.1	750	3.26	3.85	4.17	4.59	4.95	5.21	5.48	5.73	6.04	6.38	6.88	7.24	8.32	5.51	1.06	0.23	3.10	1.00	1.00	1.00
0.1	1000	3.95	4.63	5.06	5.55	5.84	6.09	6.37	6.64	6.91	7.21	7.68	8.16	8.91	6.38	1.04	0.09	3.25	1.00	1.00	1.00
0.1	1500	5.35	5.97	6.45	6.93	7.22	7.48	7.75	8.04	8.31	8.63	9.15	9.55	10.20	7.77	1.05	0.09	3.05	1.00	1.00	1.00
0.1	2000	6.56	7.40	7.72	8.18	8.53	8.78	8.99	9.23	9.51	9.90	10.36	10.76	11.40	9.02	1.02	0.04	3.13	1.00	1.00	1.00
0.2	100	-0.55	0.18	0.52	0.97	1.27	1.50	1.76	2.04	2.38	2.77	3.41	4.12	5.62	1.91	1.21	0.64	4.16	0.24	0.42	0.54
0.2	150	0.13	0.73	1.06	1.43	1.76	2.02	2.27	2.52	2.83	3.22	3.98	4.66	6.03	2.41	1.21	0.96	5.09	0.38	0.63	0.74
0.2	200	0.36	1.09	1.44	1.81	2.14	2.40	2.63	2.94	3.29	3.68	4.24	4.90	6.27	2.78	1.20	0.83	5.11	0.53	0.75	0.84
0.2	250	0.88	1.52	1.88	2.18	2.46	2.72	3.01	3.26	3.55	3.95	4.53	5.01	6.26	3.10	1.11	0.68	4.36	0.66	0.87	0.93
0.2	500	2.07	2.68	3.04	3.58	3.86	4.14	4.38	4.65	4.92	5.27	5.79	6.23	7.38	4.41	1.06	0.22	3.30	0.96	0.99	1.00
0.2	750	3.29	3.85	4.15	4.53	4.89	5.13	5.44	5.68	5.99	6.30	6.84	7.23	8.22	5.46	1.06	0.29	3.21	1.00	1.00	1.00
0.2	1000	3.90	4.62	5.01	5.46	5.77	6.04	6.29	6.57	6.87	7.17	7.58	8.03	9.11	6.32	1.04	0.11	3.20	1.00	1.00	1.00
0.2	1500	5.22	5.99	6.33	6.82	7.16	7.42	7.67	7.92	8.20	8.56	9.07	9.42	10.31	7.69	1.05	0.11	3.09	1.00	1.00	1.00
0.2	2000	6.44	7.29	7.63	8.11	8.41	8.68	8.93	9.14	9.41	9.77	10.24	10.67	11.34	8.93	1.02	0.03	3.09	1.00	1.00	1.00
0.2	2500	7.57	8.22	8.54	9.06	9.40	9.73	10.00	10.25	10.56	10.90	11.37	11.72	12.50	9.99	1.07	0.06	2.95	1.00	1.00	1.00
0.2	5000	11.67	12.40	12.74	13.17	13.51	13.81	14.10	14.38	14.69	15.05	15.53	15.95	16.54	14.11	1.08	0.04	2.94	1.00	1.00	1.00
0.2	7500	14.96	15.61	15.97	16.43	16.73	17.01	17.24	17.48	17.82	18.16	18.61	19.05	19.96	17.29	1.05	0.24	3.31	1.00	1.00	1.00
0.2	10000	17.52	18.19	18.58	19.00	19.35	19.62	19.90	20.20	20.50	20.94	21.41	21.80	22.57	19.95	1.10	0.11	2.80	1.00	1.00	1.00
0.2	15000	22.10	22.67	23.13	23.57	23.88	24.15	24.44	24.70	25.06	25.43	25.89	26.21	26.81	24.46	1.08	0.03	3.12	1.00	1.00	1.00
0.2	20000	25.92	26.54	26.91	27.39	27.72	28.06	28.29	28.50	28.78	29.17	29.65	29.98	30.60	28.27	1.04	0.04	2.93	1.00	1.00	1.00
0.2	30000	26.08	26.58	26.89	27.35	27.69	28.02	28.29	28.53	28.83	29.15	29.63	30.05	30.85	28.28	1.07	0.41	4.36	1.00	1.00	1.00
0.2	40000	25.92	26.54	26.91	27.39	27.72	28.06	28.29	28.50	28.78	29.17	29.68	29.99	30.71	28.29	1.12	1.52	19.64	1.00	1.00	1.00

Table 9 (cont.): Finite sample behavior of the z_β2 - parameter test concerning the non-observable interactions in the case of r≥0 and β2=1 and additional explanatory variables

		percentiles																p-value		
r	T	0.01	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99	mean	std.	skew.	kurt.	0.01	0.05	0.1
0.3	100	-0.45	0.14	0.47	0.93	1.23	1.49	1.73	2.00	2.34	2.77	3.50	4.21	5.62	1.89	1.23	0.69	3.96	0.24	0.41	0.53
0.3	150	0.13	0.66	0.98	1.39	1.68	1.95	2.30	2.51	2.82	3.20	3.90	4.71	5.88	2.37	1.20	0.79	3.99	0.38	0.60	0.72
0.3	200	0.36	1.08	1.34	1.78	2.09	2.34	2.60	2.87	3.22	3.63	4.25	4.88	6.27	2.74	1.22	0.94	5.96	0.51	0.74	0.83
0.3	250	0.87	1.46	1.76	2.11	2.38	2.68	2.97	3.22	3.52	3.88	4.48	4.94	6.33	3.05	1.12	0.69	4.33	0.64	0.85	0.93
0.3	500	1.89	2.61	2.94	3.47	3.77	4.03	4.33	4.59	4.85	5.18	5.74	6.14	7.38	4.34	1.08	0.24	3.27	0.95	0.99	1.00
0.3	750	3.13	3.70	3.99	4.43	4.80	5.03	5.34	5.61	5.89	6.23	6.77	7.18	8.28	5.37	1.07	0.31	3.24	1.00	1.00	1.00
0.3	1000	3.76	4.47	4.83	5.34	5.66	5.93	6.20	6.46	6.75	7.07	7.50	7.95	8.76	6.21	1.04	0.09	3.08	1.00	1.00	1.00
0.3	1500	5.05	5.85	6.24	6.69	7.02	7.30	7.53	7.77	8.07	8.47	8.95	9.31	10.11	7.56	1.05	0.09	3.06	1.00	1.00	1.00
0.3	2000	6.34	7.13	7.47	7.98	8.24	8.50	8.75	9.00	9.24	9.63	10.10	10.50	11.17	8.77	1.02	0.03	3.10	1.00	1.00	1.00
0.4	100	-0.62	0.07	0.41	0.90	1.18	1.43	1.72	2.00	2.38	2.79	3.47	4.43	5.86	1.88	1.28	0.78	4.06	0.24	0.41	0.53
0.4	150	0.08	0.63	0.94	1.36	1.64	1.89	2.21	2.45	2.78	3.19	3.87	4.64	6.11	2.33	1.22	0.87	4.41	0.36	0.58	0.70
0.4	200	0.31	0.93	1.29	1.69	2.02	2.29	2.55	2.83	3.15	3.56	4.30	4.82	6.37	2.69	1.25	1.05	6.17	0.48	0.72	0.81
0.4	250	0.75	1.36	1.65	2.05	2.33	2.59	2.90	3.18	3.49	3.81	4.41	5.04	6.46	2.99	1.18	1.25	8.83	0.61	0.83	0.90
0.4	500	1.91	2.46	2.82	3.36	3.69	3.94	4.21	4.48	4.75	5.09	5.69	6.12	7.23	4.25	1.10	0.30	3.36	0.94	0.99	0.99
0.4	750	2.98	3.54	3.85	4.30	4.64	4.91	5.22	5.51	5.78	6.11	6.69	7.07	8.24	5.25	1.10	0.35	3.27	1.00	1.00	1.00
0.4	1000	3.63	4.36	4.71	5.19	5.53	5.78	6.03	6.29	6.62	6.95	7.39	7.86	8.72	6.06	1.06	0.12	3.13	1.00	1.00	1.00
0.4	1500	4.81	5.60	6.03	6.52	6.85	7.08	7.38	7.61	7.90	8.31	8.76	9.18	10.04	7.39	1.07	0.09	3.11	1.00	1.00	1.00
0.4	2000	6.05	6.92	7.30	7.75	8.02	8.29	8.54	8.80	9.05	9.44	9.94	10.33	11.04	8.57	1.03	0.03	3.15	1.00	1.00	1.00
0.5	100	-0.98	0.04	0.40	0.81	1.14	1.39	1.64	1.94	2.29	2.72	3.44	4.34	5.54	1.82	1.28	0.62	3.83	0.23	0.39	0.50
0.5	150	0.00	0.60	0.87	1.25	1.57	1.82	2.14	2.44	2.74	3.17	3.92	4.77	6.84	2.31	1.32	1.17	5.76	0.35	0.56	0.67
0.5	200	0.13	0.84	1.19	1.58	1.96	2.20	2.50	2.76	3.10	3.55	4.25	5.01	6.96	2.63	1.28	1.06	6.18	0.48	0.70	0.78
0.5	250	0.64	1.24	1.58	1.96	2.24	2.51	2.80	3.08	3.42	3.84	4.43	4.99	6.51	2.92	1.18	0.77	4.37	0.57	0.80	0.89
0.5	500	1.81	2.34	2.69	3.19	3.53	3.84	4.07	4.32	4.67	5.03	5.62	6.03	7.12	4.13	1.13	0.36	3.35	0.91	0.98	0.99
0.5	750	2.79	3.34	3.67	4.12	4.47	4.74	5.07	5.32	5.65	6.00	6.58	7.02	8.05	5.10	1.12	0.37	3.43	0.99	1.00	1.00
0.5	1000	3.36	4.14	4.52	5.01	5.33	5.58	5.84	6.09	6.45	6.77	7.31	7.71	8.65	5.88	1.08	0.15	3.16	1.00	1.00	1.00
0.5	1500	4.66	5.35	5.76	6.29	6.59	6.86	7.12	7.38	7.69	8.09	8.59	9.03	9.67	7.16	1.09	0.12	3.02	1.00	1.00	1.00
0.5	2000	5.95	6.63	7.01	7.49	7.77	8.02	8.28	8.54	8.79	9.22	9.68	10.10	10.75	8.31	1.04	0.05	3.17	1.00	1.00	1.00
0.5	2500	6.65	7.43	7.92	8.34	8.74	8.99	9.32	9.58	9.91	10.25	10.72	11.11	11.89	9.31	1.11	0.05	2.95	1.00	1.00	1.00
0.5	5000	10.54	11.38	11.74	12.13	12.50	12.79	13.09	13.35	13.68	14.09	14.61	14.99	15.66	13.11	1.12	0.09	2.94	1.00	1.00	1.00
0.5	7500	13.70	14.35	14.75	15.17	15.49	15.76	16.02	16.29	16.61	16.97	17.42	17.88	18.72	16.07	1.07	0.21	3.17	1.00	1.00	1.00
0.5	10000	15.98	16.78	17.17	17.64	17.91	18.21	18.46	18.78	19.10	19.52	20.06	20.46	21.18	18.55	1.12	0.15	2.88	1.00	1.00	1.00

REVSTAT – Statistical Journal
Volume 7, Number 1, April 2009, 87–104

FILTERS FOR SHORT NONSTATIONARY SEQUENCES: THE ANALYSIS OF THE BUSINESS CYCLE

Author: D.S.G. Pollock
– Department of Economics, University of Leicester, United Kingdom
d.s.g.pollock@le.ac.uk

Abstract:

- This paper gives an account of some techniques of linear filtering which can be used for extracting the business cycle from economic data sequences of limited duration. It is argued that there can be no definitive definition of the business cycle. Both the definition of the business cycle and the methods that are used to extract it must be adapted to the purposes of the analysis; and different definitions may be appropriate to different eras.

Key-Words:

- *linear filters; spectral analysis; business cycles.*

AMS Subject Classification:

- 62M15, 93E11, 91B84.

1. INTRODUCTION

In recent years, there has been a renewed interest amongst economists in the business cycle. However, compared with the economic fluctuations of the nineteenth century, the business cycle in modern western economies has been a tenuous affair. For many years, minor fluctuations have been carried on the backs of strongly rising trends in national income. Their amplitudes have been so small in relative terms that they have rarely resulted in absolute reductions in the levels of aggregate income. Usually, they have succeeded only in slowing its upward progress.

Faced with this tenuous phenomenon, modern analysts have also had difficulties in reaching a consensus on how to define the business cycle and in agreeing on which methods should be used to extract it from macroeconomic data sequences. Thus, the difficulties have been both methodological and technical. This paper will deal with both of these aspects, albeit that the emphasis will be on technical matters.

It seems that many of the methodological difficulties are rooted in the tendency of economists to objectify the business cycle. If there is no doubt concerning the objective reality of a phenomenon, then it seems that it must be capable of a precise and an unequivocal definition.

However, the opinion that is offered in this paper is that it is fruitless to seek a definitive definition of the business cycle. The definition needs to be adapted to the purposes of the analysis in question; and it is arguable that it should also be influenced by the behaviour of the economy in the era that is studied.

It is also argued that a clear understanding of the business cycle can be achieved only in the light of its spectral analysis. However, the spectral approach entails considerable technical difficulties. The classical theory of statistical Fourier analysis deals with stationary stochastic sequences of unlimited duration. This accords well with the nature of the trigonometrical functions on which spectral analysis is based. In business cycle analysis, one is faced, by contrast, with macroeconomic sequences that are of strictly limited durations and that are liable to be strongly trended.

In order to apply the methods of spectral analysis to the macroeconomic data, two problems must be addressed. First, the data must be reduced to stationarity by an appropriate method of detrending. There are various ways of proceeding; and a judicious choice must be made. Then, there is the short duration of the data, which poses the problem acutely of how one should treat the ends of the sample.

One way of dealing with the end-of-sample problem is to create a circular sequence from the detrended data. By travelling around the circle indefinitely, the infinite periodic extension of the data sequence is generated, which is the essential object of an analysis that employs the discrete Fourier transform.

Such an analysis is liable to be undermined whenever there are radical disjunctions in the periodic extension at the points where the end of one replication joins the beginning of the next. Therefore, a successful Fourier analysis depends upon a careful detrending of the data. It seems that it was the neglect of this fact that led one renowned analyst to declare that spectral analysis was inappropriate to economic data. (See Granger 1966.)

2. INTERACTION OF TREND AND BUSINESS CYCLE

The business cycle has no fixed duration. In a Fourier analysis, it can be represented as a composite of sinusoidal motions of various frequencies that fall within some bandwidth. We shall consider one modern convention that defines the exact extent of this bandwidth; but it seems more appropriate that it should be determined in the light of the data.

If they are not allowed to overlap, it may be crucial to know where the low frequency range of the trend is deemed to end and where the higher range of the business cycle should begin. However, in this section, we shall avoid the issue by assuming that the business cycle is of a fixed frequency and that the trend is a simple exponential function.

In that case, the trend can be described by the function $T(t) = \exp\{rt\}$, where $r > 0$ is constant rate of growth. The business cycle, which serves to modulate the trend, is described by an exponentiated cosine function $C(t) = \exp\{\gamma \cos(\omega t)\}$. The product of the two functions, which can regarded as a model of the trajectory of aggregate income, is

$$Y(t) = \beta \exp\big\{rt + \gamma \cos(\omega t)\big\} . \tag{2.1}$$

The resulting business cycles, which are depicted in Figure 1, have an asymmetric appearance. Their contractions are of lesser duration than their expansions; and they become shorter as the growth rate r increases.

Eventually, when the rate exceeds a certain value, the periods of contraction will disappear and, in place of the local minima, there will be only points of inflection. In fact, the condition for the existence of local minima is that $\omega\gamma > r$, which is to say that the product of the amplitude of the cycles and their angular velocity must exceed the growth rate of the trend.

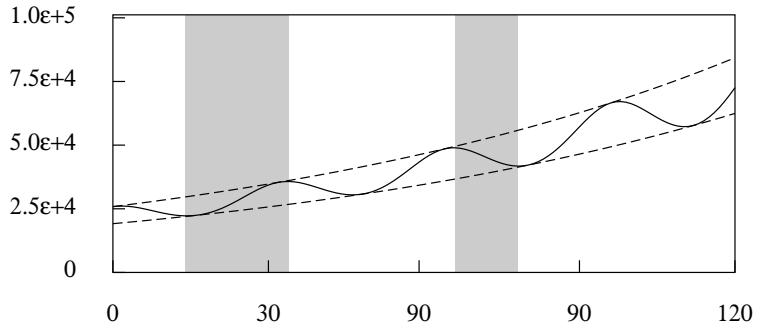

Figure 1: The function $Y(t) = \beta \exp\{rt + \gamma \cos(\omega t)\}$ as a model of the business cycle. Observe that, when $r > 0$, the duration of an expansion exceeds the duration of a contraction.

Next, we take logarithms of the data to obtain a model, represented in Figure 2, that has additive trend and cyclical components. This gives

$$\ln\big\{Y(t)\big\} \;=\; y(t) \;=\; \mu + r\,t + \gamma\cos(\omega\,t)\ , \tag{2.2}$$

where $\mu = \ln\{\beta\}$. Since logs effect a monotonic transformation, there is no displacement of the local maxima and minima. However, the amplitude of the fluctuations around the trend, which has become linear in the logs, is now constant.

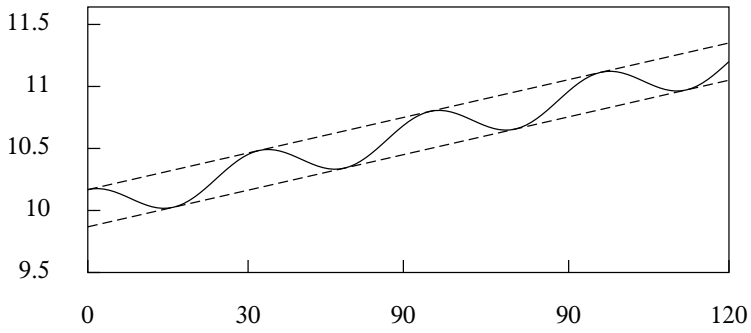

Figure 2: The function $\ln\{Y(t)\} = \ln\{\beta\} + rt + \gamma\cos(\omega t)$ representing the logarithmic business cycle data. The duration of the expansions and the contractions are not affected by the transformation.

The final step is to create a stationary function by eliminating the trend. There are two equivalent ways of doing this in the context of the schematic model. On the one hand, the linear trend $\xi(t) = \mu + r\,t$ can be subtracted from $y(t)$ to create the pure business cycle $\gamma\cos(\omega t)$.

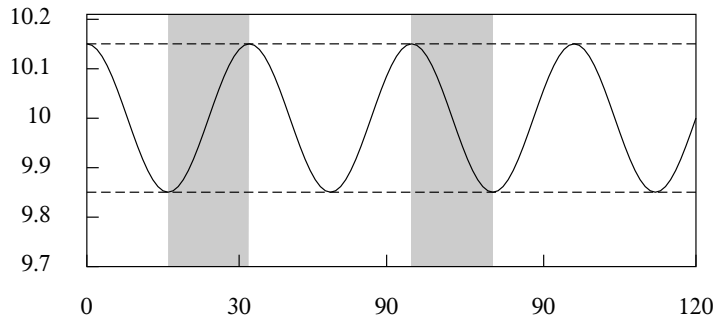

Figure 3: The function $\mu + \gamma\cos(\omega t)$ representing the detrended business cycle. The duration of the expansions and the contractions are equal.

Alternatively, the function $y(t)$ can be differentiated to give $dy(t)/dt = r - \gamma\,\omega\sin(\omega t)$. When the latter is adjusted by subtracting the growth rate r, by dividing by ω and by displacing its phase by $-\pi/2$ radians — which entails replacing the argument t by $t-\pi/2$ — we obtain the function $\gamma\cos(\omega t)$ again. Through the process of detrending, the phases of expansion and contraction acquire equal duration, and the asymmetry of the business cycle vanishes.

There is an enduring division of opinion, in the literature of economics, on whether we should be looking at the turning points and phase durations of the original data or at those of the detrended data. The task of finding the turning points is often a concern of analysts who wish to make international comparisons of the timing of the business cycle.

However, since the business cycle is a low-frequency component of the data, it is difficult to find the turning points with great accuracy. In fact, the pinnacles and pits that are declared to be the turning points often seem to be the products of whatever high-frequency components happen to remain in the data after they have been subjected to a process of seasonal adjustment.

If the objective is to compare the turning points of the cycles, then the trends should be eliminated from the data. The countries that might be compared are liable to be growing at differing rates. From the trended data, it will appear that those with higher rates of growth have shorter recessions with delayed onsets, and this can be misleading.

The various indices of an expanding economy will also grow at diverse rates. Unless they are reduced to a common basis by eliminating their trends, their fluctuations cannot be compared easily. Amongst such indices will be the percentage rate of unemployment, which constitutes a trend-stationary sequence. It would be difficult to collate the turning points in this index with those within a rapidly growing series of aggregate income, which might not exhibit any absolute reductions in its level. A trenchant opinion to the contrary, which opposes the practice of detrending the data for the purposes of describing the business cycle, has been offered by Harding and Pagan (2003).

3. BANDPASS DEFINITION OF THE BUSINESS CYCLE

The modern definition of the business cycle that has been alluded to in the previous section is that of a quasi cyclical motion comprising sinusoidal elements that have durations of no less than one-and-a-half years and not exceeding eight years.

This definition has been proposed by Baxter and King (1999) who have declared that it was the definition adopted by Burns and Mitchell (1947) in their study of the economic fluctuations in the U.S. in the late nineteenth century and in the early twentieth century. However, it is doubtful whether Burns and Mitchell were so firm in their definition of what constitutes the business cycle. It seems, instead, that they were merely speaking of what they had discerned in their data.

The definition in question suggests that the data should be filtered in order to extract the components that fall within the stated range, which is described as the pass band. Given a doubly infinite data sequence, this objective would be fulfilled, in theory, by an ideal bandpass filter comprising a doubly infinite sequence of coefficients.

The ideal bandpass filter that transmits all elements within the frequency range $[\alpha, \beta]$ and blocks all others has the following frequency response:

$$\psi(\omega) = \begin{cases} 1, & \text{if } |\omega| \in (\alpha, \beta)\,, \\ 0, & \text{otherwise}\,. \end{cases} \tag{3.1}$$

The coefficients of the corresponding time-domain filter are obtained by applying an inverse Fourier transform to this response to give

$$\psi_k = \int_{\alpha}^{\beta} e^{\mathrm{i}k\omega}\, d\omega \;=\; \frac{1}{\pi k}\left\{\sin(\beta k) - \sin(\alpha k)\right\}\,. \tag{3.2}$$

In practice, all data sequences are finite, and it is impossible to apply a filter that has an infinite number of coefficients. However, a practical filter may be obtained by selecting a limited number of the central coefficients of an ideal infinite-sample filter. In the case of a truncated filter based on $2q+1$ central coefficients, the elements of the filtered sequence are given by

$$\begin{aligned} x_t \;=\; & \psi_q\, y_{t-q} + \psi_{q-1}\, y_{t-q+1} + \cdots + \psi_1\, y_{t-1} + \psi_0\, y_t + \psi_1\, y_{t+1} \\ & + \cdots + \psi_{q-1}\, y_{t+q-1} + \psi_q\, y_{t+q}\,. \end{aligned} \tag{3.3}$$

Given a sample $y_0, y_1, ..., y_{T-1}$ of T data points, only $T - 2q$ processed values $x_q, x_{q+1}..., x_{T-q-1}$ are available, since the filter cannot reach the ends of the sample, unless it is extrapolated.

If the coefficients of the truncated bandpass or highpass filter are adjusted so that they sum to zero, then the z-transform polynomial $\psi(z)$ of the coefficient sequence will contain two roots of unit value. The adjustments may be made by subtracting $\sum_k \psi_k/(2q+1)$ from each coefficient. The sum of the adjusted coefficients is $\psi(1) = 0$, from which it follows that $1 - z$ is a factor of $\psi(z)$. The condition of symmetry, which is that $\psi(z) = \psi(z^{-1})$, implies that $1 - z^{-1}$ is also a factor. Thus the polynomial contains the factor

$$(1 - z)(1 - z^{-1}) = -z^{-1}(1 - z)^2 , \tag{3.4}$$

within which $\nabla^2(z) = (1 - z)^2$ corresponds to a twofold differencing operator.

Since it incorporates the factor $\nabla^2(z)$, the effect of applying the filter to a data sequence with a linear trend will be to produce an untrended sequence with a zero mean. The effect of applying it to a sequence with a quadratic trend will be to produce an untrended sequence with a nonzero mean.

The usual effect of the truncation will be to cause a considerable spectral leakage. Thus, if the filter is applied to trended data, then it is liable to transmit some powerful low-frequency elements that will give rise to cycles of high amplitudes within the filtered output. The divergence of the frequency response function from the ideal specification of (3.1) is illustrated in Figure 4.

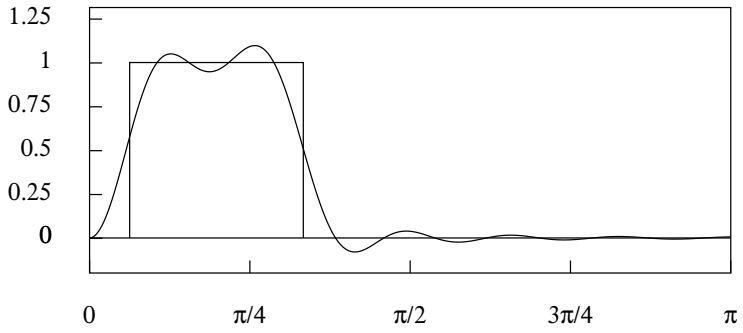

Figure 4: The frequency response of the truncated bandpass filter of 25 coefficients superimposed upon the ideal frequency response. The lower cut-off point is at $\pi/15$ radians (11.25°), corresponding to a period of 6 quarters, and the upper cut-off point is at $\pi/3$ radians (60°), corresponding to a period of the 32 quarters.

An indication of the effect of the truncated filter is provided by its application to a quarterly sequence of the logarithms of consumption in the U.K. that is illustrated in Figure 5. The filtered sequence is in Figure 6, where the loss of the data from the ends is indicated by the vertical lines.

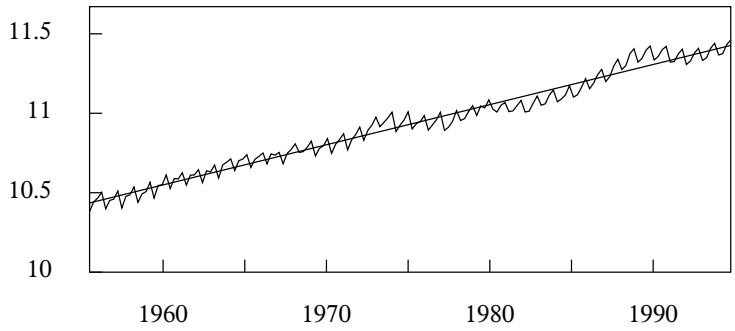

Figure 5: The quarterly sequence of the logarithms of consumption in the U.K., for the years 1955 to 1994, together with a linear trend interpolated by least-squares regression.

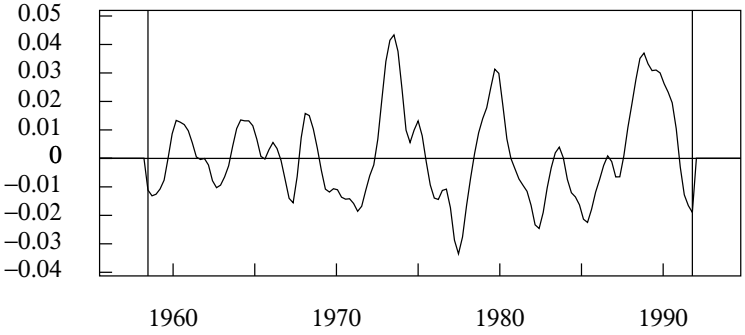

Figure 6: The sequence derived by applying the truncated bandpass filter of 25 coefficients to the quarterly logarithmic data on U.K. Consumption.

An alternative filter that is designed to reach the ends of the sample has been proposed by Christiano and Fitzgerald, (2003). The filter is described by the equation

$$\begin{aligned} x_t &= A y_0 + \psi_t y_0 + \cdots + \psi_1 y_{t-1} + \psi_0 y_t + \psi_1 y_{t+1} \\ &\quad + \cdots + \psi_{T-1-t} y_{T-1} + B y_{T-1} \ . \end{aligned} \tag{3.5}$$

This equation comprises the entire data sequence $y_0, ..., y_{T-1}$; and the value of t determines which of the coefficients of the infinite-sample filter are entailed in producing the current output. Thus, the value of x_0 is generated by looking forwards to the end of the sample, whereas the value of x_{T-1} is generated by looking backwards to the beginning of the sample.

If the process generating the data is stationary and of zero mean, then it is appropriate to set $A = B = 0$, which is tantamount to approximating the extra-sample elements by zeros. In the case of a data sequence that appears to follow a first-order random walk, it has been proposed to set A and B to the values of the sums of the coefficients that lie beyond the span of the data on either side.

Since the filter coefficients must sum to zero, it follows that

$$(3.6) \qquad A = -\left(\frac{1}{2}\psi_0 + \psi_1 + \cdots + \psi_t\right) \quad \text{and} \quad B = -\left(\frac{1}{2}\psi_0 + \psi_1 + \cdots + \psi_{T-t-1}\right).$$

The effect is tantamount to extending the sample at either end by constant sequences comprising the first and the last sample values respectively.

For data that have the appearance of having been generated by a first-order random walk with a constant drift, it is appropriate to extract a linear trend before filtering the residual sequence. In fact, this has proved to be the usual practice in most circumstances.

It has been proposed to subtract from the data a linear function $f(t) = \alpha + \beta t$ interpolated through the first and the final data points, such that $\alpha = y_0$ and $\beta = (y_{T-1} - y_0)/T$. In that case, there should be $A = B = 0$. This procedure is appropriate to seasonally adjusted data. For data that manifest strong seasonal fluctuations, such as the U.K. consumption data, a line can be fitted by least squares through the data points of the first and the final years. Figure 7, shows the effect of the application of the filter to the U.K. data adjusted in this manner.

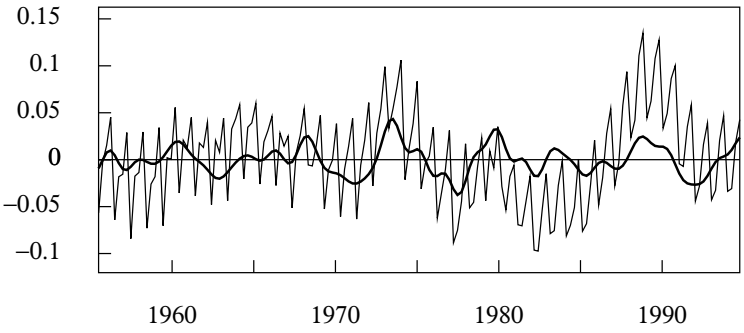

Figure 7: The sequence derived by applying the bandpass filter of Christiano and Fitzgerald to the quarterly logarithmic data on U.K. Consumption.

The filtered sequence of Figure 7 has much the same profile in its middle section as does the sequence of Figure 6, which is derived by applying truncated bandpass filter. (The difference in the scale of the two diagrams tends to conceal this similarity.) However, in comparing filtered sequence to the adjusted data, it seems fair to say that it fails adequately to represent the prominent low-frequency fluctuations. It is also beset by some noisy high-frequency fluctuations that would not normally be regarded as part of the business cycle.

4. POLYNOMIAL DETRENDING

The problems besetting the filtered sequence can be highlighted with reference to the periodogram of the residuals that are obtained by interpolating a polynomial trend line thorough the logarithmic data. Therefore, it is appropriate, at this juncture, to derive a formula for polynomial regression.

Therefore, let $L_T = [e_1, e_2, ..., e_{T-1}, 0]$ be the matrix version of the lag operator, which is formed from the identity matrix $I_T = [e_0, e_1, e_2, ..., e_{T-1}]$ of order T by deleting the leading column and by appending a column of zeros to the end of the array. The matrix that takes the p-th difference of a vector of order T is

$$\nabla_T^p = (I - L_T)^p \ . \tag{4.1}$$

We may partition this matrix so that $\nabla_T^p = [Q_*, Q]'$, where Q_*' has p rows. If y is a vector of T elements, then

$$\nabla_T^p y = \begin{bmatrix} Q_*' \\ Q' \end{bmatrix} y = \begin{bmatrix} g_* \\ g \end{bmatrix}; \tag{4.2}$$

and g_* is liable to be discarded, whereas g will be regarded as the vector of the p-th differences of the data.

The inverse matrix, which corresponds to the summation operator, is partitioned conformably to give $\nabla_T^{-p} = [S_*, S]$. It follows that

$$\begin{bmatrix} S_* & S \end{bmatrix} \begin{bmatrix} Q_*' \\ Q' \end{bmatrix} = S_* Q_*' + S Q' = I_T \ , \tag{4.3}$$

and that

$$\begin{bmatrix} Q_*' \\ Q' \end{bmatrix} \begin{bmatrix} S_* & S \end{bmatrix} = \begin{bmatrix} Q_*' S_* & Q_*' S \\ Q' S_* & Q' S \end{bmatrix} = \begin{bmatrix} I_p & 0 \\ 0 & I_{T-p} \end{bmatrix}. \tag{4.4}$$

If g_* is available, then y can be recovered from g via $y = S_* g_* + S g$.

The lower-triangular Toeplitz matrix $\nabla_T^{-p} = [S_*, S]$ is completely characterised by its leading column. The elements of that column are the ordinates of a polynomial of degree $p-1$, of which the argument is the row index $t = 0, 1, ..., T-1$. Moreover, the leading p columns of the matrix ∇_T^{-p}, which constitute the submatrix S_*, provide a basis for all polynomials of degree $p-1$ that are defined on the integer points $t = 0, 1, ..., T-1$.

A polynomial of degree $p-1$, represented by its ordinates in the vector f, can be interpolated through the data by minimising the criterion

$$(y - f)'(y - f) = (y - S_* f_*)'(y - S_* f_*) \tag{4.5}$$

with respect to f_*. The resulting values are

$$f_* = (S_*'S_*)^{-1}S_*'y \quad \text{and} \quad f = S_*(S_*'S_*)^{-1}S_*'y\ . \tag{4.6}$$

An alternative representation of the estimated polynomial is available, which is provided by the identity

$$S_*(S_*'S_*)^{-1}S_*' = I - Q(Q'Q)^{-1}Q'\ . \tag{4.7}$$

It follows that the polynomial fitted to the data by least-squares regression can be written as

$$f = y - Q(Q'Q)^{-1}Q'y\ . \tag{4.8}$$

A more general method of curve fitting, which embeds polynomial regression as a special case, is one that involves the minimisation of a combination of two sums of squares. Let f denote the vector of fitted values. Then, the criterion for finding the vector is to minimise

$$L = (y-f)'(y-f) + f'Q\Lambda Q'f\ . \tag{4.9}$$

The first term penalises departures of the resulting curve from the data, whereas the second term imposes a penalty for a lack of smoothness in the curve. The second term comprises $d = Q'f$, which is the vector of p-th-order differences of f. The matrix Λ serves to generalise the overall measure of the curvature of the function that has the elements of f as its sampled ordinates, and it serves to regulate the penalty for roughness, which may vary over the sample.

Differentiating L with respect to f and setting the result to zero, in accordance with the first-order conditions for a minimum, gives

$$y - f = Q\Lambda Q'f = Q\Lambda d\ . \tag{4.10}$$

Multiplying the equation by Q' gives $Q'(y-f) = Q'y - d = Q'Q\Lambda d$, whence $\Lambda d = (\Lambda^{-1} + Q'Q)^{-1}Q'y$. Putting this into the equation $f = y - Q\Lambda d$ gives

$$f = y - Q(\Lambda^{-1} + Q'Q)^{-1}Q'y\ . \tag{4.11}$$

If $\Lambda^{-1} = 0$ in (4.11), and if Q' is the matrix version of the twofold difference operator, then the least-squares interpolator of a linear function is derived in the form equation (4.8). The sequence of regression residuals will be given by the vector $r = Q(Q'Q)^{-1}Q'y$; and it is notable that these residuals contain exactly the same information as the vector $g = Q'y$ of the twofold differences of the data. However, whereas the low-frequency structure would be barely visible in the periodogram of the differenced data, it will be fully evident in the periodogram of the residuals of a polynomial regression.

The periodogram of the residual sequence obtained from a linear detrending of the logarithmic consumption data is presented in Figure 8. Superimposed upon the figure is a highlighted band that spans the interval $[\pi/16, \pi/3]$, which corresponds to the nominal pass band of the filters applied in the previous section.

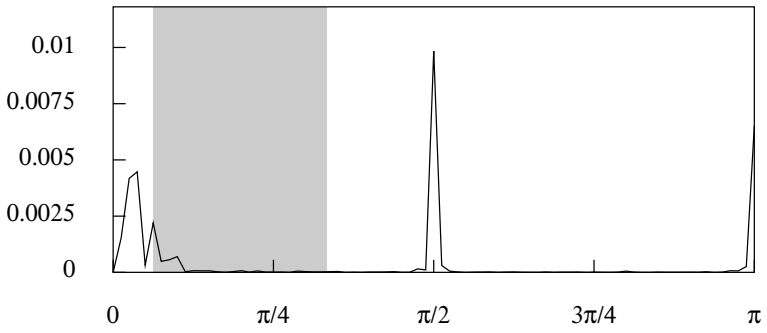

Figure 8: The periodogram of the residual sequence obtained from the linear detrending of the logarithmic consumption data. A band, with a lower bound of $\pi/16$ radians and an upper bound of $\pi/3$ radians, is masking the periodogram.

Within this periodogram, the spectral structure extending from zero frequency up to $\pi/8$ belongs to the business cycle. The prominent spikes located at the frequency $\pi/2$ and at the limiting Nyquist frequency of π are property of the seasonal fluctuations. Elsewhere in the periodogram, there are wide dead spaces, which are punctuated by the spectral traces of minor elements of noise. The highlighted pass band omits much of the information that might be used in synthesising the business cycle.

5. SYNTHESIS OF THE BUSINESS CYCLE

To many economists, it seems implausible that the trend of a macroeconomic index, which is the product of events within the social realm, should be modelled by polynomial, which may be described as a deterministic function. A contrary opinion is represented in this paper. We deny the objective reality of the trend. Instead, we consider it to be the product of our subjective perception of the data. From this point of view, a polynomial function can often serve as a firm benchmark against which to measure the fluctuations of the index. Thus, the linear trend that we have interpolated through the logarithms of the consumption data provides the benchmark of constant exponential growth.

It is from the residuals of a log-linear detrending of the consumption data that we wish to extract the business cycle. The appropriate method is to extract

the Fourier components of the residual sequence that lie within the relevant frequency band. Reference to Figure 8 suggests that this band should stretch from zero up to the frequency of $\pi/8$ radians per quarter, which corresponds to a cycle with a duration of 4 years. In Figure 9, the sequence that is synthesised from these Fourier ordinates has been superimposed upon the sequence of the residuals of the linear detrending.

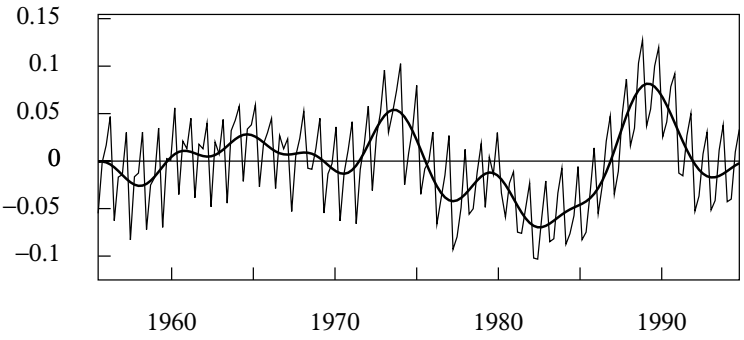

Figure 9: The residual sequence from fitting a quadratic trend to the logarithmic consumption data. The interpolated line, which represents the business cycle, has been synthesised from the Fourier ordinates in the frequency interval $[0, \pi/8]$.

To provide a symbolic representation of the method, we may denote the matrix of the discrete Fourier transform and its inverse by

$$\begin{aligned} U &= T^{-1/2}\big[\exp\{-\mathrm{i}\,2\pi tj/T\};\ t,j=0,...,T-1\big]\ , \\ \bar{U} &= T^{-1/2}\big[\exp\{\mathrm{i}\,2\pi tj/T\};\ t,j=0,...,T-1\big]\ . \end{aligned} \tag{5.1}$$

Then, the residual vector $r = Q(Q'Q)^{-1}Q'y$ and its Fourier transform ρ are represented by

$$r = T^{1/2}\,\bar{U}\rho \quad \longleftrightarrow \quad \rho = T^{-1/2}\,U r\ . \tag{5.2}$$

Let J be a matrix of which the elements are zeros apart from a string of units on the diagonal, which serve to select from ρ the requisite Fourier ordinates within the band $[0, \pi/8]$. Then, the filtered vector that represents the business cycle is given by

$$x = T^{1/2}\,\bar{U}J\rho = \{\bar{U}JU\}\,r = \Psi\, r\ . \tag{5.3}$$

Here, $\bar{U}JU = \Psi = \big[\psi^{\circ}_{|i-j|};\ i,j=0,...,T-1\big]$ is a circulant matrix of the filter coefficients that would result from wrapping the infinite sequence of the ideal bandpass coefficients around a circle of circumference T and adding the overlying elements. Thus

$$\psi^{\circ}_k = \sum_{q=-\infty}^{\infty} \psi_{qT+k}\ . \tag{5.4}$$

Applying the wrapped filter to the finite data sequence via a circular convolution is equivalent to applying the original filter to an infinite periodic extension of the data sequence. In practice, the wrapped coefficients would be obtained from the Fourier transform of the vector of the diagonal elements of the matrix J.

The Fourier method can also be exploited to create a sequence that represents a combination of the trend and the business cycle. There are various ways of proceeding. One of them is to add the vector x to that of the linear or polynomial trend that has generated the sequence of residuals. An alternative method is to obtain the trend/cycle component by subtracting its complement from the data vector.

The complement of the trend/cycle component is a stationary component. Since a Fourier method can be applied only to a stationary vector, we are constrained to work with the vector $g = Q'y$, obtained by taking the twofold differences of the data.

Since the twofold differencing entails the loss of two points, the vector g may be supplemented by a point at the beginning and a point at the end. The resulting vector may be denoted by q. The relevant Fourier ordinates are extracted by applying the selection matrix $I-J$ to the transformed vector $\gamma = Uq$. Thereafter, they need to be reinflated to compensate for the differencing operation.

The frequency response of the twofold difference operator, which is obtained be setting $z = \exp\{-\mathrm{i}\omega\}$ in equation (3.4), is

$$f(\omega) = 2 - 2\cos(\omega) \ , \tag{5.5}$$

and that of the anti-differencing operation is the inverse $1/f(\omega)$. The Fourier ordinates of a differenced vector will be reinflated by pre-multiplying their vector by the diagonal matrix $V = \mathrm{diag}\{v_0, v_1, ..., v_{T-1}\}$, which comprises the values $v_j = 1/f(\omega_j);\ j = 0, ..., T-1$, where $\omega_j = 2\pi j/T$.

The matrix that is to be applied to the Fourier ordinates of the differenced data is therefore $H = V(I - J)$. The resulting vector is transformed back to the time domain via the matrix $\bar{U}$ to produce the vector that is to be subtracted from the data vector y. The resulting estimate of the trend/cycle component is

$$z = y - \bar{U}HUq \ . \tag{5.6}$$

This is represented in Figure 10.

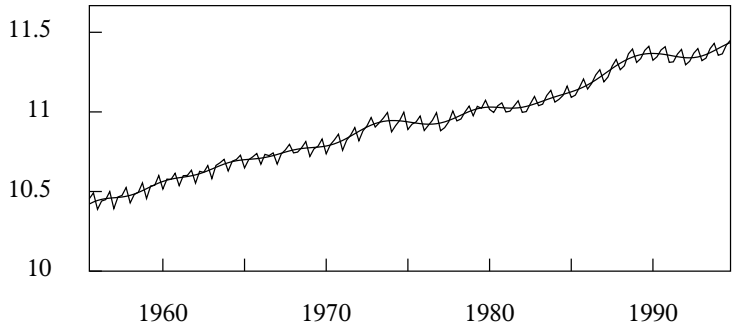

Figure 10: The trend/cycle component of U.K. Consumption determined by the Fourier method, superimposed on the logarithmic data.

6. MORE FLEXIBLE METHODS OF DETRENDING

Methods of detrending may be required that are more flexible than the polynomial interpolations that we have we considered so far. For a start, there is a need to minimise the disjunctions that occur in the periodic extension of the data sequence where the end of one replication joins the beginning of the next. This purpose can be served by a weighted version of a least-squares polynomial regression. If extra weight is given to the data points at the beginning and the end of the sample, then the interpolated line can be constrained pass through their midst; and, thereby, a major disjunction can be avoided.

The more general method of trend estimation that is represented by equation (4.11) can also be deployed. By setting $\Lambda^{-1} = \lambda^{-1}I$, a familiar filtering device is obtained that has been attributed by economists to Hodrick and Prescott (1980, 1997). In fact, an earlier exposition this filter was provided by Leser (1961), and its essential details can be found in a paper of Whittaker (1923).

The effect of the Hodrick–Prescott (H–P) filter depends upon the value of the smoothing parameter λ. As the value of the parameter increases, the vector f converges upon that of a linear trend. As the value of λ tends to zero, f converges to the data vector y. The effect of using the more flexible H–P trend in place of a linear trend is to generate estimates of the business cycle fluctuations that have lesser amplitudes and a greater regularity.

The enhanced regularity of the fluctuations is a consequence of the removal from the residual sequence of a substantial proportion of the fluctuations of lowest frequency, which can cause wide deviations from the line. This enhancement might be regarded as a spurious. However, it can be argued that such low-frequency fluctuations are liable to escape the attention of many economic agents, which is a reason for excluding them from a representation of the business cycle.

Whereas the H–P filter employs a globally constant value for the λ, it is possible to vary this parameter over the course of the sample. This will allow the trend to absorb the structural breaks or disturbances that might occasionally interrupt the steady progress of the economy. If it can be made to absorb the structural breaks, then the trend will not be thrown off course for long; and, therefore, it should serve as a benchmark against which to measure the cyclical variations when the economy resumes its normal progress. At best, the residual sequence will serve to indicate how the economy might have behaved in the absence of the break.

Figure 11 shows a trend function that has been fitted, using a variable smoothing parameter, to the logarithms of a sequence of annual data on real U.K. gross domestic product that runs from 1873 to 2001. Only the breaks after the ends of the first and second world wars have been accommodated, leaving the disruptions of the 1929 recession to be expressed in the residual sequence. The effect has been achieved by attributing a greatly reduced value to the smoothing parameter in the vicinity of the post-war breaks. In the regions that are marked by shaded bands, the smoothing parameter has been given a value of 5. Elsewhere, it has been given a high value of 100,000, which results in trend segments that are virtually linear.

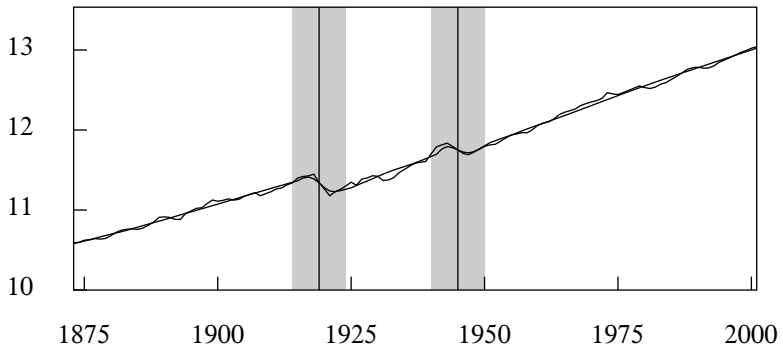

Figure 11: The logarithms of annual U.K. real GDP from 1873 to 2001 with an interpolated trend. The trend is estimated via a filter with a variable smoothing parameter.

This example serves to illustrate the contention that the trend and the accompanying business cycle are best regarded as subjective concepts. The intention of the example is to remove from the residual sequence — and, therefore, from the representation of business cycle — the effects of two major economic disruptions. For the purpose of emphasising the extent of these disruptions, the contrary approach of fitting a stiff polynomial trend line through the data should be followed.

REFERENCES

[1] Baxter, M. and King, R.G. (1999). Measuring business cycles: approximate band-pass filters for economic time series, *Review of Economics and Statistics*, **81**, 575–593.

[2] Burns, A.M. and Mitchell, W.C. (1947). *Measuring Business Cycles*, National Bureau of Economic Research, New York.

[3] Christiano, L.J. and Fitzgerald, T.J. (2003). The band-pass filter, *International Economic Review*, **44**, 435–465.

[4] Granger, C.W.J. (1966). The typical spectral shape of an economic variable, *Econometrica*, **34**, 150–161.

[5] Harding, D. and Pagan, A. (2002). Dissecting the cycle: a methodological investigation, *Journal of Monetary Economics*, **49**, 365–381.

[6] Hodrick, R.J. and Prescott, E.C. (1980). *Postwar U.S. business cycles: an empirical investigation*, Working Paper, Carnegie–Mellon University, Pittsburgh, Pennsylvania.

[7] Hodrick, R.J. and Prescott, E.C. (1997). Postwar U.S. business cycles: an empirical investigation, *Journal of Money, Credit and Banking*, **29**, 1–16.

[8] Leser, C.E.V. (1961). A simple method of trend construction, *Journal of the Royal Statistical Society, Series B*, **23**, 91–107.

[9] Whittaker, E.T. (1923). On a new method of graduation, *Proceedings of the Royal Society of Edinburgh*, **44**, 77–83.

REVSTAT – Statistical Journal
Volume 7, Number 1, April 2009, 105–117

PARAMETER ESTIMATION FOR INAR PROCESSES BASED ON HIGH-ORDER STATISTICS

Authors: Isabel Silva
– Departamento de Engenharia Civil e CEC,
Universidade do Porto, Portugal
ims@fe.up.pt

M. Eduarda Silva
– Faculdade de Economia, Universidade do Porto,
UI&D Matemática e Aplicações da Universidade de Aveiro, Portugal
mesilva@fep.up.pt

Abstract:

- The high-order statistics (moments and cumulants of order higher than two) have been widely applied in several fields, specially in problems where it is conjectured a lack of Gaussianity and/or non-linearity. Since the INteger-valued AutoRegressive, INAR, processes are non-Gaussian, the high-order statistics can provide additional information that allows a better characterization of these processes. Thus, an estimation method for the parameters of an INAR process, based on Least Squares for the third-order moments is proposed. The results of a Monte Carlo study to investigate the performance of the estimator are presented and the method is applied to a set of real data.

Key-Words:

- *INAR process; estimation; high-order statistics.*

AMS Subject Classification:

- 62M10, 62F10.

1. INTRODUCTION

In the recent past, the high-order statistics (HOS) have been widely applied in several fields. By HOS it is meant the moments and cumulants of order higher than two, in the time domain, and the corresponding multidimensional Fourier transform (polyspectrum), in the frequency domain. In this work, the time domain approach is considered. The HOS comprise information about stochastic processes such as the degree of nonlinearity and deviations from Gaussianity that is not contained in the second-order statistics.

Let $\{X_t\}$ be a k-th-order stationary stochastic process. The k-th-order joint moment of $X_t, X_{t+s_1}, ..., X_{t+s_{k-1}}$, for $s_1, ..., s_{k-1} \in \mathbb{R}$, is a function of $k-1$ variables defined by

$$\mu_X(s_1, ..., s_{k-1}) = \mathrm{E}[X_t\, X_{t+s_1} ... X_{t+s_{k-1}}] \ ,$$

with $\mu_X = \mathrm{E}[X_t]$. For a stationary stochastic process, the moments have the following symmetry properties:

$$\mu_X(m) = \mu_X(-m) \ , \qquad m > 0 \ ,$$
$$\mu_X(m,n) = \mu_X(n,m) = \mu_X(-n, m-n) = \mu_X(n-m, -m) \ , \qquad m, n > 0 \ .$$

Then, it follows that the third-order moments over the entire plane may be obtained from the values of the third-order moments over the infinite wedge bounded by the straight lines $m = 0$ and $m = n$, $m, n > 0$.

Recently, the integer-valued autoregressive process has been proposed in the literature to model time series of counts. The p-th-order integer-valued autoregressive, INAR(p), process is defined as a discrete time non-negative integer-valued stochastic process, $\{X_t\}$, that satisfies the following equation (Latour, 1998):

$$X_t = \alpha_1 \circ X_{t-1} + \alpha_2 \circ X_{t-2} + \cdots + \alpha_p \circ X_{t-p} + e_t \ , \tag{1.1}$$

where

1. $\{e_t\}$, designated the innovation process, is a sequence of independent and identically distributed (i.i.d.) non-negative integer-valued random variables with $\mathrm{E}[e_t] = \mu_e$, $\mathrm{Var}[e_t] = \sigma_e^2$ and $\mathrm{E}[e_t^3] = \gamma_e$;
2. the symbol $\circ$ represents the thinning operation (Steutel and Van Harn, 1979; Gauthier and Latour, 1994), defined by

$$\alpha_i \circ X_{t-i} = \sum_{j=1}^{X_{t-i}} Y_{i,j}, \qquad \text{for} \ \ i = 1, ..., p \ ,$$

where $\{Y_{i,j}\}$, designated the counting series, is a set of i.i.d. non-negative integer-valued random variables such that $\mathrm{E}[Y_{i,j}] = \alpha_i$, $\mathrm{Var}[Y_{i,j}] = \sigma_i^2$ and $\mathrm{E}[Y_{i,j}^3] = \gamma_i$. All the counting series are assumed independent of $\{e_t\}$;

3. $0 \leq \alpha_i < 1$, $i = 1, ..., p-1$, and $0 < \alpha_p < 1$. Note that the stationarity condition for the INAR(p) process is that $\sum_{k=1}^{p} \alpha_k < 1$.

A special case is the Poisson INAR process with binomial thinning operation, where $\{e_t\}$ has a Poisson distribution with parameter λ and the counting series, $\{Y_{i,j}\}$, are a set of Bernoulli random variables with $P(Y_{i,j}=1) = 1 - P(Y_{i,j}=0) = \alpha_i$.

Since the INAR models are non-Gaussian, the HOS can provide additional information in the characterization of these processes. Thus, an estimation method for the parameters of an INAR model that uses HOS is proposed in this work. This approach applies the Least Squares estimation method to minimize the errors between the third-order moment of the observations and of the fitted model.

This work is organized as follows: in Section 2 the third-order characterization of INAR(p) models is provided and the proposed Least Squares Estimation method based on HOS (LS_HOS) is described. In Section 3 the results of a simulation study to assess the small sample properties of the proposed estimator are given and the method is applied to a set of observations concerning the number of plants within the industrial sector in Section 4. Finally, some remarks are presented in Section 5.

2. PARAMETER ESTIMATION BASED ON HOS

2.1. Third-order characterization of INAR(p) models

The third-order characterization, in terms of moments and cumulants, of INAR models has been obtained by Silva and Oliveira (2004, 2005) and Silva (2005). In particular, the third-order moments of an INAR(p) process, defined by (1.1), satisfy a set of Yule–Walker type equations similar to those satisfied by the bilinear process, that can be written as:

$$\begin{aligned} \mu_X(0,0) \;=\; & \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p} \alpha_i\,\alpha_j\,\alpha_k\,\mu_X(i-j,i-k) \\ & +\; 3\sum_{i=1}^{p}\sum_{j=1}^{p} \alpha_j\,{\sigma_i}^2\,\mu_X(i-j) \;+\; 3\,\mu_X(\sigma_e^2+{\mu_e}^2)\sum_{i=1}^{p}\alpha_i \\ & +\; 3\,\mu_e\sum_{i=1}^{p}\sum_{j=1}^{p}\alpha_i\,\alpha_j\,\mu_X(i-j) \;+\; 3\,\mu_X\,\mu_e\sum_{i=1}^{p}{\sigma_i}^2 \\ & +\; \mu_X\sum_{i=1}^{p}(\gamma_i - 3\,\alpha_i\,{\sigma_i}^2 - \alpha_i^3) \;+\; \gamma_e\ , \end{aligned} \tag{2.1}$$

$$\mu_X(0,k) \;=\; \sum_{i=1}^{p} \alpha_i\, \mu_X(0,k-i) \;+\; \mu_e\, \mu_X(0)\;, \qquad k>0\;, \tag{2.2}$$

$$\begin{aligned} \mu_X(k,k) \;=\; & \sum_{i=1}^{p}\sum_{j=1}^{p} \alpha_i\,\alpha_j\, \mu_X(k-i,k-j) \;+\; \sum_{i=1}^{p} {\sigma_i}^2\, \mu_X(k-i) \\ & +\; 2\,\mu_e\, \mu_X(k) - \mu_X({\mu_e}^2 - {\sigma_e}^2)\;, \qquad k>0\;, \end{aligned} \tag{2.3}$$

$$\mu_X(k,m) \;=\; \sum_{i=1}^{p} \alpha_i\, \mu_X(k,m-i) \;+\; \mu_e\, \mu_X(k)\;, \qquad m>k>0\;, \tag{2.4}$$

where

$$\mu_X(0) \;=\; \sum_{i=1}^{p} \alpha_i\, \mu_X(i) \;+\; \mu_e\, \mu_X + V_p\;, \tag{2.5}$$

is the second-order moment of $\{X_t\}$, with

$$V_p \;=\; {\sigma_e}^2 + \mu_X \sum_{i=1}^{p} {\sigma_i}^2\;,$$

which represents the variance of the one-step-ahead prediction error (Silva, 2005).

These equations indicate that the INAR processes have a non-linear structure, therefore the first- and second-order moments are not sufficient to describe the dependence structure of the process. In the next section, is described an estimation method for the parameters of an INAR(p) process that uses the additional information provided by the HOS.

2.2. Least squares estimation based on HOS

Let $\{x_1, x_2, ..., x_n\}$ be a realization of a non-negative integer-valued stationary stochastic process with third-order moments $\mu(0,k),\ k>0$. The approximating model considered is an INAR(p) process (order known) with parameters $\alpha_1, ..., \alpha_p$, μ_e, σ_e^2 and third-order moments $\mu_X(0,k)$, $k>0$, satisfying (2.2), which can be represented in the following matrix form

$$\boldsymbol{\mu}_{3,X} = \mathbf{M}_{3,X}\,\boldsymbol{\alpha} + \mu_e\,\mu_X(0)\,\mathbf{1}_p\;, \tag{2.6}$$

where $\boldsymbol{\mu}_{3,X}$ is defined as

$$\boldsymbol{\mu}_{3,X} = \left[\; \mu_X(0,1) \;\cdots\; \mu_X(0,p) \;\right]^{\mathrm{T}}\,,$$

$\mathbf{M}_{3,X}$ is the $p\times p$ non-symmetric Toeplitz matrix of the third-order moments of the INAR(p) process

$$\mathbf{M}_{3,X} = \begin{bmatrix} \mu_X(0,0) & \mu_X(1,1) & \cdots & \mu_X(p-1,p-1) \\ \mu_X(0,1) & \mu_X(0,0) & \cdots & \mu_X(p-2,p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \mu_X(0,p-1) & \mu_X(0,p-2) & \cdots & \mu_X(0,0) \end{bmatrix},$$

with $\mu_X(\cdot,\cdot)$ given in (2.1) to (2.4), $\boldsymbol{\alpha} = [\,\alpha_1 \;\cdots\; \alpha_p\,]^{\mathrm{T}}$ is the vector of coefficients, $\mu_X(0)$ is the second-order moment of the INAR(p) process given in (2.5) and $\mathbf{1}_p$ is a $p\times 1$ vector of ones.

Defining

$$\mathbf{H} = \begin{bmatrix}\mathbf{M}_{3,X} & \mu_X(0)\mathbf{1}_p\end{bmatrix} \qquad \text{and} \qquad \boldsymbol{\theta} = \begin{bmatrix}\alpha_1 & \cdots & \alpha_p & \mu_e\end{bmatrix}^{\mathrm{T}},$$

equation (2.6) can be rewritten as

$$\boldsymbol{\mu}_{3,X} = \mathbf{H}\,\boldsymbol{\theta}\ ,$$

suggesting that $\boldsymbol{\theta}$ may be estimated by least squares, i.e., minimizing the squared error between the third-order moments of the fitted INAR(p) model, $\boldsymbol{\mu}_{3,X}$, and the third-order moments of the data,

$$\boldsymbol{\mu}_3 \;=\; \begin{bmatrix}\mu(0,1) & \cdots & \mu(0,p)\end{bmatrix}^{\mathrm{T}}.$$

Thus, $\hat{\boldsymbol{\theta}}$, the Least Squares estimator of $\boldsymbol{\theta}$ based on HOS (LS_HOS) satisfies

$$\hat{\boldsymbol{\theta}} \;=\; \min_{\boldsymbol{\theta}}\big\{L^*(\boldsymbol{\theta})\big\}$$

where

$$L^*(\boldsymbol{\theta}) \;=\; (\boldsymbol{\mu}_3 - \mathbf{H}\,\boldsymbol{\theta})^{\mathrm{T}}\,(\boldsymbol{\mu}_3 - \mathbf{H}\,\boldsymbol{\theta})\ .$$

In practice, the estimator is calculated by substituting the moments in $\boldsymbol{\mu}_3$ and $\mathbf{H}$ by their sample counterparts, using the usual estimators of the moments

$$\hat{\mu}_X(0) = \frac{1}{N}\sum_{t=1}^{N} {X_t}^2, \quad \hat{\mu}_X(0,k) = \frac{1}{N}\sum_{t=1}^{N-k} {X_t}^2\, X_{t+k}, \quad \hat{\mu}_X(k,k) = \frac{1}{N}\sum_{t=1}^{N-k} X_t\, {X_{t+k}}^2\ .$$

Thus,

$$\hat{\boldsymbol{\theta}} \;=\; \min_{\boldsymbol{\theta}}\big\{\hat{L}^*(\boldsymbol{\theta})\big\} \;=\; \min_{\boldsymbol{\theta}}\big\{(\hat{\boldsymbol{\mu}}_3 - \hat{\mathbf{H}}\,\boldsymbol{\theta})^{\mathrm{T}}\,(\hat{\boldsymbol{\mu}}_3 - \hat{\mathbf{H}}\,\boldsymbol{\theta})\big\}\ .$$

Note that an estimator for σ_e^2 can be obtained by

$$\hat{\sigma}_e^2 \;=\; \hat{V}_p - \overline{X}\sum_{i=1}^{p}\hat{\sigma}_i^2\ ,$$

where $\overline{X}$ is the sample mean of the observations, $\hat{\sigma}_i^2$ is an estimator of the counting series variance for the i-th thinning operation, $\alpha_i \circ X_{t-i}$, $i = 1, ..., p$, and $\hat{V}_p = \hat{R}(0) - \sum_{i=1}^{p} \hat{\alpha}_i \hat{R}(i)$, with $\hat{R}(i) = \frac{1}{N} \sum_{t=1}^{N-i} (X_t - \overline{X})(X_{t+i} - \overline{X})$, representing the sample autocovariance function. The estimation of $\hat{\sigma}_i^2$ depends on the distribution of the counting series, for instance, in the case of the binomial thinning operation (when the counting series are Bernoulli distributed), $\hat{\sigma}_i^2 = \hat{\alpha}_i(1 - \hat{\alpha}_i)$, for $i = 1, ..., p$.

The asymptotic distribution of the LS_HOS estimator depends on the sixth-order moments and cumulants of the processes, and therefore is too complex and not useful in practice. So, the finite sample properties of the estimator are investigated by a simulation study, which results are presented in the next section.

3. MONTE CARLO RESULTS

The aim of the simulation study presented in this section is twofold: to examine the small sample properties of the estimator previously described and compare its performance with other estimation methods for the parameters of an INAR process.

Thus, 1000 realizations of Poisson INAR(p) processes ($e_t \sim \mathcal{P}o(\lambda)$) with binomial thinning operation are generated, for $p = 0, ..., 3$. The sample size, N, and parameters values considered are:

- $N = 50$, 200, 500 and 1000 observations,
- $\lambda \in \{1.0, 3.0\}$,
- for $p = 1$, $\alpha_1 \in \{0.1, 0.4, 0.6, 0.9\}$,
- for $p = 2$, $(\alpha_1, \alpha_2) \in \big\{(0.1, 0.6), (0.6, 0.1), (0.3, 0.4), (0.4, 0.3), (0.1, 0.1), (0.4, 0.4)\big\}$,
- for $p = 3$, $(\alpha_1, \alpha_2, \alpha_3) \in \big\{(0.1, 0.1, 0.4), (0.1, 0.4, 0.1), (0.4, 0.1, 0.1), (0.3, 0.3, 0.3)\big\}$.

For each realization, the estimation methods used to obtain $\hat{\boldsymbol{\theta}} = [\hat{\alpha}_1, ..., \hat{\alpha}_p, \hat{\mu}_e]^{\mathrm{T}}$ are Yule–Walker estimation (YW), Conditional Least Squares estimation (CLS), Whittle estimation (WHT) and unconstrained and constrained Least Squares estimation based on HOS (LS_HOS and LS_HOS_C). For a detailed description of the YW, CLS and WHT estimation methods see Silva (2005). The constraints considered are $0 < \alpha_i < 1$, $i = 1, ..., p$, $\sum_{i=1}^{p} \alpha_i < 1$ and ${\sigma_e}^2 > 0$. The initial values of the iterative methods are the YW estimates.

The unconstrained minimizations necessary in the methods CLS, WHT and LS_HOS are performed through the MATLAB function *fminunc*, which finds a minimum of a scalar unconstrained multivariable function by using the BFGS

Quasi-Newton method with a mixed quadratic and cubic line search procedure (MathWorks (2004)). The constrained minimization of the method is accomplished by the MATLAB function *fmincon*, which finds a minimum of a scalar constrained nonlinear multivariable function by using a Sequential Quadratic Programming method (MathWorks (2004)).

For each case, the following sample statistical measures are evaluated:

- mean bias: $\overline{\text{Bias}}(\hat{\theta}_i) = \dfrac{1}{N}\sum_{j=1}^{N}(\hat{\theta}_i^{(j)} - \theta_i)\,,$
- variance: $\text{Var}(\hat{\theta}_i) = \dfrac{1}{N-1}\sum_{j=1}^{N}(\hat{\theta}_i^{(j)} - \overline{\theta}_i)^2\,,$
- mean square error: $\text{MSE}(\hat{\theta}_i) = \dfrac{1}{N-1}\sum_{j=1}^{N}(\hat{\theta}_i^{(j)} - \theta_i)^2\,,$

where $\hat{\theta}_i^{(j)}$ represents the parameters estimates, $\hat{\alpha}_1, ..., \hat{\alpha}_p, \hat{\mu}_e$, in the j-th repetition, for $j = 1, ..., N = 1000$, and $\overline{\theta}_i = \frac{1}{N}\sum_{j=1}^{N}\hat{\theta}_i^{(j)}$ is its sample mean.

With respect to the small sample properties of the LS_HOS and LS_HOS_C estimators, the following conclusions can be drawn from the analysis of all the simulations. In general, the sample bias, variance and mean square error decrease as the sample size increases, indicating that the distribution of the estimators is consistent and symmetric. However, for a small sample size there is evidence of departure from symmetry in the marginal distributions, specially for values of the parameters near the non-stationary region.

Table 1 presents numerical results for two INAR(1) processes, with parameter values $\boldsymbol{\theta} = (\alpha_1, \lambda) = (0.1, 3.0)$ and $\boldsymbol{\theta} = (0.6, 3.0)$, respectively, and two different sample sizes: $N = 50$ and 500. As expected, the results for the unconstrained and constrained estimations only differ when the value of the coefficient is near of the non-admissible region, specially when it presents a small value ($\alpha_1 = 0.1$).

Figure 1 presents the boxplots of the sample bias for the parameter estimates of an INAR(2) processes, with parameter values $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \lambda) = (0.6, 0.1, 1.0)$, obtained by LS_HOS and LS_HOS_C, for the four different sample sizes. As can be seen, the boxplots indicate that the marginal distributions of the estimators are, generally, symmetric.

When the several estimation methods are compared it is found that the LS_HOS provides similar results, in terms of the smallest values of sample bias, variance and mean square error, to the other methods. It is also verified that, in general, the proportion of non-admissible estimates of the methods is less for LS_HOS, followed by WHT and CLS. The results show that, in general, the sample mean bias of $\hat{\alpha}_i$ is negative, indicating that α_i is underestimated, while λ is overestimated, since the sample mean bias of the parameter estimate is positive.

Table 1: Sample statistical measures for the parameters estimates of Poisson INAR(1) processes.

Measure	N	$\theta = (\alpha_1, \lambda)$	$\hat{\theta}_1 = \hat{\alpha}_1$		$\hat{\theta}_2 = \hat{\lambda}$	
			LS_HOS	LS_HOS_C	LS_HOS	LS_HOS_C
$\overline{\text{Bias}}(\hat{\theta}_i)$	50	(0.1, 3.0)	−0.0411	−0.0041	0.0949	−0.0897
		(0.6, 3.0)	−0.0893	−0.0826	0.5208	0.4611
	500	(0.1, 3.0)	−0.0019	−0.0003	0.0029	−0.0050
		(0.6, 3.0)	−0.0094	−0.0099	0.0532	0.0584
$\text{Var}(\hat{\theta}_i)$	50	(0.1, 3.0)	0.0190	0.0099	0.2623	0.1566
		(0.6, 3.0)	0.0158	0.0149	0.9229	0.8841
	500	(0.1, 3.0)	0.0021	0.0020	0.0286	0.0270
		(0.6, 3.0)	0.0014	0.0014	0.0787	0.0772
$\text{MSE}(\hat{\theta}_i)$	50	(0.1, 3.0)	0.0206	0.0100	0.2711	0.1645
		(0.6, 3.0)	0.0238	0.0217	1.1933	1.0958
	500	(0.1, 3.0)	0.0021	0.0020	0.0286	0.0270
		(0.6, 3.0)	0.0015	0.0015	0.0815	0.0805

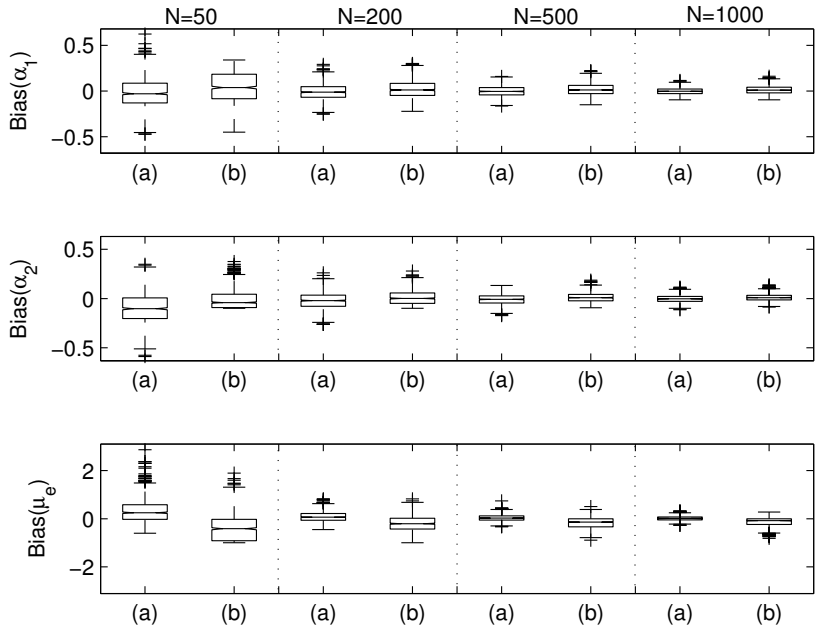

Figure 1: Boxplots of the sample bias for the estimates obtained by (a): LS_HOS and (b): LS_HOS_C, in 1000 realizations of the Poisson INAR(2) model: $X_t = 0.6 \circ X_{t-1} + 0.1 \circ X_{t-2} + e_t$, where $e_t \sim \mathcal{P}o(1)$.

In order to illustrate some of these conclusions, Figure 2 shows the boxplots of the sample bias for the estimates obtained from 50 and 200 observations of the INAR(1) process with parameter values $\boldsymbol{\theta} = (\alpha_1, \lambda) = (0.9, 1.0)$. Note that the value of α_1 is near the non-stationary region, however, even for $N = 50$ observations the LS_HOS estimates presents the best results.

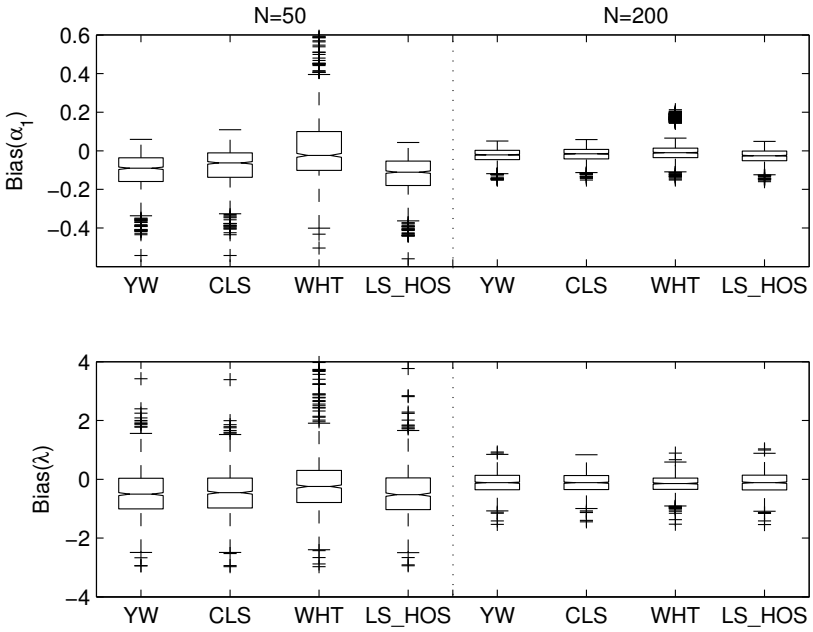

Figure 2: Boxplots of the sample bias for the estimates obtained in 1000 realizations of 50 and 200 observations of the Poisson INAR(1) model: $X_t = 0.9 \circ X_{t-1} + e_t$, where $e_t \sim \mathcal{P}o(1)$.

4. APPLICATION TO REAL DATA

In this section, the proposed estimation method is applied to a real dataset concerning the number of Swedish mechanical paper and pulp mills, from 1921 to 1981 (see Figure 3). This dataset was used by Brännäs (1995) and Brännäs and Hellström (2001), and these authors fitted an INAR(1) process to this dataset using some explanatory variables (the industrial gross profit margin and GNP). Here, an INAR(1) process, with binomial thinning operation, where the innovations are i.i.d. random variables with mean μ_e and variance σ_e^2 is considered. Since the mean of the data is 20.40 and its variance is 155.16, a Poisson innovation process is not assumed but then the method does not require that or any other assumption on the distribution of the innovations.

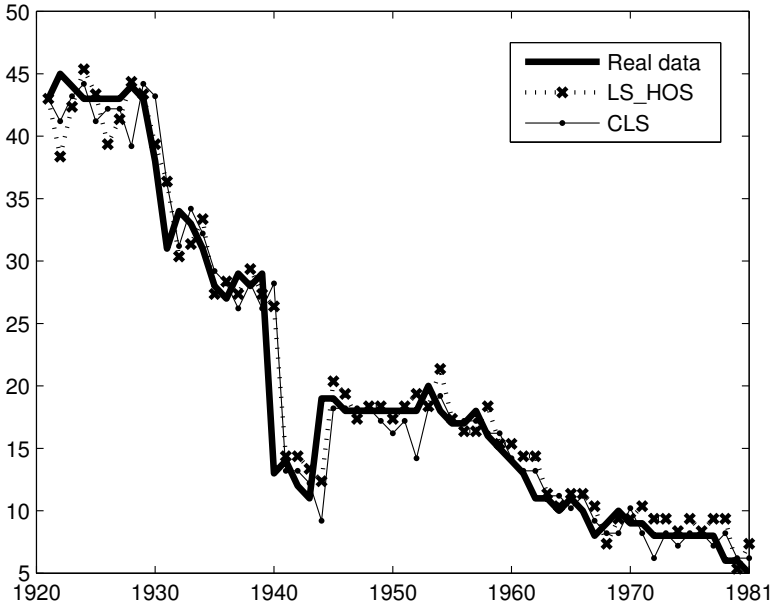

Figure 3: The number of Swedish mechanical paper and pulp mills, from 1921 to 1981 (Brännäs (1995) and Brännäs and Hellström (2001)), and the fitted values considering the LS_HOS and CLS estimates.

Table 2 presents the parameter estimates obtained by CLS and LS_HOS estimation methods. The fit of both models, based on LS_HOS and CLS estimates, are also shown in Figure 3. As can be observed, the two fits are very similar and similar to the dataset. The mean square errors (MSE) between the observations and the fitted values are also exhibited in Table 2. It can be seen that the MSE is slightly smaller for the LS_HOS fit than for CLS fit.

Table 2: The parameter estimates of the number of Swedish mechanical paper and pulp mills, from 1921 to 1981.

Method	$\hat{\alpha}$	$\hat{\mu}_e$	$\hat{\sigma}_e^2$	MSE	$\hat{\mu}_x$	$\hat{\sigma}_x^2$
CLS	0.9591	0.2017	15.2268	8.5494	4.9315	192.2764
LS_HOS	0.9269	1.3635	19.2253	7.4465	18.6525	145.4513

The last two columns of the Table 2 present the mean and variance of the estimated models:

$$\hat{\mu}_x = \frac{\hat{\mu}_e}{1-\hat{\alpha}_1} \qquad \text{and} \qquad \hat{\sigma}_x^2 = \frac{\hat{\mu}_e\,\hat{\alpha}_1 + \hat{\sigma}_e^2}{1-\hat{\alpha}_1^2}\ .$$

It is noticeable that the model estimated by LS_HOS presents mean and variance closer to the sample values.

The goodness-of-fit of both fitted models is investigated by the residuals. The analysis of the sample autocorrelation and sample partial autocorrelation functions, as well as the usual tests of randomness, do not reject the hypothesis of uncorrelated random variables for the residual series from both fitted models.

5. FINAL REMARKS

The principal advantage of HOS is the capability to detect and characterize the deviations from Gaussianity and non-linearity of the processes. Thus in this work a new estimation method for the parameters of INAR processes based on HOS is proposed. This method uses the Least Squares estimation to minimize the errors between the third-order moment of the observations and of the fitted model. Note that this estimation method does not assume any particular discrete distribution to the counting series and to the innovation process. A Monte Carlo study indicates that this estimation method provides good results in small samples, in terms of sample bias, variance and mean square error. Moreover, when used in the context of a non-Poisson real dataset the LS_HOS estimates provide a model with mean, variance and autocorrelations closer to the sample values.

ACKNOWLEDGMENTS

For the first author, this work reports research developed under financial support provided by "FCT – Fundação para a Ciência e Tecnologia", Portugal.

REFERENCES

[1] Brännäs, K. (1995). Explanatory variables in the AR(1) count data model, *Umeå Economic Studies*, **381**.

[2] Brännäs, K. and Hellström, J. (2001). Generalized integer-valued autoregression, *Econometric Reviews*, **20**(4), 425–443.

[3] Gauthier, G. and Latour, A. (1994). Convergence forte des estimateurs des paramètres d'un processus GENAR(p), *Annales des Sciences Mathématiques du Québec*, **18**(1), 49–71.

[4] Latour, A. (1998). Existence and stochastic structure of a non-negative integer-valued autoregressive process, *Journal of Time Series Analysis*, **19**(4), 439–455.

[5] Mathworks (2004). *Optimization Toolbox User's Guide For MATLAB*, Available from `http://www.mathworks.com/access/helpdesk/help/pdf_doc/optim/optim_tb.pdf`

[6] Silva, I. (2005). *Contributions to the analysis of discrete-valued time series*, unpublished PhD Thesis, Universidade do Porto, Portugal.

[7] Silva, I. and Oliveira, V.L. (2004). Difference equations for the higher-order moments and cumulants of the INAR(1) model, *Journal of Time Series Analysis*, **25**(3), 317–333.

[8] Silva, I. and Oliveira, V.L. (2005). Difference equations for the higher-order moments and cumulants of the INAR(p) model, *Journal of Time Series Analysis*, **26**(1), 17–36.

[9] Steutel, F.W. and Van Harn, K. (1979). Discrete analogues of self-decomposability and stability, *The Annals of Probability*, **7**(5), 893–899.

REVSTAT – Statistical Journal
Volume 7, Number 1, April 2009, 119–134

FORECASTING IN INAR(1) MODEL

Authors: Nélia Silva
– Departamento de Matemática, UI&D Matemática e Aplicações, Universidade de Aveiro, Portugal
neliasilva@ua.pt

Isabel Pereira
– Departamento de Matemática, UI&D Matemática e Aplicações, Universidade de Aveiro, Portugal
isabel.pereira@ua.pt

M. Eduarda Silva
– Faculdade de Economia da Universidade do Porto, UI&D Matemática e Aplicações da Universidade de Aveiro, Portugal
mesilva@fep.up.pt

Abstract:

- In this work we consider the problem of forecasting integer-valued time series, modelled by the INAR(1) process introduced by McKenzie (1985) and Al-Osh and Alzaid (1987). The theoretical properties and practical applications of INAR and related processes have been discussed extensively in the literature but there is still some discussion on the problem of producing coherent, i.e. integer-valued, predictions. Here Bayesian methodology is used to obtain point predictions as well as confidence intervals for future values of the process. The predictions thus obtained are compared with their classic counterparts. The proposed approaches are illustrated with a simulation study and a real example.

Key-Words:

- *INAR models; Bayesian prediction; integer prediction; Markov Chain Monte Carlo algorithm.*

1. INTRODUCTION

In applications we are frequently faced with time series whose characteristics are not compatible with a continuous modelling approach. Discrete variate time series occur in many contexts, often as counts of events or individuals in consecutive intervals or at consecutive points in time. Examples of these are the number of costumers waiting to be served, the daily number of absent workers in a firm, the number of busy lines in a telephone network noted every hour, the number of accidents in a manufacturing plant each month, etc. Several models that take the discreteness of the data explicitly into account have been developed in the literature. Cox (1981) proposed dividing them into two categories: observation-driven and parameter-driven models. MacDonald and Zucchini (1997), Cameron and Trivedi (1998) and the review by McKenzie (2003) provide an excellent overview of the literature in this area.

In this work we are interested in a special class of observation-driven models, the so-called integer-valued autoregressive (INAR) process introduced by McKenzie (1985) and Al-Osh and Alzaid (1987). The theoretical properties and practical applications of INAR and related processes have been discussed extensively in the literature. Silva *et al.* (2005) consider independent replications of count time series modelled by INAR(1) and proposed several estimation methods using the classical and Bayesian approaches in time and frequency domains. Nevertheless, there is still little consensus on which processes or model classes are best used in practice in contrast to the role played by the Box-Jenkins Gaussian ARMA methodology for continuous variables. This is partly due to the lack of reliable techniques for estimation, testing and prediction. In particular, the lack of forecasting methods that are coherent in the sense of producing only integer-valued predictions, seems to render useless the effort of using more complex models.

Usually forecasts are obtained from the conditional expectations, which have the optimality property but rarely will generate integer values. In order to produce coherent forecasts Freeland and McCabe (2003) use the median of the k-step-ahead conditional distribution to emphasize the intention of preserving the integer structure of the data in generating the forecasts. McCabe and Martin (2005) develop a general methodology for producing coherent predictions of low count data. In contrast to the usual applications of the model INAR(1), in which the arrival process is usually Poisson, they allow the arrivals to follow any distribution in the integer class. The forecasts are based on an estimate of the k-step–ahead predictive probability mass function. To eliminate unwanted values from the conditioning set of the predictive function, Bayesian methods are used. Jung and Tremayne (2006) extend some of the ideas used by Freeland and McCabe in higher order dependence structure by proposing a computer intensive method for generating coherent, integer out-of-sample predictions, particularly

obtaining the h-step-ahead predictor for the INAR(2). They propose a Monte Carlo approach using bootstrap methods to estimate the sampling distributions as a mean of generating one and multi-step ahead forecasts which respect the integer structure of the data.

The purpose of this paper is to obtain coherent forecasts for the Poisson INAR(1) process. Bayesian methodology is used to obtain point predictions as well as credibility intervals for future values of the process which are compared with their classic counterpart.

The remainder of the paper is divided into four main sections. Section 2 provides the theoretical results in order to obtain the point forecasts. Section 3 presents methods for producing confidence intervals or highest posterior predictive density intervals for forecasts. In Section 4 we conduct a simulation study to compare the performance of the classical and Bayesian approaches, considering point and interval predictions. Section 5 gives an example of forecasting a count data series using PoINAR(1) model. The data are the number of claimants receiving wage loss benefits due to injuries from burns, supplied by the Workers Compensation Board of the Province of British Columbia, Canada. The proposed methodology presented in this work is applied to this data set and compared with classical inference and forecasting procedures from Freeland (1998).

2. POINT PREDICTION

Consider a non negative integer-valued random variable X and $\alpha \in [0,1]$, the generalized thinning operation, hereafter denoted by '$\circ$', is defined as

$$\alpha \circ X = \sum_{j=1}^{X} Y_j \ , \tag{2.1}$$

where $\{Y_j\}$, $j = 1, ..., X$, is a sequence of independent and identically distributed non-negative integer-valued random variables, independent of X, with finite mean α and variance σ^2. This sequence is called the counting series of $\alpha \circ X$. When $\{Y_j\}$ is a sequence of Bernoulli random variables, the thinning operation is called binomial thinning operation and was defined by Steutel and van Harn (1979).

The well-known INAR(1) process $\{X_t;\ t = 0, \pm 1, \pm 2, ...\}$ is defined on the discrete support $\mathbb{N}_0$ by the equation

$$X_t = \alpha \circ X_{t-1} + \epsilon_t \ ,$$

where $0 < \alpha < 1$, $\{\epsilon_t\}$ is a sequence of independent and identically distributed integer-valued random variables, with $E[\epsilon_t] = \mu_\epsilon$ and $\mathrm{Var}[\epsilon_t] = \sigma^2_\epsilon$.

In this paper we consider only Poisson INAR(1) process, i.e., $\{\epsilon_t\}$ is a sequence of independent Poisson distributed variables with parameter λ, independent of all counting series $\{Y_j\}$. Note that, assuming $\epsilon_t \frown Po(\lambda)$ it is straightforward to show that $X_t \frown Po\big(\lambda/(1-\alpha)\big)$. The Poisson INAR(1) process will henceforth be denoted PoINAR(1).

Given that we have observed the series up through time n, i.e., $\mathbf{x}_n = (x_1, x_2, ..., x_n)$ is known, the most common procedure for constructing predictions in time series models is to use conditional expectations. The following theorem states and important result in this context.

Theorem 2.1 (Freeland, 1998, pp. 30)**.** *The moment generation function of X_{n+h} given X_n is*

$$\varphi_{X_{n+h}|X_n}(s) \;=\; \big[\alpha^h e^s + (1-\alpha^h)\big]^{X_n} \exp\left[\lambda\,\frac{1-\alpha^h}{1-\alpha}\,(e^s-1)\right]. \tag{2.2}$$

Expression (2.2) shows that the distribution of $X_{n+h}|X_n$ is a convolution of a binomial distribution with parameters α^h and X_n and a Poisson distribution with parameter $\lambda(1-\alpha^h)/(1-\alpha)$. That is, the probability function of $X_{n+h}|X_n$ is given by

$$\begin{aligned} f(x_{n+h}|x_n) \;&=\; P\big(X_{n+h}=x_{n+h}\,|\,X_n=x_n\big) \\ &=\; \exp\left\{-\lambda\,\frac{1-\alpha^h}{1-\alpha}\right\} \sum_{i=0}^{M_h} \frac{1}{(x_{n+h}-i)!} \\ &\quad\times \left(\lambda\,\frac{1-\alpha^h}{1-\alpha}\right)^{x_{n+h}-i} \binom{x_n}{i} (\alpha^h)^i\,(1-\alpha^h)^{x_n-i}\,, \qquad x_{n+h}=0,1,...\,, \end{aligned} \tag{2.3}$$

where $M_h = \min(X_{n+h}, X_n)$. Consequently, we have the following corollary:

Corollary 2.1. *The INAR(1) model satisfies the properties*

a) $E\big[X_{n+h}|X_n\big] = \alpha^h\left[X_n - \dfrac{\lambda}{1-\alpha}\right] + \dfrac{\lambda}{1-\alpha}\,, \quad h=1,2,3,...,$

b) $\mathrm{Var}\big[X_{n+h}|X_n\big] = \alpha^h(1-\alpha^h)\,X_n + \lambda\,\dfrac{1-\alpha^h}{1-\alpha}\,, \quad h=1,2,3,...,$

c) *As $h \to +\infty$, $X_{n+h}|X_n$ is a Poisson distribution with parameter $\lambda/(1-\alpha)$.*

So, we can conclude that for α constant,

$$\lim_{h\to+\infty} E\big[X_{n+h}|X_n\big] \;=\; \lim_{h\to+\infty} \mathrm{Var}\big[X_{n+h}|X_n\big] \;=\; \lambda/(1-\alpha)\;,$$

i.e., as $h \to \infty$ and $0 < \alpha < 1$, the mean and the variance of $X_{n+h}|X_n$ remain equal and approach the mean of the process.

2.1. Classical methodology

The h-step-ahead predictor based on the conditional expectation of INAR(1),

$$\hat{X}_{n+h}|\mathbf{x}_n = E\big[X_{n+h}|X_n\big] = \alpha^h\left[X_n - \frac{\lambda}{1-\alpha}\right] + \frac{\lambda}{1-\alpha}\ , \quad h = 1, 2, 3, ... \tag{2.4}$$

was obtained by Brännäs (1994) and Freeland and McCabe (2003), but it will hardly produce integer-valued forecasts. In order to obtain coherent predictions for X_{n+h} Freeland and McCabe (2003) suggest using the value which minimizes the expected absolute error given the sample, i.e., the value that minimizes $E\big[|X_{n+h} - \hat{X}_{n+h}|\,\big|\,X_n\big]$. So, they concluded that $\hat{X}_{n+h} = \hat{m}_{n+h}$ is the median of the h-step-ahead conditional distribution $f(x_{n+h}|x_n)$.

2.2. Bayesian methodology

The Bayesian predictive probability function is based on the assumption that, both, the future observation, X_{n+h} and the vector of unknown parameters $\boldsymbol{\theta} = (\alpha, \lambda)$ are random. As we know the information about $\boldsymbol{\theta}$ is given by the observed sample $\mathbf{x}_n$ and quantified in the posterior predictive, $\pi(\boldsymbol{\theta}|\mathbf{x}_n)$.

Definition 2.1. Let $\boldsymbol{\theta} \in \Theta$ be the vector of unknown parameters. The h-step-ahead Bayesian posterior predictive distribution is given by

$$\begin{aligned} f(x_{n+h}|\mathbf{x}_n) &= \int_\Theta f(x_{n+h};\boldsymbol{\theta}|\mathbf{x}_n)\, d\boldsymbol{\theta} \\ &= \int_\Theta f(x_{n+h}|\mathbf{x}_n;\boldsymbol{\theta})\, \pi(\boldsymbol{\theta}|\mathbf{x}_n)\, d\boldsymbol{\theta}\ , \end{aligned} \tag{2.5}$$

where $\pi(\boldsymbol{\theta}|\mathbf{x}_n)$ is the posterior probability function of $\boldsymbol{\theta}$ and $f(x_{n+h}|\mathbf{x}_n;\boldsymbol{\theta})$ is the predictive distribution (classical) given by (2.3).

The h-step-ahead predictive distribution of $X_{n+h}|\mathbf{x}_n$ given by expression (2.5) can be viewed as having all information about the future values. Once $f(x_{n+h}|\mathbf{x}_n)$ is obtained, the Bayesian h-step-ahead predictor can be given by the expected valued, the median or the mode of X_{n+h} given $\mathbf{x}_n$.

Since beta and gamma are conjugate of binomial and Poisson distributions, respectively, we use them for prior distributions of the parameters to INAR(1) model, $\alpha \frown Beta(a, b)$, $a, b > 0$ and $\lambda \frown Gamma(c, d)$, $c, d > 0$. Considering independence between α and λ, the prior distribution of (α, λ) is given by

$$p(\alpha, \lambda) \propto \lambda^{c-1} \exp(-d\lambda)\, \alpha^{a-1}(1-\alpha)^{b-1}\ , \qquad \lambda > 0, \quad 0 < \alpha < 1\ , \tag{2.6}$$

where a, b, c and d are known parameters. Note that, as $a \to 0$, $b \to 0$, $c \to 0$ and $d \to 0$ we have a vague prior distribution.

The posterior distribution of (α, λ) can be written as

$$\begin{aligned} p(\alpha,\lambda|\mathbf{x}_n) \;&\propto\; L(\mathbf{x}_n,\alpha,\lambda|x_1)\; p(\lambda,\alpha) \\ &=\; \exp\Big[-\big(d+(n-1)\big)\,\lambda\Big]\,\lambda^{c-1}\,\alpha^{a-1}(1-\alpha)^{b-1} \\ &\quad\times\; \prod_{t=2}^{n}\sum_{i=0}^{M_t}\frac{\lambda^{x_t-i}}{(x_t-i)!}\binom{x_{t-1}}{i}\alpha^i(1-\alpha)^{x_{t-1}-i}\;, \end{aligned}$$

where $L(\mathbf{x}_n|x_1)$ is the conditional likelihood function and $M_t = \min(X_t, X_{t-1})$.

Consequently for the PoINAR(1) model, the Bayesian predictive function of X_{n+h} given $\mathbf{x}_n$ is given by

$$\begin{aligned} f(x_{n+h}|\mathbf{x}_n) \;\propto\; &\int_\alpha\int_\lambda \sum_{i=0}^{M_h}\binom{x_n}{i}(\alpha^h)^i\,(1-\alpha^h)^{x_n-i}\,\frac{1}{(x_{n+h}-i)!} \\ &\times\; \exp\!\left(-\lambda\,\frac{1-\alpha^h}{1-\alpha}\right)\left(\lambda\,\frac{1-\alpha^h}{1-\alpha}\right)^{x_{n+h}-i}\exp\big[-(d+n)\,\lambda\big]\,\lambda^{c-1} \\ &\times\; \alpha^{a-1}(1-\alpha)^{b-1}\prod_{t=2}^{n}\sum_{i=0}^{M_t}\frac{\lambda^{x_t-i}}{(x_t-i)!}\binom{x_{t-1}}{i}\alpha^i(1-\alpha)^{x_{t-1}-i}\,d\alpha\,d\lambda\;. \end{aligned} \tag{2.7}$$

The complexity of $f(x_{n+h}|\mathbf{x}_n)$ does not allow us to work with it directly. In order to estimate X_{n+h}, we can adapt to the integer case the Tanner (1996) composition method. That is, to sample $(X_{n+h,1}, X_{n+h,2}, ..., X_{n+h,m})$, we can use the following algorithm:

Algorithm 2.1

1. From the sample $(X_1, X_2, ..., X_n)$, calculate (through the classical method) a starting estimate for α, denoted by α_0;

2. Using the adaptive rejection Metropolis sampling (ARMS) within Gibbs methodology (see Gilks *et al.*, 1995), calculate from the full conditional distributions of the parameters α and λ, a sample $(\alpha_1, \lambda_1), (\alpha_2, \lambda_2), ..., (\alpha_m, \lambda_m)$;

3. For each i $(i = 1, ..., m)$ sample $X_{n+h,i}$ from $f(x_{n+h}|x_n, \alpha_i, \lambda_i)$, using the inverse transform method adapted to integer variables, that is,

(**a**) sample u from uniform $U(0, 1)$,

(**b**) calculate the least integer-valued s: $\sum_{i=0}^{s} f(x_{n+h}|x_n, \alpha_i, \lambda_i) \geq u$,

(**c**) consider $X_{n+h,i} = s$.

After sampling $X_{n+h,1}, X_{n+h,2}, ..., X_{n+h,m}$, the h-step-ahead predictor of X_{n+h}, can be calculated from the sample mean ($\hat{X}_{n+h}$), the median ($\hat{m}_{n+h}$) or the mode ($\hat{mo}_{n+h}$).

But we can also calculate $E(X_{n+h}|\mathbf{x}_n)$ using an appropriate property of mathematical expectation. As we know $E\big[g(X_{n+h})|\mathbf{x}_n\big] = E\big[E[g(X_{n+h})|\mathbf{x}_n, \boldsymbol{\theta}]\,|\,\mathbf{x}_n\big]$; thus,

$$\begin{aligned} E(X_{n+h}|\mathbf{x}_n) &= E\Big[E(X_{n+h}|\boldsymbol{\theta}, \mathbf{x}_n)\,\Big|\,\mathbf{x}_n\Big] \\ &= E\left[\alpha^h X_n + \frac{1-\alpha^h}{1-\alpha}\,\lambda\,\middle|\,\mathbf{x}_n\right] \\ &= X_n\,E\big[\alpha^h|\mathbf{x}_n\big] + E\left[\frac{1-\alpha^h}{1-\alpha}\,\lambda\,\middle|\,\mathbf{x}_n\right]. \end{aligned}$$

These expected values can be estimated through Markov Chain Monte Carlo (MCMC) algorithms. We perform Metropolis algorithm in conjunction with Adaptive Rejection Sampling Method (ARMS) in order to sample values from full conditional distributions of α and λ; let them be noted by $(\alpha_1, \alpha_2, ..., \alpha_m)$, $(\lambda_1, \lambda_2, ..., \lambda_m)$, respectively (see Silva *et al.*, 2005). We have

$$\hat{E}\big[\alpha^h|\mathbf{x}_n\big] = \frac{1}{m}\sum_{i=1}^{m}\alpha_i^h\,,$$

$$\hat{E}\left[\frac{1-\alpha^h}{1-\alpha}\,\lambda\,\middle|\,\mathbf{x}_n\right] = \frac{1}{m}\sum_{i=1}^{m}\frac{1-\alpha_i^h}{1-\alpha_i}\,\lambda_i\,.$$

Consequently the predictor can be written as

$$\tilde{X}_{n+h} = X_n\left(\frac{1}{m}\sum_{i=1}^{m}\alpha_i^h\right) + \left(\frac{1}{m}\sum_{i=1}^{m}\frac{1-\alpha_i^h}{1-\alpha_i}\,\lambda_i\right). \tag{2.8}$$

3. INTERVAL PREDICTION

In this section we obtain interval predictions for the h-step-ahead observation, using the classical framework and Bayesian methodology.

3.1. Classical methodology

A confidence interval for the predictor $\hat{X}_{n+h}$, can be calculated through the probability function of the h-step-ahead prediction error, given by

$$e_{n+h}|\mathbf{x}_n = X_{n+h}|\mathbf{x}_n - \hat{X}_{n+h}|\mathbf{x}_n\,.$$

Replacing $\hat{X}_{n+h}|\mathbf{x}_n$ given by (2.4), we obtain

$$e_{n+h}|\mathbf{x}_n \;=\; X_{n+h} - \alpha^h x_n - \lambda\,\frac{1-\alpha^h}{1-\alpha}\;.$$

Since $e_{n+h}|\mathbf{x}_n$ is a function of discrete random variable X_{n+h}, we have

$$e_{n+h}|\mathbf{x}_n = \; k - \alpha^h x_n - \lambda\,\frac{1-\alpha^h}{1-\alpha}\;, \qquad k = 0, 1, 2, \ldots\;.$$

So,

$$\begin{aligned} & P\left(e_{n+h} = k - \alpha^h x_n - \lambda\,\frac{1-\alpha^h}{1-\alpha}\,\Big|\,\mathbf{x}_n\right) \;= \\ & \quad = \; P\Big(X_{n+h} = k \,|\, X_n = x_n\Big) \\ & \quad = \; \exp\left\{-\lambda\,\frac{1-\alpha^h}{1-\alpha}\right\} \sum_{i=0}^{M_h} \frac{1}{(k-i)!}\left(\lambda\,\frac{1-\alpha^h}{1-\alpha}\right)^{k-i}\binom{x_n}{i}(\alpha^h)^i\,(1-\alpha^h)^{x_n-i}\,. \end{aligned} \tag{3.1}$$

From the expression (3.1) we obtain a γ level confidence interval for X_{n+h}

$$\left(\hat{X}_{n+h} + e_{t_1}, \hat{X}_{n+h} + e_{t_2}\right), \tag{3.2}$$

where $\hat{X}_{n+h}$ is given by (2.4), e_{t_1} is the largest value of e_{n+h} such as $P(e_{n+h} \le e_{t_1})$ $\le (1-\gamma)/2$ and e_{t_2} is the smallest value of e_{n+h} such as $P(e_{n+h} \le e_{t_2}) \ge (1+\gamma)/2$.

3.2. Bayesian methodology

We propose an adaptive generalization of the method used to obtain Highest Posterior Density (HPD) intervals of the model parameters, in which we consider the predictive distribution instead of the posterior.

Definition 3.1. A $100\gamma\%$ predictive interval for X_{n+h} is given by

$$P\big(X_L \le X_{n+h} \le X_R\big) \;=\; \sum_{x_{n+h}=X_L}^{X_R} f(x_{n+h}|\mathbf{x}_n)\;.$$

However, since $f(x_{n+h}|\mathbf{x}_n)$ is not always symmetric[1], the intervals with a maximum posterior predictive probability are more desirable than predictive intervals (Chen *et al.*, 2000).

[1]We made a previous study with some samples from PoINAR(1) and we verified that many were neither symmetric nor unimodal.

Definition 3.2. $R(\gamma) = (X_L, X_R)$ is a $100\gamma\%$ HPD interval for X_{n+h} if

$$P\big(X_L \leq X_{n+h} \leq X_R\big) \;=\; \sum_{x_{n+h}=X_L}^{X_R} f(x_{n+h}|\mathbf{x}_n) \;\geq\; K_\gamma \;, \tag{3.3}$$

where K_γ is the largest constant such that $P\big[X_{n+h} \in R(\gamma)\big] \geq \gamma$.

Due to the complexity of the predictive probability function given by (2.7) it is not possible to calculate the exact HPD interval for X_{n+h}; we can give an approximation for $R(\gamma)$ by using the Chen and Shao (1999) algorithm, because this method does not require the knowledge of the closed form of $f(x_{n+h}|\mathbf{x}_n)$. The Chen and Shao algorithm can be described as:

Algorithm 3.1

1. Obtain an MCMC sample $(X_{n+h,1}, X_{n+h,2}, ..., X_{n+h,m})$ (Algorithm 2.1);
2. Consider $\big(X_{(n+h,1)} \leq X_{(n+h,2)} \leq ... \leq X_{(n+h,m)}\big)$;
3. Compute the $100\gamma\%$ credible intervals

$$R_i(\gamma) = \big(X_{(n+h,i)}, X_{(n+h,i+[m\gamma])}\big) \;, \qquad 1 \leq i \leq m - [m\gamma] \;,$$

   where $[m\gamma]$ is integer part of $m\gamma$;
4. The $100\gamma\%$ HPD interval to X_{n+h} is the one, denoted by $\hat{R}(\gamma)$, with the smallest amplitude among all credible intervals.

Under certain regularity conditions, $\hat{R}(\gamma) \to R(\gamma)$ a.s. as $n \to \infty$, where $R(\gamma)$ is defined in (3.3) (Chen *et al.*, 2000).

Sometimes we obtain more than one interval. For this situation, we consider for $\hat{R}(\gamma)$ the interval with greater absolute frequency, among the intervals width smaller with. When the interval is still not unique we take the one with the smallest lower limit.

4. A SIMULATION STUDY

For the simulation study we consider samples with size $n = 40, 90, 190$ generated by INAR(1) models with the parameters values $\alpha = 0.2, 0.5, 0.8$ and $\lambda = 1, 3$, and considering the hyperparameters $a = b = c = d = 10^{-4}$.

4.1. Point prediction

From the various simulated samples we conclude that large values of α and λ are related with high dispersion values. Consequently the increase in α and λ provides large values of $|x_{n+h} - x_n|$, $h > 1$. As we can see in Figure 1, independently of the prediction methodology used, the forecasts performance depends on two basic aspects: one is the difference between x_n and x_{n+h}, $h > 1$; the other is the approximation between x_n and $\hat{\lambda}/(1-\hat{\alpha})$, in particular when h increases (note that $\hat{E}(X_{n+h}|X_n) \to \hat{\lambda}/(1-\hat{\alpha})$, $h \to \infty$). These findings are illustrated in Table 1 where point predictions for 10 steps ahead for a particular model are given. The table includes the h-step ahead simulated and predicted values and the square of the deviations between x_{190} and x_{190+h}, $h = 1, ..., 10$. The last line contains the classical limiting distribution.

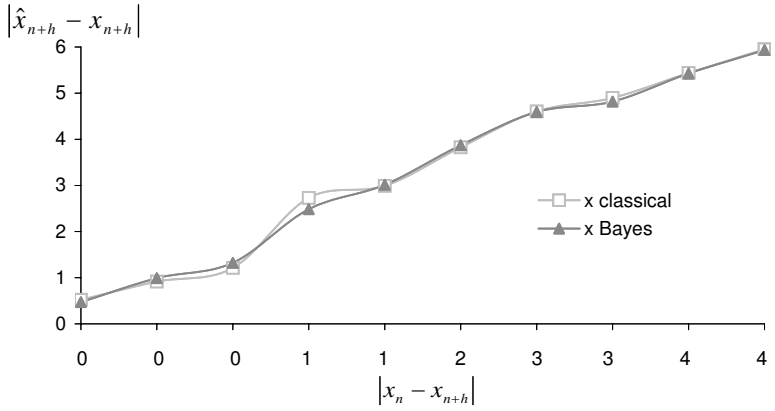

Figure 1: Values of $|\hat{x}_{n+h} - x_{n+h}|$ for a PoINAR(1) sample with $\alpha = 0.8$, $\lambda = 3$, $n = 190$ and $h = 1, 2, ..., 10$.

To confront classical and Bayesian methodologies we use the mean square error (MSE) to compare means, the mean absolute deviation (MAD) to compare medians and the "everything or nothing" lost function (FPTN), given by $1/n \sum I(x_{n+h})$ where

$$I(x_{n+h}) = \begin{cases} 1 & \text{if } |\hat{x}_{n+h} - x_{n+h}| > \delta\,, \\ 0 & \text{if } |\hat{x}_{n+h} - x_{n+h}| \leq \delta\,, \end{cases}$$

to compare modes. In this situation we consider $\delta = 1$ since we have integer values.

Table 2 shows the MSE, MAD and FPTN values from 10 one-step-ahead predictions. Values of MSE($\tilde{X}_{n+h}$) are obtained considering the Bayesian predictors given by (2.8) (values of MSE($\hat{X}_{n+h}$) are similar). Values of MAD and FPTN were calculated, respectively, through medians and modes. The indices "C" or "B" indicate which methodology is used (classical or Bayesian, respectively).

Table 1: Point predictions considering two samples of size $n = 190$ with parameters $(\lambda = 1,\ \alpha = 0.2,\ x_{190} = 0)$ and $(\lambda = 3,\ \alpha = 0.8,\ x_{190} = 16)$, respectively.

$(\lambda = 1, \quad \alpha = 0.2; \quad x_{190} = 0)$						
h	x_{190+h}	jump	classical approach		Bayesian approach	
			$\hat{x}_{190+h}$	$(x_{190+h} - \hat{x}_{190+h})^2$	$\hat{x}_{190+h}$	$(x_{190+h} - \hat{x}_{190+h})^2$
1	2	2	1.068	0.869	1.090	0.828
2	0	0	1.247	1.555	1.292	1.670
3	0	0	1.277	1.631	1.340	1.796
4	0	0	1.282	1.643	1.342	1.801
5	5	5	1.283	13.816	1.288	13.779
6	0	0	1.283	1.646	1.302	1.695
7	1	1	1.283	0.080	1.210	0.044
8	1	1	1.283	0.080	1.348	0.121
9	1	1	1.283	0.080	1.248	0.061
10	2	2	1.283	0.514	1.298	0.493
∞			1.283			

$(\lambda = 3, \quad \alpha = 0.8; \quad x_{190} = 16)$						
h	x_{190+h}	jump	classical approach		Bayesian approach	
			$\hat{x}_{190+h}$	$(x_{190+h} - \hat{x}_{190+h})^2$	$\hat{x}_{190+h}$	$(x_{190+h} - \hat{x}_{190+h})^2$
1	16	0	15.477	0.274	15.530	0.221
2	16	0	15.084	0.839	15.010	0.980
3	16	0	14.787	1.471	14.678	1.748
4	20	4	14.564	29.550	14.574	29.441
5	19	3	14.396	21.197	14.408	21.086
6	17	1	14.270	7.453	14.516	6.170
7	18	2	14.175	14.631	14.128	14.992
8	19	3	14.103	23.981	14.182	23.213
9	20	4	14.049	35.414	14.066	35.212
10	17	1	14.008	8.952	13.986	9.084
∞			13.884			

As we can see, when $\alpha = 0.8$ Bayesian methodology provides smaller values than classical methodology, so the Bayesian predictions seems to have a better performance than classical predictions.

In order to study and compare the estimates given by the sample mean, sample median and sample mode we use the minimum absolute percentual error (MAPE), given by

$$1/H \sum_{h=1}^{H} |\hat{X}_{n+h} - X_{n+h}|/X_{n+h} \ ,$$

where H represents the number of predictions realized. This criteria does not benefit any measure (mean, median or mode) in particular. The results are

Table 2: Values of MSE, MADE and FPTN considering 10 one-step-ahead predictions for the model $x_t = \alpha \circ x_{t-1} + \epsilon_t$, $\epsilon_t \frown \mathrm{P}(3)$ and sample sizes 40, 90 and 190.

	α	0.2			0.8		
	n	40	90	190	40	90	190
MSE	$\hat{X}_{n+h,\mathrm{C}}$	6.68	2.01	**5.56**	12.98	3.82	16.17
	$\hat{X}_{n+h,\mathrm{B}}$	**6.46**	**1.93**	6.06	**5.22**	**3.05**	**3.57**
MAD	$\hat{m}_{n+h,\mathrm{C}}$	2.27	1.18	**2.00**	3.18	1.45	3.81
	$\tilde{m}_{n+h,\mathrm{B}}$	**2.00**	1.18	2.05	**1.82**	**1.27**	**1.68**
FPTN	$\hat{mo}_{n+h,\mathrm{C}}$	**0.45**	0.45	**0.45**	0.73	0.55	1.00
	$\tilde{mo}_{n+h,\mathrm{B}}$	0.55	**0.36**	0.64	**0.36**	**0.36**	**0.55**

illustrated in Table 3 for three samples with sizes 40, 90 and 190 of the model $x_t = \alpha \circ x_{t-1} + \epsilon_t$, $\epsilon_t \frown P(3)$. MAPE minimum is always obtained with Bayesian approach. Similar results are obtained for $\lambda = 1$.

Table 3: Values of MAPE considering 10 one-step-ahead predictions for the model $x_t = \alpha \circ x_{t-1} + \epsilon_t$, $\epsilon_t \frown \mathrm{P}(3)$ and sample sizes 40, 90 and 190. The indices "C" or "B" indicate which methodology is used (classical or Bayesian, respectively).

α	0.2			0.8		
n	40	90	190	40	90	190
$\hat{X}_{n+h,\mathrm{C}}$	0.714	0.573	0.870	0.707	0.564	0.919
$\tilde{X}_{n+h,\mathrm{B}}$	0.178	0.137	0.260	0.110	0.109	0.0811
$\hat{m}_{n+h,\mathrm{C}}$	0.652	0.588	0.631	0.606	0.561	0.831
$\tilde{m}_{n+h,\mathrm{B}}$	0.187	0.120	0.209	0.115	0.103	0.091
$\hat{mo}_{n+h,\mathrm{C}}$	0.619	0.464	0.929	0.625	0.506	0.831
$\tilde{mo}_{n+h,\mathrm{B}}$	0.187	0.127	0.231	0.086	0.125	0.098

4.2. Interval prediction

Prediction intervals for future observations were calculated using expression (3.2) for classical methodology and Chen and Shao algorithm for Bayesian methodology. The performance of the intervals (with 95% of confidence or credibility) obtained by each approach is measured by the amplitudes and coverage probabilities, from 100 replicates. The simulation results for the case $\lambda = 1, 3$, $\gamma = 0.95$ and $n = 100$, are presented in Tables 4 and 5 respectively.

Table 4: Coverage probability estimates and mean amplitudes of the intervals for the h-step-ahead future values, in INAR(1) model with $n = 100$ and $\lambda = 1$.

h	$\alpha = 0.2$				$\alpha = 0.8$			
	cov. prob. estimates		mean amplitude		cov. prob. estimates		mean amplitude	
	classical	Bayes.	classical	Bayes.	classical	Bayes.	classical	Bayes.
1	0.43	0.99	2.56	4.17	0.96	0.98	6.01	4.72
2	0.32	0.98	2.74	4.22	0.95	0.94	7.79	6.18
3	0.36	1.00	2.75	4.22	0.96	0.92	8.20	6.87
4	0.36	0.99	2.75	4.16	0.94	0.92	8.81	7.35
5	0.34	1.00	2.75	4.19	0.96	0.93	9.05	7.57
6	0.38	0.99	2.75	4.21	0.97	0.90	9.43	7.83
7	0.29	0.98	2.75	4.19	0.98	0.92	9.55	7.85
8	0.28	0.99	2.75	4.22	0.94	0.91	9.49	8.04
9	0.38	0.99	2.75	4.22	0.94	0.90	9.78	8.00
10	0.34	1.00	2.75	4.15	0.97	0.92	9.78	8.11

Table 5: Coverage probability estimates and mean amplitudes of the intervals for the h-step-ahead future values, in INAR(1) model with $n = 100$ and $\lambda = 3$.

h	$\alpha = 0.2$				$\alpha = 0.8$			
	cov. prob. estimates		mean amplitude		cov. prob. estimates		mean amplitude	
	classical	Bayes.	classical	Bayes.	classical	Bayes.	classical	Bayes.
1	0.68	0.96	6.34	7.34	0.97	0.96	10.18	9.53
2	0.78	0.99	7.16	7.79	0.96	0.94	12.63	12.00
3	0.98	0.99	8.00	7.88	0.95	0.93	13.89	13.29
4	0.99	0.94	8.00	7.77	0.95	0.89	14.81	14.31
5	0.98	0.98	8.00	7.78	0.95	0.87	15.36	14.67
6	0.96	1.00	8.00	7.86	0.96	0.92	15.58	14.92
7	0.96	0.99	8.00	7.79	0.96	0.91	15.78	14.97
8	0.95	0.99	8.00	7.79	0.92	0.94	15.83	15.30
9	0.96	0.96	8.00	7.77	0.93	0.93	15.95	15.47
10	0.93	0.98	8.00	7.84	0.96	0.91	15.97	15.54

Tables 4 and 5 indicate that when $\alpha = 0.2$ the Bayesian intervals have large coverage probability, specially when $\lambda = 1$; it can be noted that, when $h > 2$ and for small values of α and λ the classic intervals have small amplitudes and they coincide with those obtained from the asymptotic distribution. Moreover, it is worthwhile to mention that when $\alpha \geq 0.5$ the mean amplitude of the prediction intervals obtained by Bayesian methodology are lower than their classic counterpart.

5. ANALYSIS OF BURN CLAIMS DATA

We apply the proposed methodology to a data set analysed by Freeland (1998) comprising 120 monthly counts of workers collecting Wage Loss Benefits (WLB) for burn injuries. All the descriptive details of the data set can be found in Freeland (1998) who concludes that PoINAR(1) is a plausible choice for modelling the data. In order to evaluate and compare the different prediction methodologies, h-step ahead forecasts ($h = 1, 2, 3, 4, 5, 6$) are produced for the time period from July to December 1994, for which we know the observed values. The point forecasts based on the mean, median and mode and the observed values are presented in Table 6. In general, it can be noted that MAPE values of classic point predictions are smaller than those of Bayesian predictions. This result is expected in view of the simulation results presented in the last section since the estimated value for alpha is 0.4.

Interval predictions for the period July to December 1994 are obtained using the two approaches proposed given by (3.2). The intervals obtained, presented in Table 6, are analogous, although the Bayesian have smaller width.

Table 6: h-step ahead predictions of monthly claims from July to December 1994.

h	year/ month	claims of	point prediction						interval prediction	
			classical			Bayesian			classical	Bayesian
			$\hat{x}$	$\hat{m}$	$\hat{m}_0$	$\hat{x}$	$\hat{m}$	$\hat{m}_0$		
1	94/07	11	7.89	8	7	7.67	7	7	(2.1,13.0)	(3.0,13.1)
2	94/08	12	8.24	8	8	8.01	8	7	(2.0,14.0)	(3.0,14.0)
3	94/09	11	8.38	8	8	8.14	8	7	(2.0,14.0)	(3.0,14.0)
4	94/10	12	8.44	8	8	8.21	8	7	(2.0,15.0)	(3.0,14.0)
5	94/11	7	8.46	8	8	8.22	8	7	(2.0,15.0)	(3.0,14.0)
6	94/12	11	8.47	8	8	8.22	8	7	(2.0,15.0)	(3.0,14.0)
		MAPE	0.26	0.27	0.27	0.27	0.27	0.32		

6. FINAL REMARKS

Forecasting low integer values of time series of counts remains an open problem. Although conditional means do not preserve coherently the integer nature of the data, it seems there is no advantage in using median or mode values of the predictive distribution. Simulations indicate that the performance of the different approaches depend on the parameters of the model and that Bayesian methodology provides the best results when MAPE statistic is used.

REFERENCES

[1] Al-Osh, M.A. and Alzaid, A.A. (1987). First-order integer-valued autoregressive (INAR(1)) process, *Journal of Time Series Analysis*, **8**, 261–275.

[2] Brännäs, K. (1994). Estimation and testing in integer-valued AR(1) models, *Umeå Economic Studies*, **335**.

[3] Cameron, A.C. and Trivedi, P.K. (1998). *Regression Analysis of Count Data*, Cambridge University Press, Cambridge.

[4] Chen, M.-H. and Shao, Q.-M. (1999). Monte Carlo estimation of Bayesian credible and HPD intervals, *Jounal of Computational and Graphical Statistics*, **8**, 69–92.

[5] Chen, M.-H.; Shao, Q.-M. and Ibrahim, J.G. (2000). *Monte Carlo Methods in Bayesian Computation*, Springer Series in Statistics.

[6] Cox, D.R. (1981). Statistical analysis of time series: some recent developments, *Scandinavian Journal of Statistics*, **8**, 93–115.

[7] Freeland, R.K. (1998). *Statistical analysis of discrete time series with application to the analysis of Worker's Compensation Claims Data*, Ph.D. Thesis, The University of British Columbia, Canada.

[8] Freeland, R.K. and McCabe, B.P.M. (2003). Forecasting discrete valued low count series, *International Journal of Forecasting*, **20**, 427–434.

[9] Gilks, W.R.; Best, N.G. and Tan, K.K.C. (1995). Adaptive rejection Metropolis sampling within Gibbs sampling, *Appl. Statist.*, **44**, 455–472.

[10] Jung, R.C. and Tremayne, A.R. (2006). Coherent forecasting in integer time series models, *International Journal of Forecasting*, **22**, 223–238.

[11] MacDonald, I.L. and Zucchini, W. (1997). *Hidden Markov and Other Models for Discrete-Valued Time Series*, Chapman and Hall, London.

[12] McCabe, B.P.M. and Martin, G.M. (2005). Bayesian predictions of low count time series, *International Journal of Forecasting*, **21**, 315–330.

[13] McKenzie, E. (1985). Some simple models for discrete variate time series, *Water Resources Bulletin*, **21**, 645–650.

[14] McKenzie, E. (2003). Discrete variate time series, *Handbook of Statistics*, **21**, 576–606.

[15] Silva, I.; Silva, M.E.; Pereira, I. and Silva, N. (2005). Replicated INAR(1) Process, *Methodology and Computing in Applied Probability*, **7**, 517–542.

[16] Steutel, F.W. and Van Harn, K. (1979). Discrete analogues of self-decomposability and stability, *The Annals of Probability*, **5**, 893–899.

[17] Tanner, M.A. (1996). *Tools for statistical inference*, 3rd ed., Springer Verlag, New York.

REVSTAT – Statistical Journal

Background

Statistical Institute of Portugal (INE), well aware of how vital a statistical culture is in understanding most phenomena in the present-day world, and of its responsibility in disseminating statistical knowledge, started the publication of the scientific statistical journal *Revista de Estatística*, in Portuguese, publishing three times a year papers containing original research results, and application studies, namely in the economic, social and demographic fields.

In 1998 it was decided to publish papers also in English. This step has been taken to achieve a larger diffusion, and to encourage foreign contributors to submit their work.

At the time, the Editorial Board was mainly composed by Portuguese university professors, being now composed by national and international university professors, and this has been the first step aimed at changing the character of *Revista de Estatística* from a national to an international scientific journal.

In 2001, the *Revista de Estatística* published three volumes special issue containing extended abstracts of the invited contributed papers presented at the 23rd European Meeting of Statisticians.

The name of the Journal has been changed to REVSTAT – STATISTICAL JOURNAL, published in English, with a prestigious international editorial board, hoping to become one more place where scientists may feel proud of publishing their research results.

— The editorial policy will focus on publishing research articles at the highest level in the domains of Probability and Statistics with emphasis on the originality and importance of the research.

— All research articles will be refereed by at least two persons, one from the Editorial Board and another, external.

— The only working language allowed will be English.

— Three volumes are scheduled for publication, one in March, one in June and the other in November.

— On average, four articles will be published per issue.

Aims and Scope

The aim of REVSTAT is to publish articles of high scientific content, in English, developing innovative statistical scientific methods and introducing original research, grounded in substantive problems.

REVSTAT covers all branches of Probability and Statistics. Surveys of important areas of research in the field are also welcome.

Abstract/indexed in

REVSTAT is expected to be abstracted/indexed at least in *Current Index to Statistics*, *Mathematical Reviews*, *Statistical Theory and Method Abstracts*, and *Zentralblatt für Mathematic*.

Instructions to Authors, special-issue editors and publishers

Papers may be submitted in two different ways:

— By sending a paper copy to the Executive Editor and one copy to one of the two Editors or Associate Editors whose opinion the author(s) would like to be taken into account, together with a postscript or a PDF file of the paper to the e-mail: revstat@fc.ul.pt.

— By sending a paper copy to the Executive Editor, together with a postscript or a PDF file of the paper to the e-mail: revstat@fc.ul.pt.

Submission of a paper means that it contains original work that has not been nor is about to be published elsewhere in any form.

Submitted manuscripts (text, tables and figures) should be typed only in black, on one side, in double spacing, with a left margin of at least 3 cm and not have more than 30 pages.

The first page should include the name, affiliation and address of the author(s) and a short abstract with the maximum of 100 words, followed by the key words up to the limit of 6, and the AMS 2000 subject classification.

Authors are obliged to write the final version of accepted papers using LaTeX, in the REVSTAT style.

This style (REVSTAT.sty), and examples file (REVSTAT.tex), which may be download to *PC Windows System* (Zip format), *Mackintosh*, *Linux* and *Solaris Systems* (StuffIt format), and *Mackintosh System* (BinHex Format), are available in the REVSTAT link of the National Statistical Institute's Website: http://www.ine.pt/revstat.html

Additional information for the authors may be obtained in the above link.

Accepted papers

Authors of accepted papers are requested to provide the LaTeX files and also a postscript (PS) or an acrobat (PDF) file of the paper to the Secretary of REVSTAT: liliana.martins@ine.pt.

Such e-mail message should include the author(s)'s name, mentioning that it has been accepted by REVSTAT.

The authors should also mention if encapsulated postscript figure files were included, and submit electronics figures separately in .tiff, .gif, .eps or .ps format. Figures must be a minimum of 300 dpi.

Also send always the final paper version to:

Maria José Carrilho
Executive Editor, REVSTAT – STATISTICAL JOURNAL
Instituto Nacional de Estatística
Av. António José de Almeida
1000-043 LISBOA
PORTUGAL

Copyright and Reprints

Upon acceptance of an article, the author(s) will be asked to transfer copyright of the article to the publisher, the INE, in order to ensure the widest possible dissemination of information, namely through the National Statistical Institute's Website (http://www.ine.pt).

After assigning the transfer copyright form, authors may use their own material in other publications provided that the REVSTAT is acknowledged as the original place of publication. The Executive Editor of the Journal must be notified in writing in advance.

Authors of articles published in the REVSTAT will be entitled to one free copy of the respective issue of the Journal and twenty-five reprints of the paper are provided free. Additional reprints may be ordered at expenses of the author(s), and prior to publication.