Statistica Sinica **16**(2006), 1313-1333

ON THE ESTIMATION AND TESTING OF TIME VARYING CONSTRAINTS IN ECONOMETRIC MODELS

Susan Orbe[1], Eva Ferreira[1] and Juan Rodriguez-Poo[2]

[1] *Universidad del País Vasco and* [2] *Universidad de Cantabria*

Abstract: In this paper, we propose a new nonparametric method to impose time varying restrictions in econometric models. The method relies on the assumption that the structural parameters evolve smoothly over time and they are required to satisfy some pre-specified constraints. Smoothing techniques are employed in order to estimate the parameter sequences. In particular, we propose a local constrained least squares method. The asymptotic properties are derived and a test procedure for the validity of the time varying constraints is derived. The methodology is applied to estimate a time varying demand system and a simulation study is performed to illustrate the estimation method as well as the test procedure.

Key words and phrases: Constrained estimators, nonparametric estimation, time varying coefficients, time varying cross restrictions.

1. Introduction

In empirical specifications of production and consumption systems it is sometimes necessary to impose additional restrictions on the structural parameters. In production theory, regardless of the parametric specification chosen for the production function, we estimate the structural parameters subject to restrictions implied by homogeneity, product exhaustion, symmetry and monotonicity (see Fried, Lovell and Schimdt (1993) and Jorgenson (2000)). Also in consumption, when a demand function is specified, the assumption of linear budget constraints leads to added homogeneity restrictions. Furthermore, some additional restrictions are suggested by economic theory, such as the Slutsky symmetry condition or the Engle condition (see Deaton and Muellbauer (1980)). Usually, once these constraints have been imposed on structural econometric models, it is of interest to test them for validity.

These issues are well known to applied econometricians, and many different procedures to both estimate and test possibly nonlinear systems incorporating linear or nonlinear constraints have been proposed in the relevant literature (see Gourieroux and Monfort (1989)). Most of the theoretical models that have been used by empirical researchers are based on a non-dynamic framework, and this might be the main reason why most of the restrictions quoted above have been considered as time invariant. However, if we have a richer source of data, then there is no reason to keep considering non-dynamic models and, therefore, time invariant restrictions.

In fact, there are many examples where it becomes natural to impose time varying restrictions. In the production setting, Diewert and Wales (1987) propose methods for introducing global curvature conditions in the context of cost function estimation. Meanwhile when estimating a translog cost function, if technical progress is allowed, then linear homogeneous restrictions become time varying. In their paper, they assume that the dual production function does not exhibit technical changes and this implies that the restrictions are time invariant. In the demand setting, consumer demand systems which are linear in variables (in logs) have restrictions that are typically time varying even when considering the elasticities time invariant. Doran and Rambaldi (1997) motivate imposing time varying constraints, instead of the usual "at the mean" one, by consistency and efficiency considerations. First, if the theory implied constraints are non-linear in the data, imposing them at the mean will lead to inconsistency. Second, even if they are linear in the explanatory variables, to impose them at every point may lead to an efficiency gain over the "at the mean" constrained estimators. Such theory-implied constraints include the Engle condition and the Slutsky symmetry condition.

In this paper, we propose a new method of incorporating and testing time varying restrictions in econometric models. Our method relies on the assumption that structural parameters evolve over time with some degree of smoothness while they are required to satisfy some pre-specified constraints.

Many alternative smoothing methods for estimating unrestricted varying coefficient models have been proposed in the literature. Hastie and Tibshirani (1993) propose an algorithm based on backfitting; Cai, Fan and Yao (2000) and Cai, Fan and Li (2000) approach the problem using local polynomials in regression and time series models. Chen and Tsay (1993) and Chen and Liu (2001) use kernels in a similar context. Robinson (1989, 1991) proposes a local least squares method. However, all previous nonparametric approaches have paid no attention to the problem of incorporating restrictions on the coefficients. Orbe, Ferreira and Rodriguez-Poo (2000, 2005) are the first to consider seasonal restrictions using Robinson's approach.

In this paper, the problem of interest is to estimate time varying parameters under cross restrictions. To the best of our knowledge, the properties of smoothing techniques in this context have not been studied so far. Our objective is twofold. First, we seek to derive a consistent nonparametric method based on smoothers that introduce the time varying restrictions on the coefficients. Second, we propose several statistics in order to test the validity of such restrictions.

The rest of the paper is organized as follows. Section 2 presents the nonparametric method developed for estimating the structural parameters of a econometric model incorporating some pre-specified constraints. We also obtain the statistical properties of these estimators under some general conditions, such as consistency and asymptotic normality. Section 3 proposes several statistics to test for the time varying restrictions. In Section 4 the restricted time varying coefficients are estimated in a data framework, where a demand system model for different categories of meat is considered. Section 5 provides the results from both parts, estimation and testing, of a simulation study, and proofs are provided in the Appendix.

2. Parameter estimation under time varying constraints

Consider a general linear model of the form

$$y_t = \beta_t^T x_t + \epsilon_t \quad t = 1, \ldots, n, \tag{1}$$

where y_t and $x_t = (x_{1t}, \ldots, x_{pt})$ are the observed values for the dependent and the p explanatory variables, respectively. The vector $\beta_t = (\beta_{1t}, \ldots, \beta_{pt})^T$ contains the coefficients associated with the variables at the tth moment and they are assumed to be an unknown but deterministic function of the time index; i.e. $\beta_t = \beta\,(t/n)$, where n is the sample size. The error terms, ϵ_t, are assumed to be independently distributed with zero mean and finite, possibly heteroskedastic, variance σ_t^2.

The restrictions suggested by economic theory are introduced into the model through the time varying constraints

$$g_{jt}\,(\beta_t) = 0 \qquad t = 1, \ldots, n; \quad j = 1, \ldots, q < p, \tag{2}$$

where q is the number of restrictions considered and n is the number of observations. The function $g_{jt}\,(\cdot)$ is a known function from $\mathbb{R}^p$ to $\mathbb{R}$. Note that we are allowing for restrictions varying with time. If this is not the case, we have $g_{jt} \equiv g_j$. In general, we are considering functions without imposing any relation between them for the different indexes; that is, between g_{jt} and g_{js} for $t \neq s$. That is, cross restrictions are allowed but not restrictions along time.

Depending on the context, model (1) can represent a production function, a cost function, a linear expenditure system or a log linear demand function; the restrictions represented in (2) can be homogeneity, Slutsky conditions or any other cross restrictions.

Now we proceed to estimate the sequence of parameter vectors $\beta_1, \ldots, \beta_n$, taking into account the set of restrictions mentioned above. We solve the following optimization problem for each $r = 1, \ldots, n$,

$$\begin{aligned} &\min \mathcal{S}(\beta_r) \\ \text{s.t.}\quad &g_{jr}(\beta_r) = 0 \text{ for } j = 1, \ldots, q, \end{aligned} \tag{3}$$

where $\mathcal{S}(\beta_r)$ is a local measure of the goodness of fit. We consider

$$\mathcal{S}(\beta_r) = \sum_{t=1}^{n} K_{rt}(y_t - \beta_r^T x_t)^2 \quad \text{for } r = 1, \ldots, n, \tag{4}$$

where K_{rt} play the role of local weights. These are defined as $(nh)^{-1}K((r - t)/nh)$, where $K(\cdot)$ is a second order kernel and h is the bandwidth. Hence, this criterion function becomes a smoothed local residual sum of squares, similar to that proposed in a likelihood context by Staniswalis (1989) and Robinson (1989). The weights control the smoothness of the parameter changes through the bandwidth h. It is of interest to point out that if $p = 1$, the direct minimization of the previous criterion without introducing kernel weights will give us a degenerate solution. On the other hand if $p > 1$, the direct minimization of (4) with $h = 0$ leads to identification problems. Thus, the introduction of the weights with $h > 0$ turns the problem into a solvable one. And, as it will be shown, this is the case also under restrictions.

According to the constraints, we focus on solving the problem when the functions $g_{jr}(\cdot)$ are linear, although they are allowed to vary with r. That is, for each period r, the considered optimization problem can be formulated as

$$\begin{aligned} &\min \mathcal{S}(\beta_r) \\ \text{s.t.} \quad &G_r\beta_r = g_r, \end{aligned} \tag{5}$$

where G_r is a $(q \times p)$ matrix and g_r is a q order vector. The Lagrangian function associated to this constrained optimization problem can be defined in a matrix notation as

$$\mathcal{S}^o(\beta_r, \lambda_r) = \mathcal{S}(\beta_r) + 2\lambda_r^T(G_r\beta_r - g_r), \tag{6}$$

where λ_r is a $(q \times 1)$ vector that contains the Lagrange multipliers and $\mathcal{S}(\beta_r)$ can be written as $\mathcal{S}(\beta_r) = (y - X\beta_r)^T W_r (y - X\beta_r)$, where $y = (y_1, \ldots, y_n)^T$ and X is the $(n \times p)$ data matrix, with its tth row given by the vector x_t. The weight matrix, W_r, is a diagonal matrix of order n with its tth element given by K_{rt}.

Let us denote by $\widehat{\beta}_r^o$ and $\widehat{\lambda}_r^o$ the solution to the constrained optimization problem

$$(\widehat{\beta}_r^o, \widehat{\lambda}_r^o) \equiv \operatorname{argmin}_{\beta_r, \lambda_r} \mathcal{S}^o(\beta_r, \lambda_r). \tag{7}$$

A closed expression for the vector containing the coefficient estimators will be obtained from the first order conditions in (6) as

$$\widehat{\beta}_r^o = \widehat{\beta}_r - \left(X^T W_r X\right)^{-1} G_r^T \left[G_r \left(X^T W_r X\right)^{-1} G_r^T\right]^{-1} (G_r\widehat{\beta}_r - g_r), \tag{8}$$

where $\widehat{\beta}_r$ represents the unconstrained estimator; that is, the one obtained when $\lambda_r = 0$. This estimator corresponds to the solution of the unrestricted problem (see Robinson (1989)), and it is computed as

$$\widehat{\beta}_r = (X^T W_r X)^{-1} X^T W_r y. \tag{9}$$

Thus, from (8) we obtain an expression for the constrained estimator $\widehat{\beta}_r^o$ in terms of its relation with the unconstrained estimator. The estimator of the Lagrange multiplier is also derived from the first order conditions.

$$\widehat{\lambda}_r^o = \left[G_r \left(X^T W_r X\right)^{-1} G_r^T\right]^{-1} (G_r \widehat{\beta}_r - g_r). \tag{10}$$

As usual, $\widehat{\lambda}_r^o$ estimates the cost of imposing a non-true restriction, so that its value increases in the same direction as the difference $G_r \widehat{\beta}_r - g_r$. Sufficient conditions for these estimators to be the unique solution of the optimization problem given in (7) are guaranteed by the assumption

(A.1) Rank $\left(W_r^{1/2} X\right) = p < nh \quad \forall r = 1, \ldots, n$, and $\text{Rank}(G_r) = q$ for $r = 1, \ldots, n$.

That is, we consider full rank local data matrix and we do not allow restrictions that can be obtained as linear combinations of others.

At this point, we would like to make some remarks about the role of the parameter h in both constrained and unconstrained estimators. As expected in a nonparametric context, the behavior of the unconstrained estimator depends crucially on the value of the smoothness parameter h. High values of h eliminate the variability of the coefficients. In fact, when this value tends to infinity, we obtain one unique coefficient per explanatory variable; which is equivalent to the time invariant OLS estimator. Small values of h allow differences between adjoining coefficients, and the estimator can therefore present more roughness.

However, this is not the case for the restricted estimator. For a better understanding of this fact let us consider two particular cases, unrealistic but useful for this illustration. First, consider the case where $q = 1$, $G_r \equiv G$, and g_r takes the value one when r is odd and zero otherwise; that is, g_r is a function that varies drastically with r. Now, if the value of h is very high, the unconstrained estimator $\widehat{\beta}_r$ will be a constant vector, but the restriction will make the constrained estimator $\widehat{\beta}_r^o$ very sharp, since it must fulfill the restriction. Specifically, the constrained estimator will be

$$\widehat{\beta}_r^o = \begin{cases} \widehat{\beta} + B(G\widehat{\beta} - 1), & \text{when } r \text{ is odd,} \\ \widehat{\beta} + BG\widehat{\beta}, & \text{when } r \text{ is even,} \end{cases}$$

where B is a matrix that does not depend on r. Therefore, it is clear that the constrained estimator can be very rough even when a large amount of smoothness is imposed through the bandwidth.

For the second case, let $q = p$, $G_r \equiv I$ and $g_r \equiv g$. In this case, if the value of h is taken to be very low, the unconstrained estimator can be very rough. However, given the imposed constraints, the constrained estimator is just $\widehat{\beta}_r^o = g$ for all r. Thus, the constrained estimator is very smooth (it is the same constant for all r) although the smoothing parameter is small.

Hence, in the nonparametric constrained setting, the size of h does not directly dictate the degree of the smoothness for the estimated coefficients. As can be read from (8), $\widehat{\beta}_r^o$ is a mixture of the unconstrained nonparametric estimator, where the role of h is the usual one in this context, and the other term that forces $\widehat{\beta}_r^o$ to fulfill the restrictions.

We now proceed to present the main asymptotic results for the constrained estimator. The following additional assumptions are needed.

(A.2) $\beta_{it} = \beta_i(t/n)$, where $\beta_i \in C^2[0,1]$ for all $i = 1, \ldots, p$.

This assumption introduces some degree of smoothness into the coefficient changes without imposing any pre-specified functional form.

(A.3) The function $K(\cdot)$ is a second order kernel with compact support $\Omega = [-1,1]$. We also assume that the Fourier transform of K_{rt} is absolutely integrable and $\int_\omega K^4(u)du$ is bounded.

This compactness assumption can be relaxed if we assume the existence of finite higher order moments. In fact both assumptions are standard in nonparametric regression literature (see Härdle (1990)).

(A.4) The observations $x_1, \ldots, x_n$ are independent realizations from a random variable $\mathbf{X} \in I\!R^p$. We define $M_t = E(x_t x_t^T)$ a p order symmetric positive definite matrix with generic element $m_{ijt} = E(x_{it}x_{jt})$ that can be decomposed as $m_{ij}(t/n) + O(n^{-1})$. The functions $m_{ij}(t/n)$ are at least twice differentiable and uniformly bounded, and all cross moments involving x_{it} are uniformly bounded up to order eight.

(A.5) The explanatory variables are statistically uncorrelated with the error term, and $E|\epsilon_t|^{2+\delta} < \infty$ for some $\delta > 0$.

These assumptions are standard in nonparametric estimation. However, **(A.4)** is weaker than the usual assumption of *i.i.d.* explanatory variables. In fact, this assumption allows for locally stationary variables, as defined in Dahlhaus (2000), and studied as explanatory variables in Orbe, Ferreira and Rodriguez-Poo (2005).

Now we present some results that resemmue those obtained in standard restricted least squares regression.

Theorem 1. *Under* **(A.1)** *to* **(A.4)**, *$h \to 0$ and $nh \to \infty$ as n tends to infinity, the asymptotic bias and variance for the constrained coefficients estimator are*

$$BIAS(\widehat{\beta}_r^o) = Q_r(g_r - G_r\beta_r) + \frac{d_k h^2}{2}(I - Q_rG_r)M_r^{-1}\beta_r'' + o(h^2), \tag{11}$$

$$V(\widehat{\beta}_r^o) = \frac{\sigma_r^2 c_k}{nh}(I - Q_r G_r)M_r^{-1} + o((nh)^{-1}), \tag{12}$$

where $c_k = \int_\Omega K^2(u)du$, $\widehat{\beta}_r''$ *is a* p *order vector that contains the second derivatives of the coefficients,* $d_k = \int_\Omega u^2 K(u)du$ *and* $Q_r = M_r^{-1}G_r^T(G_r M_r^{-1} G_r^T)^{-1}$.

Note that the result coming from the previous theorem puts the well known conclusions for restricted parametric regression into the nonparametric framework. Thus, the bias shown in (11) has two terms. The first one, as in the standard parametric case, appears when the restriction is not true. The second term is due to the smoothness. This last term vanishes when h tends to zero, whereas the first part of the bias can be zero, if the restriction is fulfilled, or nonzero if the imposed restriction is wrong. The variance term is dominated by the smoothness although, as it can be observed in (12), the restriction plays a role in the leading term. Thus, for small sample sizes, the variance of the unrestricted estimator is going to be larger than the restricted one, uniformly in h. This is exactly the same as in the parametric case.

Therefore, when the restricted model is estimated, the parameter h is selected under this null hypothesis. According to the previous findings it is then possible to use, as a data driven method, cross validation techniques. For this type of generalized ridge regression estimators, as stated by Li (1986), cross validation works for any estimator that converges at a smaller rate than root n. This is clear in our case. We can also use plug-in or penalized methods, where the unique difference is that the mean square error and the projection matrix have now a different expression than in the unrestricted model.

Finally, it is worth noting that the constrained estimator ($\widehat{\beta}_r^o$) is consistent, and presents a smaller bias than the unconstrained one ($\widehat{\beta}_r$). For the variance term, regardless of the constraint, the constrained estimator provides a smaller variance than the unconstrained one. These results are also similar to those obtained in classic regression analysis. The formal statements can be read from the following corollary, where we compare bias and variance for the two estimators.

Corollary 1. *Given the assumptions in Theorem* 1

$$V(\widehat{\beta}_r) - V(\widehat{\beta}_r^o) = \frac{\sigma_r^2 c_k}{nh} M_r^{-1} G_r^T (G_r M_r^{-1} G_r^T)^{-1} G_r \text{ is positive semidefinite.}$$

In addition, if $G_r \beta_r = g_r$, *then*

$$BIAS(\widehat{\beta}_r) - BIAS(\widehat{\beta}_r^o) = \frac{d_k h^2}{2} M_r^{-1} G_r^T (G_r M_r^{-1} G_r^T)^{-1} G_r \beta''.$$

Next, we obtain the asymptotic distribution for the constrained estimator.

Theorem 2. *Given* **(*A*.1)** *to* **(*A*.5)**, *if* $h = o(n^{-1/5})$ *and* $nh \to \infty$, *then under the null hypothesis* $H_o : G_r\beta_r = g_r$, *the asymptotic distribution of the constrained estimator is*

$$\sqrt{nh}(\widehat{\beta}_r^o - \beta_r) \xrightarrow{d} N\Big(0, c_k\sigma_r^2[I - Q_rG_r]M_r^{-1}\Big) \tag{13}$$

as n *tends to infinity.*

The unknown terms in the distribution can be consistently estimated as

$$\widehat{Q}_r = \widehat{M}_r^{-1}G_r^T(G_r\widehat{M}_r^{-1}G_r^T)^{-1} \quad \textit{where} \quad \widehat{M}_r = \sum_{t=1}^{n} K_{rt}x_tx_t^T \tag{14}$$

$$\hat{\sigma}_r^2 = \frac{\sum_{t=1}^{n} K_{rt}(y_t - \widehat{\beta}_{1r}^0 x_{1t} - \widehat{\beta}_{2r}^0 x_{2t} - \cdots - \widehat{\beta}_{pr}^0 x_{pt})^2}{\sum_{t=1}^{n} K_{rt}}. \tag{15}$$

Note that the rate of convergence of the estimator is $\sqrt{nh}$ and, given the assumption for h, this is slower than the optimal parametric rate $\sqrt{n}$. This is a typical result in nonparametric regression estimation. We pay a price for the flexibility of coefficient changes. Within the nonparametric framework, the assumption for h allows us to attain the optimal rate of convergence. Moreover, it must be noticed that there is no curse of dimensionality in this context. Let us recall that the model is linear in the explanatory variables and the coefficients are the parameters to be estimated nonparametrically, where the smoothing procedure only considers time variation. That is, with respect to the kernel methodology, the estimation is univariate.

So far, we have derived the consistency of the constrained estimator and given its asymptotic distribution when the constraints are fulfilled. It is a natural interest to test for this condition, and that will be the aim of the next section.

3. Hypothesis Testing

In this section we derive several statistics to test the validity of the constraints. These statistics will be natural adaptations of the standard ones in econometric literature (see Gourieroux and Monfort (1989)). However, in order to compute the critical regions, the whole theory needs to be revisited to account for the nonparametric structure of both the restricted and unrestricted parameter estimators.

We first consider a pointwise test; that is, we construct a statistic useful for testing the restriction at a fixed time value. Fix an arbitrary r, and consider the hypothesis,

$$H_0 : \quad G_r\beta_r = g_r.$$

The statistics proposed are based on the classic Wald (Wald (1943)), Score (Rao (1947)), Hausman (Hausman (1978)) and Likelihood ratio test statistics. The versions of these statistics in this framework are

(i) The Wald statistic

$$\varepsilon_r^W = \frac{nh}{c_k\sigma_r^2}(G_r\widehat{\beta}_r - g_r)^T[G_rM_r^{-1}G_r^T]^{-1}(G_r\widehat{\beta}_r - g_r); \tag{16}$$

(ii) The Score statistic

$$\varepsilon_r^S = \frac{nh}{c_k\sigma_r^2}\widehat{\lambda}_r^{oT}[G_rM_r^{-1}\ G_r^T]\widehat{\lambda}_r^o; \tag{17}$$

(iii) The Hausman statistic

$$\varepsilon_r^H = \frac{nh}{c_k\sigma_r^2}(\widehat{\beta}_r^o - \widehat{\beta}_r)^T M_r(\widehat{\beta}_r^o - \widehat{\beta}_r); \tag{18}$$

(iv) The Likelihood ratio statistic

$$\varepsilon_r^R = \frac{nh}{c_k\sigma_r^2}(\mathcal{S}(\widehat{\beta}_r) - \mathcal{S}^o(\widehat{\beta}_r^o, \widehat{\lambda}_r^o)). \tag{19}$$

The tests are defined through the critical region for these statistics. This requires the asymptotic distribution for the Lagrange multiplier estimator, as established in the next theorem.

Theorem 3. *Under* **(*A*.1)** *to* **(*A*.5)**, *if $h \to 0$ and $nh \to \infty$, the asymptotic mean and variance of $\widehat{\lambda}_r^o$ are*

$$E(\widehat{\lambda}_r^o) = 2(G_rM_r^{-1}G_r^T)^{-1}(G_r\beta_r - g_r) + o(h^2), \tag{20}$$

$$V(\widehat{\lambda}_r^o) = \frac{\sigma_r^2 c_k}{nh}(G_rM_r^{-1}G_r^T)^{-1} + o((nh)^{-1}). \tag{21}$$

Furthermore, if $h = o(n^{-1/5})$ and $G_r\beta_r = g_r$ holds,

$$\sqrt{nh}\ \widehat{\lambda}_r^o \xrightarrow{d} N\Big(0,\ c_k\sigma_r^2[G_rM_r^{-1}G_r^T]^{-1}\Big) \tag{22}$$

as n tends to infinity.

This result will allow us to prove the equivalence of the statistics defined above, and the consistency of the tests defined through their critical regions, as stated in the theorem below.

Theorem 4. *Under* **(*A*.1)** *to* **(*A*.5)**, *if $h = o(n^{-1/5})$ and $nh \to \infty$, then*

(i) *the four statistics* ε_r^W, ε_r^S, ε_r^H *or* ε_r^R *are asymptotically equivalent almost surely*;
(ii) *the test defined by the critical region* $\{\varepsilon_r \geq \chi^2_{1-\alpha}(q)\}$, *where* ε_r *can be any of the statistics above, is consistent and has asymptotic level* α, *as* $n \to \infty$.

We have derived a pointwise test that can test the restrictions in the model for a fixed time value. The next theorem states the asymptotic distribution of a statistic defined as the sum of the pointwise statistics at separate locations, in order to test overall structure.

Theorem 5. *Assume* **(*A*.1)** *to* **(*A*.5)**, *and that* $h = o(n^{-1/5})$ *and* $nh \to \infty$ *as* $n \to \infty$. *Then, for* ε_r *defined as in the previous theorem and under the null hypothesis,*

$$\left\{\sum_{r=r_1}^{r_k} \varepsilon_r\right\} \xrightarrow{d} \chi^2(kq),$$

where $r_1, \ldots, r_k$ *are locations. Separed by order* nh *points.*

4. An Application: Estimation of a Demand System under Several Time Varying Constraints

The purpose of this section is to illustrate the performance of the proposed methodology. We consider a log linear demand system (Deaton and Muellbauer (1980)) to model the demand for different categories of meat in the U.S. The dataset used consists of yearly observations for the period 1970-1998, collected from the U.S. Economic Research Service. The meat categories considered are beef, pork, lamb and poultry, and for each category we use per capita consumption (q_{it}) and retail prices (p_{it}). Following Murray (1984) we assume that the consumer's utility function is weakly separable between meat and all other commodity groupings, so we have taken the total expenditure per year on meat for the income variable, E_t. This meat demand system has the following structural equations

$$\begin{aligned}
\ln q_{1t} &= \alpha_{1t} + \delta_{1t} \ln p_{1t} + \sum_{j\neq 1} \beta_{1jt} \ln p_{jt} + \gamma_{1t} \ln E_t + u_{1t} \\
\ln q_{2t} &= \alpha_{2t} + \delta_{2t} \ln p_{2t} + \sum_{j\neq 2} \beta_{2jt} \ln p_{jt} + \gamma_{2t} \ln E_t + u_{2t} \\
\ln q_{3t} &= \alpha_{3t} + \delta_{3t} \ln p_{3t} + \sum_{j\neq 3} \beta_{3jt} \ln p_{jt} + \gamma_{3t} \ln E_t + u_{3t} \qquad (23) \\
\ln q_{4t} &= \alpha_{4t} + \delta_{4t} \ln p_{4t} + \sum_{j\neq 4} \beta_{4jt} \ln p_{jt} + \gamma_{4t} \ln E_t + u_{4t} \\
& j = 1,2,3,4, \quad t = 1,\ldots,T,
\end{aligned}$$

where j denotes the meat category, t the time period, q_{jt} the consumption of the jth category of meat during the tth year, p_{jt} is the retail price of the jth category of meat during the tth year and E_t stands for the income at time t. The errors u_{jt} are assumed to be *i.i.d.* Given the model's log-linear structure, the coefficients δ_{jt} correspond to the time-varying own price elasticity, β_{jit} the time varying cross price elasticity and γ_{jt} the income elasticity. According to economic theory (Pollak and Wales (1992)) the coefficients of the structural equations must fulfill next restrictions for each period:

- The Engel condition

$$\sum_{j=1}^{4} w_{jt}\gamma_j = 1 \quad \forall t; \tag{24}$$

- The Slutsky symmetry conditions

$$\frac{\beta_{ij}}{w_{it}} + \gamma_j = \frac{\beta_{ji}}{w_{jt}} + \gamma_i \quad j, i = 1, \ldots, 4 \quad j \neq i \quad \forall t; \tag{25}$$

- The Homogeneity condition

$$\sum_{i=1}^{4} \beta_{ij} + \gamma_j = 0 \quad j = 1, \ldots 4. \tag{26}$$

Here w_{jt} is the weight $w_{jt} = p_{jt}q_{jt} / \sum_{i=1}^{4} p_{it}q_{it}$. Note that the first two constraints are time varying whereas the homogeneity restriction is time invariant. Moreover, the first two equations introduce cross restrictions relating the different structural equations while the homogeneity condition imposes constraints among coefficients corresponding to the same equation.

The elasticities for the different meat categories are estimated using the restricted nonparametric estimator proposed in Section 2, taking into account the restrictions in (24), (25) and (26), and using the estimation algorithm in Orbe,Ferreira and Rodriguez-Poo (2003). The kernel is Epanechnikov's and the bandwidths for each equation, h_j, have been selected according to the Generalized Cross Validation criterion.

Figure 1 shows the income elasticities for each meat category throughout the sample time period and, as expected, all income elasticities are positive. It can be observed that beef seems to be the preferred meat category for U.S. consumers during the whole sample period. Pork appears to be the closest substitute for beef, at least in a significant part of the sample period. Finally, lamb and poultry present the lowest income elasticities and their tendency has shifted upwards significantly by the end of the period of study.

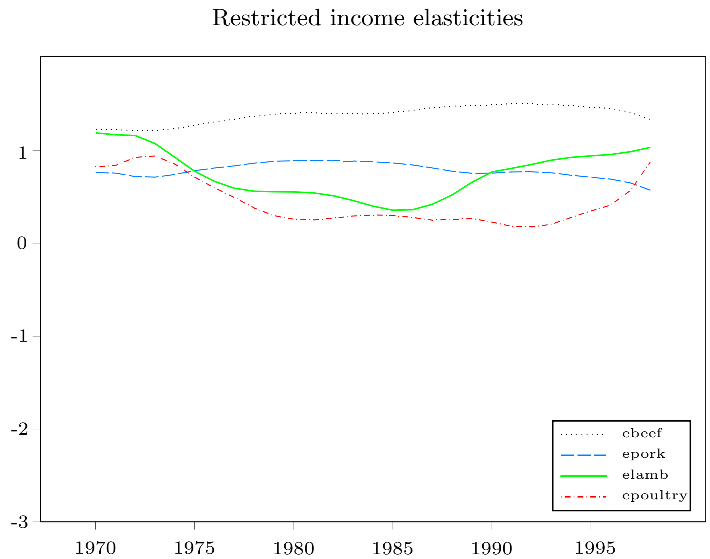

Figure 1. Restricted income elasticities for all meat categories. The y-axis represents elasticities and the x-axis the time index.

Figure 2 shows the own price elasticities for each demand equations in (23). In general they present negative signs, as expected, although they have changed over time. It can be observed that in the poultry demand equation, the own price elasticity takes positive values in a short time period, coinciding with a change of thinking in response to an increasing health consciousness (see "Food Consumption, Prices and Expenditures, 1970-97", Economic Research Service, USDA).

With respect to the cross price elasticities, the results are also quite reasonable. Lamb prices do not have a significant influence over the consumption of the other meat types, but the influence holds in the other direction. Beef demand appears to be sensitive to changes in poultry price, although looking at the estimated values, the symmetric behavior over time between poultry and beef demand elasticities makes it difficult to state whether they are substitutes or complements. The demand for beef and pork do not appear to be very sensitive to changes in prices of each other. Beef's price cross elasticity with respect to poultry changes over time. At the beginning of the sample the elasticity sign is positive, and pork and poultry could be interpreted as substitute goods. The same occurs after the pork industry crisis when an aggressive campaign announced that "pork is a white meat".

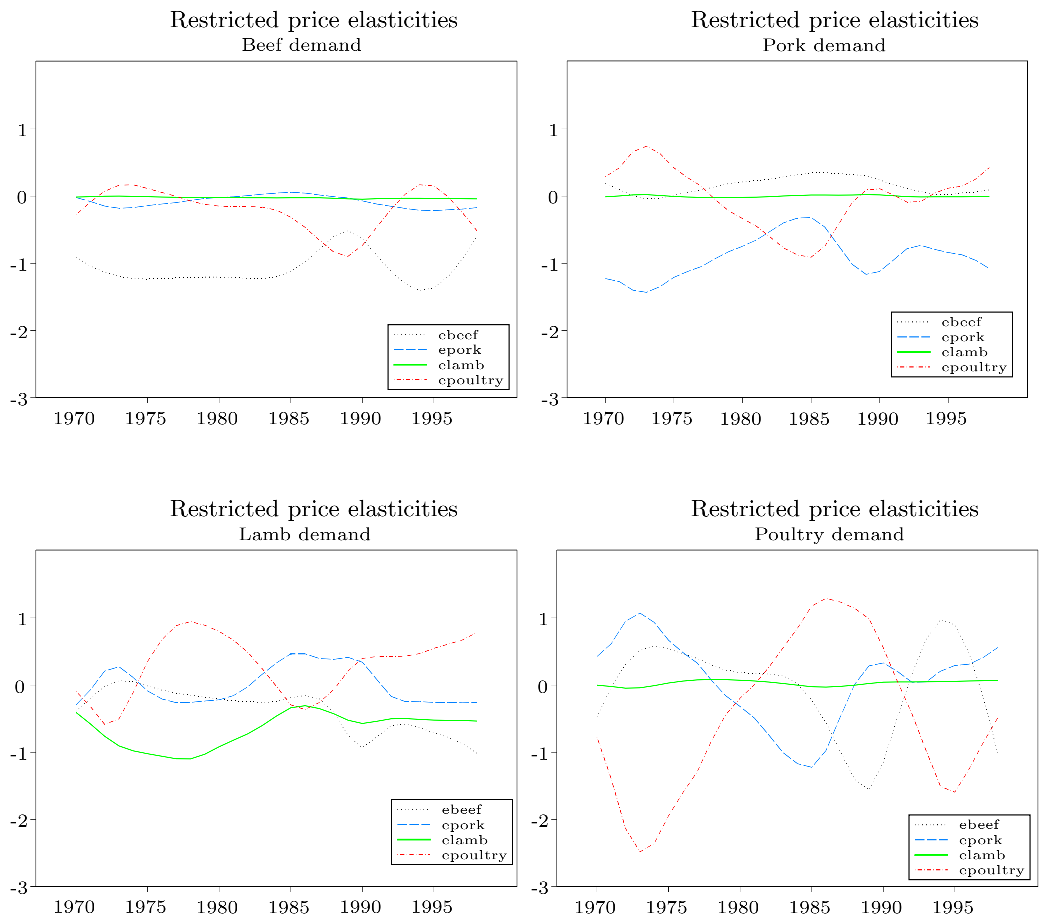

Figure 2. Restricted price elasticities for the beef demand equation (l. h. upper corner), pork demand equation (r. h. upper corner), lamb demand equation (l. h. lower corner) and poultry demand equation (r. h. lower corner). The y-axis represents elasticities and the x-axis the time index.

It could be useful to test for the validity of the restrictions. We have used the statistics presented in Section 3, and the conclusion leads to the rejection of the null. However, there are different reasons why this test is not very useful for this case. First, the sample is short. Second, we assume that the consumer's utility function is weakly separable between meat categories and all other commodity groups. This is a questionable assumption since there might be other commodities (e.g., fish and eggs) that could also be considered in this group. Moreover, we do not present the unrestricted estimators since this model only makes sense when the restrictions hold. Thus, this example is useful to for checking the ability of the method to estimate the time varying parameters under cross restrictions, and a more detailed analysis of the empirical properties of the method is done in next section.

5. A Simulation Study

In this section a simulation study is performed to show, for several sample sizes, the empirical accuracy of the estimation procedure and the performance of the test statistics. In order to do this, we simulate a Cobb-Douglas production function

$$P_t = L_t^{\beta_{1t}} K_t^{\beta_{2t}} e^{u_t},$$

where P denotes production, L labor, K capital and u is the error term. The linearized model to be estimated is

$$\ln P_t = \beta_{1t} \ln K_t + \beta_{2t} \ln L_t + u_t. \tag{27}$$

The coefficients β_{1t} and β_{2t} have the usual interpretation, labor and capital elasticities (ceteris paribus). Thus the restriction $\beta_{1t} + \beta_{2t} = 1$ implies constant scale economies for a given period t. Note that this time varying framework allows to estimate and check weather the restriction is fulfilled for a subsample of periods and, therefore, to consider a log-linear production function with variable scale economies as described in Fried, Lovell and Schimdt (1993).

For the simulation study, the explanatory variables (L, K) are generated independently $N(50, \sigma^2 = 6)$, and the error term u_t in the linearized model is $N(0, \sigma^2 = 0.05)$. Note the explanatory variables in the final model are $\ln L$, and $\ln K$, and this explains the scale of the variances.

The selected function for the first coefficient is $\beta(\tau) = 1.3 - \cos(\tau)$, $\tau \in (0, 1)$. The coefficients for each sample are computed from the function, avoiding the extremes of $\beta_{1t} = 1.3 - cos(t + 25/(n + 50))$, for $t = 1, \ldots, n$, and for $\beta_{2t} = 1 - \beta_{1t}$.

Three sample sizes were considered: $n = 100$, 500 and 1,000. For each sample, 1,000 replications were generated and the smoothing parameters were selected by generalized cross validation.

The results for the first objective of this study, the accuracy of the estimation process, are summarized in Table 1 and Figure 3. Table 1 presents the lower quartile, median and upper quartile of the empirical distribution for the mean squared error $MSE_b(\beta_i) = 1/n \sum_{t=1}^{n} (\beta_{it} - \hat{\beta}_{it})^2$, for $b = 1, \ldots, 1,000$ replications, when the unconstrained estimator is considered, or $MSE_b(\beta_i) = 1/n \sum_{t=1}^{n} (\beta_{it} - \hat{\beta}_{it}^{o})^2$ when using the constrained estimator. The boxplot of each distribution is presented in Figure 3.

As can be observed, the results support the theoretical results. First, as expected since the restriction is corrected, the restricted estimator leads to a smaller

MSE. Moreover, as long as the sample size increases, the MSE decreases, as do the differences between the restricted and the unrestricted estimator. This is natural since both estimators are consistant. Finally, it is important to note the accuracy of the estimator even with a sample size of $n = 100$.

Table 1. Statistics for the mean squared error.

Stat.	upper quartile			median			lower quartile		
Est.	$n\!=\!100$	$n\!=\!500$	$n\!=\!1,000$	$n\!=\!100$	$n\!=\!500$	$n\!=\!1,000$	$n\!=\!100$	$n\!=\!500$	$n\!=\!1,000$
$\hat{\beta}_1$	0.01933	0.00932	0.00340	0.01770	0.00597	0.00288	0.01604	0.00363	0.00242
$\hat{\beta}_2$	0.01931	0.00928	0.00337	0.01757	0.00596	0.00286	0.01594	0.00364	0.00241
$\hat{\beta}_1^o,\ \hat{\beta}_2^o$	0.01535	0.00927	0.00310	0.01332	0.00566	0.00263	0.01154	0.00337	0.00221

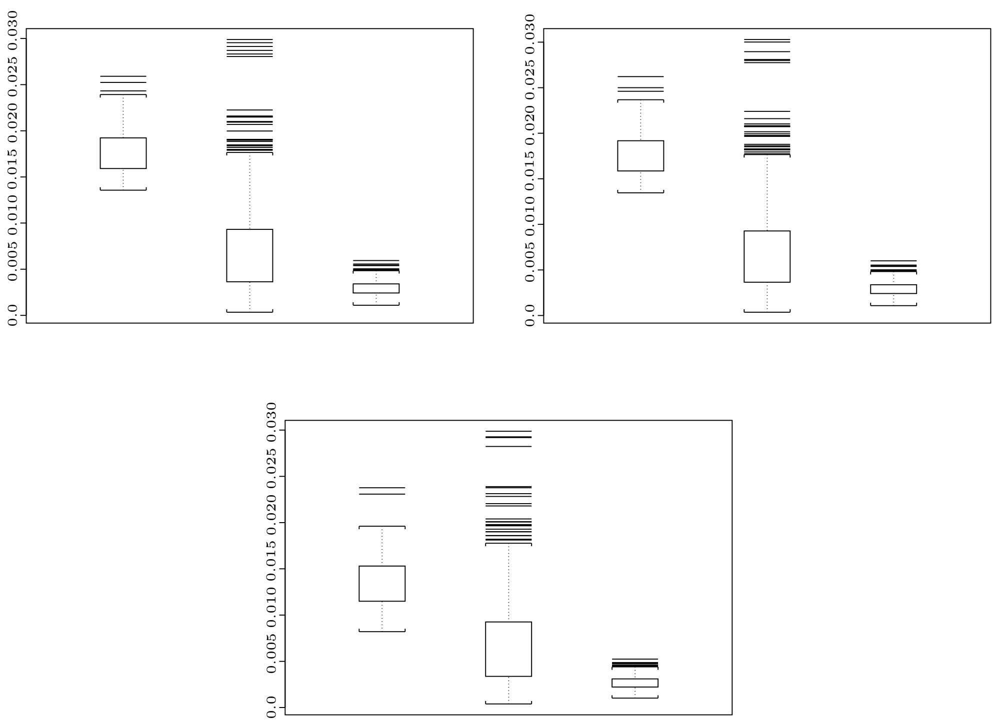

Figure 3. The left upper box shows the boxplots for the distribution of $\hat{\beta}_1$'s MSE, with sample sizes $n = 100, 500, 1,000$ in left to right order. The right upper box shows the same for $\hat{\beta}_2$ and the down box for both (since the MSE are equal) the restricted estimators $\hat{\beta}_1^o$ and $\hat{\beta}_2^o$.

For the second objective, the performance of the four test statistics, were computed and compared with the corresponding asymptotic χ^2 distribution. Table 2 summarizes the results. Although the statistics are asymptotically equivalent, it is clear that their performances in finite samples are very different. In all cases, Wald and Hausman statistics provide the worst results, while the score statistic appears to be the best.

Table 2. Empirical sizes.

α=0.05	Wald	Score	Hausman	L.Ratio
$n = 100$	0.279	0.070	0.311	0.190
$n = 500$	0.212	0.068	0.234	0.186
$n = 1,000$	0.083	0.039	0.116	0.089

Hence, the simulation study supports the practical validity of the estimation process. For the testing part, it is clear that the method should be revised and compared with others. It is possible that statistics based on the maximum rather than on the sum, as the one used here, would lead to better results. That is, more statistics can be used and compared in a practical setting. However, the main objective of this section has been to demonstrate the theoretical properties studied in the previous sections and, therefore, this further study is relegated to further research.

Acknowledgement

This research has been supported by the Ministerio de Ciencia y Tecnología, Universidad del País Vasco and Fundación BBVA, through the projects SEJ2005-05549/ECON, SEJ2005-08269; 9/UPV0003.321-13631/2001 and 1/BBVA 00038. 16421/2004, respectively.

Appendix

In order to prove the main results we need some additional lemmas.

Lemma 1. *Under **(A.1)** to **(A.6)**, and if $h \to 0$ and $nh \to \infty$ as n tends to infinity, then the asymptotic bias of the unconstrained estimator is*

$$BIAS(\widehat{\beta}_r) = \frac{d_k h^2}{2} M_r^{-1} \beta_r'' + o(h^2), \qquad (28)$$

where $\widehat{\beta}_r''$ is a p order vector that contains the second derivatives of the coefficients, and $d_k = \int_\Omega u^2 K(u) du$. The variance is

$$V(\widehat{\beta}_r) = \frac{\sigma_r^2 c_k}{nh} M_r^{-1} + o((nh)^{-1}), \qquad (29)$$

where $c_k = \int_\Omega K^2(u) du$.

The proof follows directly from Orbe, Ferreira and Rodriguez-Poo (2005), Theorem 1, for the particular case when the vector λ in that theorem is taken to be zero.

Lemma 2. *Under* **(*A*.1)** *to* **(*A*.6)**, *if $h = o(n^{-1/5})$ and $nh \to \infty$ when n tends to infinity, then*

$$\sqrt{nh}(\widehat{\beta}_r - \beta_r) \xrightarrow{d} N(0,\ c_k \sigma_r^2\ M_r^{-1}). \tag{30}$$

The result follows directly from Orbe, Ferreira and Rodriguez-Poo (2005), Theorem 2.

Proof of Theorem 1. First, consider that $\widehat{\beta}_r^o = \widehat{\beta}_r - \tilde{Q}_r(G_r\widehat{\beta}_r - g_r) = (I - \tilde{Q}_r G_r)\widehat{\beta}_r + \tilde{Q}_r g_r$ where $\tilde{Q}_r = (X^T W_r X)^{-1} G_r^T [G_r (X^T W_r X)^{-1} G_r^T]^{-1}$.

Given the results in Lemma 1, the asymptotic mean of $\widehat{\beta}_r^o$ is

$$\begin{aligned} E(\widehat{\beta}_r^o) &= (I - Q_r G_r)\Big[\beta_r + \frac{d_k h^2}{2} M_r^{-1}\beta_r'' + o(h^2)\Big] + Q_r g_r \\ &= \beta_r - Q_r(G_r\beta_r - g_r) + \frac{d_k h^2}{2}(I - Q_r G_r) M_r^{-1}\beta_r'' + o(h^2), \end{aligned}$$

where $Q_r = M_r^{-1} G_r^T \left[G_r M_r^{-1} G_r^T\right]^{-1}$, and $\tilde{Q}_r \xrightarrow{p} Q_r$.

Following similar arguments and taking into account that $Q_r G_r$ is an idempotent matrix, the asymptotic variance of the constrained estimator is

$$\begin{aligned} V(\widehat{\beta}_r^o) &= (I - Q_r G_r) V(\widehat{\beta}_r)(I - Q_r G_r)^T \\ &= \frac{c_k \sigma_r^2}{nh}(I - Q_r G_r) M_r^{-1} (I - Q_r G_r)^T \\ &= \frac{c_k \sigma_r^2}{nh}(I - Q_r G_r) M_r^{-1}. \end{aligned}$$

Proof of Corollary 1. The proof of this result is straightforward if Lemma 1 and Theorem 1 are applied.

Proof of Corollary 2. From the relation between $\widehat{\beta}_r^o$ and $\widehat{\beta}_r$, using the result of Lemma 2, the asymptotic distribution of the constrained estimator is immediate using Cramer's theorem:

$$\sqrt{nh}(\widehat{\beta}_r^o - \beta_r) \xrightarrow{d} N\Big(0,\ c_k \sigma_r^2 [M_r^{-1} - M_r^{-1} G_r^T (G_r M_r^{-1} G_r^T)^{-1} G_r M_r^{-1}]\Big).$$

The consistency for $\widehat{M_r}$ can be found in Orbe, Ferreira and Rodriguez-Poo (2005) (Lemma 1). Next, we prove the consistency for the estimator of the error variance analyzing the absolute value of the difference

$$E|\hat{\sigma}_r^2 - \sigma_r^2| \le E\sum_{t=1}^{n} CK_{rt}x_t^T(\beta_t - \hat{\beta}_r^o)(\beta_t - \hat{\beta}_r^o)^T x_t + E|\sum_{t=1}^{n} CK_{rt}(\varepsilon_t^2 - \sigma_r^2)| \\ +2CE|\sum_{t=1}^{n} K_{rt}x_t^T(\beta_t - \hat{\beta}_r^o)\varepsilon_t|,$$

since $(\sum K_{rt})^{-1} \le C$ for some nonzero constant C.

Using the Cauchy-Schwartz inequality, the cross term is bounded by the other terms. Based on the results of Theorem 1, the first term tends to zero. The second term also tends to zero using Kintchine's Theorem and, finally, using the inequality of Markov, the consistency is proved.

Proof of Theorem 3. The first result is shown as follows. Based on the relation between $\widehat{\lambda}_r^o$ and $\widehat{\beta}_r$, given in expression (10), we can write

$$\widehat{\lambda}_r^o = [G_r(X^TW_rX)^{-1}G_r^T]^{-1}(G_r\widehat{\beta}_r - G_r\beta_r) + [G_r(X^TW_rX)^{-1}G_r^T]^{-1}(G_r\beta_r - g_r).$$

The asymptotic mean of $\widehat{\lambda}_r^o$ is

$$E(\widehat{\lambda}_r^o) = \left[G_rM_r^{-1}G_r^T\right]^{-1} G_r\ BIAS(\widehat{\beta}_r) + \left[G_rM_r^{-1}G_r^T\right]^{-1}(G_r\beta_r - g_r).$$

Replacing the bias of $\widehat{\beta}_r$ and simplifying terms we have that $E(\widehat{\lambda}_r^o) = 2[G_rM_r^{-1}G_r^T]^{-1}(G_r\beta_r - g_r)$ On the other hand, the asymptotic variance is obtained as

$$V(\widehat{\lambda}_r^o) = [G_rM_r^{-1}G_r^T]^{-1}G_rV(\widehat{\beta}_r)G_r^T[G_rM_r^{-1}G_r^T]^{-1} = \frac{c_k\sigma_r^2}{nh}[G_rM_r^{-1}G_r^T]^{-1}.$$

Next, based on the asymptotic distribution of $\widehat{\beta}_r$ and the relation given in (10), the asymptotic distribution of $\widehat{\lambda}_r^o$, under the null hypothesis $H_o : G_r\beta_r = g_r$, is

$$\sqrt{nh}\ \widehat{\lambda}_r^o \xrightarrow{d} N\Big(0, c_k\ \sigma_r^2[G_rM_r^{-1}G_r^T]^{-1}\Big). \tag{31}$$

Proof of Theorem 4. If $A - B \to 0$ a.s., we say A is asymptotically equivalent to B and write $A \asymp B$.

First we show the asymptotic equivalence between ε_r^R and ε_r^H. In matrix notation we write the optimization functions as

$$\begin{aligned}\mathcal{S}(\widehat{\beta}_r) &= y^TW_ry - y^TW_rX\widehat{\beta}_r - \widehat{\beta}_r^TX^TW_ry + X^TW_rX\widehat{\beta}_r \\ &= (y - X\widehat{\beta}_r)^TW_r(y - X\widehat{\beta}_r),\end{aligned} \tag{32}$$

$$\begin{aligned}\mathcal{S}^o(\widehat{\beta}_r^o, \widehat{\lambda}_r^o) &= y^TW_ry - y^TW_rX\widehat{\beta}_r - \widehat{\beta}_r^TX^TW_ry + X^TW_rX\widehat{\beta}_r \\ &\quad +2\hat{\lambda}_r^{oT}(G_r\widehat{\beta}_r^o - g_r) = (y - X\widehat{\beta}_r^o)^TW_r(y - X\widehat{\beta}_r^o),\end{aligned} \tag{33}$$

given that the constrained estimator satisfies the constraint, $G_r\widehat{\beta}_r^o = g_r$. Then we modify (33) accordingly:

$$\begin{aligned}\mathcal{S}^o(\widehat{\beta}_r^o, \widehat{\lambda}_r^o) &= (y - X\widehat{\beta}_r + X\widehat{\beta}_r - X\widehat{\beta}_r^o)^T W_r (y - X\widehat{\beta}_r + X\widehat{\beta}_r - X\widehat{\beta}_r^o) \\ &= (y - X\widehat{\beta}_r)^T W_r (y - X\widehat{\beta}_r) + (X\widehat{\beta}_r - X\widehat{\beta}_r^o)^T W_r (X\widehat{\beta}_r - X\widehat{\beta}_r^o) \\ &\quad -2(y - X\widehat{\beta}_r)^T W_r (X\widehat{\beta}_r - X\widehat{\beta}_r^o). \end{aligned} \tag{34}$$

Based on the normal equation of the unconstrained optimization problem, we set

$$\begin{aligned}(y - X\widehat{\beta}_r)^T W_r (X\widehat{\beta}_r - X\widehat{\beta}_r^o) &= (y - X\widehat{\beta}_r)^T W_r X (\widehat{\beta}_r - \widehat{\beta}_r^o) \\ &= (X^T W_r y - X^T W_r X\widehat{\beta}_r)^T (\widehat{\beta}_r - \widehat{\beta}_r^o) \\ &= \vec{0}^T (\widehat{\beta}_r - \widehat{\beta}_r^o) = 0, \end{aligned}$$

$$\mathcal{S}^o(\widehat{\beta}_r^o, \widehat{\lambda}_r^o) = (y - X\widehat{\beta}_r)^T W_r (y - X\widehat{\beta}_r) + (X\widehat{\beta}_r - X\widehat{\beta}_r^o)^T W_r (X\widehat{\beta}_r - X\widehat{\beta}_r^o), \tag{35}$$

with the difference between (35) and (32)

$$\mathcal{S}^o(\widehat{\beta}_r^o, \widehat{\lambda}_r^o) - \mathcal{S}(\widehat{\beta}_r) = (\widehat{\beta}_r - \widehat{\beta}_r^o)^T X^T W_r X (\widehat{\beta}_r - \widehat{\beta}_r^o).$$

Thus, taking into account that $X^T W_r X \longrightarrow M_r$ *a.s.* we have,

$$\varepsilon_r^R = \frac{nh}{c_k \sigma_r^2} (\mathcal{S}^o(\widehat{\beta}_r^o, \widehat{\lambda}_r^o) - \mathcal{S}(\widehat{\beta}_r)) \asymp \frac{nh}{c_k \sigma_r^2} (\widehat{\beta}_r^o - \widehat{\beta}_r)^T M_r (\widehat{\beta}_r^o - \widehat{\beta}_r) = \varepsilon_r^H.$$

We now state the asymptotic equivalence between ε_r^H and ε_r^W. This relation is obtained by merely replacing the term $\widehat{\beta}_r^o - \widehat{\beta}_r$ from expression (8) in the expression for ε_r^H:

$$\begin{aligned}\varepsilon_r^H &= \frac{nh}{c_k \sigma_r^2} (\widehat{\beta}_r^o - \widehat{\beta}_r)^T M_r (\widehat{\beta}_r^o - \widehat{\beta}_r) \\ &\asymp \frac{nh}{c_k \sigma_r^2} (G_r\widehat{\beta}_r - g_r)^T [G_r M_r^{-1} G_r^T]^{-1} (G_r\widehat{\beta}_r - g_r) = \varepsilon_r^W. \end{aligned}$$

Finally, using the expression of $\widehat{\lambda}_r^o$ given in (10), we obtain the asymptotic equivalence of ε_r^W and ε_r^S:

$$\begin{aligned}\varepsilon_r^W &= \frac{nh}{c_k \sigma_r^2} (G_r\widehat{\beta}_r - g_r)^T [G_r M_r^{-1} G_r^T]^{-1} (G_r\widehat{\beta}_r - g_r) \\ &\asymp \frac{nh}{c_k \sigma_r^2} \hat{\lambda}^{oT} [G_r M_r^{-1} G_r]^{-1} \widehat{\lambda}_r^o = \varepsilon_r^S. \end{aligned}$$

Once asymptotic equivalence is established between the different test statistics, we derive only the asymptotic distribution for the Wald test statistic.

From the asymptotic normal distribution of $\sqrt{nh}(\widehat{\beta}_r - \beta_r)$ given at (30), and since G_r is a full rank non random matrix, it is straightforward to show that $\sqrt{nh}(G_r\widehat{\beta}_r - G_r\beta_r) \xrightarrow{d} N(0, c_k\sigma_r^2 G_r M_r^{-1} G_r^T)$. Under the null hypothesis, $G_r\beta_r = g_r$, $\sqrt{nh}(G_r\widehat{\beta}_r - g_r) \xrightarrow{d} N(0, c_k\sigma_r^2 G_r M_r^{-1} G_r^T)$. Therefore,

$$\varepsilon_r^W = \frac{nh}{c_k\sigma_r^2}(G_r\widehat{\beta}_r - g_r)^T[G_r M_r^{-1} G_r^T]^{-1}(G_r\widehat{\beta}_r - g_r) \xrightarrow{d} \chi^2(q),$$

and this result means that the test with the critical region defined as $\{\varepsilon_r^W \geq \chi^2_{1-\alpha}(q)\}$ has an asymptotic level equal to α.

We show now that this test is consistent. In fact, if $\beta_r^o \notin \Theta^o = \{\beta_r : G_r\beta_r = g_r\}$, the unrestricted estimator $\widehat{\beta}_r$ converges to β_r^o and $G_r\widehat{\beta}_r - g_r$ converges to a nonzero vector. Hence, the statistic ε_r^W converges to $+\infty$ and $P_\beta(\varepsilon_r^W \geq \chi^2_{1-\alpha}(q)) \rightarrow 1$.

Proof of Theorem 5. The proof of this result is straightforward since the locations are selected such that the different terms in the sum are independent (see Härdle (1990, p.100), for similar statistics). Then, taking the asymptotic distribution, the result follows from Theorem 4.

References

Cai, Z., Fan, J. and Li, R. (2000). Efficient estimation and inferences for varying-coefficients models. *J. Amer. Statist. Assoc.* **95**, 888-902.

Cai, Z., Fan, J. and Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series. *J. Amer. Statist. Assoc.* **451**, 941-956.

Chen, R. and Tsay, R. (1993). Functional-coefficient autoregressive model. *J. Amer. Statist. Assoc.* **88**, 298-308.

Chen, R. and Liu, L. (2001). Functional coefficient autoregressive models: estimation and test of hypotheses. *J. Time Ser. Anal.* **22**, 151-173.

Deaton, A. and Muellbauer, J. (1980). *Economics and Consumer Behavior.* Cambridge University Press, Cambridge.

Dahlhaus, R. (2000). A likelihood approximation for locally stationary processes. *Ann. Statist.* **28**, 1762-1794.

Diewert, W. E. and Wales, T. J. (1987). Flexible functional forms and global curvature conditions. *Econometrica* **55**, 43-68.

Doran, H. W. and Rambaldi, A. N. (1997). Applying linear time-varying constraints to econometric models: with an application to demand systems. *J. Econometrics* **79**, 83-95.

Fried, H. O., Lovell, C. A. K. and Schimdt, S. S. (1993) *The Measurement of Productive Efficiency.* Cambridge University Press, Cambridge.

Gourieroux, C. and Monfort, A. (1989). *Statistics and Econometric Models.* Cambridge University Press, Cambridge.

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

Hastie, T. and Tibshirani, R. (1993). Varying coefficient models. *J. Roy. Statist. Soc. Ser. B* **55**, 757-796.

Hausman, J. (1978). Specification tests in econometrics. *Econometrica* **46**, 1251-1272.

Jorgenson, D. W. (2000). *Econometric Modeling of Producer Behavior.* MIT Press, Cambridge.

Li, K. C. (1986). Asymptotic optimality of C_L and generalized cross-validation in ridge regression with application to spline smoothing. *Ann. Statist.* **14**, 1011-1112.

Murray, J. (1984). Retail demand for meat in Australia: An utility theory approach. *The Economic Record* **60**, 45-56.

Orbe, S., Ferreira, E. and Rodriguez-Poo, J. M. (2000). A nonparametric method to estimate time varying coefficients under seasonal constraints. *J. Nonparametr. Statist.* **12**, 779-806.

Orbe, S., Ferreira, E. and Rodriguez-Poo, J. M. (2003). An algorithm to estimate time-varying parameters SURE models under different types of restrictions. *Comput. Statist. Data Anal.* **42**, 363-383.

Orbe, S., Ferreira, E. and Rodriguez-Poo, J. M. (2005). Nonparametric estimation of time varying parameters under shape restrictions. *J. Econometrics* **126**, 53-77.

Pollak, R. A. and Wales, T. J. (1992). *Demand System Specification and Estimation* Oxford University Press, New York.

Rao, C. R. (1947). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Proc. Cambridge Philosophical Society* **44**, 50-57.

Robinson, P. M. (1989). Nonparametric estimation of time varying parameters. In *Statistics Analysis and Forecasting of Economic Structural Change* (Edited by P. Hackled), 253-264. Springer-Verlag, Berlin.

Robinson, P. M. (1991). Time-varying nonlinear regression. In *Statistics Analysis and Forecasting of Economic Structural Change* (Edited by P. Hackl), 179-190. Springer-Verlag, Berlin.

Staniswalis, J. G. (1989). The kernel estimate of a regression function in likelihood-based models. *J. Amer. Statist. Assoc.* **84**, 276-283.

Wald, A. (1943). Test of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc.* **54**, 426-482.

Departamento de Econometría y Estadística, Universidad del País Vasco, Avenida Lehendakari 83, 48015 Bilbao, Spain.

E-mail: susan.orbe@ehu.es

Departamento de Econometría y Estadística, Universidad del País Vasco, Avenida Lehendakari 83, 48015 Bilbao, Spain.

E-mail: eva.ferreira@ehu.es

Departamento de Economía, Universidad de Cantabria, Avenida de los Castros s/n, 39005 Santander, Spain.

E-mail: rodrigjm@unican.es

(Received October 2004; accepted February 2006)