Nonparametric estimation of time-varying parameters under shape restrictions

Susan Orbe*
Departamento de Econometría y Estadística
Universidad del País Vasco - Spain

Eva Ferreira
Departamento de Econometría y Estadística
Universidad del País Vasco - Spain

Juan Rodríguez-Póo
Departamento de Economía
Universidad de Cantabria - Spain

Summary

In this paper we propose a new method to estimate nonparametrically a time varying parameter model when some qualitative information from outside data (e.g. seasonality) is available. In this framework we make two main contributions. First, the resulting estimator is shown to belong to the class of generalized ridge estimators and under some conditions its rate of convergence is optimal within its smoothness class. Furthermore, if the outside data information is fullfilled by the underlying model, the estimator shows efficiency gains in small sample sizes. Second, for the implementation process, since the estimation procedure envolves the computation of the inverse of a high order matrix we provide an algorithm that avoids this computation and, also, a data-driven method is derived to select the control parameters. The practical performance of the method is demonstrated in a simulation study and in an application to the demand of soft drinks in Canada.

*All correspondence should be directed to Susan Orbe. Dpto. de Econometría y Estadística. Universidad del País Vasco. Avda. Lehendakari Aguirre, 83. 48015 Bilbao, Spain. This work was supported by *Dirección General de Enseñanza Superior del Ministerio de Educación y Ciencia* under research grant PB98-0149, and by the *Universidad del País Vasco* under research grant UPV 038.321-HA129/99.

Keywords: Nonparametric regression, Kernel estimators, Time varying coefficients, Seasonality, Bandwidth selection

1 Introduction

Since their introduction by Cooley and Prescott (1973, 1976) and Rosenberg (1972), time-varying parameters regression models have been used extensively in empirical works. These models permit the regression coefficients to evolve over time, so they can be applied to time series models with parameter instability. We will consider models where there is a linear relationship between the dependent variable and a set of p explanatory variables. Thus, we can write,

$$y_t = \beta_t^T x_t + u_t \qquad t = 1, \ldots, n; \tag{1}$$

where $\{(y_t, x_t), t = 1, \ldots, n\}$ are respectively the observed values for the dependent and the explanatory variables, β_t is a p-vector of time-varying parameters that needs to be estimated, and the errors, u_t, are considered to be identically distributed with zero mean with finite variance σ^2.

In order to solve the estimation problem, and according to the time path structure assumed for the parameters, three main alternative approaches have been conducted in the literature. First, parameters are allowed to vary across subsets of observations within the sample in a deterministic way. Examples of such models include general systematically varying parameter models, seasonal models and switching regression models. A second class of models is that where the parameters are assumed to be stochastic and they can be thought of as being generated by a stationary stochastic process. Finally, a third class of models consists of those where the stochastic parameters are generated by a nonstationary process. Chow (1984), Harvey (1989), and Nicholls and Pagan (1985) present detailed reviews of these approaches. The first framework will be the one considered in this paper. Based on the assumption that the regression coefficients are smooth functions of the time index, Gallant and Fuller (1973) estimate them nonparametrically using piecewise cubic splines. Robinson (1989, 1991) does a similar work using kernel methods.

Although smoothness provides sufficient conditions both to identify and estimate a time varying coefficient model, there might exist other qualitative features in the time path that one would like to model through the parameters behavior. For example, many important economic variables such as ouput, consumption and employment exhibit seasonal patterns. Unfortunately, seasonally adjusted data contain little or no information about seasonal variation in the parameters and also, as Wallis (1974) and Sims (1974) point out, the use of this type of data in dynamic statistical models can lead to serious misspecification errors. On these grounds it would be of interest to include seasonal variation in the coefficients time shape.

Motivated by the previous reasoning, in this paper, we are interested in the nonparametric estimation of the regression coefficients, under the presence of seasonality or any other type of shape constraints. Among the existent techniques that allow for constraints in nonparametric models (see Mammen, Marron, Turlach and Wand, 1998), those that incorporate this information in a constrained optimization fashion appear to be the most appropiated to solve this problem. Rodríguez-Póo (1999) and Ferreira, Núñez-Antón and Rodríguez-Póo (2000) propose several alternative criterion functions. In our setting, the optimization problem is a modification of the (smoothed) local likelihood approach followed by Robinson (1989), where the constraints account for the restrictions in the time varying path for the coefficients. The resulting estimator will be the solution of this constrained optimization problem and it turns out that it can be included within the class of generalized ridge estimators. This fact presents at least two advantages. First, it enables to derive an efficient algorithm to make the estimation process feasible without a high computing cost and second, it makes the analytical study of the statistical properties easier. Furthermore, the method will encompass a great variety of time-varying parameter models: systematically varying seasonal coefficients models, the bayesian seasonal smoothness priors estimator of Gersowitz and MacKinnon (1978) and finally, the estimator proposed in Robinson (1989).

The rest of the article will go as follows. Section 2 presents the estimation procedure and the algorithm. In Section 3 the asymptotic bounds of the estimator are derived as well as its asymptotic distribution. From these results we obtain the conditions in the control parameters for the consistency of the estimators. These results are the extension to the p-dimensional case of those obtained by Orbe et al (2000). A data-driven method to select the control parameters is given in Section 4. Finally, in Section 5 the performance of the methodology is analyzed through a Monte Carlo study. The results from the simulations indicate that the estimator procedure is able to detect very different cases in terms of smoothness and seasonality. Furthermore, we apply the methodology to the analysis of the demand of soft drinks in Canada. The results are consistent with those coming from other empirical works. The Appendix has the proofs of the main results stated in Sections 2 and 3.

2 The constrained nonparametric estimator

In the model introduced in the previous section we additionally consider the following assumptions:

(A.1) (Smoothness) Each component of β_t, $\beta_{it} = \beta_i\,(t/n)$ is a smooth function such that $\beta_i \in C^2[0,1]$ for all $i = 1, \ldots, p$.

To account for the seasonal restriction, first define s_n as the number of the observations per season, such that s_n/n remains constant as n increases. To work with a simpler notation, we will drop the subscript in s_n from now on.

(A.2) (Seasonality) For some values ρ_i and for each of the p sequences

$$n^{-1} \sum_{t=s+1}^{n} \left(\beta_{it} - \beta_{i(t-s)}\right)^2 \leq \rho_i, \tag{2}$$

Note that one implication of assumption (A.1) is that there exist some constants ρ_i^* such that $n^{-1} \sum_{t=s+1}^{n} \left(\beta_{it} - \beta_{i(t-s)}\right)^2 \leq \rho_i^*$ for each $i = 1, \ldots, p$. Hence ρ_i controls the degree of seasonality; i.e., if $\rho_i \geq \rho_i^*$ no seasonality is imposed at all, but as far as $\rho_i < \rho_i^*$ we are imposing stronger seasonal patterns.

The estimates of all parameters $\{\beta_{it}\}_{t=1}^{n}$ will result as the solution to the following optimization problem:

$$\begin{gathered}
\min \sum_{r=1}^{n} \sum_{t=1}^{n} K_{rt}(y_t - \beta_{1r}x_{1t} - \ldots - \beta_{pr}x_{pt})^2 \\
s.t. \begin{cases} n^{-1} \sum_{t=s+1}^{n} \left(\beta_{1t} - \beta_{1t-s}\right)^2 \leq \rho_1 \\ \qquad \vdots \\ n^{-1} \sum_{t=s+1}^{n} \left(\beta_{pt} - \beta_{pt-s}\right)^2 \leq \rho_p \end{cases}
\end{gathered} \tag{3}$$

where $K_{rt} = (nh)^{-1}K((r-t)/nh)$. For the function $K(\cdot)$ we assume the following:

(A.3) The function $K(\cdot)$ is a second order kernel with compact support $\Omega = [-1,1]$. We also assume that its Fourier is absolutely integrable and that $\int_\Omega u^4 K(u)du$, $\int_\Omega K^4(u)du$, are strictly positive and finite.

A direct application of the natural extension of Proposition 1 in Orbe et al (2000) to the p dimensional case allows us to write in matrix notation the optimization problem referred in (3) in a closed form as,

$$\min_{\beta} (Y - X\beta)^T W (Y - X\beta) + \beta^T A_\lambda \beta \tag{4}$$

where $Y = i_n \otimes [Y_1 \ldots Y_n]^T$ and $X = [I_n \otimes x]$, being i_n a unit vector of order n and I_n a n order identity matrix. The data matrix x is of order $n \times p$. The vector of coefficients β is such that $\beta^T = (\beta_1^T \ldots \beta_n^T)$. The weight matrix W is of order n^2 with the following diagonal structure: $W = diag(w_1 \ldots w_n)$, where each submatrix w_r is $diag(K_{r1} \ldots K_{rn})$.

The term $\beta^T A_\lambda \beta$ takes into account the p seasonal restrictions. The matrix A_λ is defined as $\lambda^\star \bar{R}^T \bar{R}$, where

$$\bar{R} = \begin{bmatrix} 1 & 0 & \overbrace{0\ldots0}^{ps-2} & -1 & 0 & \ldots & \ldots & 0 \\ 0 & 1 & 0 & \overbrace{0\ldots0}^{ps-2} & -1 & 0 & \ldots & 0 \\ \vdots & \ddots & & & \ddots & & & \vdots \\ 0 & & \ldots & \ldots & \ldots & 1 & \overbrace{0\ldots0}^{ps-1} & -1 \end{bmatrix}$$

and $\lambda^\star = I_{n-s} \otimes \lambda$. The vector λ contains the different seasonal control parameters $\lambda = [\lambda_1 \ldots \lambda_p]^T$. Now, once the values of h and $\{\lambda_i\}_{i=1}^p$ are fixed, the resulting estimator must solve the following system of equations

$$\left[X^T W X + A_\lambda\right] \widehat{\beta}_\lambda = X^T W Y \tag{5}$$

and, if in addition we assume

(A.4) $Rank(w_r^{1/2} x) = p < n \quad \forall h \quad \forall r = 1, \ldots, n,$

then, the estimator has the closed form

$$\widehat{\beta}_\lambda = \left[X^T W X + A_\lambda\right]^{-1} X^T W Y. \tag{6}$$

This expression shows the relationship between $\widehat{\beta}_\lambda$ and the class of generalized ridge estimators. These estimators present a great variety of optimality properties (see Li, 1986), and this ridge structure will enable us to compute the parameters in a feasible way. Furthermore, note that in order to compute the estimators of the coefficients (6), we need to compute the inverse of $\left[X^T W X + A_\lambda\right]$ whose order is np that difficults the practical estimation of the model. To overcome this problem we provide an algorithm that instead of computing this inverse it only needs the calculation of several inverses of order p. The derivation of the algorithm is based on the same idea than the one developed in Lütkepohl and Herwatz (1996) and it consists basically in the decomposition of the system of normal equations given in (5) and its use to compute the estimators in a recursive form. We present the main steps for the implementation of the algorithm.

1. Construct $\mathcal{Z}_r = \begin{cases} x^T w_r x + \triangle_p & \text{if } r = 1, \ldots, s \\ x^T w_r x + 2\triangle_p & \text{if } r = s+1, \ldots, n-s \\ x^T w_r x + \triangle_p & \text{if } r = n-s+1, \ldots, n \end{cases}$

   where $\triangle_p = diag\{\lambda_1 \ldots \lambda_p\}$ and $x^T w_r x$ is a p order matrix with $\sum_{t=1}^n K_{rt} x_{it} x_{jt}$ as generic (ij) term.

   Construct also the vector of dimension p, $\Upsilon_r = \left[\sum_{t=1}^n K_{rt} x_{1t} y_t \ldots \sum_{t=1}^n K_{rt} x_{pt} y_t \right]^T$

2. For $j = 1, \ldots, s$ compute:

   $a_{0,j} = \mathcal{Z}_j^{-1}; \ldots;$

   $a_{t,j} = \left[\mathcal{Z}_{2t+j} - \triangle_p \cdot a_{(t-1),j} \cdot \triangle_p\right]^{-1}; \ldots;$

   $a_{\left[\frac{n-s}{s}\right],j} = \left[\mathcal{Z}_{[n-s]+j} - \triangle_p \cdot a_{[n-2s],j} \cdot \triangle_p\right]^{-1}$

3. For each $j = 1, \ldots, s$, calculate;

   $b_{0,j} = a_{0,j} \cdot \Upsilon_j; \ldots;$

   $b_{t,j} = a_{t,j} \cdot \Upsilon_{ts+j} + a_{t,j} \cdot \triangle_p \cdot b_{(t-1),j}; \ldots;$

   $b_{\left[\frac{n-s}{s}\right],j} = a_{n-s/s,j} \cdot \Upsilon_{n-s+j} + a_{n-s/s,j} \cdot \triangle_p \cdot b_{n-2/s,j}$

4. Finally, for each $j = 1, \ldots, s$ solve:

   $c_{\left[\frac{n-s}{s}\right],j} = b_{\left[\frac{n-s}{s}\right],j} = \widehat{\beta}_{\lambda j+n-s};$

   $c_{\left[\frac{n-2s}{s}\right],j} = a_{\left[\frac{n-2s}{s}\right],j} \cdot \triangle_p \cdot c_{\left[\frac{n-s}{s}\right],j} + b_{\left[\frac{n-2s}{s}\right],j}; \ldots;$

   $c_{1,j} = a_{1,j} \cdot \triangle_p \cdot c_{2,j} + b_{1,j} = \widehat{\beta}_{\lambda j+s};$

   $c_{0,j} = a_{0,j} \cdot \triangle_p \cdot c_{1,j} + b_{0,j} = \widehat{\beta}_{\lambda j}$

The last step provides the coefficients corresponding to each exogenous variable associated to the j-th moment of time. That is, $\widehat{\beta}_{\lambda j}$ contains the p coefficients of the j-th time. This form allows us to emphasize the differences between the estimator proposed by Robinson (1989,1991) and our estimator. In fact, if we set $\lambda_1 = \cdots = \lambda_p = 0$ then we obtain the estimator as the one proposed by Robinson,

$$\widehat{\beta} = \left[X^T W X\right]^{-1} X^T W Y. \tag{7}$$

Therefore, the seasonal restrictions introduce cross restrictions among parameters at different time periods, being this the reason to do the estimation procedure in a global way. Each estimated coefficient is a function of the whole sample through the smoothing parameter h and the seasonal penalty parameters vector λ. If we only had the smoothness restriction, the procedure would become easier and, for each time value, the vector of coefficients could be computed separately from the rest.

In order to understand the influence of the different control parameters, h and λ, on the estimator, we analyze some simple cases. As far as the components in λ increase, the seasonal constraints will have a higher influence since the estimation process will penalize heavily the differences between the values for the same seasonal coefficients. Under this setting and if the smoothing parameter is zero, the resulting estimator turns out to be the standard OLS estimator for the regression parameters in a seasonal dummy linear regression model. Otherwise, if some smoothness is required, we end with the same result than in Gersovitz and Mackinnon (1978), who define this context as "smooth seasonality". On the other side, where no seasonal restriction is imposed, maintaining the smoothness assumptions recovers the estimator proposed in Robinson (1989), as mentioned before.

3 Asymptotic Results

In order to measure the accuracy of the estimation procedure, we need to study some error measure. A measure could be the usual average of the squared differences between the data and the true values for the regression curve. However, since our main interest is to provide good estimators for the coefficients in the regression curve, we propose to use a measure of discrepancy between the coefficient estimators and their true values. Concretely, we will use the mean average square error defined by the following expression:

$$\begin{aligned} MASE(\widehat{\beta}_\lambda) &= (np)^{-1}\sum_{i=1}^{p}\sum_{t=1}^{n} E(\widehat{\beta}_{\lambda it}-\beta_{it})^2 \\ &= (np)^{-1}\textstyle\sum_{i=1}^{p}\sum_{t=1}^{n}\left\{(E(\widehat{\beta}_{\lambda it}-\beta_{it}))^2+E(\widehat{\beta}_{\lambda it}-E(\widehat{\beta}_{\lambda it}))^2\right\} \\ &= S^2(\widehat{\beta}_\lambda)+V(\widehat{\beta}_\lambda) \end{aligned}$$

where $\widehat{\beta}_{\lambda it}$ denotes the estimate of the coefficient β_{it}.

Previous to other assumptions we introduce some additional notation. Let us define M as the symmetric matrix of order p whose elements are $m_{ij} = E(x_{it}x_{jt})$. $\mathcal{M}$ is the matrix $I_n \otimes M$ and $\mathcal{M}_\lambda = (\mathcal{M}+A_\lambda)^{-1}$. $\mathcal{S}$ will be $(\mathcal{S}_1^T \ldots \mathcal{S}_n^T)^T$, where the i-th element of the r-th subvector, $\mathcal{S}_r$, of $\mathcal{S}$ is $\mathcal{S}_{ri} = \sum_{j=1}^{p} m_{ij}\beta''_{jr}$. We also consider the functionals of the kernel, $c_k = \int K^2(u)du$, $c_4 = \int K^4(u)du$ and $d_k = \int u^2K(u)du$.

Now, in order to allow for some dependence on the data generating process let us assume

(A.5) The random variables $\{x_t\}$ and $\{u_t\}$ are statistically independent and, the sequence $\Phi_t = \{u_t, x_{1t} \ldots, x_{pt}\}_{t \geq 1} \subset \Re \times \Re^p$ is α-mixing. That is, for any n, $k \in Z^+$ and for any pair of sets $A \in \sigma(\Phi_1, \ldots, \Phi_k)$ and $B \in \sigma(\Phi_{k+n} \Phi_{k+n+1} \ldots)$, there exists a sequence of constants $\alpha(n)$ tending to zero such that $|P(A \cap B) - P(A)P(B)| \leq \alpha(n)$.

(A.6) For all t and for all $i, j = 1, \ldots, p$:

$$0 < m_{ij} < \infty \quad i, j = 1, \ldots, p$$

$$0 < E(x_{it}^2 x_{jt}^2) < \infty \quad i, j = 1, \ldots, p$$

and furthermore, M is positive definite.

(A.7) The sequence of constants in the α-mixing condition is such that $n^{-2} \sum_{k=0}^{n} k\alpha(k) = o(n^{-6/5})$.

(A.8) $E|u_t|^{2+\delta} < \infty$ for some $\delta > 0$.

We proceed to state the main results related to the consistency of the estimator and its asymptotic distribution.

Theorem 1 *Given model (1) and, under the assumptions (A.1) to (A.7), we have that*

$$\begin{aligned} S^2(\widehat{\beta}_\lambda) &= \frac{d_k^2 h^4}{4}(np)^{-1}\mathcal{S}^T \mathcal{M}_\lambda^2 \mathcal{S} + (np)^{-1}\beta^T A_\lambda \mathcal{M}_\lambda^2 A_\lambda \beta + \\ & \quad \frac{d_k h^2}{2}(np)^{-1}\beta^T A_\lambda \mathcal{M}_\lambda^2 \mathcal{S} + o((\sum_{j=1}^{p} \lambda_j + h^2)^2) \\ &= O((\sum_{j=1}^{p} \lambda_j + h^2)^2) \end{aligned}$$

$$V(\widehat{\beta}_\lambda) = \frac{\sigma^2 c_k}{nh}\frac{1}{np} tr \mathcal{M}_\lambda \mathcal{M} \mathcal{M}_\lambda + o((nh)^{-1}) = O((nh)^{-1})$$

when $h \longrightarrow 0$, $\lambda \longrightarrow 0$ *and* $nh \longrightarrow \infty$, *as* $n \longrightarrow \infty$.

The first consequence is the consistency of the estimator. Corollary 1 shows the rate of convergence of the mean average squared error. Note that under some specific rates for the control parameters, the proposed nonparametric estimator achieves a rate of convergence that was shown by Stone (1982) to be optimal for this smoothness class. Corollary 2 states that, if the underlying time-path of the coefficients follows a periodic pattern, the proposed estimator is more efficient than the unrestricted nonparametric estimator proposed by Robinson.

Corollary 1 *Under the same assumptions than in Theorem 1 and if, in addition, $h \sim n^{-1/5}$ and for each j, $\lambda_j = O(n^{-2/5})$, then*

$$MASE(\hat{\beta}_\lambda) = O(n^{-4/5})$$

Corollary 2 *Under the same assumptions than in Theorem 1 and if, for each j, $\beta_j(\cdot)$ is a periodic function of period s/n, then*

$$MASE(\widehat{\beta}_\lambda) < MASE(\widehat{\beta})$$

for all $h > 0$ and any $\lambda_j > 0 \quad j = 1, \ldots, p$.

The asymptotic distribution of the estimator is provided in Theorem 2, while Theorem 3 states the consistency of the variance estimator for the error term.

Theorem 2 *Under the assumptions (A.1) to (A.8) and if in addition $h = o(n^{-1/5})$, the asymptotic distribution for the proposed estimator is*

$$(nh)^{1/2}\left(\widehat{\beta}_\lambda(\tau) - \beta(\tau)\right) \xrightarrow{d} N\left[0, \sigma^2 c_k M^{-1}\right] \tag{8}$$

Theorem 3 *Under the same assumptions than in Theorem 1, $\hat{\sigma}^2_\lambda = n^{-1}\sum_{t=1}^{n}(y_t - x_t^T\widehat{\beta}_{\lambda t})^2$ is a consistent estimator of σ^2.*

All these results are the extension of those proved in Orbe et al (2000) to the context of p explanatory variables. Although the results are quite similar, the proofs need to be revised since, unfortunately, they cannot be derived straightforwardly from the univariate case. At this point, we would like to remark that there is no course of dimensionality because the model is additive.

4 Selection of the control parameters

Once we have the desired properties of the estimator, we must approach the practical problem of selecting the control parameters, which becomes an uneasy task when, as in this case, we have several parameters to choose. To our knowledge, there exist at least three possible ways to select the control parameters in a nonparametric estimation setting. Leave-one-out techniques, plug-in methods and penalized residual sum of squares. Härdle (1990) and Wand and Jones (1995) provide detailed discussions of each. In our context the leave-one-out procedures are computationally intensive and the plug-in methods can be applied only if the expressions for the optimum control parameters are known. Unfortunately this is not the case. On these grounds we consider a method based on a penalized residual sum of squares. Thus we propose to choose the different control parameters solving the following optimization problem,

$$\min_{(h,\lambda)} n^{-1}\sum_{t=1}^{n}(y_t - \hat{\beta}_{\lambda 1t}x_{1t} - \ldots - \hat{\beta}_{\lambda pt}x_{pt})^2 \cdot p(h,\lambda). \tag{9}$$

As it can be remarked from other procedures that can be included within the class of penalty methods (Rice criterion or Generalized Cross-Validation) the penalty term is a function of the so called projection or hat matrix; that is, the matrix that allows us to write the estimates as a linear combination of the data. In this context such a matrix is not available so we propose an ad-hoc method that will be described along the following lines. First, we consider a penalty function additive in the control parameters; that is, $p(h,\lambda) = p(h) + p(\lambda_1) + \ldots + p(\lambda_p)$[1]. We select the smoothness parameter penalty function ($p(h)$) under no seasonality restrictions and each seasonal parameter penalty function ($p(\lambda_i)$) will be selected under no smoothness restriction.

Smoothness penalty: Under the assumption of no seasonality, $\lambda = 0$, we construct an auxiliary regression for each r moment of time as:

$$y_t(r) = \beta_{1r}x_{1t}(r) + \ldots + \beta_{pr}x_{pt}(r) + u_t(r) \quad t = 1,\ldots,n \tag{10}$$

where $y_t(r) = k_{rt}y_t$, $x_{jt}(r) = k_{rt}x_{jt}$ and $u_t(r) = k_{rt}u_t$ for $j = 1,\ldots p$. In this context we define the projection matrix as $H^r(h) = w_r^{1/2}x(x'w_r x)^{-1}x'w_r^{1/2}$ for all r. Given that n time observations are available we define the matrix $H(h)$ as the block diagonal matrix $diag\{H^r(h)\}_{r=1}^{n}$, which trace is np and it does not depend on the smoothness parameter. We propose the following penalty function for the smoothness parameter:

$$p(h) = \frac{1}{[(n^2h)^{-1}tr(I - H(h))]^2}. \tag{11}$$

Seasonal penalty: Under the assumption of no smoothness, $h = 0$, and given the j-th exogenous variable, we consider the next auxiliary regression for the ℓ-th season:

$$y_{\ell i} = \alpha_{j\ell i}x_{j\ell i} + u_{\ell i} \quad i = 1,\ldots,f = n/s \tag{12}$$

subject to the seasonal constraint $\alpha'_{j\ell}R'_f R_f \alpha_{j\ell} \leq \delta_j$, with $\delta_j \geq 0$. $y_{\ell i}$ stands for the i-th element of the $f = n/s$ order vector y_ℓ, that contains the observations belonging to the ℓ-th season, $y_\ell = (y_{\ell 1}\ldots y_{\ell f})^T$. The vectors $x_{j\ell}$ and $\alpha_{j\ell}$ and their elements are defined similary. The value $\alpha_{j\ell i}$ is

[1]Although we use the same notation for all penalty functions, the context will clarify which is considered and we think that this makes the notation simpler.

the coefficient associated to the i-th observation in season ℓ of the j-th explanatory variable. The matrix R_f collects the seasonal structure imposed among the coefficients.

$$R_f = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 \\ 0 & 1 & -1 & \dots & 0 \\ & \ddots & \ddots & & \\ 0 & & & 1 & -1 \end{bmatrix} \tag{13}$$

Then, given the j-th variable and the ℓ-th season we define the projection matrix associated to the auxiliary regression (12) as:

$$H_\ell(\lambda_j) = diag(x_{j\ell}) \left[diag(x_{j\ell}) + \lambda_j R_f' R_f\right]^{-1} diag(x_{j\ell})$$

and by adding the seasons we have

$$H_\ell^{\lambda_j} = \sum_{\ell=1}^{s} H_\ell(\lambda_j) = \sum_{\ell=1}^{s} diag(x_{j\ell}) \left[diag(x_{j\ell}) + \lambda_j R_f' R_f\right]^{-1} diag(x_{j\ell}).$$

We propose the penalty function for the j-th parameter that regulates the imposed seasonality on the sequence of parameters corresponding to the j-th exogenous variable as:

$$p(\lambda_j) = \frac{1}{\left[n^{-1} tr(I - H(\lambda_j))\right]^2}. \tag{14}$$

5 A simulation study

The purpose of this section is to analyze the practical performance of the estimation procedure. That is, the objective is not only to check the adequacy of the estimator but we also want to analyze the data driven method of selecting the control parameters as well. In order to do so, we have performed several situations for this analysis and, in all cases, the general model for the data generating process is:

$$y_t = \beta_{1t}x_{1t} + \beta_{2t}x_{2t} + \beta_{3t}x_{3t} + u_t \quad u_t \sim IN(0, \sigma^2 = 3000) \quad t = 1, \dots, n \tag{15}$$

We have considered four different situations for the coefficients: constant, strict seasonal pattern, smooth with no seasonality and coefficients presenting a mixture between smoothness and a seasonal pattern. From each specification we have generated 100 replications of sample size $n = 180$. Since we consider monthly data, the overall period corresponds to 15 years. The selection of the control parameters has been made using the method described in Section 4. For the sake of simplicity, we only present the results for the last case. In fact, this is the case where the method must detect both, smoothness and seasonality. For the rest of the cases the results are similar and they can be obtained from the authors upon request. In the selection of the control parameters we have considered a grid of possible parameters $(h, \lambda_1, \lambda_2, \lambda_3)$; where h goes from 0 to one, with intervals of length 0.1 and λ can take values between 0 and 10^6. In the case of the mixture of both effects, we have generated the explanatory variables as $x_1 \sim U(20, 80)$, $x_2 \sim U(5, 36)$ and $x_3 \sim U(0, 46)$. The coefficients are defined as:

$$\beta_{1t} = 10 * \sin(\pi t/n) + S_{1t}$$
$$\beta_{2t} = 10 * \cos(\pi t/n) + S_{2t}$$
$$\beta_{3t} = 10 * \sin(2\pi t/n) + S_{3t},$$

and the seasonal structure is constructed as follows:

t	S_{1t}	S_{2t}	S_{3t}
$t = 12k + 1$	20	40	150
$t = 12k + 2$	40	70	30
$t = 12k + 3$	10	90	70
$t = 12k + 4$	60	20	90
$t = 12k + 5$	90	60	50
$t = 12k + 6$	30	10	20

t	S_{1t}	S_{2t}	S_{3t}
$t = 12k + 7$	20	80	110
$t = 12k + 8$	50	30	70
$t = 12k + 9$	70	110	100
$t = 12k + 10$	100	70	10
$t = 12k + 11$	10	40	50
$t = 13k$	130	10	10

for $k = 0, \ldots, 14$. Thus, the structure of the coefficients has both components of interest since they vary smoothly and present a seasonal pattern.

Figure 1: First estimated coefficient: $\hat{\beta}_1$

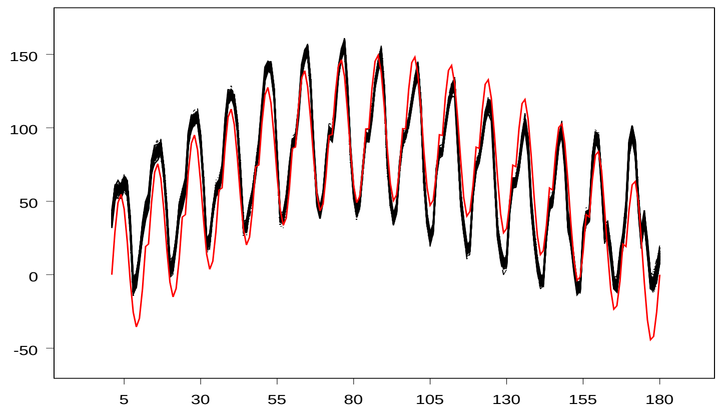


Figure 2: Second estimated coefficient: $\hat{\beta}_2$

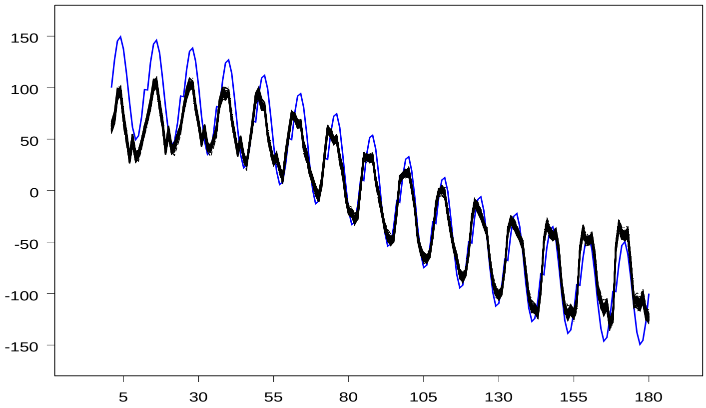
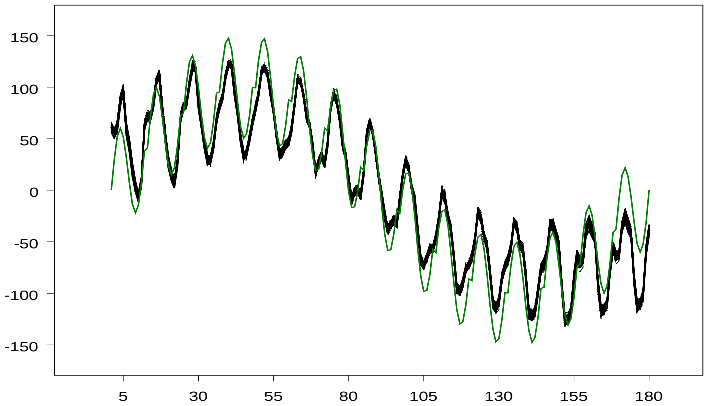


Figure 3: Third estimated coefficient: $\hat{\beta}_3$



As it can be observed from the figures above, the methodology seems to perform well under a mixed situation where both, smoothness and seasonality, are present. As mentioned above, we have obtained similar results for the rest of cases studied and therefore, the method works properly under very different type of situations. However, we have not derived theoretical results for this proposal and the results must be taken with caution. In any case, the results motivate further research in this topic.

To check the perfomance of the methodology in a real framework, we estimate the demand of soft drinks in Canada, a dataset also analysed in Gersovitz and Mackinnon (1978) and Hylleberg (1986). We have 180 monthly seasonaly unadjusted observations from January 1959 to December 1973. The model to be estimated is:

$$\log(C/P)_t \quad = \quad \beta_{1t} + \beta_{2t}\log(SDP/CP)_t + \beta_{3t}\log(FP/CP)_t + \beta_{4t}\log(EXP/P)_t +$$

$$+ \quad \beta_{5t} \log((P - P25)/P)_t + \epsilon_t \tag{16}$$

where:

- C: volumen index of soft drink production, 1961=100.
- P: total noninstitutional population.
- $P25$: total noninstitutional population, aged 25 and over.
- SDP: consumer price index for soft drinks.
- FP: consumer price index for food.
- CP: consumer price index for all items excluding food.
- $QEXP$: total personal expenditure on nondurable goods in constant dollars, quartely data.
- EXP: interpolation of $QEXP$. If the month is the middle one, EXP is equal to $QEXP$; otherwise, EXP is equal to two-third of $QEXP$ plus one-third of $QEXP$ for the adjacent quarter.
- ϵ_t are zero mean errors, independently distributed with common variance.

We have applied the estimation procedure to the data allowing for a possible seasonal pattern in all coefficients. The method has selected a zero value for the smoothness parameter and the maximum degree of seasonality in all coefficients. This means that a dummy specification provides a good model for this analysis and this is the same conclusion than the one obtained in Gersovitz and Mackinnon (1978) and Hylleberg (1986).

Next figures show the estimated coefficients and last figure shows the obtained adjustment.

Figure 4: Estimated coefficient: $\hat{\beta}_{1t}$

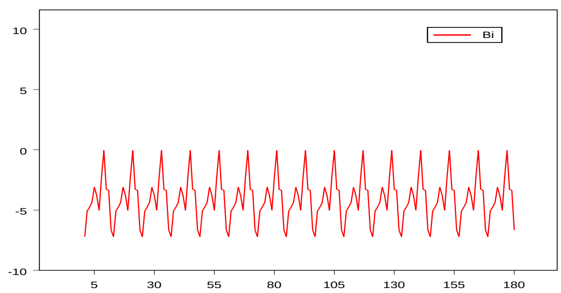


Figure 5: Estimated coefficient: $\hat{\beta}_{2t}$

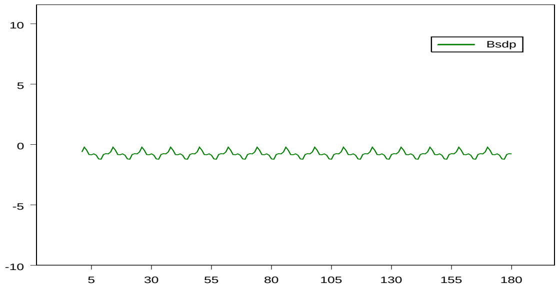


Figure 6: Estimated coefficient: $\hat{\beta}_{3t}$

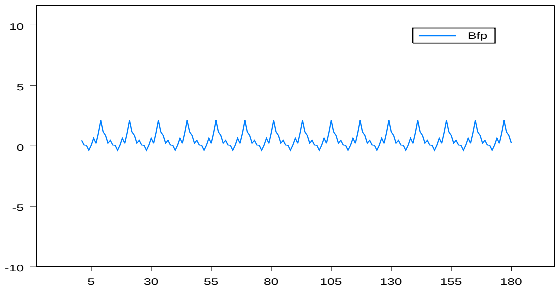


Figure 7: Estimated coefficient: $\hat{\beta}_{4t}$

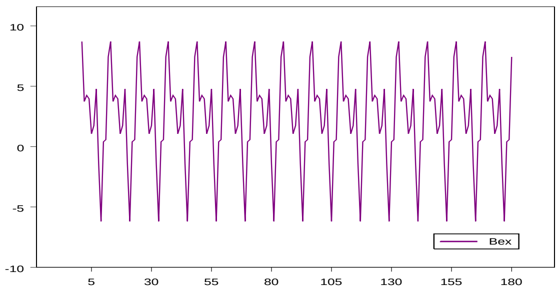


Figure 8: Estimated coefficient: $\hat{\beta}_{5t}$

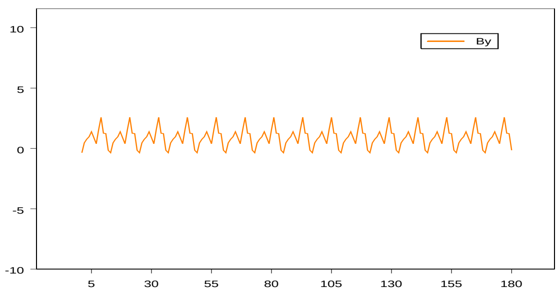
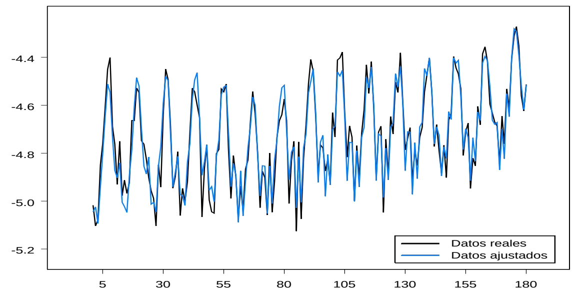


Figure 9: Real data vs adjusted data



APPENDIX

For the proofs of the results stated in Section 3, we will state the following lemmas.

Lemma 1 *Under assumptions (A.3), (A.4) and (A.6), and if the explanatory variables are independent random variables, we have that, as* $n \longrightarrow \infty$, $h \longrightarrow 0$ *and* $nh \longrightarrow \infty$,

$$(nh)^{-1} \sum_t K_t(\tau) x_t x_t^T \longrightarrow M \quad a.s. \tag{17}$$

$$(nh)^{-1} \sum_t K_t^2(\tau) x_t x_t^T \longrightarrow c_k M \quad a.s. \tag{18}$$

where $K_t(\tau) = K((n\tau - t)/nh)$.

Proof:

For (17):

Since the sums in expression (17) are matrices, we will prove the convergence of each element to the corresponding one in the limit. Let us define $Z_{ijt} = K_t(\tau) x_{it} x_{jt}$. Then, $\{Z_{ijt}\}_{t=1}^n$ is a sequence of independent random variables with mean $E(Z_{ijt}) = K_t(\tau) m_{ij}$ and variance $V(Z_{ijt}) = K_t^2(\tau) V(x_{it} x_{jt})$. From assumption (A.6), it follows that $\sum_{t=1}^{\infty} V(Z_{ijt})/t^2 < \infty$ and from this we obtain that $\overline{Z}_{ijt} - E(\overline{Z}_{ijt}) \xrightarrow{a.s.} 0$. Now, taking into account the kernel structure of $K(\cdot)$, the average over the nh elements is $\overline{Z}_{ijt} = (nh)^{-1} \sum_t K_t(\tau) x_{it} x_{jt}$, the mean is $E(\overline{Z}_{ijt}) = m_{ij} (nh)^{-1} \sum_t K_t(\tau)$, and in consequence, $E(\overline{Z}_{ijt}) = m_{ij} + O((nh)^{-1})$. Therefore, the result holds.

For (18):

Let us define $Z_{ijt} = K_t^2(\tau) x_{it} x_{jt}$, that has mean equal to $E(Z_{ijt}) = K_t^2(\tau) m_{ij}$ and variance $V(Z_{ijt}) = K_t^4(\tau) V(x_{it} x_{jt})$ Now, the result is obtained in a similar way than for (17). $\bullet$

The lemma that will be proved now states the particular case of Theorem 1 when the explanatory variables and the error term are considered to have an independent structure.

Lemma 2 *Under the same set of assumptions than in Theorem 1, where we consider the particular case of $\alpha(n) = 0$ rather than assumption (A.7), then the corresponding asymptotic bias and variance expressions in the mean average squared error are:*

$$\begin{aligned} S^2(\widehat{\beta}_\lambda) &= \frac{d_k^2 h^4}{4}(np)^{-1}\mathcal{S}^T\mathcal{M}_\lambda^2\mathcal{S} + (np)^{-1}\beta^T A_\lambda \mathcal{M}_\lambda^2 A_\lambda \beta + \\ & \quad \frac{d_k h^2}{2}(np)^{-1}\beta^T A_\lambda \mathcal{M}_\lambda^2 \mathcal{S} + o((\sum_{j=1}^p \lambda_j + h^2)^2) \\ &= O((\sum_{j=1}^p \lambda_j + h^2)^2) \end{aligned}$$

$$\begin{aligned} V(\widehat{\beta}_\lambda) &= \frac{\sigma^2 c_k}{nh}\frac{1}{np} tr \mathcal{M}_\lambda \mathcal{M} \mathcal{M}_\lambda + o((nh)^{-1}) \\ &= O((nh)^{-1}) \end{aligned}$$

Proof: First, note that the mean average square error can be written as: $MASE(\widehat{\beta}_\lambda) = (np)^{-1} tr \bar{E}\left[(\widehat{\beta}_\lambda - \beta)^T(\widehat{\beta}_\lambda - \beta)\right]$, where $\widehat{\beta}_\lambda = \left[\widehat{\beta}_{\lambda 1}^T \ldots \widehat{\beta}_{\lambda n}^T\right]^T = \left[X^T W X + A_\lambda\right]^{-1} X^T W Y$.

In order to avoid the problem of dealing with a random expression in the denominator of the $MASE$, we define a modified $MASE$ as

$$MASE^*(\widehat{\beta}_\lambda) = (np)^{-1} tr E\left[W^*(\widehat{\beta}_\lambda - \beta)(\widehat{\beta}_\lambda - \beta)^T W^{*T}\right]$$

where $W^* = [\mathcal{M} + A_\lambda]^{-1}\left[X^T W X + A_\lambda\right]$ and the rest are as defined in Section 3.

Now, we will study the rate of convergence of $MASE^*(\widehat{\beta}_\lambda)$. The process is tedious because of the special structure of the estimator and we will try to detail the main steps, while repeated arguments will be omited. First, we split the expression $W^*(\widehat{\beta}_\lambda - \beta)$ as:

$$\begin{aligned} & W^*(\widehat{\beta}_\lambda - \beta) = [\mathcal{M} + A_\lambda]^{-1}\left(X^T W Y - X^T W X \beta\right) - [\mathcal{M} + A_\lambda]^{-1} A_\lambda \beta = \\ & [\mathcal{M} + A_\lambda]^{-1}\mathcal{B} + [\mathcal{M} + A_\lambda]^{-1} X^T W U - [\mathcal{M} + A_\lambda]^{-1} A_\lambda \beta \end{aligned} \tag{19}$$

where $\mathcal{B} = \left[\mathcal{B}_1^T \ \ldots \mathcal{B}_n^T\right]^T$, being $\mathcal{B}_r$ a p order subvector with i-th element $\sum_{t=1}^n \sum_{j=1}^p K_{rt} x_{it} x_{jt}(\beta_{jt} - \beta_{jr})$. The vector U has a similar structure than Y and it is defined as $U = i_n \otimes [u_1 \ldots u_n]^T$.

Taking into account the expression (19), we compute the $MASE^*$ separating it in the following terms where $\mathcal{M}_\lambda = [\mathcal{M} + A_\lambda]^{-1}$.

$$\frac{1}{np} tr \mathcal{M}_\lambda E\left[\mathcal{B}\mathcal{B}^T\right]\mathcal{M}_\lambda + \tag{20}$$

$$\frac{1}{np} 2 tr \mathcal{M}^* E\left[\mathcal{B}\right]\beta^T A_\lambda \mathcal{M}_\lambda + \tag{21}$$

$$\frac{1}{np} tr\mathcal{M}_\lambda A_\lambda \beta\beta^T A_\lambda \mathcal{M}_\lambda + \tag{22}$$

$$\frac{1}{np} tr\mathcal{M}_\lambda E\left[X^T W U U^T W X\right]\mathcal{M}_\lambda, \tag{23}$$

where the first three terms account for the bias and the fourth corresponds to the variance.

To study the previous sums let us analyze the matrix $E\left[\mathcal{B}\mathcal{B}^T\right]$. The i-th diagonal element of one arbitrary submatrix $\mathcal{B}_r\mathcal{B}_r^T$ in $\mathcal{B}\mathcal{B}^T$ can be decomposed in

$$\sum_j^p \sum_t^n K_{rt}^2 x_{it}^2 x_{jt}^2 (\beta_{jt} - \beta_{jr})^2 + \tag{24}$$

$$\sum_{j\neq j'} \sum_t K_{rt}^2 x_{it}^2 x_{jt} x_{j't} (\beta_{jt} - \beta_{jr})(\beta_{j't} - \beta_{j'r}) + \tag{25}$$

$$\sum_{j\neq j'} \sum_{t\neq s} K_{rt} K_{rs} x_{it} x_{jt} x_{is} x_{j's} (\beta_{jt} - \beta_{jr})(\beta_{j's} - \beta_{j'r}) + \tag{26}$$

$$\sum_{j} \sum_{t\neq s} K_{rt} K_{rs} x_{it} x_{jt} x_{is} x_{js} (\beta_{jt} - \beta_{jr})(\beta_{js} - \beta_{jr}). \tag{27}$$

Clearly, the terms (24) and (25) are negligible with respect to the rest. The term (27) converges to $(h^4 d_k^2/4)\sum_j (m_{ij}\beta''_{jr})^2$ and the term (26) to $(d_k^2 h^4/2)\sum_{j\neq j'} m_{ij}\beta''_{jr} m_{ij'}\beta''_{j'r}$ for $j \neq j'$. So, the leading term in the i-th diagonal element of $E\left[\mathcal{B}_r\mathcal{B}_r^T\right]$ is given by $(d_k^2 h^4/4)\sum_{j,j'=1}^p m_{ij} m_{ij'}\beta''_{jr}\beta''_{j'r}$. With respect the (ii') nondiagonal elements of $\mathcal{B}_r\mathcal{B}_r^T$, the asymptotic expression for the mean is $(d_k^2 h^4/4)\sum_{j,j'=1}^p m_{ij} m_{i'j'}\beta''_{jr}\beta''_{j'r}$. Thus, $E\left[\mathcal{B}_r\mathcal{B}_r^T\right]$ can be expressed asymptotically as $(d_k^2 h^4/4)\mathcal{S}_r\mathcal{S}_r^T$, where we recall that $\mathcal{S}_r$ is a p order vector whose i-th component is $\sum_{j=1}^p m_{ij}\beta''_{jr}$. Following the same steps in the elements of the nondiagonal matrix $E\left[\mathcal{B}_r\mathcal{B}_{r'}^T\right]$, we have that the mean in this case is $E\left[\mathcal{B}\mathcal{B}^T\right] = (d_k^2 h^4/4)\mathcal{S}\mathcal{S}^T$, where $\mathcal{S}$ is given by $\mathcal{S} = \left[\mathcal{S}_1^T \ldots \mathcal{S}_n^T\right]^T$. For (23) we must compute $E\left[X^T W U U^T W X\right]$, which has n square submatrices of order p. Given the independence between the error term and the explanatory variables and using Lemma 2 this mean is equal to $(\sigma^2 c_k/nh)\mathcal{M}$.

The term, (21), depends on $E\left[\mathcal{B}\right]$. This vector has subvectors of order p and, a generic r-th element of the i-th subvector is $\sum_j^p \sum_t^n K_{rt} x_{it} x_{jt} (\beta_{jt} - \beta_{jr})$. In the limit, the mean is $(d_k h^2/2)\ \sum_{j=1}^p m_{ij}\ \beta''_{jr}$, so, $E\left[\mathcal{B}\right] = (d_k h^2/2)\ \mathcal{S}$.

Combining all previous results and after some standard algebra, the resulting $MASE^*(\cdot)$ is:

$$\begin{aligned} MASE^*(\widehat{\beta}_\lambda) &= \frac{d_k^2 h^4}{4}\frac{1}{np}\mathcal{S}^T\mathcal{M}_\lambda^2\mathcal{S} + \frac{1}{np}\beta^T A_\lambda \mathcal{M}_\lambda^2 A_\lambda \beta \\ &+ \frac{d_k h^2}{2}\frac{1}{np}\beta^T A_\lambda \mathcal{M}_\lambda^2 \mathcal{S} + \frac{\sigma^2 c_k}{nh}\frac{1}{np} tr\mathcal{M}_\lambda \mathcal{M}\mathcal{M}_\lambda. \end{aligned}$$

From this, the three first terms are the leading terms stated for the bias and the fourth is the leading term for the variance. Now, it is left to obtain the rates of convergence of the terms from expression above. Since $(1/np)\mathcal{S}^T\mathcal{M}_\lambda^2\mathcal{S}$ is finite, the first term is of order $O(h^4)$.

For the second term, we can check the order if we reorganize its elements adecuately. Let be $\tilde{\beta} = (\tilde{b}_1^T \ldots \tilde{b}_p^T)$ the reorganized vector. Each subvector $\tilde{b}_j$, contains the n coefficients associated to the j-th explanatory variable, $\tilde{b}_j = (\beta_{j1} \ldots \beta_{jn})$. Given the new structure of $\tilde{\beta}$, we organize the rest of matrices in the same way. Let be $\tilde{A}_\lambda = diag(\lambda_1 \ldots \lambda_p) \otimes A_n$ and $\tilde{\mathcal{M}}_\lambda = \left[(M \otimes I_n) + \tilde{A}_\lambda\right]^{-1}$ where $M = E(x_t x_t^T)$. Under the new notation we can write

$$\frac{1}{np}\beta^T A_\lambda \mathcal{M}_\lambda^2 A_\lambda \beta = \frac{1}{np}\tilde{\beta}^T \tilde{A}_\lambda \tilde{\mathcal{M}}_\lambda^2 \tilde{A}_\lambda \tilde{\beta} = \frac{1}{np}\sum_{i=1}^{p}\sum_{j=1}^{p} \lambda_i \lambda_j \tilde{\beta}_j^T A \tilde{\mathcal{M}}_{ij} A \tilde{\beta}_i$$

where $\tilde{\mathcal{M}}_{ij}$ is the ij-th submatrix of dimension $n \times n$ of the whole matrix $\tilde{\mathcal{M}}_\lambda^2$. Now, it is easy to check from this last expression that the order of this term is $O((\lambda_1 + \ldots + \lambda_p)^2)$. The order of the third term is bounded by the two above ones and, since $(np)^{-1} tr \mathcal{M}_\lambda \mathcal{M} \mathcal{M} \lambda$ is bounded uniformly in n, the fourth term of $MASE^*$ is of order $O((nh)^{-1})$.

Finally, we must show that the use of $MASE^*$ instead of $MASE$ does not affect to the rate of convergence. From Theorem 6.1 in Vieu (1991), $MASE^*(\widehat{\beta}_\lambda) - MASE(\widehat{\beta}_\lambda) = o(MASE(\widehat{\beta}_\lambda))$ if $\left[X^T W X + A_\lambda\right]$ converges almost surely to $[\mathcal{M} + A_\lambda]$, and this result comes from Lemma 1 •

Lemma 3 *Under the same assumptions than in Theorem 1, we have that*

$$MASE(\widehat{\beta}_\lambda) - \overline{MASE}(\widehat{\beta}_\lambda) = o(\overline{MASE}(\widehat{\beta}_\lambda))$$

where $\overline{MASE}(\cdot)$ is defined as the mean average square error considered in the context of independent variables considered in Lemma 2. Therefore, the term $MASE(\cdot)$ corresponds to the mean average square error in the context of α-mixing dependence.

Proof: Looking at the expressions from (20) to (22), the different terms are $E\left[\mathcal{B}\mathcal{B}^T\right]$, $E\left[X^T W U U^T W X\right]$ and $E\left[\mathcal{B}\right]$. We compare these mathematical expectations with the corresponding to the independent case, for which we will use the notation $\bar{E}$.

We begin with $E\left[\mathcal{B}\mathcal{B}^T\right]$. The terms (24) and (25) are still negligible since $E(x_{it}^2 x_{jt}^2)$ and $E(x_{it}^2 x_{jt} x_{j't})$ are bounded. Now, the difference between the sum of the terms (27) and (26) with respect to the analogous expression in the independent case is:

$$\sum_j \sum_{t \neq s} K_{rt} K_{rs} (\beta_{jt} - \beta_{jr})(\beta_{js} - \beta_{jr}) \left[E(x_{it} x_{jt} x_{is} x_{js}) - \bar{E}(x_{it} x_{jt} x_{is} x_{js})\right] +$$
$$\sum_{j \neq j'} \sum_{t \neq s} K_{rt} K_{rs} (\beta_{jt} - \beta_{jr})(\beta_{j's} - \beta_{j'r}) \left[E(x_{it} x_{jt} x_{is} x_{j's}) - \bar{E}(x_{it} x_{jt} x_{is} x_{j's})\right]$$

For the first term, an application of Proposition 1 in Hart and Vieu (1990) shows that it can be bounded by $C(nh)^{-2} \sum_{t \neq s} \alpha(|t-s|)$. The constant C comes from the product between a finite and positive constant and a term C_o, such that $|K_{rt}x_{it}x_{jt}(\beta_{jt}-\beta_{jr})|\,|K_{rs}x_{is}x_{js}(\beta_{js}-\beta_{js})| \leq C_o < \infty$. Now, by assumption $(A.7)$ we have that this bound can be expressed as,

$$2C(nh)^{-2}\sum\nolimits_{t\leq s}\alpha(t-s) = 2C(nh)^{-2}\sum\nolimits_{l=1}^{n-1} l\alpha(l) = 2Ch^{-2}o(n^{-6/5}) \qquad (28)$$

Following similar steps, the second term is bounded by $C'(nh)^{-2}\sum\nolimits_{t\neq s}\alpha(|t-s|)$, where the constant C'_o corresponds to the product of some positive constant and C'_o such that

$$|K_{rt}x_{it}x_{jt}(\beta_{jt}-\beta_{jr})|\,|K_{rs}x_{is}x_{i's}(\beta_{j's}-\beta_{j's})| \leq C'_o < \infty$$

In consequence, we obtain the order in (28). For the nondiagonal elements of $E(\mathcal{B}\mathcal{B}^T)$ the same order is obtained in a similar way.

The difference corresponding to the expression (23) is $E\left[X^TWUU^TWX\right]-\bar{E}\left[X^TWUU^TWX\right]$ and it has as generic ij element, $\sum_{t\neq s} KK_{rs}\left[E(x_{it}x_{js}u_tu_s)-\bar{E}(x_{it}x_{js}u_tu_s)\right]$. For the last difference, note that the i-th element of the r-th subvector of $E\left[\mathcal{B}\right]-\bar{E}\left[\mathcal{B}\right]$ is $\sum_t K_{rt}(\beta_{1t}-\beta_{1r})\left[E(x_{it}x_{1t})-\bar{E}(x_{it}x_{1t})\right]+\ldots+\sum_t K_{rt}(\beta_{pt}-\beta_{pr})\left[E(x_{it}x_{pt})-\bar{E}(x_{it}x_{pt})\right]$.

For all these elements we obtain the same order as in (28). All together with the fact that $h = O(n^{-1/5})$ leads to $MASE(\hat{\beta}_\lambda)-\overline{MASE}(\hat{\beta}_\lambda) = o(h^4)$. Therefore, the leading order of convergence is the same for both measure. •

Proof of Theorem 1: The proof is now straightforward using Lemmas 1, 2 and 3.

Proof of Corollary 2: Under strict seasonality then $\beta^TA_\lambda = 0$ and the unique term remaining in the bias is $d_k^2h^4(4np)^{-1}\mathcal{S}^TM_\lambda^2\mathcal{S}$, which is smaller than the one corresponding to the case $\lambda = 0$. The reason is that, with positive values of λ_j, A_λ is a semidefinite positive matrix and so is $\mathcal{S}M^2\mathcal{S}-\mathcal{S}M_\lambda^2\mathcal{S}$. Since the variance is always smaller for positive values of the seasonal parameters, the statement holds. •

Proof of Theorem 2: The proof is based on Theorem 3.2 of White and Domowitz (1984). First, we obtain the asymptotic normality for the estimator $\widehat{\beta}_\lambda$ corresponding to the r-th time period when $\lambda = 0$, $\hat{\beta}_r = \left[\widehat{\beta}_{r1}\ldots\widehat{\beta}_{rp}\right]^T$.

To keep the same notation than in Theorem 3.2 of White and Domowitz, we define $\widetilde{y}_t = \widetilde{x}_t^T\beta_r+e_t$ where $\widetilde{y}_t = K_{rt}^{1/2}y_t$, $\widetilde{x}_t^T = (K_{rt}^{1/2}x_{1t}\ldots K_{rt}^{1/2}x_{pt})$ and $e_t = K_{rt}^{1/2}u_t$. Then, under assumptions (A.1) to (A.8) the mentioned theorem provides that

$$\sqrt{n}\bar{A}_n\bar{B}_n^{-1/2}(\widehat{\beta}_r-\beta_r) \longrightarrow N(0,I_p)$$

where $\bar{A}_n$ and $\bar{B}_n$ are defined by

$$\bar{A}_n = 2n^{-1}\sum_{t=1}^{n} E\left[\tilde{x}_t\tilde{x}_t^T\right] = 2n^{-1}\sum_{t=1}^{n} E(K_{rt}x_tx_t^T)$$

$$
\begin{aligned}
\bar{B}_n &= 4n^{-1}\sum_{t=1}^{n} E\left[e_t^2 \tilde{x}_t \tilde{x}_t^T\right] + 4n^{-1}\sum_{l=1}^{n-1}\sum_{t=l+1}^{n} E\left[e_t e_{t-l}\left(\tilde{x}_t \tilde{x}_{t-l}^T + \tilde{x}_{t-l}\tilde{x}_t^T\right)\right] \\
&= 4n^{-1}\sum_{l=1}^{n} K_{rt} E(u_t^2) E(x_t x_t^T) + II
\end{aligned}
$$

The limit of $n\bar{A}_n$ is $2M$ and the limit of the first term in $n^2 h \bar{B}_n$ is $4\sigma^2 c_k M$. The second term II of $\bar{B}_n$ is the sum of $(n-1)$ matrices which depend on the number of lags, l. Each ij-th element in those matrices is given by $\sum_{t=l+1}^{n} E(K_{rt}K_{rt-l}u_t u_{t-l} x_{it} x_{jt-l})$.

The use of Proposition 1 of Hart and Vieu (1990) allows us again, to relate this term with the analogous in the independent case and we obtain

$$
\sum_{t=1+l}^{n}\left[E(K_{rt}K_{rt-l}u_t u_{t-l} x_{it} x_{jt-l}) - \bar{E}(K_{rt}K_{rt-l}u_t u_{t-l} x_{it} x_{jt-l})\right] = o(h^4)
$$

where $\bar{E}(\cdot)$ denotes the mathematical mean under the assumption of independence.

Hence, we have that lim $\sqrt{nh}\bar{A}_n^{-1}\bar{B}_n\bar{A}_n^{-1} = \sigma^2 c_k M$ and therefore we can write

$$
\sqrt{nh}(\widehat{\beta}(\tau) - \beta(\tau)) \longrightarrow N(0, \sigma^2 c_k M^{-1})
$$

The asymptotic distribution of $\widehat{\beta}_\lambda$ is inmediate, since if $h = o(n^{-1/5})$, the bias is negligible and the seasonal parameters do not affect the rate of convergence. •

Proof of Theorem 3: The convergence in probability is deduced by the inequality of Markov once we show that $E|\hat{\sigma}_\lambda^2 - \sigma^2|$ converges to zero. This expectation can be bounded by

$$
\begin{gathered}
E|\hat{\sigma}_\lambda^2 - \sigma^2| \le n^{-1} E\sum_t x_t^T(\beta_t - \widehat{\beta}_{\lambda t})(\beta_t - \widehat{\beta}_{\lambda t})^T x_t \\
+E|n^{-1}\sum(u_t^2 - \sigma^2)| + n^{-1}E|\sum_t x_t^T(\beta_t - \widehat{\beta}_{\lambda t})u_t|.
\end{gathered}
$$

Theorem 1 provides the convergence to zero of the first term. For the second term we can apply Kintchine's Theorem. Finally, the Cauchy-Swchartz's inequality bounds the cross term by the other two ones, and the result is given. •

References

Chow, G.C. (1984), *Random and Changing Coefficients Models* in Handbook of Econometrics (Z. Griliches and M.D. Intriligator). New York: North Holland, vol 2, 1213-1245.

Cooley, T. F. and E. C. Prescott (1973), An Adaptive Regression Model, *International Economic Review*, **14**, 364-371.

Cooley, T. F. and E. C. Prescott (1976), Estimation in the Presence of Stochastic Parameter Variation, *Econometrica*, **44**, 167-184.

Ferreira, E., V. Núñez-Anton and J. M. Rodríguez-Póo (2000), Semiparametric Approaches to Signal Extraction Problems in Economic Time Series*Computational Statistics and Data Analysis*, **33**, 315-333.

Gallant, A. and W. Fuller (1973), Fitting Segmented Polynomial Regression Models Whose Points Have to be Estimated, *Journal of American Statistical Association* **68**, (1973), 144-147.

Gersovitz, M. and J. Mackinnon, Seasonality in Regression: An Application of Smoothness Priors, *Journal of American Statistical Association* **73**, (1978), 264-273.

Härdle, W. (1990), *Applied Nonparametric Regression*, Cambridge: Cambridge University Press.

Hart, J. and P. Vieu (1990), Data-Driven Bandwidth Choice for Density Estimation based on Dependent Data, *Annals of Statistics* **18**, 873-890.

Harvey, A. C. (1989), *Forecasting Structural Time Series and the Kalman Filter*, Cambridge: Cambridge University.

Hylleberg, S. (1986), *Seasonality in Regression*, London: Academic Press.

Li, K. (1986), Asymptotic Optimality for c_p, c_l and Cross-Validation in Ridge Regression with Applications to Splines, *Annals of Statistics* **14**, 1101-1112.

Lütkepohl, H. and H. Herwatz (1996), Specification of Varying Coefficients Time Series Models vis Generalized Flexible Least Squares, *Journal of Econometrics* **70**, 261-290.

Mammen, E., J. S. Marron, B.A. Turlach and M.P. Wand (1998), A General Framework for Constrained Smoothing. Preprint.

Nicholls, D. F. and A. Pagan, (1985), Varying Coefficient Regression, in *Handbook of Statistics*, Vol. 5, eds. E. J. Hannan, P.R. Krishnaiah, and M. M. Rao, Amsterdam: North Holland, 413-449.

Orbe, S., E. Ferreira and J. M. Rodríguez-Póo (2000), A Nonparametric Method to Estimate Time Varying Coefficients, *Journal of Nonparametric Statistics*, **12**, 779-806.

Robinson, P.M. (1989), *Nonparametric Estimation of Time Varying Parameters* in Statistics Analysis and Forecasting of Economic Structural Change (P. Hackl), Berlin: Springer-Verlag, pp. 253-264.

Robinson, P.M. (1991), *Time Varying Non Linear Regression* in Statistics Analysis and Forecasting of Economic Structural Change (P. Hackl), Berlin: Springer-Verlag, pp. 179-190.

Rodriguez-Poo, J.M. (1999), Constrained Smoothing Splines, *Econometric Theory*, **15**, 114-138.

Rosenberg, B. (1972), The Estimation of Stationary Stochastic Regression Parameters Re-examined, *Journal of the American Statistical Association*, **67**, 650-654.

Sims, C. A. (1974), Seasonality in Regression, *Journal of the American Statistical Association*, **69**, 618-626.

Stone, C. (1982), Optimal Global Rates of Convergence for Nonparametric Regression, *Annals of Statistics* , **12**, 1104-1153.

Vieu P. (1991), Quadratic Errors for Nonparametric Estimates under Dependence, *Journal of Multivariate Analysis*, **39**, 324-347.

Wallis, K.F. (1974), Seasonal Adjustment and Relations Between Variables, *Journal of the American Statistical Association*, **69**, 18-31.

Wand, M.P. and M.C. Jones (1995), *Kernel Smoothing*, London: Chapman and Hall.

White, H. and I. Domowitz (1984), Nonlinear Regression with Dependent Observations, *Econometrica* **52**, 143-161.