Manuscript submitted to *The Econometrics Journal*, pp. 1–21.

Nonparametric panel data regression with parametric cross-sectional dependence

Alexandra Soberon[†] and Juan M. Rodriguez-Poo[†] and Peter M. Robinson[‡]

[†] *Department of Economics, Universidad de Cantabria, Avda. Los Castros 56, Santander, 39005, Spain.*
E-mail: soberonap@unican.es, rodrigjm@unican.es

[‡] *Department of Economics, London School of Economics, Houghton Street, Holborn, London, WC2A 2AE, UK.*
E-mail: p.m.robinson@lse.ac.uk

Summary In this paper, we consider efficiency improvement in a nonparametric panel data model with cross-sectional dependence. A Generalized Least Squares (GLS)-type estimator is proposed by taking into account this dependence structure. Parameterizing the cross-sectional dependence, a local linear estimator is shown to be dominated by this type of GLS estimator. Also, possible gains in terms of rate of convergence are studied. Asymptotically optimal bandwidth choice is justified. To assess the finite sample performance of the proposed estimators, a Monte Carlo study is carried out. Further, some empirical applications are conducted with the aim of analyzing the implications of the European Monetary Union for its member countries.

Keywords: *local linear estimation; panel data; cross-sectional dependence; generalized least squares; optimal bandwidth; pseudo maximum likelihood estimation.*

1. INTRODUCTION

Traditionally, panel data models assume independence across individuals. However, economic agents (regions, states or countries, among others) are typically interdependent due to externalities, spill-overs, or the presence of common shocks, among other reasons. Therefore, ignoring this type of dependence may be inappropriate since standard estimation procedures can lead to inefficient estimators as it was shown in Phillips and Sul (2003), Hsiao and Tahmiscioglu (2008), and the references therein. Further, the question of efficiency improvements using the correlation structure emerges naturally in this setting, independently whether the dependence is allowed in either time or cross-sectional dimension (or both).

For all these reasons, the question of how to characterize cross-sectional dependence has received considerable attention in recent years, emerging two prominent strands in the literature. On the one hand, the spatial econometric approach assumes that the correlation structure can be modeled through a pre-specified spatial weight matrix that may depend on either the geographic locations of the cross-sectional units or more general economic variables. Thus, spatial processes such as the well-known spatial autoregressive (SAR) or spatial moving average (SMA) models are very popular in this approach. See Cliff and Ord (1972) or Arbia (2006), among others. On the other hand, the residual multifactor approach states that the correlation structure can be characterized by the presence of a finite number of unobserved common factors that affect all individuals with different intensities. See Kelejian and Prucha (1999), Andrews (2005), Pesaran (2006), or Bai (2009), for example.

However, parametric panel data models are usually subject to strong assumptions about the functional forms that are hardly justified by Economic Theory. In this situation, the risk of misspecification of the functional form is high and the resulting estimators are often inconsistent, invalidating the subsequent statistical inference. To overcome it, nonparametric and semiparametric panel data models are increasingly popular in last decades given their powerful capability of handling complex empirical problems. See Su and Ullah (2011) and Rodriguez-Poo and Soberon (2017), among others, for an intensive review in this regard.

Nevertheless, despite the relevance of the dependency issue among individuals, new proposals to take into account this type of correlation in this nonparametric framework are scarce both theoretically and empirically. Assuming a factorial approach for the cross-sectional dependence, Su and Jin (2012) and Huang (2013), among others, propose to estimate first the unknown function using standard nonparametric techniques and improving the efficiency of the resulting estimator by extending the common correlated effect estimator (CCE) in Pesaran (2006) to the nonparametric setting. However, this pre-specified dependence structure may be inappropriate in several empirical studies. Alternatively, Robinson (2012) and Lee and Robinson (2015) assume an unknown structure of the cross-sectional dependence and show that a simple Nadaraya-Watson estimate is dominated by a Generalized Least Squares (GLS)-type one in efficiency terms under some conditions on the rate at which the cross-sectional dimension, N, is allowed to growth with the time series length, T (see Theorems 6 and 7 in Lee and Robinson (2015), for example).

In this paper we consider efficiency improvements in nonparametric panel data models with cross-sectional dependence. Firstly, we propose to estimate the unknown function through a local linear nonparametric method. Secondly, we improve the efficiency of the previous estimator by parameterizing the cross-sectional dependence structure. Assuming that both N and T are large, we are able to show that the proposed GLS-type estimator dominates the standard local linear estimator. In particular, we obtain similar efficiency results to those obtained in Robinson (2012) or Lee and Robinson (2015) without having to impose any restriction about the growing rates among N and T as they need, but at the price of assuming a pre-specified parametric form for the cross-sectional dependence. Note that through this paper we consider the local linear fitting (see Fan and Gijbels (1995) and Ruppert and Wand (1994) for a detailed discussion about their main features). However, the proposed technique can be easily extended to other nonparametric estimates such as the Nadaraya-Watson.

To assess the finite sample performance of the proposed estimators, a Monte Carlo study is carried out. Further, some empirical applications are conducted with the aim of analyzing the implications of the European Monetary Union (EMU) for its member countries. Specifically, we conduct two empirical semiparametric applications that extend our estimation technique to the partially linear setting. In the first one, the impact of the ECB's monetary policy on the house price index of Eurozone countries is analyzed. In the second one, the role of the wage flexibility of a particular country is studied as an instrument of adjustment against asymmetric shocks.

The paper is structured as follows. Section 2 describes the basic model and the list of assumptions required to establish the asymptotic normality of the local linear estimator. In Section 3, a more efficient nonparametric estimator is defined under cross-sectional dependence. In Section 4, an optimal asymptotical bandwidth is reported. Section 5 reports the empirical applications for some partially linear models. All proofs of the theorems and the Monte Carlo study can be found in the Online Supplementary material.

2. MODEL SPECIFICATION AND ESTIMATION TECHNIQUE

We consider the nonparametric panel data model

$$y_{it} = m(z_t) + \mu_i + v_{it}, \qquad i = 1, \ldots, N; \qquad t = 1, \ldots, T, \tag{2.1}$$

where y_{it} and the q-dimensional vector of explanatory variables z_t are observed, $m(\cdot)$ is a nonparametric function that needs to be estimated, μ_i is the unknown nonstochastic individual heterogeneity, and v_{it} are unobservable zero-mean random variables, uncorrelated across time but possibly correlated over the cross section. Also, v_{it} is independent of z_s for any t and s, when $s \neq t$.

While model (2.1) is of practical interest in itself, it can be more broadly motivated from a semiparametric model which involves also explanatory variables that vary along individuals and time. For example, if y_{it} denotes the deflacted house price index, z_t the interest rate set by the European Central Bank, and x_{it} is a set of explanatory variables (such as GDP per capita, population and unemployment rate), the semiparametric model to consider is such as

$$y_{it} = x_{it}'\beta + m(z_t) + \mu_i + v_{it}, \qquad i = 1, \ldots, N; \qquad t = 1, \ldots, T. \tag{2.2}$$

Therefore, we firstly consider the estimation of (2.1). Later, we show how the proposed methodology is straightforward extended to model (2.2) and several empirical applications are considered.

As already pointed out in previous works, direct estimation of $m(\cdot)$ through the use of standard nonparametric regression leads to biased estimation of $m(\cdot)$, due to the presence of μ_i. In order to overcome this situation, we follow Robinson (2012) and Lee and Robinson (2015) and impose the following restriction

$$\sum_{i=1}^{N} \mu_i = 0, \tag{2.3}$$

where the model (2.1) can be rewritten as

$$\overline{y}_{At} = m(z_t) + \overline{v}_{At}, \tag{2.4}$$

the A subscript denotes averaging and the cross-sectional means are defined as

$$\overline{y}_{At} = \frac{1}{N}\sum_{i=1}^{N} y_{it}, \qquad \overline{v}_{At} = \frac{1}{N}\sum_{i=1}^{N} v_{it}.$$

Following Fan and Gijbels (1995), Ruppert and Wand (1994), and Zhan-Qian (1996), among others, and defining Z_z as a $T \times (q+1)$ matrix of the form

$$Z_z = \begin{bmatrix} 1 & (z_1 - z)' \\ \vdots & \vdots \\ 1 & (z_T - z)' \end{bmatrix}$$

and $W_z = diag(h^{-q}K((z_1 - z)/h), \ldots, h^{-q}K((z_T - z)/h))$ as a $T \times T$ matrix for which

$$K(u) = \prod_{\ell=1}^{q} k(u_\ell), \qquad u = (u_1, u_2, \ldots, u_q)',$$

where $k(\cdot)$ a univariate kernel function and h a positive bandwidth. Assuming that

$Z_z' W_z Z_z$ is nonsingular, the corresponding local weighted linear least-squares estimator for $m(z)$ is defined as

$$\widetilde{m}(z;h) = e_1' \left(Z_z' W_z Z_z\right)^{-1} Z_z' W_z \overline{y}_A, \tag{2.5}$$

where e_1 is the $(q+1) \times 1$ vector having 1 in the first entry and all other entries 0, $\overline{y}_A = \left(\overline{y}_{A1}, \ldots, \overline{y}_{AT}\right)'$ is a T-dimensional vector. Of course, a general diagonal or non-diagonal bandwidth matrix could be employed, but for the sake of simplicity, a single scalar bandwidth is used.

In order to establish the asymptotic properties of $\widetilde{m}(z;h)$ and other estimators proposed in the following sections, some notations are needed. Let $\mathcal{H}_m(z) = \partial m(z)/\partial z \partial z'$ be the Hessian matrix of $m(\cdot)$, we define $\Phi(z) = tr(\mathcal{H}_m(z))$ as the bias measure of the estimator. We let $\mathcal{Z} \in I\!\!R^q$ be the support of z_t. Furthermore, the following assumptions are necessary.

Assumption 2.1. *The idiosyncratic errors v_{it} are such that:*

- *For all i, t, $E(v_{it}) = 0$.*
- *For all i, j, t, there exist finite constants ω_{ij} such that $E(v_{it}v_{jt}) = \omega_{ij}$, and for all i, j, t, s, $E(v_{it}v_{js}) = 0$ when $t \neq s$.*

Assumption 2.2. *Let $\{z_t\}_{t=1,\ldots,T}$ be a set of i.i.d. $I\!\!R^q$-random variables, where z_t has continuous probability density function (pdf), $f(z)$, that is twice continuously differentiable with bounded second order derivatives in a neighborhood of $z \in int(\mathcal{Z})$. Also, z_t is independent of v_{is} for any t and s, when $s \neq t$.*

Assumption 2.3. *In a neighborhood of $z \in int(\mathcal{Z})$, all second-order derivatives of $m(\cdot)$ are bounded and uniformly continuous.*

Assumption 2.4. *K is a product kernel such that $K(u) = \prod_{\ell=1}^{q} k(u_\ell)$, where $k(\cdot)$ is a compactly supported bounded kernel such that $\int k(u)du = 1$, $\int uu'k(u)du = \mu_2(K)I_q$, and $\int k^2(u)du = R(K)$, where $\mu_2(K) \neq 0$ and $R(K) \neq 0$ are scalars and I_q is the $q \times q$ identity matrix. In addition, all odd-order moments of k vanish, that is, $\int u_1^{\iota_1} \ldots u_q^{\iota_q} k(u)du = 0$ for all nonnegative integers $\iota_1, \ldots, \iota_q$ such that their sum is odd.*

Assumption 2.5. *h is a positive bandwidth that, as $T \to \infty$, $h + (Th^q)^{-1} \to 0$.*

Assumption 2.2 characterizes the idiosyncratic component. Assumption 2.1 is a rather strong assumption in a context of time series variables. However, our interest in this paper is to analyze the impact of cross-sectional dependence in terms of the efficiency of the local linear estimator, so we opt to impose this condition. In this way, we are able to isolate the impact of this dependence. Under suitable weak dependence conditions in the time dimension we expected to get similar results. Assumptions 2.3-2.5 are basically smoothness and boundedness conditions. Assumptions 2.4-2.5 are standard in the local linear literature, where Assumption 2.5 is related to the bandwidth condition for smoothing techniques.

The difficulty of obtaining an exact expression of the MSE is well known. Then, to assess the speed at which $\widetilde{m}(z;h)$ converges, we study the following conditional MSE

$$MSE(\widetilde{m}(z;h)|\mathbb{Z}) = E[\left(\widetilde{m}(z;h) - m(z)\right)^2|\mathbb{Z}],$$

where $\mathbb{Z} = (z_1, \ldots, z_T)$ is the observed covariate vector. The following two theorems are essentially restatements of earlier results, so no proofs are given.

THEOREM 2.1. *Assume conditions 2.1-2.5 hold. For z in the interior of $\mathcal{Z}$ and $f(z) > 0$, as $T \to \infty$,*

$$MSE(\widetilde{m}(z;h)|\mathbb{Z}) \sim \left(\frac{h^2 \mu_2^q(K)}{2} \right)^2 \Phi^2(z) + \frac{R^q(K)}{Th^q f(z)} \nu_N,$$

where $\nu_N = N^{-2} \imath_N' E(v_{\cdot t} v_{\cdot t}') \imath_N$, $v_{\cdot t} = (v_{1t}, \ldots, v_{Nt})'$ is a N-dimensional vector, and $\imath_N$ is a N-dimensional vector of ones.

Theorem 2.1 contains rather standard results in this literature (see Theorem 2.2 in Ruppert and Wand (1994)) about the conditional mean and variance of $\widetilde{m}(z;h)$. Nevertheless, the variance term exhibits a new element which reflects the strength of the cross-sectional dependence, that is, ν_N. Further, the results of Theorem 2.1 are valid for both N fixed and large.

With the aim of obtaining the asymptotic distribution of $\widetilde{m}(z;h)$, the following additional assumption is needed.

ASSUMPTION 2.6. *For some $\varepsilon > 0$, $E[|v_{it}|^{(2+\varepsilon)}]$ exists and is bounded.*

Using the above condition, we obtain the following asymptotic distribution for the local linear estimator, for which a similar proof scheme as in Cai and Li (2008) has been followed in order to check the Lyapounov condition.

THEOREM 2.2. *Assume conditions 2.1-2.6 hold. For z in the interior of $\mathcal{Z}$ and $f(z) > 0$, as $T \to \infty$,*

$$\sqrt{Th^q} \nu_N^{-1/2} \big(\widetilde{m}(z;h) - m(z) - b(z;h) \big) \xrightarrow{d} N\big(0, v(z)\big)$$

where

$$\begin{aligned} b(z;h) &= \frac{h^2}{2} \mu_2^q(K) \Phi(z) + o_p(h^2), \\ v(z) &= R^q(K)/f(z). \end{aligned}$$

This theorem shows that $\widetilde{m}(z;h)$ is consistent and asymptotically normal with a rate of convergence which depends on the rate of increase, if any, of ν_N. Furthermore, as it is expected from the nonparametric literature, the local linear estimator exhibits a bias reduction with respect to the Nadaraya-Watson estimator, but the variance term is of the same order as the obtained in Lee and Robinson (2015). Therefore, as it is usual when cross-sectional dependence is allowed, an alternative estimator with better asymptotic properties in terms of variance-reduction can be obtained by taking into account the information of the correlation matrix, i.e., ν_N. In the following section, an alternative technique is developed in this sense.

3. EFFICIENT NONPARAMETRIC ESTIMATION

In this section, we propose an alternative nonparametric estimator that is more efficient than the local linear estimator (2.5) taking into account the potential cross-sectional dependence. With this aim, the following assumption about the random error is imposed.

Assumption 3.1. *Let $v_{\cdot t}$ be i.i.d. with zero mean vector and covariance matrix $\sigma_0^2\Omega$, where σ_0^2 is a positive scalar and Ω is a $N \times N$ positive definite matrix.*

Hence, under Assumption 3.1 it can be shown that $\nu_N = N^{-2}\sigma_0^2 \imath_N' \Omega \imath_N$. Then, as N increases, we get $\nu_N = O(N^{-1})$ and the rate of convergence of $\widetilde{m}(z;h)$ will be $\sqrt{NTh^q}$, which is analogous to the common weak dependence assumption in time series. On its part, boundedness ω_{ij} implies $\nu_N = O(1)$ and the rate of convergence of $\widetilde{m}(z;h)$ will be $\sqrt{Th^q}$, allowing "long-range cross-sectional dependence". (See Robinson (2012) for a deeper discussion).

In this situation, developing more efficient estimators requires $m(\cdot)$ to be identified in a different way from that in Section 2. Then, following the spirit in Robinson (2012) we rewrite (2.1) as

$$y_{\cdot t} = \imath_N m^{(\varpi)}(z_t) + \mu^{(\varpi)} + v_{\cdot t}, \qquad t = 1, \ldots, T, \tag{3.6}$$

where $y_{\cdot t} = (y_{1t}, \ldots, y_{Nt})'$, $\mu^{\varpi} = (\mu_1^{\varpi}, \ldots, \mu_N^{\varpi})'$, and $v_{\cdot t} = (v_{1t}, \ldots, v_{Nt})'$ are N-dimensional vectors, and ϖ represents a vector of weights such as $\varpi = (\varpi_1, \ldots, \varpi_N)'$.

To identify the parameters of interest in (3.6) and given that condition (2.3) was arbitrary, it is imposed

$$\varpi' \imath_N = 1, \tag{3.7}$$

$$\varpi' \mu^{(\varpi)} = 0. \tag{3.8}$$

Premultiplying by ϖ both sides of (3.6) and imposing conditions (3.7)-(3.8), it is obtained

$$\varpi' y_{\cdot t} = m^{(\varpi)}(z_t) + \varpi' v_{\cdot t}. \tag{3.9}$$

Comparing (2.4) and (3.9) it can be noted that there is a vertical shift between $m^{(\varpi)}(\cdot)$ identified by (3.7)-(3.8) and $m(\cdot)$ identified by (2.3) such as $m^{(\varpi)}(z) - m(z) = \varpi'\mu$ for all z. See Robinson (2012) for further details. Further, we choose ϖ to minimize variance. In place of ν_N, under Assumption 3.1 we have $\nu_{N\varpi} = Var(\varpi' v_{\cdot t}) = \sigma_0^2 \varpi' \Omega \varpi$, and deduce the optimal $\varpi = \varpi^*$ subject to (3.7) obtaining

$$\varpi^* = \arg\min_{\varpi} \nu_{N\varpi} = \left(\imath_N' \Omega^{-1} \imath_N\right)^{-1} \Omega^{-1} \imath_N. \tag{3.10}$$

Replacing (3.10) in (3.9) and following a similar procedure as above, an optimal GLS local linear estimator for the unknown function can be proposed such as

$$\widetilde{m}^{(\varpi)}(z;h) = e_1'(Z_z' W_z Z_z)^{-1} Z_z' W_z Y' \varpi^*, \tag{3.11}$$

where $Y = (y_{1\cdot}, \ldots, y_{N\cdot})'$ is a $(N \times T)$ matrix and $y_{i\cdot} = (y_{i1}, \ldots, y_{iT})'$.

However, this estimator is infeasible given that ϖ^* depends on Ω which is generally unknown. To overcome this problem, we propose to parameterize the cross-sectional dependence defining $\Omega = \Omega(\theta_0)$, where θ_0 is a r-dimensional parameter vector known only to lie in a given compact subset Θ of $\mathbb{R}^r$. Note that investing in a correct parametric model for $\Omega(\theta_0)$ can be very appealing for several reasons. On the one hand, it can lead to better finite sample properties of the resulting estimates. On the other hand, this approach allows to avoid restrictive assumptions about the growing rates among N and T, as it was required in Robinson (2012) or Lee and Robinson (2015). Finally, it enables to cover several situations of interest. More precisely, when the cross-sectional dependence is described by SAR(1) models we get

$$\Omega(\theta_0) = (I_N - \theta_0 W_N)^{-1}(I_N - \theta_0 W_N)^{-1'}, \tag{3.12}$$

where $\theta_0 \in (-1, 1)$, I_N is a $N \times N$ identity matrix, and W_N is a $N \times N$ row- and column-normalized weight matrix. Alternatively, if the dependence is represented by means of SMA(1) models we get

$$\Omega(\theta_0) = (I_N + \theta_0 W_N)(I_N + \theta_0 W_N)'. \tag{3.13}$$

Note that if $\Omega(\theta_0)$ is a known function of θ_0, and θ_0 is known, $\Omega(\theta_0)$ is just a case of the known Ω in the infeasible estimates of Robinson (2012) and Lee and Robinson (2015). Nevertheless, the difficulty of this procedure stems from the fact that $\Omega(\cdot)$ is a known function, but θ_0 and σ_0^2 are unknown. To overcome it, we propose to replace these terms by suitable estimators that can be obtained by the Pseudo-Maximum Likelihood Estimation (PMLE).

Denote by θ and σ^2 any admissible values of θ_0 and σ_0^2, respectively, and let $\Omega(\theta)$ be the $N \times N$ matrix with (i, j)th element $\omega_{ij}(\theta)$. For any $\theta \in \Theta$, we define the approximate Gaussian pseudo log-likelihood as

$$lnL_{NT}(\theta, \sigma^2) = -\frac{NT}{2}ln(2\pi) - \frac{NT}{2}ln\sigma^2 - Tln|\Omega(\theta)|^{1/2} - \frac{1}{2\sigma^2}\sum_{t=1}^{T} \widetilde{v}'_{\cdot t}\Omega^{-1}(\theta)\widetilde{v}_{\cdot t}, \tag{3.14}$$

where $\widetilde{v}_{\cdot t} = (\widetilde{v}_{1t}, \dots, \widetilde{v}_{Nt})'$ is a N-dimensional vector of transformed error terms whose it-th element is of the form $\widetilde{v}_{it} = y_{it} - \overline{y}_{iA} + \overline{y}_{AA} - \widetilde{m}(z_t; h)$, where $\overline{y}_{iA} = \frac{1}{T}\sum_{t=1}^{T} y_{it}$ and $\overline{y}_{AA} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{it}$. Note that this transformation was already proposed in Robinson (2012) and it is very appealing since it provides a solution for the error term that is valid when $T \to \infty$ and N is either small or large.

Hence, our estimates of θ_0 and σ_0^2 minimize $lnL_{NT}(\theta, \sigma^2)$. For given θ, (3.14) is minimised with respect to σ^2 obtaining

$$\overline{\sigma}^2(\theta) = \frac{1}{NT}\sum_{t=1}^{T} \widetilde{v}'_{\cdot t}\Omega^{-1}(\theta)\widetilde{v}_{\cdot t}, \tag{3.15}$$

and the concentrated function is such as

$$Q(\theta) = \frac{NT}{2}(ln(2\pi) + 1) + \frac{NT}{2}ln\overline{\sigma}^2(\theta) + \frac{T}{2}ln|\Omega(\theta)|. \tag{3.16}$$

In this situation, we define the Gaussian pseudo-maximum-likelihood estimators of θ_0 and σ_0^2 as

$$\widehat{\theta} = \arg\min_{\theta \in \Theta} Q(\theta), \tag{3.17}$$

$$\widehat{\sigma}^2 = \overline{\sigma}^2(\widehat{\theta}). \tag{3.18}$$

Then, replacing (3.17) in (3.10), the corresponding optimal feasible GLS (FGLS) local linear estimator of $m^{(\varpi)}(\cdot)$ is of the form

$$\widehat{m}^{(\varpi)}(z; h; \widehat{\theta}) = e_1'(Z_z' W_z Z_z)^{-1} Z_z' W_z Y' \varpi^*(\widehat{\theta}), \tag{3.19}$$

where $\varpi^*(\widehat{\theta}) = \left(\imath_N'\Omega^{-1}(\widehat{\theta})\imath_N\right)^{-1}\Omega^{-1}(\widehat{\theta})\imath_N$.

For a real matrix A, we denote by $\|A\|$ the spectral norm of A, i.e., the square root of the largest eigenvalue of $A'A$. To analyze the main asymptotic properties of this estimator, the following additional assumptions are necessary.

ASSUMPTION 3.2. *For a neighbourhood $\mathcal{N}$ of θ_0, $\overline{\lim_{N\to\infty}} \sup_{\mathcal{N}}(\|\Omega(\theta)\| + \|\Omega^{-1}(\theta)\|) < \infty$.*

ASSUMPTION 3.3. *Let (r) be the r-th derivative with respect to θ_r and assuming that $\Omega(\theta)$ is differentiable for a neighbourhood $\mathcal{N}$ of θ_0, $\overline{\lim_{N\to\infty}} \sup_{\mathcal{N}} \|\Omega^{(r)}(\theta)\| < \infty$.*

Given the several specifications of the $\Omega(\cdot)$ and the relative simplicity of Assumptions 3.2 and 3.3, it is interest in checking these conditions under more primitive conditions. With this aim, we take as benchmark the SMA model defined previously, where $\theta \in \Theta = \{\theta : |\theta| \leq 1 - \epsilon\}$ is assumed for $\epsilon > 0$ sufficiently small and $\|W_N\| \leq 1$. In order to check the weak dependence and invertibility condition in Assumption 3.2, we can write

$$\|\Omega(\theta)\| \leq \|(I_N - \theta W_N)^{-1}\|^2 \leq \left(1 - \sum_{r=0}^{\infty} |\theta|^r \|W_N\|^r\right)^{-2} \leq (1 - |\theta|)^{-2} \leq 1/\epsilon^2 < \infty,$$

given that, for normalization, $\|W_N\| \leq 1$ implies $\|W_N\| = 1$, if $W_N\imath_N = \imath_N$. Under similar reasoning,

$$\|\Omega(\theta)\|^{-1} \leq \|(I_N - \theta W_N\|^2 \leq (1 - |\theta|)^2 \leq \epsilon^2 < \infty,$$

so it is proved that Assumption 3.2 holds. Similar results can be obtained whether $\Omega(\theta)$ is parameterized as a SMA(1) or naturally factored as $\Omega(\theta) = BB'$, where B is a known matrix function of θ. Therefore, Assumption 3.2 effectively upper- and lower-bounds variances and mildly limits the extend of cross-sectional dependence, independently of the type of dependence specification. Under cross-sectional uncorrelatedness, this condition reduces to $\overline{\lim_{N\to\infty}} \sup(\omega_{ii} + \omega_{ii}^{-1}) < \infty$.

We focus now on the validity of Assumption 3.3. Remember that $\Omega(\theta)$ is differentiable over the compact support Θ. Then, we can write

$$\begin{aligned}\|\Omega^{(j)}(\theta)\| &\leq \|(I_N - \theta W_N)^{-1}W_N\Omega(\theta)\| + \|\Omega(\theta)W_N'(I_N - \theta W_N)'\| \\ &\leq 2\|(I_N - \theta W_N)^{-1}\|\|\Omega(\theta)\| \leq 2\left(1 - \sum_{r=0}^{\infty} |\theta|^r \|W_N\|^r\right)^{-1} = 2/\epsilon < \infty,\end{aligned}$$

so Assumption 3.3 is guaranteed.

In order to analyze the consistency of $\widehat{\lambda} = (\widehat{\theta}, \widehat{\sigma}^2)'$, we denote $G(\theta) = \Omega(\theta)^{-1}\Omega^{(1)}(\theta)$, where $\Omega^{(1)}(\theta)$ is the first-order derivative with respect to θ, and $\pi_{NT}(\theta) = \frac{T}{2}tr\{G(\theta)G(\theta)'\}$. Also, the following assumption is required.

ASSUMPTION 3.4. *Either (a) the limit of π_{NT} is nonsingular for each possible θ in Θ; or (b)*

$$\lim_{N\to\infty} \left(\frac{1}{N}|\sigma_0^2\Omega^{-1'}\Omega^{-1}| - \frac{1}{N}ln|\overline{\sigma}^2(\theta)\Omega^{-1}(\theta)'\Omega^{-1}(\theta)|\right) \neq 0 \quad for \quad \theta \neq \theta_0.$$

Assumption 3.4 is a condition for the nonsingularity of the limiting information matrix $\Sigma_{NT}(\lambda_0)$, where $\Sigma_{NT}(\lambda_0) = -E\left[\frac{1}{NT}\frac{\partial^2 lnL_{NT}(\lambda_0)}{\partial\lambda\partial\lambda'}\right]$. For local identification, a sufficient condition (but not necessary) is that the information matrix $\Sigma_{NT}(\lambda_0)$ is nonsingular and $-E\Big(\frac{1}{NT}\frac{\partial^2 lnL_{NT}(\lambda_0)}{\partial\lambda\partial\lambda'}\Big)$ has full rank for any λ in some neighborhood $\mathcal{N}(\lambda_0)$ of λ_0 (see Rothenberg (1971)). However, when $\lim_{T\to\infty} E[\pi_{NT}(\theta)]$ is singular, global identification can still be obtained from Assumption 3.4(b) imposing a condition on the variance structure of the model.

THEOREM 3.1. *Assume conditions 2.2-2.6 and 3.1-3.4 hold, $\lambda_0 = (\theta_0', \sigma_0^2)'$ is identified and*

$$\widehat{\lambda} - \lambda_0 \xrightarrow{p} 0.$$

Considering now the asymptotic distribution of the PMLE estimators, we obtain the following result. We denote $\Omega_{NT}(\lambda_0) = E\Big[\frac{1}{NT}\frac{\partial lnL_{NT}(\lambda_0)}{\partial\lambda} \times \frac{\partial lnL_{NT}(\lambda_0)}{\partial\lambda}\Big] - \Sigma_{NT}(\lambda_0)$.

THEOREM 3.2. *Assume conditions 2.2-2.6 and 3.1-3.4 hold, as N and T tends to infinity,*

$$\sqrt{NT}(\widehat{\lambda} - \lambda_0) \xrightarrow{d} N\big(0, \Sigma(\lambda_0)^{-1} + \Sigma(\lambda_0)^{-1}\Omega(\lambda_0)\Sigma(\lambda_0)^{-1}\big),$$

where $\Sigma(\lambda_0) = -\lim_{N\to\infty} E\Big(\frac{1}{NT}\frac{\partial^2 lnL_{NT}(\lambda_0)}{\partial\lambda\partial\lambda'}\Big)$ *and* $\Omega(\lambda_0) = \lim_{N\to\infty}\Omega_{NT}(\lambda_0)$.

Unlike what happens in other studies as Lee (2004) or Yu et al. (2008) among others, this asymptotic result is valid regardless of the distribution of the error term since it does not depend on any term related to the third order moment of the error.

Considering now the asymptotic distribution of the optimal GLS local linear estimator, the following result is obtained.

THEOREM 3.3. *Assume conditions 2.2-2.6 and 3.1-3.3 hold. For z in the interior of $\mathcal{Z}$ and $f(z) > 0$, as N and T tend to infinity,*

$$\sqrt{Th^q}\nu^{(\varpi)-1/2}(\theta_0)\Big(\widetilde{m}^{(\varpi)}(z;h;\theta_0) - m^{(\varpi)}(z) - \frac{h^2}{2}\mu_2^q(K)\Phi^{(\varpi)}(z) + o_p(h^2)\Big) \xrightarrow{d} N\bigg(0, \frac{\sigma_\eta^2 R^q(K)}{f(z)}\bigg),$$

where $m^{(\varpi)}(z) = m(z) + \varpi'\mu$, $\nu^{(\varpi)}(\theta_0) = \big(\imath_N'\Omega^{-1}(\theta_0)\imath_N\big)^{-1}$, *and* $\Phi^{(\varpi)}(z) = tr(\mathcal{H}_m^{(\varpi)}(z))$, *for* $\mathcal{H}_m^{(\varpi)}(z)$ *being the Hessian matrix of* $m^{(\varpi)}(\cdot)$.

Finally, in order to show the efficiency gains of the FGLS procedure proposed in this paper, it is necessary to prove the asymptotic equivalence between $\widehat{m}^{(\varpi)}(z;h;\widehat{\theta})$ and $\widetilde{m}^{(\varpi)}(z;h;\theta_0)$ under the consistency of $\widehat{\theta}$.

THEOREM 3.4. *Assume conditions 2.2-2.6 and 3.1-3.4 hold for $z \in I\!R^q$. As N and T tends to infinity,*

$$\widehat{m}^{(\varpi)}(z;h;\widehat{\theta}) - \widetilde{m}^{(\varpi)}(z;h;\theta_0) = o_p\bigg(\frac{1}{\sqrt{NTh^q}}\bigg),$$

provided that $|\mu_i|$ is bounded by a constant.

Looking at the results in Theorem 3.3, we could be noted that the optimal GLS estimator is consistent an asymptotically normal, with a rate of convergence of $\sqrt{Th^q}\nu^{(\varpi)-1/2}(\theta_0)$.

Therefore, the convergence of $m^{(\varpi)}(\cdot)$ depends on the rate of increase, if any, of $\nu_N^{(\varpi)}(\theta_0)$. In addition, from Theorem 3.4 it can be highlighted that the optimal FGLS and GLS estimators are asymptotically equivalent when both N and T are large. Comparing the results in Theorems 2.2 and 3.3-3.4, the efficiency improvement of this new procedure is corroborated since $\nu^{(\varpi)}(\theta_0) < \nu_N(\theta_0)$, unless $\Omega(\theta_0)$ has an eigenvector $\imath_N$. Note that this phenomena was already pointed out in Robinson (2012), where this particular situation is analyzed in factor and spatial autoregressive models. Further, the rate of convergence of $m^{(\varpi)}(z)$ is faster than $m(z)$ if $\nu_N(\theta_0)/\nu^{(\varpi)}(\theta_0) \to 0$.

To sum up, all these results are in line with those obtained in Robinson (2012) and Lee and Robinson (2015). Therefore, it is shown that the methodology developed in this paper enables us to overcome the usual restriction in this literature about the growth rate among N and T, but at the price of assuming a fully parametric structure of the covariance matrix.

4. OPTIMAL BANDWIDTH CHOICE

As it is clear from previous sections, the bandwidth term, h, plays a crucial role in the estimation of $m(\cdot)$. Choosing a large h, the variance of our estimator will be reduced but at the cost of enlarging bias. In order to solve this trade-off, h should be chosen to minimize a certain distance measure. One such measure is the conditional mean integrated square error (MISE) defined as

$$MISE(\widehat{m}^{(\varpi)}(z;h;\widehat{\theta})|\mathbb{Z}) = E\int_0^1 E[\big(\widehat{m}(u;h;\widehat{\theta}) - m(u)\big)^2|\mathbb{Z}]\psi(u)du,$$

where $\psi(u)$ is a weight function chosen to ensure that the integral converges.

More precisely, it is easy to show that the bandwidth that minimizes the asymptotic conditional MISE of $\widehat{m}^{(\varpi)}(z;h;\widehat{\theta})$ is

$$h_{opt} = \left(\frac{qR^q(K)\sigma_0^2\nu^{(\varpi)}(\theta_0)}{T\mu_2^{2q}(K)f(z)\{\Phi^{(\varpi)}(z)\}^2}\right)^{1/(q+4)}. \tag{4.20}$$

However, the optimal bandwidth (4.20) cannot be computed directly in practice. Although $R(K)$ and $\mu_2(K)$ can be trivially calculated, $\Phi^{(\varpi)}(\cdot)$, $f(\cdot)$, θ_0, and σ_0^2 are unknown. In order to consistently estimate both $\Phi^{(\varpi)}(\cdot)$ and $f(\cdot)$, standard nonparametric techniques can be used. (See Gasser et al. (1991), for example). Meanwhile, to provide a consistent estimator for θ_0 and σ_0^2, we propose to use the Gaussian pseudo log-likelihood estimates obtained in (3.18) and (3.19).

Therefore, the feasible optimal bandwidth is

$$\widehat{h}_{opt} = \left(\frac{qR^q(K)\nu^{(\varpi)}(\widehat{\theta})\widehat{\sigma}^2}{T\mu_2^{2q}(K)\widehat{f}(z)\{\widehat{\Phi}^{(\varpi)}(z)\}^2}\right)^{1/(q+4)}, \tag{4.21}$$

where $\nu^{(\varpi)}(\widehat{\theta}) = \big(\imath_N'\Omega^{-1}(\widehat{\theta})\imath_N\big)^{-1}$.

In order to show that the optimal bandwidth and the feasible one are asymptotically equivalent, the following additional assumption is needed.

Assumption 4.1. *As T tends to infinity,*

$$\widehat{f}(z) - f(z) = O_p\big(\|\Omega(\theta_0)\|^{-1}\|\Omega(\widehat{\theta}) - \Omega(\theta_0)\|\big),$$
$$\{\widehat{\Phi}^{(\varpi)}(z)\}^2 - \{\Phi^{(\varpi)}(z)\}^2 = O_p\big(\|\Omega(\theta_0)\|^{-1}\|\Omega(\widehat{\theta}) - \Omega(\theta_0)\|\big).$$

Assumption 4.1 is written in an nonprimitive way but it is needed to guarantee that the effect of estimating bias is negligible. However, we believe that it is not going to be a problem since our main interest is to analyze the impact of cross-sectional dependence and $\Phi^{(\varpi)}(z)$ is not affected by that.

THEOREM 4.1. *Assume conditions 2.2-2.5, 3.1-3.4, and 4.1 hold, as T tends to infinity,*

$$\frac{\widehat{h}_{opt}}{h_{opt}} \xrightarrow{p} 1.$$

Therefore, in Theorem 4.1 it is shown that despite what is obtained in Lee and Robinson (2015) and Robinson (2012), in this particular case any requirement on the relative rates of N and T is necessary to show the asymptotically equivalence of the optimal bandwidth.

5. EMPIRICAL APPLICATIONS

In this section we apply our estimation methodology in two real data sets with the aim of showing the empirical viability of this methodology. In both cases we consider partially linear regression models with the aim of showing that the estimating procedure proposed in this paper can be easily extended to that particular setting. In the first one, we analyze the impact of the low interest rate policy carried out by the ECB on the house price index of Eurozone countries since the outbreak of the financial crisis in 2007-2008. In the second one, the wage flexibility of a particular country is studied as an instrument of adjustment against asymmetric shocks using the Spanish case as a relevant example.

The Europe's unification in the Economic and Monetary Union (EMU) has radically changed the economic scenario of the Eurozone countries and nowadays there is still an ongoing debate about the cost and benefits of joining it. See Mélitz (1997) and Tavlas (1993, 1994), among others, for a detailed discussion. On the one hand, the theory of Optimum Currency Areas (OCA) states that the single currency eliminates the risk of the exchange rate and reduces transaction costs among member countries by favoring intra-community trade. On the other hand, detractors of EMU argue that the loss of the exchange and monetary policy in favor of the ECB reduces the scope of the governments to deal against adverse asymmetric shocks. Thus, the Eurozone countries have to resort to another type of adjustment instrument to cope with these difficulties.

The growing importance of the euro in international trades and the increasing trade activities which result from adopting the currency clearly shows that benefits will outweigh costs. However, the global financial crisis in 2007-2008 has acted as an asymmetric shock which has exacerbated a structural problem of competitiveness between the Eurozone countries and the long and medium term stability of the EMU has been called into question. In this situation, analyzing the impact of the loss of the monetary policy among the EMU's member countries is a crucial issue. With this aim, we are going to consider two empirical application: i) to analyze the impact of the ECB's monetary policy on the house price index of the Eurozone countries; ii) to study the role of the wage flexibility in Spain as an adjustment instrument against asymmetric shocks. In both situations, the regression model to estimate is a partially linear regression model of the following form

$$y_{it} = x_{it}'\beta + m(z_t) + \mu_i + v_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{5.22}$$

where x_{it} and z_t are $p \times 1$ and $q \times 1$ vectors of explanatory variables, β is a p-dimensional vector of unknown parameters and $m(\cdot)$ is the unknown smooth function. This model is particularly suitable when x_{it} contains categorical variables, and is often used when the overall number of explanatory variables are large in order to avoid the curse of dimensionality of the fully nonparametric specification. The estimation of this model has received much attention. See Robinson (1988) or Fan and Li (1999).

In order to control for the fixed effects, we impose restriction (2.3) obtaining

$$\overline{y}_{At} = \overline{x}'_{At}\beta + m(z_t) + \overline{v}_{At}, \tag{5.23}$$

where $\overline{x}_{At} = N^{-1}\sum_{i=1}^{N} x_{it}$. Following the Robinson (1988)'s approach to consistently estimate the unknown parameters and unknown function in the above expression, we take the conditional expectations on z_t of (5.23). Subtracting the resulting expression from (5.23), we premultiply each term by the density function in order to avoid the random denominator problem. Then, the resulting expression is such as

$$\overline{y}_{At}f(z_t) - E(\overline{y}_{At}|z_t)f(z_t) = [\overline{x}_{At}f(z_t) - E(\overline{x}_{At}|z_t)f(z_t)]'\beta + \overline{v}_{At}f(z_t), \tag{5.24}$$

where $f(z_t)$ is the density function described previously in Assumption 2.2.

In order to obtain a feasible estimator for β, we propose to estimate the above conditional expectations using some nonparametric methods. Hence, the consistent estimators of $f(z_t)$, $E(\overline{x}_{At}|z_t)f(z_t)$, and $E(\overline{y}_{At}|z_t)f(z_t)$ are $\widehat{f}(z_t) = (1/Ta^q)\sum_{t=1}^{T} K((z_t - z)/a)$, $\widehat{E}(\overline{x}_{At}|z_t)\widehat{f}(z_t) = (1/Ta^q)\sum_{t=1}^{T} \overline{x}_{At}K((z_t - z)/a)$, and $\widehat{E}(\overline{y}_{At}|z_t)\widehat{f}(z_t) = (1/Ta^q)\sum_{t=1}^{T} \overline{y}_{At} \times K((z_t - z)/a)$, respectively, where a is the bandwidth term and $K(\cdot)$ is a kernel function defined as in (2.5).

Let $\widehat{\ddot{x}}_{At} = \overline{x}_{At}\widehat{f}(z_t) - \widehat{E}(\overline{x}_{At}|z_t)\widehat{f}(z_t)$ and $\widehat{\ddot{y}}_{At} = \overline{y}_{At}\widehat{f}(z_t) - \widehat{E}(\overline{y}_{At}|z_t)\widehat{f}(z_t)$. Suppose that $\left(\sum_{t=1}^{T} \widehat{\ddot{x}}_{At}\widehat{\ddot{x}}'_{At}\right)$ is nonsingular, the resulting estimator for β is of the form

$$\widetilde{\beta} = \left(\sum_{t=1}^{T} \widehat{\ddot{x}}_{At}\widehat{\ddot{x}}'_{At}\right)^{-1} \sum_{t=1}^{T} \widehat{\ddot{x}}_{At}\widehat{\ddot{y}}_{At}. \tag{5.25}$$

Considering now the estimation of $m(\cdot)$, we replace $\widetilde{\beta}$ in (5.23) and propose the following local weighted linear least-squares estimator following a similar procedure as in (2.5),

$$\widetilde{m}(z;h) = e'_1(Z'_zW_zZ_z)^{-1}Z'_zW_z(\overline{y}_A - \overline{x}_A\widetilde{\beta}), \tag{5.26}$$

where $\overline{x}_A = (\overline{x}_{A1}, \ldots, \overline{x}_{AT})'$ is a $T \times p$ matrix.

Let $\mathcal{K}_\upsilon$ be the class of kernels and $\mathcal{G}^\alpha_\varsigma$ the function class defined as Definitions 1 and 2 in Robinson (1988), respectively. Kernels in $\mathcal{K}_\upsilon$ are of order υ, and the functions in $\mathcal{G}^\alpha_\varsigma$ are ς-times partially differentiable with a Lipschitz-continuous remainder, α controls the moment properties of the remainder. In particular, the functions in $\mathcal{G}^\alpha_\varsigma$ are bounded. To prove the main asymptotic properties of the estimator of β, we adapt the following assumptions from Li and Stengos (1996) and Qian and Wang (2012).

Assumption 5.1. *$f \in \mathcal{G}\varsigma^\alpha$ for some constant $\varsigma \geq 1$, $m \in \mathcal{G}^{4+\varepsilon}_\upsilon$, $E(x_{it}|z_t) \in \mathcal{G}^{(4+\varepsilon)}_\upsilon$ for some $\varepsilon > 0$ and positive integer υ with $\varsigma < \upsilon \leq \varsigma + 1$.*

Assumption 5.2. *$k \in \mathcal{K}_\upsilon$, and as N and T tends to infinity, $NTa^{2q-4} \to \infty$ and $NTa^{4\upsilon}$.*

These assumptions are fairly standard in the semiparametric literature, but some remarks are necessary. On the one hand, Assumption 5.1 is stronger than that is made in Li and Stengos (1996). In particular, we assume that the density function is bounded and at least first-order partially differentiable with a Lipschitz-continuous remainder since this is required for estimating the nonlinear part. However, this stronger condition is not necessary for the asymptotic properties of $\widetilde{\beta}$, only for $\widetilde{m}(\cdot)$. On the other hand, Assumption 5.2 is slightly weaker than that is assumed in Li and Stengos (1996), which requires $NTa^{2q} \to \infty$. From Assumption 5.2, $4\upsilon > 2q - 4$ or $\upsilon > q/2 - 2$. Hence, a second order kernel ($\upsilon = 2$) can be used if $q < 6$. See Qian and Wang (2012) for a deeper discussion.

THEOREM 5.1. *Under Assumptions 2.1-2.2 and 5.1-5.2, as N and T goes to infinity,*

$$\sqrt{T}\nu_N^{-1/2}\big(\widetilde{\beta} - \beta\big) \xrightarrow{d} N(0, \Gamma^{-1}\Psi\Gamma^{-1}),$$

where $\Gamma = T^{-1}\sum_{t=1}^{T} E\big[\ddot{x}_{At}\ddot{x}'_{At} f^2(z_t)\big]$, $\Psi = T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T} E\big[\ddot{x}_{At}\ddot{x}'_{As} f^2(z_t) f^2(z_s)\big]$, *and* $\nu_N = N^{-2}\sigma_0^2 \imath'_N \Omega(\theta_0)\imath_N$. *Further, we can consistently estimate Γ and Ψ by plugging in the estimates for each term.*

Let $\nu_x = \ddot{x}_{At}\ddot{x}'_{As}$ and assuming that $\nu_N = O_p(N^{-1})$ for both $t = s$ and $t \neq s$, Theorem 5.1 contains two possible situations. If there is no cross-sectional dependence in X_{it}, $\nu_x = O(1)$ and the convergence in distribution of $\widetilde{\beta}$ is $\sqrt{NT}$. On the contrary, if the explanatory variables exhibit some kind of cross-sectional dependence, $\nu_x = O(N^{-1})$ and the rate of convergence is $\sqrt{T}$. Note that whether we are interested in the main asymptotic properties of the nonparametric estimator $\widetilde{m}(z; h)$, we obtain exactly the same results as in Theorem 2.2 using the consistency results of $\widetilde{\beta}$ and following a similar proof scheme as in that theorem.

Note that in the semiparametric framework the bandwidth selection issue requires extra attention, given that the estimation procedure requires bandwidths specified in each step. For the fully parametric part, $(1 + p)$-sets of bandwidth are required, each of dimension q, that can be calculated via cross-validation functions. Later, the bandwidth selection procedure proposed in Section 4 can be used for the nonparametric part. See Henderson and Parmeter (2015) for a more deeper discussion.

Considering the more efficient estimation, we can rewrite (5.22) in vectorial form and given the conditions (3.7)-(3.8) it is obtained

$$\varpi' y_{\cdot t} = \varpi' x_{\cdot t}\beta + m^{(\varpi)}(z_t) + \varpi' v_{\cdot t}, \tag{5.27}$$

where $x_{\cdot t} = (x_{1t}, \ldots, x_{Nt})'$ is a $N \times p$ matrix. Following a similar procedure as in (3.10) to obtain the optimal $\varpi^* = \varpi$, we use the PMLE procedure to obtain the consistent estimators for θ_0 and σ_0^2 by replacing the transformed errors by $\widetilde{v}_{it} = y_{it} - \overline{y}_{iA} + \overline{y}_{AA} - \widetilde{m}(z_t) - (x_{it} - \overline{x}_{iA} + \overline{x}_{AA})'\widetilde{\beta}$.

Let $\widehat{\ddot{x}}^{(\varpi)} = \varpi^*(\widehat{\theta})' x_{\cdot t}\widehat{f}(z_t) - \widehat{E}(\varpi^*(\widehat{\theta})' x_{\cdot t}|z_t)\widehat{f}(z_t)$ and $\widehat{\ddot{y}}^{(\varpi)} = \varpi^*(\widehat{\theta})' y_{\cdot t}\widehat{f}(z_t) - \widehat{E}(\varpi^*(\widehat{\theta})' y_{\cdot t}|z_t)\widehat{f}(z_t)$. Assuming that $\left(\sum_{t=1}^{T}\widehat{\ddot{x}}_{At}^{(\varpi)'}\widehat{\ddot{x}}_{At}^{(\varpi)}\right)$ is nonsingular, the feasible estimator proposed for β is of the form

$$\widehat{\beta} = \left(\sum_{t=1}^{T}\widehat{\ddot{x}}_{At}^{(\varpi)'}\widehat{\ddot{x}}_{At}^{(\varpi)}\right)^{-1}\sum_{t=1}^{T}\widehat{\ddot{x}}_{At}^{(\varpi)'}\widehat{\ddot{y}}_{At}^{(\varpi)}. \tag{5.28}$$

Similarly, the resulting feasible local weighted linear least-squares estimator is such as

$$\widehat{m}^{(\varpi)}(z;h) = e_1'(Z_z'W_zZ_z)^{-1}Z_z'W_z\Big(Y'\varpi^*(\widehat{\theta}) - \sum_{d=1}^{p} X_d'\varpi^*(\widehat{\theta})\widehat{\beta}_d\Big), \tag{5.29}$$

where X_d is a $N \times T$ matrix, for $d = 1,\ldots,p$.

THEOREM 5.2. *Under Assumptions 2.2, 3.1-3.4, and 5.1-5.2 as N and T tends to infinity,*

$$\sqrt{T}\nu^{(\varpi)-1/2}(\theta_0)\big(\widehat{\beta} - \beta\big) \xrightarrow{d} N(0, \Gamma_\varpi^{-1}\Psi_\varpi\Gamma_\varpi^{-1}),$$

where $\Gamma_\varpi = T^{-1}\sum_{t=1}^{T} E[\ddot{x}_{\cdot t}'\varpi\varpi'\ddot{x}_{\cdot t}f^2(z_t)]$, $\Psi_\varpi = T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T} E[\ddot{x}_{\cdot t}'\varpi\varpi'\ddot{x}_{\cdot s}f^2(z_t)f^2(z_s)]$, *and* $\nu^{(\varpi)}(\theta_0) = (\imath_N'\Omega^{-1}(\theta_0)\imath_N)^{-1}$. *Further, we can consistently estimate Γ_ϖ and Ψ_ϖ by plugging in the estimates for each term.*

Again this theorem contains two possible situations. If X_{it} exhibits some kind of cross-sectional dependence, the rate of convergence of this estimator is $\sqrt{T}$ and if there is no cross-sectional dependence on X_{it} this rate is $\sqrt{NT}$. Furthermore, the efficiency improvement of $\widehat{\beta}$ depends again on the behavior of ν_N and $\nu^{(\varpi)}(\theta_0)$ and it is corroborated if $\nu^{(\varpi)}(\theta_0) < \nu_N$. Finally, considering the main asymptotic properties of the nonparametric estimator $\widehat{m}(z;h)$ and using the consistency results of $\widehat{\beta}$, similar results as in Theorems 3.3-3.4 are obtained.

5.1. House price index for Eurozone countries

The financial crisis originated after the sub-prime crisis in the United States (US) in the early 2007 as well as the role played within it by the fluctuations in the housing price has refreshed the debate about the suitable policy response. The EMU's case is specially interesting given that the global crisis has impacted heterogeneously on the property markets of the Eurozone countries.

In countries with their own national currency and monetary policy (such as the US, the United Kingdom or New Zealand, for example) the central bank can increase interest rates to slow down the growth rate of housing prices. However, in the EMU the ECB sets interest rates with the primary interest of maintaining price stability in all member countries. That means that monetary policy cannot longer be used as an instrument of adjustment by these countries to deal with asymmetric negative sector and/or country shocks.

In this framework, the aim of this study is to analyze the determinants of housing prices in the Eurozone countries. Note that predicting changes in housing prices is not an easy task. Apart from microeconomic factors (such as the set of characteristics of the house or the environment in which it is located), certain macroeconomic factors such as GDP or interest rate, among others, play a critical role since they determine the structure of aggregate demand. Thus, we are interested in the following issues: First, to identify the effect of the more relevant macroeconomic factors over that price level; Second, to analyze whether the low interest rate policy carried out by the ECB since the outbreak of the financial crisis in 2007 has a non-linear effect over the housing price; Third, to improve the precision of these estimators taking into account the cross-sectional dependence, as it was established previously. To the best of our knowledge, that is a totally new approach in the analysis of the housing price determinants.

Following what it is established in Muellbauer and Murphy (1997) and trying to incorporate flexibility in the regression model, we propose the following semi-parametric model

$$hpi_{it} = \beta_1 gdppc_{it} + \beta_2 pop_{it} + \beta_3 unem_{it} + m(rate_t) + \mu_i + v_{it}, \tag{5.30}$$

where $i \in \{1, \ldots, N\}$ denotes the country, $t \in \{1, \ldots, T\}$ is the number of observations per country, $\beta = (\beta_1, \ldots, \beta_3)$ is the unknown parameters vector of interest, and $m(\cdot)$ is an unknown function to estimate. Further, it is assumed that the cross-sectional dependence is described by SMA(1) models. Not every country of the EMU has data on real house prices over a long time period. Then, the data used in this study includes information on quarterly terms for $N = 12$ countries (i.e., Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Portugal, and Spain) from $1999(I)$ to $2019(IV)$, so $T = 84$.

The data used in this study come from two main sources. The interest rate over main refinancing operations (*rate*) is obtained from ECB. The data for deflacted house price index (*hpi*), total population (*pop*), GDP per capita (*gdppc*), and total unemployment rate (*unem*) are obtained from the OECD Statistics database. The variables *gdppc* and *pop* are taken in logarithms.

With the aim of consistently estimate the semi-parametric regression model (5.30), we follow a similar estimation procedure as in (5.22). For the sake of simplicity, the bandwidths (a and h) have been chosen following the Silverman rule-of-thumb. Of course, the feasible optimal bandwidth justified in Theorem 4.1 could be used instead.

The estimated results are collected in Table 1, where SP denotes the semi-parametric results when cross-sectional dependence is ignored and SP.FGLS is the semi-parametric results when this dependence is taken into account. Further, for the sake of comparison we include the OLS estimates for the fully parametric specification of equation (5.30) with a quadratic term for *rate*.

Analyzing the OLS results of Table 1, it can be noted that GDP per capita is highly significant and has the expected positive sign indicating that changes in income are strongly positively related to changes in house prices. However, the coefficient estimates for population and unemployment rate are the opposite as it was expected. If we compare the OLS results with the corresponding for SP it can be observed that the signs and the significance of the linear variables change considerably. All these results corroborate that these OLS estimators may be subjected to some misspecification problems.

If we analyze the semiparametric results increases in the unemployment rate has the expected negative sign in house prices, whereas population has a strong positive relationship to house prices. Interestingly, in this semiparametric specification it is obtained a highly significant and negative sign for the GDP per capita. Furthermore, as it was expected, greater differences are not detected when we compare that results with the corresponding for SP.FGLS.

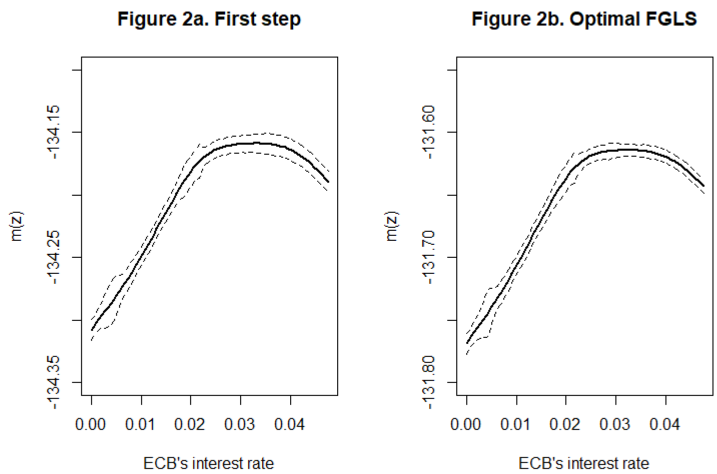
In Figure 1, the estimated curves for the interest rate set by the ECB are plotted. Specifically, in Figure 1a it is detected a positive relationship between the interest rate an the HPI up to certain point after which there seems to be a slightly negative trend. Meanwhile, Figure 1b provides more precise results corroborating that result. Hence, we can conclude that when we control for both nonlinear effect of interest rate and

Table 1. Results for the HPI determinants.

	OLS	SP	SP.FGLS
gdppc	1.3829***	-0.9749***	-0.9347***
	(0.0524)	(0.1700)	(0.1677)
pop	-1.4024***	14.9025	14.6009***
	(0.0589)	(1.2799)	(1.2624)
unem	0.1908***	-4.3925***	-4.3533***
	(0.3623)	(0.4186)	(0.4129)
rate	23.6902***	.	.
	(1.6493)	.	.
rate^2	-396.5664***	.	.
	(29.2446)	.	.
$\widehat{\theta}$	.	.	0.1034
$\widehat{\sigma}^2$	.	.	0.1109

Note: Standard errors are in parenthesis. *** indicates statistical significance at the 1% level.

Figure 1. Nonparametric estimates for the ECB's interest rate.



Note: Thick line denotes the estimates curve while dotted line is the 95% pointwise confidence interval.

cross-sectional dependence among the EMU's countries we get a nonlinear negative relationship, as it was expected from the literature.

To sum up, all these results corroborate that the interest rate set by the ECB has a nonlinear effect over the HPI. Further, the policy of low interest rates applied by the ECB with the aim of reactivating the economies of the EMU's member countries seems to have the desired effect when we control for the dependence among these countries.

5.2. Regional adjustment and wage flexibility

As it has been pointed out previously, one of the main disadvantages of belonging to a monetary union is the loss of both the monetary policy and the exchange rates as adjustment instruments. Furthermore, despite the efforts made in recent years towards convergence among the EMU's member countries, the EMU is still characterized by the existence of structural regional differences in income per capita and unemployment rate as well as the limited interregional labor mobility. In this framework, wage flexibility becomes one of the main adjustment mechanisms of the Eurozone countries to deal with negative asymmetric shocks through lower labour cost, see Abraham (1996) for further details.

With the aim of analyzing the regional wage flexibility as an instrument of adjustment against the global crisis that took place in $2007-2008$, we carry out an empirical analysis of wage flexibility analyzing the particular case of the Spanish labour market. Taking as benchmark the model of wage bargaining between unions and employees as in Abraham (1996) and trying to incorporate some flexibility in the regression model, we propose the following estimated equation

$$\dot{wage}_{it} = \beta_1 \dot{unem}_{it} + \beta_2 \dot{wage}_t + m(\dot{prod}_t) + \mu_i + v_{it}, \tag{5.31}$$

where $\dot{wage}_{it}$ denotes the real wage growth in region i and period t, $\dot{unem}_{it}$ is the regional unemployment growth rate, $\dot{wage}_t$ is the average national real wage growth, and $\dot{prod}_t$ is the evolution of the growth in national labour productivity measured with a proxy defined as the growth of the national GDP per capita. All variables except unemployment are in logs. Also, β_1 and β_2 are the structural parameters to estimate, $\dot{prod}_t$ has an influence on $\dot{wage}_{it}$ of unknown form, and v_{it} are the random errors generated following SAR(1) models as in (3.12). For this study, we consider quarterly data about the Spanish regions from $2000(II)$ to $2020(II)$, so $N = 17$ and $T = 81$. All variables are taken from the regional accounts recorded in the Spanish National Bureau of Statistics (INE).

Thus, the novelty of the approach that we propose here is two-fold. First, the use of semi-parametric techniques to incorporate flexibility in the wage regression model to analyze. Most of the research on wage flexibility resort to fully parametric techniques, but they can be subjected to some misspecification problems as it was noted previously. See Abraham (1996) and Guisan and Aguayo (2007), among others. Second, to improve the precision of these estimators we apply the methodology proposed in the previous sections to take into account the cross-sectional dependence for estimates. To the best of our knowledge, this approach is totally new but very necessary. The likelihood that regional wage flexibility can be influenced by some unobserved common features among these regions is very high, and ignoring this fact can lead to misleading inference.

Further, as in the previous application, we resort to the partially linear approach proposed for (5.22) to estimate (5.31) and use the Silverman rule-of-thumb to choose the bandwidth. The estimated results are collected in Table 2.

Analyzing the results of Table 2, it can be highlighted that the coefficient related to the national wage growth is not statistically different from 1 in the three estimated equations. Hence, there seems to be evidence that the regional wage growth in Spain is fairly homogeneous with respect to the average wage growth. In this context, the Spanish labor market is characterized by a low degree of salary flexibility, where salaries show a low capacity to respond to the employment situation of each individual. Regarding the regional unemployment growth rate, in the three estimations it is obtained a negative

Table 2. Results for the regional wage flexibility.

	OLS	SP	SP.FGLS
$\dot{wage}$	1.0899***	1.0872***	1.0874***
	(0.0108)	(0.0103)	(0.0103)
$\dot{unem}$	-0.0058***	-0.0062***	-0.0062
	(0.0009)	(0.0009)	(0.0009)
$\dot{prod}$	-0.1706***	.	.
	(0.0695)	.	.
$\dot{prod}^2$	-0.5395***	.	.
	(0.3850)	.	.
$\widehat{\theta}$	.	.	0.0060
$\widehat{\sigma}^2$	.	.	0.1100

Note: Standard errors are in parenthesis. *** indicates statistical significance at the 1% level.

and highly significant relationship with regional wages, as it was expected. Furthermore, the OLS results seem to corroborate a nonlinear relationship between the regional wages growth and the evolution of the national productivity. As it can be seen from the first column of Table 2, increases in the national productivity seems to have a positive and significatively effect on regional wages up to certain threshold from which the effect becomes negative.

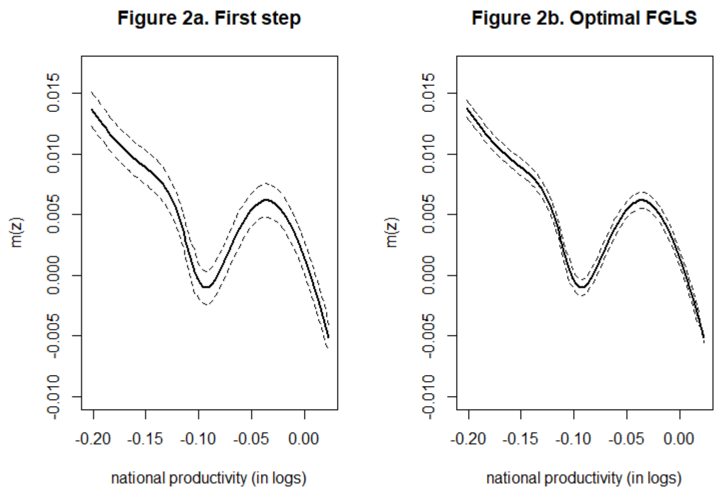
With the aim of identifying more precisely the nonlinear relationship between the national productivity and the regional wages, the estimated curves are plotted in Figure 2. Analyzing Figure 2a, it can be seen that there seems to be a negative relationship between national productivity and regional wages until certain point after which there is a U-shaped relationship. Note that this result is very far from the quadratic relationship that was assumed for the OLS. Similar results are obtained when the cross-sectional dependence is incorporated in Figure 2b, where more precise estimates and confidence band can be observed.

Therefore, these results corroborates that imposing a pre-specified nonlinear function for the national productivity in this type of studies can lead to misleading results about regional labour markets in the EMU, but ignoring the cross-sectional dependence is not a great issue. This is not a surprising result since we are analyzing different Spanish regions. Completely different results are expected if we extend this study to different European regions.

6. ACKNOWLEDGEMENTS

The authors gratefully acknowledge financial support from the Programa Estatal de Generación de Conocimiento y Fortalecimiento Científico y Tecnológico del sistema de I+D+i y del Programa Estatal de I+D+i Orientada a los Retos de la Sociedad/Spanish Ministry of Science and Innovation. Ref. PID2019-105986GB-C22. In addition, this work is part of the Research Project APIE 1/2015-17: "New methods for the empirical analysis

Figure 2. Nonparametric estimates for the national productivity.



Note: Thick line denotes the estimates curve while dotted line is the 95% pointwise confidence interval.

of financial markets" of the Santander Financial Institute (SANFI) of UCEIF Foundation resolved by the University of Cantabria and funded with sponsorship from Banco Santander.

REFERENCES

Abraham, F. (1996). Regional adjustment and wage flexibility in the European Union. *Regional Science and Urban Economics 26*, 51–75.

Andrews, D. (2005). Cross-section regression with common shocks. *Econometrica 73*, 1551–1585.

Arbia, G. (2006). *Spatial Econometrics: Statistical Foundations and Applications to Regional Analysis*. Springer-Verlag, Berlin.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica 77*, 1229–1279.

Cai, Z. and Q. Li (2008). Nonparametric estimation of varying coefficient dynamic panel data models. *Econometric Theory 24*, 1321–1342.

Cliff, A. and C. Ord (1972). Testing for spatial autocorrelation among regression residuals. *Geographical Analysis 4*, 267–284.

Fan, J. and I. Gijbels (1995). *Local polynomial modelling and its applications*. Chapman & Hall.

Fan, J. and Q. Li (1999). Root-n-consistent estimation of partially linear time series models. *Journal of Nonparametric Statistics 11*, 251–269.

Gasser, T., A. Kneip, and W. Kohler (1991). A flexible and fast method for automatic smoothing. *Journal of the American Statistical Association 86*, 647–652.

Guisan, M. and E. Aguayo (2007). Wages, productivity and human capital in the European Union: econometric models and comparision with the USA 1985-2005. *Applied Econometrics and International Development 7*, 43–56.

Henderson, D. and C. Parmeter (2015). *Applied nonparametric econometrics*. Cambridge University Press.

Hsiao, C. and A. Tahmiscioglu (2008). Estimation of dynamic panel data models with both individual and time-specific effects. *Journal of Statistical and Planning Inference 138*, 2698–2721.

Huang, X. (2013). Nonparametric estimation in large panels with cross-sectional dependence. *Econometric Reviews 32*, 754–777.

Kelejian, H. and I. Prucha (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review 40*, 509–533.

Lee, J. and P. M. Robinson (2015). Panel nonparametric regression with fixed effects. *Journal of Econometrics 188*, 346–362.

Lee, L. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica 72*, 1899–1925.

Li, Q. and T. Stengos (1996). Semiparametric estimation of partially linear panel data models. *Econometrica 71*, 389–397.

Mélitz, J. (1997). The evidence about the cost and benefits of EMU. *Swedish Economic Policy Review 4*, 359–410.

Muellbauer, J. and A. Murphy (1997). Booms and busts in the UK housing market. *Economic Journal 107*, 1701–1727.

Pesaran, M. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica 74*, 967–1012.

Phillips, P. and D. Sul (2003). Dynamic panel estimation and homogeneity testing under cross section dependence. *Econometrics Journal 6*, 217–259.

Qian, J. and L. Wang (2012). Estimating semiparametric panel data models by marginal integration. *Journal of Econometrics 167*, 483–493.

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

Robinson, P. M. (2012). Nonparametric trending regression with cross-sectional dependence. *Journal of Econometrics 169*, 4–14.

Rodriguez-Poo, J. M. and A. Soberon (2017). Nonparametric and semiparametric panel data models: recend developments. *Journal of Economic Surveys 31*, 923–960.

Rothenberg, T. (1971). Identification in parametric models. *Econometrica 39*, 577–591.

Ruppert, D. and M. P. Wand (1994). Multivariate locally weighted least squares regression. *The Annals of Statistics 22*(3), 1346–1370.

Su, L. and S. Jin (2012). Sieve estimation of panel data models with cross section dependence. *Journal of Econometrics 169*, 34–47.

Su, L. and A. Ullah (2011). Nonparametric and semiparametric panel econometric models: estimation and testing. In A. Ullah and D. E. Giles (Eds.), *Handbook of Empirical Economics and Finance*, pp. 455–497. New York: Taylor and Francis Group.

Tavlas, G. (1993). The new theory of optimum currency areas. *World Economic 16*, 663–685.

Tavlas, G. (1994). The theory of monetary integration. *Open Economic Review 5*, 221–230.

Yu, J., R. Jong, and L. Lee (2008). Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and t are large. *Journal of Econometrics 146*, 118–134.

Zhan-Qian, L. (1996). Multivariate locally weighted polynomial fitting and partial derivative estimation. *Journal of Multivariate Analysis 59*, 187–205.

7. SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Online Supplement