Efficient and rate optimal estimation and inference in nonparametric varying coefficient panel data models

Alexandra Soberón,*
Departament of Economics, University of Cantabria
Winfried Stute
Mathematical Institute, University of Giessen
and
Juan M. Rodriguez-Poo
Department of Economics, University of Cantabria and University of Genève

December 23, 2015

Abstract

This paper is concerned with the efficient and rate optimal estimation of a panel data varying coefficient model when standard distributional assumptions cannot be justified. A two step nonparametric estimator is proposed. In the first step, a pairwise differencing transformation is taken to overcome the presence of unobserved individual effects. In the second step, a backfitting algorithm is implemented to incorporate the information from the error covariance matrix. Further, in order to obtain consistent estimators of the asymptotic variance-covariance matrix an alternative root-n consistent estimation of the variance components is proposed. The main interest of this new technique is that no additional assumptions of functional forms are needed to obtain the estimates. Also note that despite to the established in this literature, in this paper it is shown that the resulting GLS-type estimator is asymptotically more efficient than the first above without imposing any closeness property within groups. As a by-product of the previous results we propose a very simple and intuitive test to detect the presence of unobserved individual heterogeneity. The proposed estimators and test statistic are augmented by simulation studies and they are illustrated in the analysis of the production efficiency of the companies in the European Union.

Keywords: asymptotic normality, moment estimator, pairwise difference, consistent test, fixed effects.

*The authors gratefully acknowledge financial support from the Programa Estatal de Fomento de la Investigación Científica y Técnica de Excelencia/Spanish Ministry of Economy and Competitiveness. Ref. ECO2013-48326-C2-2-P. In addition, this work is part of the Research Project APIE 1/2015-17: "New methods for the empirical analysis 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.

1 Introduction

This paper is concerned with the efficient and rate optimal nonparametric estimation of a fixed effects panel data varying coefficient model. Traditionally, in the context of pure nonparametric regression in panel data models, many authors (see, among others Henderson and Ullah (2005), Lin and Carroll (2000), Ruckstuhl et al. (2000), Su and Ullah (2007), and Zhou et al. (2010)) have been interested in exploiting the variance-covariance matrix of the idiosyncratic errors to obtain efficiency gains.

All these studies can be somehow interpreted as extensions from the standard GLS literature to the nonparametric regression case. Surprisingly, unlike what happens in the parametric case, Ruckstuhl et al. (2000) and Lin and Carroll (2000) find that ignoring the correlation structure within groups entirely and "pretending" that the data are really independent will result in more efficient estimators than incorporating the true correlation structure, i.e., the so-called "working independence" approach. On the contrary, Zhou et al. (2010) show that this result is not always true. They find that when the design points of the nonparametric component have a closeness property within groups, it is possible to obtain a nonparametric estimator with the same asymptotic bias as the local linear estimator but smaller asymptotic covariance matrix. Thus, the question of whether it is possible to construct an estimator with an asymptotic mean square error uniformly lower than the traditional one is still open. Furthermore, in all cases, the authors assume a random effect setting just to guarantee that direct nonparametric estimators of the quantities of interest are indeed consistent. To our knowledge, in the presence of fixed effects, the aim has been rather more modest and oriented in most part of cases to obtain consistent estimators in the quantities of interest after applying some profile likelihood techniques or some differencing transformations. Issues like efficiency or even rate optimality have not been a matter of concern within this area of research.

At the same time, in the nonparametric literature there has been an increasing interest in varying coefficient models. This growing interest is due to several reasons: First, varying coefficient models encompass a great variety of econometric models such as partially linear models; second, they mitigate the so-called "curse of dimensionality"; finally, they have been justified on the grounds of economic theory (see Chamberlain (1992)). In the context of panel data varying coefficient models some authors have already investigated the problem of defining estimators with optimal rates of convergence (see Su and Ullah (2011) and Chen et al. (2013), for some surveys). Unfortunately, to our knowledge none of these papers obtain estimators that are asymptotically efficient.

In this paper, we introduce a two-step estimator that turns out to be asymptotically efficient and it exhibits an optimal nonparametric rate of convergence. In the first step a pairwise differencing transformation is proposed to deal with the presence of fixed effects, and in the second step a backfitting algorithm is implemented. Furthermore, in order to obtain consistent estimators of the variance-covariance matrix, based on Wu et al. (2012), a root-n consistent estimator of the variance components is proposed. The main interest of this new technique is that, when analyzing the statistical properties of the resulting estimators, there is no need to add further distributional assumptions on both the idiosyncratic error term and the unobserved individual heterogeneity. Also, it is pointed out that unlike the findings in Ruckstuhl et al. (2000), Lin and Carroll (2000), and Zhou et al. (2010), in this paper it is shown that the resulting GLS-type estimator is asymptotically more efficient than the estimator proposed previously without using any closeness property within groups.

As a by-product of the previous estimation technique we propose a very simple and intuitive test to detect the presence of individual heterogeneity. As it is well-known in the literature, the presence of this unobserved term rends standard OLS panel data estimators at least inefficient and inconsistent. To the best of our knowledge, this result is completely new in the panel data framework when distributional assumptions of the error cannot be justified. Finally, to illustrate the usefulness of the results of this paper we first focus on the finite sample properties of the proposed estimators and test statistic. Later, we study the production efficiency of the European Union (EU)'s companies to establish whether the marginal productivity of labor and capital depend on their level of research and development (R&D) expenditure.

The rest of the paper is organized as follows: In Section 2 we set up the model of interest and the estimation procedure. Also we study the main asymptotic properties. In Section 3 we propose a more efficient estimation procedure and show the basic properties of the resulting estimator. In Section 4 the estimation procedure for the variance components is discussed and a test statistic is proposed. In Section 5 we apply our results to a production efficiency study and compare the estimators and test statistic considered via a Monte Carlo simulation. Section 6 provides a summary of the paper. The proofs of the main results are collected in the Appendix.

2 Econometric model and preliminary estimation

Assume that data are available from a varying coefficient panel data model of the form

$$Y_{it} = X_{it}^{\top} m(Z_{it}) + b_i + v_{it}, \qquad i = 1, \cdots, N; \quad t = 1, \cdots, T, \tag{2.1}$$

where Y_{it} denotes the response variable of the individual i in the period t, Z_{it} and X_{it} are vectors of covariates of dimension $q \times 1$ and $d \times 1$, respectively, and $m(\cdot)$ is a $d \times 1$ vector of unknown functions to estimate. The relationship between X_{it} and Y_{it} described by (2.1) contains an unknown individual effect b_i that is different for each individual i but, for the same individual, is constant along time. Further, these quantities are perturbed by the idiosyncratic error term v_{it} that is assumed to be independent and identically distributed ($i.i.d.$) with zero mean and finite variance, σ_v^2, and that is independent of X_{it} and Z_{it} for all i and t. It is also assumed throughout that b_i is $i.i.d.$ with zero mean and finite variance, σ_b^2. In addition, b_i and v_{it} are independent with each other for all i and t.

Note that we are not willing to impose any condition on the statistical relationship between b_i and the covariates of the model. To deal with the unobserved individual effects, a standard solution is to resort to the well-known differencing techniques, where the first differences and the within transformation are the most popular. Under the assumptions of strict exogeneity, Wooldridge (2003) points out that the first differences estimator is preferred in standard parametric panel data models if the error term exhibits a strong positive serial correlation. In all other situations, the within estimator performs better for model (2.1). However, although the within transformation solves the problem of the fixed effects we need a T-dimensional kernel weight to obtain asymptotically unbiased estimators but a price is paid in terms of slower rate of convergence. This problem is severe when the time dimension is large or when there is a large number of covariates, see Rodriguez-Poo and Soberon (2015) and references therein for further details.

In this situation, we propose to overcome this problem with a pairwise differencing transformation. Specifically, this alternative transformation is very appealing in this framework since it enables us to remove the individual effects and obtain nonparametric estimators that almost achieve the optimal rate of convergence of this type of problems. Inspired by Stromberg et al. (2000) and Honoré and Powell (2005), the pairwise-difference transformation implies subtracting from time t of (2.1) time s, for $s \neq t$, yielding

$$Y_{it} - Y_{is} = X_{it}^{\top} m(Z_{it}) - X_{is}^{\top} m(Z_{is}) - v_{it} - v_{is}, \quad i = 1, \cdots, N; \quad t, s = 1, \cdots, T, \quad s \neq t. \tag{2.2}$$

Note that instead of working with a sample size of NT observations, the dimension of the transformed regression model is $NT(T-1)$ because differences for all $s \neq t$ are considered.

Let $\widetilde{Y}_{its} = Y_{it} - Y_{is}$ and $\widetilde{v}_{its} = v_{it} - v_{is}$. The pairwise differencing transformation (2.2) can be rewritten as

$$\widetilde{Y}_{its} = X_{it}^{\top} m(Z_{it}) - X_{is}^{\top} m(Z_{is}) + \widetilde{v}_{its}, \quad i = 1, \cdots, N; \quad t, s = 1, \cdots, T, \quad s \neq t. \tag{2.3}$$

Direct estimation of $m(\cdot)$ in (2.3) through any standard nonparametric technique ends up in a non-negligible asymptotic bias and, in order to avoid this, we define a kernel weight that controls the distance between any Z_{it}, Z_{is}. See for example Rodriguez-Poo and Soberon (2015).

A fairly simple technique to estimate $m(\cdot)$ and its derivatives can be, in the case $d = q = 1$, to approximate $m(\cdot)$ through a Taylor expansion in any $z \in A$, where A is a compact subset in a non-empty interior of $I\!R$, obtaining

$$\begin{aligned} X_{it} m(Z_{it}) - X_{is} m(Z_{is}) &\approx (X_{it} - X_{is})\, m(z) + [X_{it}(Z_{it} - z) - X_{is}(Z_{is} - z)]\, m'(z) \\ &+ \frac{1}{2}\left[X_{it}(Z_{it} - z)^2 - X_{is}(Z_{is} - z)^2\right] m''(z) + \cdots \\ &+ \frac{1}{p!}\left[X_{it}(Z_{it} - z)^p - X_{is}(Z_{is} - z)^p\right] m^{(p)}(z) \\ &\equiv \sum_{\lambda=0}^{p} \beta_{\lambda}\left[X_{it}(Z_{it} - z)^{\lambda} - X_{is}(Z_{is} - z)^{\lambda}\right]. \end{aligned} \tag{2.4}$$

This suggests that we estimate $m(z), m'(z), \cdots, m^{(p)}(z)$ by regressing $\widetilde{Y}_{its}$ on the terms of the right-hand side of (2.4) with kernel weights.

Consider the simplest case, $p = 0$, then $\beta = m(z)$ can be estimated by minimizing the following criterion function,

$$\sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s \neq t}^{T} \left(\widetilde{Y}_{its} - \widetilde{X}_{its}^{\top} \beta\right)^2 K_{H_1}(Z_{it} - z) K_{H_1}(Z_{is} - z), \tag{2.5}$$

see for example Fan and Gijbels (1995) and Ruppert and Wand (1994), where $\widetilde{X}_{its} = X_{it} - X_{is}$, H_1 is a $q \times q$ symmetric positive-definite bandwidth matrix and, for each u, K a multivariate kernel such that

$$\int K(u) du = 1 \quad \text{and} \quad K_{H_1}(u) = H_1^{-1/2} K(H_1^{-1/2} u).$$

Let $\widehat{\beta}$ be the minimizer of (2.5). It is equal to

$$\widehat{m}(z; H_1) = S_{\widetilde{X}\widetilde{X}}^{-1}(z) S_{\widetilde{X}\widetilde{Y}}(z), \tag{2.6}$$

where

$$S_{\widetilde{X}\widetilde{X}}(z) = \frac{1}{NT(T-1)} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s \neq t}^{T} K_{H_1}(Z_{it} - z) K_{H_1}(Z_{is} - z) \widetilde{X}_{its} \widetilde{X}_{its}^{\top},$$

and

$$S_{\widetilde{X}\widetilde{Y}}(z) = \frac{1}{NT(T-1)} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s \neq t}^{T} K_{H_1}(Z_{it} - z) K_{H_1}(Z_{is} - z) \widetilde{X}_{its} \widetilde{Y}_{its}.$$

In order to show the asymptotic properties of this estimator, let us consider some assumptions. Under the data generating process defined in (2.1), we also assume the following:

Assumption 2.1 *Let $f_{Z_{1t}}(\cdot)$ be the probability density function of Z_{1t}. All density functions are continuously differentiable in all their arguments and they are bounded away from zero in any point of their support.*

Assumption 2.2 *Let $\|A\| = \sqrt{tr(A^{\top}A)}$, $E\left[\|\widetilde{X}_{its}\widetilde{X}_{its}^{\top}\|^2 | Z_{it} = z_1, Z_{is} = z_2\right]$ is bounded and uniformly continuous in its support. In addition, the matrix function $E\left[\widetilde{X}_{its}\widetilde{X}_{its}^{\top} | Z_{it} = z_1, Z_{is} = z_2\right]$ is bounded and uniformly continuous in its support, $\mathcal{Z}$.*

Assumption 2.3 *The matrix $E\left[\widetilde{X}_{its}\widetilde{X}_{its}^{\top} | Z_{it} = z_1, Z_{is} = z_2\right]$ is positive definite for any interior point of (z_1, z_2) in the support of $f_{Z_{it},Z_{is}}$.*

Assumption 2.4 *Let z be an interior point in the support of $f_{Z_{1t}}$. All second-order derivatives of $m_1(\cdot), m_2(\cdot), \cdots, m_d(\cdot)$ are bounded and uniformly continuous.*

Assumption 2.5 *The q-variate Kernel functions K are compactly supported and bounded such that $\int uu^{\top}K(u)du = \mu_2(K)I$ and $\int K^2(u)du = R(K)$, where $\mu_2(K) \neq 0$ and $R(K) \neq 0$ are scalars and I is the $q \times q$ identity matrix. In addition, all odd-order moments of K vanish, that is $\int u_1^{\iota_1} \cdots u_q^{\iota_q} K(u)du = 0$, for all nonnegative integers $\iota_1, \cdots, \iota_q$ such that their sum is odd.*

Assumption 2.6 *The bandwidth matrix H_1 is symmetric and strictly definite positive. Also, as $N \to \infty$ each entry of the matrix tends to zero in such a way that $N|H_1| \to \infty$.*

Assumption 2.7 *For some $\delta > 0$, the following functions $E\left[|\widetilde{X}_{its}v_{it}|^{2+\delta} | Z_{it} = z_1, Z_{is} = z_2\right]$, and $E\left[|\widetilde{X}_{its}v_{is}|^{2+\delta} | Z_{it} = z_1, Z_{is} = z_2\right]$ are bounded and uniformly continuous in any point of their support.*

As the reader can notice, Assumptions 2.1, 2.2 and 2.4 are basically smoothness and boundedness conditions on the density function and moments functionals. Assumption 2.3 is a generalization of the well-known rank condition of parametric models that guarantees that $m(\cdot)$ is identified. Also, Assumptions 2.5 and 2.6 are standard in the literature of the local linear regression for the kernel function and bandwidth matrix. Finally, Assumption (2.7) is required for the Central Limit Theorem. Specifically, it enables us to show that the Lyapunov condition holds.

Under these assumptions, we obtain the following result for $\widehat{m}(z;H_1)$.

Theorem 2.1 *Under Assumptions 2.1-2.7, as N tends to infinity and T is fixed*

$$\sqrt{N|H_1|}\left(\widehat{m}(z;H_1)-m(z)-B(z;H_1)\right) \quad \xrightarrow{d} \quad \mathcal{N}(0,V(z;H_1)),$$

where

$$\begin{aligned} B(z;H_1) &= \mu_2(K)\left[diag_d\left(tr(H_1D_f(z)D_{m_r}(z))\right)\imath_d f^{-1}_{Z_{it},Z_{is}}(z,z)+\frac{1}{2}diag_d\left(tr(H_1\mathcal{H}_{m_r}(z))\right)\imath_d\right] \\ V(z;H_1) &= 2\sigma_v^2R^2(K)\mathcal{B}^{-1}_{\widetilde{X}\widetilde{X}}(z,z), \end{aligned}$$

for $r=1,\cdots,d$, D_{m_r} is the first-order derivative vector of the rth component of $m(\cdot)$, $\mathcal{H}_{m_r}(z)$ the Hessian matrix, $D_f(z)$ the first-order derivative vector of the density function, and

$$\mathcal{B}_{\widetilde{X}\widetilde{X}}(z,z)=E\left[\widetilde{X}_{its}\widetilde{X}^{\top}_{its}|Z_{it}=z,Z_{is}=z\right]f_{Z_{it},Z_{is}}(z,z).$$

In addition, $diag_d(tr(H_1\mathcal{H}_{m_r}(z)))$ and $diag_d(tr(H_1D_f(z)D_{m_r}(z)))$ stand for a diagonal matrix of elements of $tr(H_1\mathcal{H}_{m_r}(z))$ and $tr(H_1D_f(z)D_{m_r}(z))$, respectively, being $\imath_d$ a $d\times 1$ unit vector.

The proof of this theorem is postponed to the Appendix.

The results shown in Theorem 2.1 are rather standard. However, there are some differences that need to be pointed out. More precisely, as far as we have more curvature in $m(\cdot)$ the bias is enlarged. On its part, the variance will be penalized when H_1 is higher and when there is sparser data near z. Nevertheless, note that the resulting estimators present at least two problems that need to be solved: First, as the reader can observe, the rate of convergence of this estimator is slower than $N|H_1|^{1/2}$, the lower rate of convergence of this type of problems. That is, the estimator is suboptimal in this sense. Second, this estimator ignores the correlation within groups that results as an outcome from the pairwise differencing transformation. This causes problems of efficiency. In the next section we give a solution to both problems.

3 Efficient and rate optimal nonparametric estimation

Let $\mathbb{X} = (X_{11}, \cdots, X_{NT}, Z_{11}, \cdots, Z_{NT})$ be the observed covariate vector. For the sake of comparison, let us denote for i and t fixed $\widetilde{v}_{it}$ as a $1 \times (T-1)$ vector that stacks the subscript s of $\widetilde{v}_{its}$ in vectorial form. Then, the variance-covariance matrix of (2.3) is of the form

$$E\left(\widetilde{v}_{it}\widetilde{v}_{it}^{\top}|\mathbb{X}\right) = \sigma_v^2\left(I_{(T-1)} + \imath_{(T-1)}\imath_{(T-1)}^{\top}\right),$$

where $I_{(T-1)}$ is a $(T-1) \times (T-1)$ identity matrix and $\imath_{(T-1)}$ is a $(T-1) \times 1$ unit vector.

Then, for all t we can build a $T(T-1) \times T(T-1)$ variance-covariance matrix V_i such that

$$V_i = E\left(\widetilde{v}_i\widetilde{v}_i^{\top}|\mathbb{X}\right) = \sigma_v^2 I_T \otimes \left(I_{(T-1)} + \imath_{(T-1)}\imath_{(T-1)}^{\top}\right), \tag{3.1}$$

where I_T is a $T \times T$ diagonal matrix, $\widetilde{v}_i = (\widetilde{v}_{i12}, \cdots, \widetilde{v}_{i1T}, \cdots, \widetilde{v}_{iT1}, \cdots, \widetilde{v}_{iT(T-1)})^{\top}$ is a $T(T-1) \times 1$ vector and $\otimes$ denotes Kronecker product.

Let $\tau = \frac{1-1/\sqrt{T}}{(T-1)}$, the square root of the inverse matrix of V_i is of the form

$$V_i^{-1/2} = \frac{1}{\sigma_v} I_T \otimes \left(I_{(T-1)} - \tau\imath_{(T-1)}\imath_{(T-1)}^{\top}\right). \tag{3.2}$$

In order to obtain spherical disturbances, we premultiply the stacked version of (2.3) for all t and s by $V_i^{-1/2}$ and hence we obtain the following expression

$$\widetilde{Y}_{its} - \tau\sum_{r\neq t}^{T}\widetilde{Y}_{itr} = (1-\tau(T-1))X_{it}^{\top}m(Z_{it}) - \left(X_{is}^{\top}m(Z_{is}) - \tau\sum_{r\neq t}^{T}X_{ir}^{\top}m(Z_{ir})\right) + \widetilde{v}_{its} - \tau\sum_{r\neq t}^{T}\widetilde{v}_{itr}. \tag{3.3}$$

Following Rodriguez-Poo and Soberon (2015) and using (3.3) we could define the estimator for $m(z) = \alpha$ as the solution to the following minimization problem

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s\neq t}^{T}\left(\widetilde{Y}_{its} - \tau\sum_{r\neq t}^{T}\widetilde{Y}_{itr} - \left(\widetilde{X}_{its} - \tau\sum_{t\neq t}^{T}\widetilde{X}_{itr}\right)^{\top}\alpha\right)^2 K_{H_2}(Z_{i1}-z, \cdots, Z_{iT}-z), \tag{3.4}$$

where $\widetilde{X}_{itr} = X_{it} - X_{ir}$. Let us denote by $\widehat{\alpha}$ the solution to this problem.

However, note that the resulting estimator is going to show an achievable rate of convergence of order $N|H_2|^{T/2}$. Just to obtain a rate of $N|H_2|^{1/2}$ that is optimal we propose an one-step backfitting algorithm that will consist in redefining the transformed dependent variable in (3.3) as

$$\widetilde{Y}_{its}^{*} - \tau\sum_{r\neq t}^{T}\widetilde{Y}_{itr}^{*} = \left(\widetilde{Y}_{its} - \tau\sum_{r\neq t}^{T}\widetilde{Y}_{itr}\right) + \left(X_{is}^{\top}m(Z_{is}) - \tau\sum_{r\neq t}^{T}X_{ir}^{\top}m(Z_{ir})\right). \tag{3.5}$$

Now, combining (3.3) and (3.5) we have the following regression model

$$\widetilde{Y}^*_{its} - \tau \sum_{r \neq t}^{T} \widetilde{Y}^*_{itr} = (1 - \tau(T-1)) X_{it}^\top m(Z_{it}) + \widetilde{v}_{its} - \tau \sum_{r \neq t}^{T} \widetilde{v}_{itr}, \quad i = 1, \cdots, N; \quad t, s = 1, \cdots, T, \quad s \neq t \tag{3.6}$$

where, as clearly seen, $m(\cdot)$ can be estimated at an optimal rate by any standard nonparametric regression technique.

Finally, note that the transformation defined in (3.5) and (3.6) is unfeasible because we need to know the values of $m(\cdot)$. To overcome this situation we propose to replace $m(\cdot)$ in (3.3) by the Nadaraya-Watson estimator proposed in Section 2. By doing so, the estimator for $m(\cdot)$ in (3.5) that we propose is the so-called GLS Nadaraya-Watson estimator that will have the form

$$\widehat{m}(z; H_2) = S_n^{-1}(z) \frac{\kappa}{NT(T-1)} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s \neq t}^{T} K_{H_2}(Z_{it} - z) X_{it} \left(\widehat{Y}_{its} - \tau \sum_{r \neq t}^{T} \widehat{Y}_{itr} \right) \tag{3.7}$$

where $\kappa = \frac{1}{1-\tau(T-1)}$,

$$\widehat{Y}_{its} - \tau \sum_{r \neq t}^{T} \widehat{Y}_{itr} = \widetilde{Y}_{its} - \tau \sum_{r \neq t}^{T} \widetilde{Y}_{itr} + \left(X_{is}^\top \widehat{m}(Z_{is}; H_1) - \tau \sum_{r \neq t}^{T} X_{ir}^\top \widehat{m}(Z_{ir}; H_1) \right),$$

and

$$S_n(z) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} K_{H_2}(Z_{it} - z) X_{it} X_{it}^\top.$$

We now study the asymptotic properties of this GLS Nadaraya-Watson estimator $\widehat{m}(z, H_2)$. In order to do it, we assume the following.

Assumption 3.1 *Let $\|A\| = \sqrt{tr(A^\top A)}$, $E\left[\|X_{it}X_{it}^\top\|^2 | Z_{it} = z_1\right]$ be bounded and uniformly continuous in its support. Furthermore, the matrix function $E\left[X_{it}X_{it}^\top | Z_{it} = z_1\right]$ shall be bounded and uniformly continuous in its support.*

Assumption 3.2 *The matrix $E\left[X_{it}X_{it}^\top | Z_{it} = z_1\right]$ is positive definite for any interior point of z_1 in the support of $f_{Z_{it}}(z_1)$.*

Assumption 3.3 *The following functions $E\left[|X_{it}v_{it}|^{2+\delta} | Z_{it} = z_1\right]$ and $E\left[|X_{it}v_{is}|^{2+\delta} | Z_{it} = z_1\right]$ are bounded and uniformly continuous in any point of their support, for some $\delta > 0$.*

Note that since we use the nonparametric estimator of Section 2 to obtain a feasible estimator of $m(\cdot)$, we need to ensure that the bias rate of the estimator $\widehat{m}(z;H_1)$ is uniform. In order to do it and following Masry (1996) we need to impose an assumption about the bandwidth H_1 and its relationship with H_2.

Assumption 3.4 *The bandwidth matrix H_2 is symmetric and strictly definite positive. Furthermore, each entry of the matrix tends to zero as N tends to infinity in such a way that $N|H_2| \to \infty$.*

Assumption 3.5 *The bandwidth matrices H_1 and H_2 must fulfill the conditions that $tr(H_1)/tr(H_2) \to 0$ and $N|H_1||H_2|/log(N) \to \infty$ as N tends to infinity.*

Under these assumptions it is then possible to show the following result for the GLS Nadaraya-Watson estimator $\widehat{m}(z;H_2)$.

Theorem 3.1 *Under Assumptions 2.1, 2.4-2.6, and 3.1-3.5, as N tends to infinity and T is fixed*

$$\sqrt{N|H_2|^{1/2}}\left(\widehat{m}(z;H_2) - m(z) - B(z;H_2)\right) \quad \xrightarrow{d} \quad \mathcal{N}(0, V(z;H_2)),$$

where

$$\begin{aligned} B(z;H_2) &= \mu_2(K)\left[diag_d\left(tr(H_2 D_f(z) D_{m_r}(z))\right)\imath_d f_{Z_{it}}(z)^{-1} + \frac{1}{2} diag_d\left(tr(H_2 \mathcal{H}_{m_r}(z))\right)\imath_d\right] \\ V(z;H_2) &= \sigma_v^2 \kappa^2 R(K) \mathcal{B}_{XX}^{-1}(z), \end{aligned}$$

where

$$\mathcal{B}_{XX}(z) = E\left[X_{it}X_{it}^\top | Z_{it} = z\right] f_{Z_{it}}(z).$$

In addition, $diag_d(tr(H_2\mathcal{H}_{m_r}(z)))$ and $diag_d(tr(H_2 D_f(z) D_{m_r}(z)))$ stand for a diagonal matrix of elements of $tr(H_2\mathcal{H}_{m_r}(z))$ and $tr(H_2 D_f(z) D_{m_r}(z))$, respectively, being $\imath_d$ a $d \times 1$ unitary vector.

The proof of this result is postponed to the Appendix.

By looking at these results we can highlight that the GLS-type nonparametric estimator that we propose, $\widehat{m}(z;H_2)$, is consistent and asymptotically normal with an optimal rate of convergence. Also, if we compare the findings of this theorem with the corresponding of Theorem 2.1 it is proved that $\widehat{m}(z;H_2)$ is asymptotically more efficient than $\widehat{m}(z;H_1)$ in the sense of having a smaller asymptotic covariance matrix. Thus, this result is different from the one obtained in Ruckstuhl et al. (2000), Lin and Carroll (2000), and Zhou et al. (2010) since we show that it is possible to obtain an asymptotic improvement of the nonparametric estimator by considering the error correlation structure.

4 Consistent estimation of variance components

According to the results provided in Theorem 3.1, if we want to make inference on $m(z)$, it is necessary to obtain a consistent estimator for σ_v^2. Furthermore, since we are not willing to assume any structure in the distribution of the idiosyncratic error term the task of defining a consistent estimator for σ_v^2 is rather cumbersome. To the best of our knowledge, Cox and Hall (2002) and Wu et al. (2012) are the only references in the literature that propose estimators of higher than second order moments of the random error when standard distributional assumptions cannot be justified. In order to obtain consistent estimators of variance components, we propose to extend the fundamental Lemma of Wu et al. (2012) from the linear mixed case to the nonparametric panel data framework, something that to our knowledge is completely new.

Let $\epsilon_{it} = b_i + v_{it}$ be defined as the sum of two independent random variables with zero mean, Wu et al. (2012) state that it is possible to define the following set of nonlinear functions

$$f_j^k(i) = \sum_{t=1}^{T} \epsilon_{it}^j \left[\sum_{t=1}^{T} \epsilon_{it} \right]^{k-j}, \qquad 1 \le j \le k,$$

that can be approximated through the following Lemma.

Lemma 4.1 *Let $a \wedge b$ and $a \vee b$ be the minimum and maximum, respectively, of two real numbers a and b, we have*

$$f_j^k(i) = \sum_{\ell=0}^{k} \sum_{r=(\ell-k+j)\vee 0}^{\ell \wedge j} \binom{j}{r} \binom{k-j}{\ell-r} \left(\sum_{t=1}^{T} v_{it}^r \right) \left(\sum_{t=1}^{T} v_{it} \right)^{\ell-r} b_i^{k-\ell} T^{k-j-\ell+r}.$$

Note that this lemma is based on simple calculus but it is very appealing for our purpose since many of the terms in the expansion of $f_m^k(i)$ will vanish when we take expectations. Thus, it provides a set of estimating equations, each of them leading to consistent estimators. Then, in order to find the efficient estimator of σ_v^2 we have to propose a suitable combination of these polynomial functions of the residuals.

If we focus our attention on the second moments of the idiosyncratic error term, σ_v^2, and we use Lemma 4.1 we obtain

$$\begin{aligned} f_2^2(i) &= \sum_{t=1}^{T} v_{it}^2 + T b_i^2 + 2 b_i \sum_{t=1}^{T} v_{it}, \\ f_1^2(i) &= T^2 b_i^2 + 2T b_i \sum_{t=1}^{T} v_{it} + \left(\sum_{t=1}^{T} v_{it} \right)^2, \end{aligned}$$

from which we can represent σ_v^2 in terms of the f's, i.e.,

$$E\left[Tf_2^2(i) - f_1^2(i)\right] = T(T-1)\sigma_v^2.$$

By averaging over $1 \leq i \leq N$ and replacing the unknown ϵ_{it} by the residuals $\widehat{\epsilon}_{it}$ the estimator of σ_v^2 has the form

$$\widehat{\sigma}_v^2 = \frac{1}{NT(T-1)} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[T\widehat{\epsilon}_{it}^2 - \left(\sum_{t=1}^{T} \widehat{\epsilon}_{it} \right)^2 \right], \tag{4.1}$$

where $\widehat{\epsilon}_{it} = Y_{it} - X_{it}^\top \widehat{m}(Z_{it}; H_1)$.

Similarly, combining this expressions in a proper way we also obtain

$$E[f_1^2(i) - f_2^2(i)] = T(T-1)\sigma_b^2$$

so that we can propose an estimator of σ_b^2 such as

$$\widehat{\sigma}_b^2 = \frac{1}{NT(T-1)} \sum_{i=1}^{N} \left[\left(\sum_{t=1}^{T} \widehat{\epsilon}_{it} \right)^2 - \sum_{t=1}^{T} \widehat{\epsilon}_{it}^2 \right]. \tag{4.2}$$

Once the estimators of the variance components have been proposed, we get the following result.

Theorem 4.1 *Under the data generating process (2.1) and assuming conditions 2.6 and 3.1-3.3 hold, when both $\gamma_v^4 = E(v_{it}^4)$ and $\gamma_b^4 = E(b_i^4)$ are finite, as $N \to \infty$ and T is fixed we have*

$$\sqrt{N}(\widehat{\sigma}_v^2 - \sigma_v^2) \quad \xrightarrow{d} \quad \mathcal{N}(0, \mu_v^2)$$

and

$$\sqrt{N}(\widehat{\sigma}_b^2 - \sigma_b^2) \quad \xrightarrow{d} \quad \mathcal{N}(0, \mu_b^2),$$

where $\mu_v^2 = \gamma_v^4 - \left(\sigma_v^2\right)^2$ *and* $\mu_b^2 = \gamma_b^4 - \left(\sigma_b^2\right)^2 + \left(\frac{4T-1}{T(T-1)}\right)\sigma_b^2\sigma_v^2 + \frac{2}{T(T-1)}(\sigma_v^2)^2$.

The proof of this result is shown in the Appendix.

Compared to the results in Wu et al. (2012) note that their asymptotic analysis is performed for both T and N tending to infinity, whereas in our case T is kept fixed. This is of special interest when looking at the asymptotic variance of $\widehat{\sigma}_b^2$. We can also point out that the estimators that we propose here for σ_v^2 and σ_b^2 are asymptotically normal and have the same limit variance as if the unknown errors were known. Indeed, in the Appendix it is proved

$$\widehat{\sigma}_v^2 - \sigma_v^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(v_{it}^2 - \sigma_v^2) + O_{\mathbb{P}}\left(\frac{\sqrt{lnN}}{\sqrt{N|H_1|}}\right) + O_{\mathbb{P}}(tr(H_1)), \tag{4.3}$$

$$\widehat{\sigma}_b^2 - \sigma_b^2 = \frac{1}{NT(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s\neq t}^{T}(\epsilon_{it}\epsilon_{is} - \sigma_b^2) + O_{\mathbb{P}}\left(\frac{\sqrt{lnN}}{\sqrt{N|H_1|}}\right) + O_{\mathbb{P}}(tr(H_1)), \tag{4.4}$$

and we assume that as N goes to infinity and T is fixed, $H_1 \to 0$ in such a way that $N|H_1| \to \infty$.

In order to make inference about both σ_v^2 and σ_b^2 it is necessary to estimate μ_v^2 and μ_b^2, respectively. As it can be also seen from Theorem 4.1 these variances depend on higher order moments (i.e., fourth order) of both random variables. Therefore, a technique is needed to provide estimators for these higher order moments.

4.1 Estimation of higher order moments

In this section we focus on the estimation of the third and fourth moments of the idiosyncratic error term and the unobserved individual heterogeneity. Focusing first on the third order moments we get from Lemma 4.1

$$\begin{aligned}
f_3^3(i) &= \sum_{t=1}^{T} v_{it}^3 + 3b_i\sum_{t=1}^{T} v_{it}^2 + 3b_i^2\sum_{t=1}^{T} v_{it} + Tb_i^3,\\
f_2^3(i) &= \left(\sum_{t=1}^{T} v_{it}^2\right)\sum_{t=1}^{T} v_{it} + 2b_i\left(\sum_{t=1}^{T} v_{it}\right)^2 + Tb_i\sum_{t=1}^{T} v_{it}^2 + 3Tb_i^2\sum_{t=1}^{T} v_{it} + T^2b_i^3,\\
f_1^3(i) &= \left(\sum_{t=1}^{T} v_{it}\right)^3 + 3Tb_i\left(\sum_{t=1}^{T} v_{it}\right)^2 + 3T^2b_i^2\sum_{t=1}^{T} v_{it} + T^3b_i^3.
\end{aligned}$$

Then, combining these results in a proper way we obtain

$$E[2f_1^3(i) + T^2f_3^3(i) - 3Tf_2^3(i)] = T(T-1)(T-2)\gamma_v^3$$

from which the estimator of γ_v^3 has the form

$$\widehat{\gamma}_v^3 = \frac{1}{NT(T-1)(T-2)}\sum_{i=1}^{N}\left[2\left(\sum_{t=1}^{T}\widehat{\epsilon}_{it}\right)^3 + T^2\sum_{t=1}^{T}\widehat{\epsilon}_{it}^3 - 3T\left(\sum_{t=1}^{T}\widehat{\epsilon}_{it}^2\right)\sum_{t=1}^{T}\widehat{\epsilon}_{it}\right]. \tag{4.5}$$

Similarly, for γ_b^3 we oobtain the following estimating equation

$$E[f_1^3(i) - 3f_2^3(i) + 2f_3^3(i)] = T(T-1)(T-2)\gamma_b^3$$

so the resulting estimator of γ_b^3 is of the form

$$\widehat{\gamma}_b^3 = \frac{1}{NT(T-1)(T-2)} \sum_{i=1}^{N} \left[\left(\sum_{t=1}^{T} \widehat{\epsilon}_{it} \right)^3 - 3 \left(\sum_{t=1}^{T} \widehat{\epsilon}_{it}^2 \right) \sum_{t=1}^{T} \widehat{\epsilon}_{it} + 2 \sum_{t=1}^{T} \widehat{\epsilon}_{it}^3 \right]. \tag{4.6}$$

Finally, analyzing the fourth order moments from Lemma 4.1 we obtain

$$\begin{aligned}
f_4^4(i) &= \sum_{t=1}^{T} v_{it}^4 + 4b_i \sum_{t=1}^{T} v_{it}^3 + 6b_i^2 \sum_{t=1}^{T} v_{it}^2 + 4b_i^3 \sum_{t=1}^{T} v_{it} + Tb_i^4, \\
f_3^4(i) &= \sum_{t=1}^{T} v_{it}^3 \sum_{t=1}^{T} v_{it} + 3Tb_i \sum_{t=1}^{T} v_{it}^2 \sum_{t=1}^{T} v_{it} + T^2 b_i \sum_{t=1}^{T} v_{it}^3 + 3b_i^2 \left(\sum_{t=1}^{T} v_{it} \right)^2 + 3Tb_i^2 \sum_{t=1}^{T} v_{it}^2 \\
&+ 4Tb_i^3 \sum_{t=1}^{T} v_{it} + T^2 b_i^4, \\
f_2^4(i) &= \sum_{t=1}^{T} v_{it}^2 \left(\sum_{t=1}^{T} v_{it} \right)^2 + 2b_i \left(\sum_{t=1}^{T} v_{it} \right)^3 + 2Tb_i \sum_{t=1}^{T} v_{it}^2 \sum_{t=1}^{T} v_{it} + 5Tb_i^2 \left(\sum_{t=1}^{T} v_{it} \right)^2 \\
&+ T^2 b_i^2 \sum_{t=1}^{T} v_{it}^2 + 4T^2 b_i^3 \sum_{t=1}^{T} v_{it} + T^3 b_i^4, \\
f_1^4(i) &= \sum_{t=1}^{T} v_{it} \left(\sum_{t=1}^{T} v_{it} \right)^3 + 4Tb_i \left(\sum_{t=1}^{T} v_{it} \right)^3 + 6T^2 b_i^2 \left(\sum_{t=1}^{T} v_{it} \right)^2 + 4T^3 b_i^3 \sum_{t=1}^{T} v_{it} + T^4 b_i^4.
\end{aligned}$$

As before, the combination of these expressions provides consistent estimators for γ_v^4 and γ_b^4. However, it is true that in this case there is a great variety of possible combinations that can lead us to inefficient estimators. Following Wu et al. (2012) the proper combination to obtain an efficient estimator of γ_v^4 is

$$E\left[(T^2 - 2T + 3)(Tf_4^4(i) - 4f_3^4(i)) + 6Tf_2^4(i) - 3f_1^4(i) - 3(2T-3)f_5^4(i) \right] = T(T-1)(T-2)(T-3)\gamma_v^3,$$

where $f_5^4(i) = \left(\sum_{t=1}^{T} \epsilon_{it}^2 \right)^2$. Therefore, the resulting estimator of γ_v^4 is

$$\begin{aligned}
\widehat{\gamma}_v^4 &= \frac{1}{NT(T-1)(T-2)(T-3)} \sum_{i=1}^{N} \left[(T^2 - 2T + 3) \left(T \sum_{t=1}^{T} \widehat{\epsilon}_{it}^4 - 4 \sum_{t=1}^{T} \widehat{\epsilon}_{it}^3 \sum_{t=1}^{T} \widehat{\epsilon}_{it} \right) \right. \\
&+ \left. 6T \sum_{t=1}^{T} \widehat{\epsilon}_{it}^2 \left(\sum_{t=1}^{T} \widehat{\epsilon}_{it} \right)^2 - 3 \left(\sum_{t=1}^{T} \widehat{\epsilon}_{it} \right)^4 - 3(2T-3) \left(\sum_{t=1}^{T} \widehat{\epsilon}_{it}^2 \right)^2 \right].
\end{aligned} \tag{4.7}$$

For the estimator of γ_b^4 we propose the following estimating equation

$$E\left[f_1^4(i) - 6f_2^4(i) + 8f_3^4(i) + 6f_4^4(i) + 3f_5^4(i) \right] = T(T-1)(T-2)(T-3)\gamma_b^4$$

so the estimator of γ_b^4 is of the form

$$\begin{aligned}\widehat{\gamma}_b^4 &= \frac{1}{NT(T-1)(T-2)(T-3)}\sum_{i=1}^{N}\left[\left(\sum_{t=1}^{T}\widehat{\epsilon}_{it}\right)^4 - 6\sum_{t=1}^{T}\widehat{\epsilon}_{it}^2\left(\sum_{t=1}^{T}\widehat{\epsilon}_{it}\right)^2 + 8\sum_{t=1}^{T}\widehat{\epsilon}_{it}^3\sum_{t=1}^{T}\widehat{\epsilon}_{it}\right. \\ &- \left. 6\sum_{t=1}^{T}\widehat{\epsilon}_{it}^4 + 3\left(\sum_{t=1}^{T}\widehat{\epsilon}_{it}^2\right)^2\right]. \qquad (4.8)\end{aligned}$$

Theorem 4.2 *Under the data generating process (2.1) and assuming conditions 2.6 and 3.1-3.3 with $\gamma_v^6 = E(v_{it}^6)$, $\gamma_b^6 = E(b_i^6)$, $\gamma_v^8 = E(v_{it}^8)$ and $\gamma_b^8 = E(b_i^8)$ finite, as $N \to \infty$ and T is fixed, we have*

$$\sqrt{N}(\widehat{\gamma}_v^3 - \gamma_v^3) \xrightarrow{d} \mathcal{N}(0, \mu_v^3) \quad , \quad \sqrt{N}(\widehat{\gamma}_b^3 - \gamma_b^3) \xrightarrow{d} \mathcal{N}(0, \mu_b^3),$$

and

$$\sqrt{N}(\widehat{\gamma}_v^4 - \gamma_v^4) \xrightarrow{d} \mathcal{N}(0, \mu_v^4) \quad , \quad \sqrt{N}(\widehat{\gamma}_b^4 - \gamma_b^4) \xrightarrow{d} \mathcal{N}(0, \mu_b^4),$$

where

$$\begin{aligned}\mu_v^3 &= \gamma_v^6 - \left(\gamma_v^3\right)^2 - 6\sigma_v^2\sigma_b^2 + 9\left(\sigma_v^2\right)^3, \\ \mu_v^4 &= \gamma_v^8 - \left(\gamma_v^4\right)^2 - 8\gamma_v^3\gamma_v^5 + 16\sigma_v^2\left(\gamma_v^3\right)^2, \\ \mu_b^3 &= \gamma_b^6 - \left(\gamma_b^3\right)^2 + \frac{9}{T}\gamma_b^4\gamma_v^2 + \frac{18}{T(T-1)}\sigma_b^2\left(\sigma_v^2\right)^2 + \frac{4}{T(T-1)(T-2)}\left(\sigma_v^2\right)^3, \\ \mu_b^4 &= \gamma_b^8 - \left(\gamma_b^4\right)^2 + \frac{16}{T}\gamma_b^6\sigma_v^2 + \frac{72}{T(T-1)}\gamma_b^4\left(\sigma_v^2\right)^2 + \frac{64}{T(T-1)(T-2)}\sigma_b^2\left(\sigma_v^2\right)^3 + \frac{7}{T(T-1)(T-2)(T-3)}\left(\sigma_v^2\right)^4.\end{aligned}$$

The proof of this result follows exactly the same lines as in the proof of Theorem 4.2 and Wu et al. (2012) and it is therefore omitted.

4.2 Testing the one-way error structure

Note also that as an interesting by-product, Theorem 4.1 enables us to propose a test for the absence of individual effects. In this sense, Breusch and Pagan (1980) propose a Lagrange multiplier (LM) test in the context of panel data linear model. However, this test is not usually reasonable since the variance of the individual effects should be either null or positive. Honda (1985) noted that problem and proposed a test statistic based on the signed square root of the LM test statistic that is valid under non-linearities. Nevertheless, Moulton and Randolph (1989) show that the asymptotic $N(0,1)$ approximation of this alternative LM statistic can be poor even in large samples.

In light of these results, to test the significance of the individual effects in (2.1) without assuming any standard distributional assumption we propose an alternative test statistic. Under the null hypothesis this test is constructed directly using the estimators of second order moments of the variance components obtained previously.

Specifically, the hypothesis of interest can be written as

$$H_0 \ : \ \sigma_b^2 = 0,$$
$$H_1 \ : \ \sigma_b^2 > 0.$$

Based on the results from Theorem 4.1, the one-sided test statistic developed here is based on the sample analogue of $\mathcal{J} = (\widehat{\sigma}_b^2 - \sigma_b^2)/\sqrt{\mu_b^2}$. Specifically, under the null hypothesis that $\sigma_b^2 = 0$, the feasible test statistic that we propose is given by

$$\widehat{\mathcal{J}} = \sqrt{\frac{NT(T-1)}{2}} \left(\frac{\widehat{\sigma}_b^2}{\widehat{\sigma}_v^2} \right). \tag{4.9}$$

Using the results of Theorem 4.1 it is straightforward to obtain the following asymptotic properties of this statistic.

Theorem 4.3 *Under the data generating process of (2.1) and assuming conditions 2.6 and 3.1-3.3 with both γ_v^4 and γ_b^4 finite, as $N \to \infty$ and T is fixed, we have*

$$\sqrt{\frac{NT(T-1)}{2}} \left(\frac{\widehat{\sigma}_b^2}{\widehat{\sigma}_v^2} \right) \xrightarrow{d} \mathcal{N}(0,1)$$

where $\widehat{\sigma}_v^2$ and $\widehat{\sigma}_b^2$ are the consistent estimators of σ_v^2 and σ_b^2, respectively.

Thus, Theorem 4.3 states that $\widehat{\mathcal{J}}$ is a consistent one-sided test statistic for testing H_0 against H_1. If $\widehat{\mathcal{J}}$ is greater than the critical values from the standard normal distribution, we reject H_0 at the corresponding significance levels.

5 Monte Carlo simulations and applications

To illustrate the feasibility and possible gains of the proposed method in this paper, we first carry out some simulation studies about the two estimators proposed in this paper and then make a comparison with the fixed effects estimator developed in Rodriguez-Poo and Soberon (2015). Later, we apply the method to analyze a real data example about the production efficiency of the EU companies.

5.1 Monte Carlo experiment

We consider the following data generating process,

$$Y_{it} = X_{it}^{\top} m(Z_{it}) + b_i + v_{it}, \quad i = 1, \cdots, N; \quad t = 1, \cdots, T, \tag{5.1}$$

where the chosen functional form is $m(Z_{it}) = \sin(Z_{it}\pi)$, while X_{it} and Z_{it} are random variables satisfying $X_{it} = 0.5X_{i(t-1)} + \xi_{it}$ and $Z_{it} = \omega_{it} + \omega_{i(t-1)}$, where ω_{it} and ξ_{it} are generated as independent and identically distributed ($i.i.d.$) uniform random variable in $[0, \pi/2]$ and Gaussian random variables $NID(0, 1)$, respectively. Also, the individual heterogeneity is computed as $b_i = \mu_i + c_0 \overline{Z}_{i\cdot}$, where c_0 is the correlation between b_i and the regressors of the model and $\overline{Z}_{i\cdot} = T^{-1} \sum_{t=1}^{T} Z_{it}$.

The finite sample performance of the proposed estimators. To estimate the model parameters and the variance components we consider the following four cases:

$$\begin{aligned}
&(1) \quad v_{it} \sim i.i.d.\ 0.5N(0,1) \quad \text{and} \quad \mu_i \sim i.i.d.\ 0.5N(0,1);\\
&(2) \quad v_{it} \sim i.i.d.\ 0.5N(0,1) \quad \text{and} \quad \mu_i \sim i.i.d.\ 0.5t(8);\\
&(3) \quad v_{it} \sim i.i.d.\ 0.5t(8) \quad \text{and} \quad \mu_i \sim i.i.d.\ 0.5N(0,1);\\
&(4) \quad v_{it} \sim i.i.d.\ 0.5t(8) \quad \text{and} \quad \mu_i \sim i.i.d.\ 0.5t(8).
\end{aligned}$$

The simulation results are based on 1000 samples of data $\{(X_{it}, Z_{it}, Y_{it}) : i = 1, \cdots, N, t = 1, \cdots, T\}$. The number of time observations T is set up at 4, while the number of cross-sections N is either 50, 100 and 150. The Gaussian kernel has been used and the bandwidth is chosen following Silverman's rule-of-thumb, i.e. $\widehat{H}_2 = \widehat{h}_2 I = \widehat{\sigma}_z (NT)^{-1/5}$, where $\widehat{\sigma}_z$ is the sample standard deviation of Z_{it}, and $\widehat{H}_1 = \widehat{h}_1 I = \widehat{\sigma}_z (NT)^{-1/3}$.

We calculate the bias, standard deviations (SD) and the root mean squared error (RMSE) of the estimators of the variance components. In addition, we also study the finite sample performances of the estimators of the nonparametric component. An estimator of $m(\cdot)$ is assessed via the square root of the averaged squared errors (SRASE) defined as

$$SRASE(\widehat{m}(z; H)) = \left[\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\widehat{m}(Z_{it}; H) - m(Z_{it}) \right)^2 \right]^{1/2}.$$

For the sake of comparison we present the mean and SD of the SRASE of the estimators considered in this paper: the pairwise estimator (PWE), the GLS estimator (GLSE), and the

Table 1. Finite sample performance of the estimators of the error variances.

		$\widehat{\sigma}_v^2$	$\widehat{\sigma}_v^3$	$\widehat{\sigma}_v^4$
$v_{it} \sim_{i.i.d.} 0.5\mathcal{N}(0,1)$,	$\varepsilon_i \sim_{i.i.d} 0.5\mathcal{N}(0,1)$			
N=50	Bias	0.0627	-0.0081	0.2972
	SD	0.0565	0.3036	0.3581
	MSE	0.0806	0.3044	0.3727
N=100	Bias	0.0396	0.0014	0.1189
	SD	0.0308	0.0627	0.3368
	MSE	0.0464	0.0645	0.3565
N=150	Bias	0.0299	0.0024	0.0798
	SD	0.0236	0.03784	0.1558
	MSE	0.0360	0.0402	0.1730
$v_{it} \sim_{i.i.d.} 0.5\mathcal{N}(0,1)$,	$\varepsilon_i \sim_{i.i.d} 0.5t(8)$			
N=50	Bias	0.0643	0.0088	0.2273
	SD	0.0561	0.1463	0.4036
	MSE	0.0807	0.1473	0.4254
N=100	Bias	0.0411	0.0038	0.1525
	SD	0.0331	0.0671	0.3703
	MSE	0.0501	0.0677	0.3999
N=150	Bias	0.0299	-0.0049	0.0895
	SD	0.0223	0.0447	0.2042
	MSE	0.0351	0.0456	0.2241
$v_{it} \sim_{i.i.d.} 0.5t(8)$,	$\varepsilon_i \sim_{i.i.d} 0.5\mathcal{N}(0,1)$			
N=50	Bias	0.0611	0.0024	0.1825
	SD	0.0664	0.1279	0.5412
	MSE	0.0775	0.1042	0.4638
N=100	Bias	0.0402	-0.0003	0.1203
	SD	0.0436	0.0876	0.3786
	MSE	0.0518	0.0709	0.3408
N=150	Bias	0.0313	0.0006	0.0728
	SD	0.0333	0.0594	0.2381
	MSE	0.0393	0.0544	0.1800
$v_{it} \sim_{i.i.d.} 0.5t(8)$,	$\varepsilon_i \sim_{i.i.d} 0.5t(8)$			
N=50	Bias	0.0642	-0.0040	0.2357
	SD	0.0657	0.1463	0.7842
	MSE	0.0807	0.1403	0.7614
N=100	Bias	0.0413	0.0034	0.1230
	SD	0.0436	0.0903	0.3688
	MSE	0.0510	0.0738	0.2731
N=150	Bias	0.0303	-0.0033	0.0795
	SD	0.0320	0.0636	0.2419
	MSE	0.0353	0.0457	0.1700

Table 2. Finite sample performance of the estimators of the error variances.

		$\widehat{\sigma}_b^2$	$\widehat{\sigma}_b^3$	$\widehat{\sigma}_b^4$
$v_{it} \sim_{i.i.d.} 0.5\mathcal{N}(0,1)$,	$\varepsilon_i \sim_{i.i.d} 0.5\mathcal{N}(0,1)$			
N=50	Bias	0.0055	0.0068	0.0126
	SD	0.0828	0.1150	0.1804
	MSE	0.0518	0.0750	0.1341
N=100	Bias	0.0031	0.0062	0.0099
	SD	0.0520	0.0769	0.1272
	MSE	0.0307	0.0528	0.0936
N=150	Bias	0.0003	-0.0004	0.0014
	SD	0.0453	0.0604	0.1009
	MSE	0.0277	0.0387	0.0703
$v_{it} \sim_{i.i.d.} 0.5\mathcal{N}(0,1)$,	$\varepsilon_i \sim_{i.i.d} 0.5t(8)$			
N=50	Bias	0.0045	0.0074	0.0186
	SD	0.1047	0.2283	0.6206
	MSE	0.0528	0.1045	0.2545
N=100	Bias	0.0035	0.0073	0.0178
	SD	0.0748	0.1499	0.3460
	MSE	0.0379	0.0765	0.1756
N=150	Bias	0.0022	0.0045	0.0121
	SD	0.0666	0.2076	0.1911
	MSE	0.0295	0.0748	0.1680
$v_{it} \sim_{i.i.d.} 0.5t(8)$,	$\varepsilon_i \sim_{i.i.d} 0.5\mathcal{N}(0,1)$			
N=50	Bias	0.0134	0.0086	0.0199
	SD	0.0819	0.1212	0.1990
	MSE	0.0564	0.0875	0.1556
N=100	Bias	0.0078	0.0039	0.0162
	SD	0.0596	0.0851	0.1309
	MSE	0.0407	0.0552	0.0996
N=150	Bias	0.0008	0.0021	0.0039
	SD	0.0510	0.0698	0.1157
	MSE	0.0296	0.0465	0.0828
$v_{it} \sim_{i.i.d.} 0.5t(8)$,	$\varepsilon_i \sim_{i.i.d} 0.5t(8)$			
N=50	Bias	0.0095	0.0011	0.0195
	SD	0.1145	0.2799	0.4995
	MSE	0.0609	0.1136	0.2724
N=100	Bias	0.0085	-0.0032	0.0184
	SD	0.0793	0.1614	0.3922
	MSE	0.0420	0.0777	0.1909
N=150	Bias	0.0068	0.0078	0.0093
	SD	0.0640	0.1392	0.3654
	MSE	0.0343	0.0638	0.1783

Table 3. Finite sample performance of the estimators of the nonparametric component.

		FEE	PWE	GLSE
$v_{it} \sim_{i.i.d.} 0.5\mathcal{N}(0,1)$,	$\varepsilon_i \sim_{i.i.d} 0.5\mathcal{N}(0,1)$			
N=50	Mean	0.4529	0.2538	0.2217
	SD	0.1393	0.0606	0.0366
	RE	1.0000	0.5978	0.5301
N=100	Mean	0.3748	0.1966	0.1722
	SD	0.1121	0.0411	0.0236
	RE	1.0000	0.5576	0.4953
N=150	Mean	0.3370	0.1731	0.1528
	SD	0.0911	0.0352	0.0200
	RE	1.0000	0.5421	0.4827
$v_{it} \sim_{i.i.d.} 0.5\mathcal{N}(0,1)$,	$\varepsilon_i \sim_{i.i.d} 0.5t(8)$			
N=50	Mean	0.4637	0.2538	0.2199
	SD	0.1623	0.0635	0.0369
	RE	1.0000	0.5880	0.5183
N=100	Mean	0.3788	0.1979	0.1715
	SD	0.1113	0.0467	0.0257
	RE	1.0000	0.5577	0.4843
N=150	Mean	0.3356	0.1705	0.1513
	SD	0.0816	0.0347	0.0211
	RE	1.0000	0.5312	0.4762
$v_{it} \sim_{i.i.d.} 0.5t(8)$,	$\varepsilon_i \sim_{i.i.d} 0.5\mathcal{N}(0,1)$			
N=50	Mean	0.4739	0.2629	0.2247
	SD	0.1549	0.0629	0.0403
	RE	1.0000	0.5990	0.5203
N=100	Mean	0.3962	0.2015	0.1762
	SD	0.1154	0.0467	0.0840
	RE	1.0000	0.5360	0.4750
N=150	Mean	0.3652	0.1815	0.1538
	SD	0.1020	0.0449	0.0225
	RE	1.0000	0.5244	0.4516
$v_{it} \sim_{i.i.d.} 0.5t(8)$,	$\varepsilon_i \sim_{i.i.d} 0.5t(8)$			
N=50	Mean	0.4717	0.2621	0.2230
	SD	0.1536	0.0653	0.0400
	RE	1.0000	0.5968	0.5159
N=100	Mean	0.4085	0.2141	0.1801
	SD	0.1315	0.0473	0.0258
	RE	1.0000	0.5607	0.4761
N=150	Mean	0.3797	0.1779	0.1535
	SD	0.1157	0.0338	0.0223
	RE	1.0000	0.5081	0.4390

fixed effects estimator (FEE) proposed in Rodriguez-Poo and Soberon (2015). These results are computed together with the relative efficiencies (RE) of the estimators proposed in this paper defined as the ratio of the SRASE of each estimator to that of the benchmark estimator (i.e., the FEE of Rodriguez-Poo and Soberon (2015)).

Some of the representative results for the estimators of the second, third and fourth moments of the individual effects and random errors are listed in Tables 1 and 2. Table 3 presents the results of the proposed nonparametric estimators.

From Tables 1 and 2, for all T, as N increases the bias and the standard deviation of the estimators of $\widehat{\sigma}_b^j$ and $\widehat{\sigma}_v^j$, for $j = 1, 2, 3$, are reduced in the four cases of study. We also get that the MSEs of these estimators are lower, as we expected from their asymptotic properties described in Section 4.

As we can see in Table 3, the three estimators perform quite well. For the three cases of study, as N increases for T fixed, these estimators are asymptotically unbiased. In addition, the standard deviation of the GLSE is smaller than that of the PWE, as we expected from their theoretical results of Sections 2 and 3. Finally, the relative efficiency of the GLSE is smaller than the corresponding of the PWE and this value is even lower as the sample size increases. Therefore, it is proved that the GLS-type estimator proposed in this paper performs asymptotically better than the estimator that ignores the error structure.

The finite sample performance of the test statistic. In order to check the significance of the individual effects, in Section 4.2 we develop an alternative test statistic. Now, to examine the finite sample performance of this test we use the same data generating process as in (5.1). For simplicity, we consider $c_0 = 0$ and focus on the case $v_{it} \sim i.i.d. 0.5N(0, \sigma_v^2)$ and $b_i \sim i.i.d. 0.5N(0, \sigma_b^2)$. Also, we fix the total variance across experiments to be $\sigma_v^2 + \sigma_b^2 = 10$ and assume that they vary according to $\rho = \sigma_b^2/(\sigma_b^2 + \sigma_v^2)$ to take values $(0.1, 0.2, 0.5)$, respectively. The results are listed in Tables 4-6.

From Table 4 we can see the estimated sizes of the $\widehat{J}$ test (i.e., the case of $c_0 = 0$) are close to the nominal sizes. From Tables 5 and 6 we observe the estimated power of the $\widehat{J}$ test. Analyzing these results we can highlight that, as we expected, the proposed test behaves well under the null hypothesis and is powerful under the alternative hypothesis. Specifically, for large N and fixed T, the power of the test increases rapidly as either the sample size increases or as the correlation between the individual effects and the regressors increases.

Table 4. Estimated size for the $\widehat{J}$ test when $\rho = 0$.

	1%	5%	10%
N=50	0.04	0.10	0.17
N=100	0.03	0.07	0.14
N=150	0.02	0.07	0.14

Table 5. Estimated power for the $\widehat{J}$ test when $\rho = 0.1$.

	1%	5%	10%
N=50	0.31	0.52	0.65
N=100	0.49	0.70	0.80
N=150	0.65	0.84	0.91

Table 6. Estimated power for the $\widehat{J}$ test when $\rho = 0.2$.

	1%	5%	10%
N=50	0.79	0.92	0.94
N=100	0.92	0.98	0.99
N=150	0.96	0.99	1.00

5.2 Applications

To demonstrate the usefulness of the estimation procedure and test proposed in this paper, now we consider the estimation of the production efficiency of the EU companies. The data used for this study are drawn from the Analyse Major Database from European Sources (AMADEUS) which contains information about the accounting and financial statements of around 10 millions of private and public companies of the main European countries. In order to hold the *i.i.d.* assumptions of the individuals, we randomly select 3.000 companies and consider those with research and development (R&D) expenses from 2008 to 2014. Therefore, after removing firms with missing values we get a final sample of 1.120 observations, i.e., 120 companies and 7 time periods.

When we consider the production efficiency, it is usual to use a Cobb-Douglas production function that assumes that the capital and labor elasticities are constant over the time. However, this is a strong assumption and we can find some empirical studies in which these elasticities vary according to other features of the companies such as the R&D expenses. As it is pointed out in Ahmad et al. (2005) among others, varying coefficient models are a natural way to extend these constant elasticities to the functional form. Also, it is a standard belief that the level of R&D has impact on the fixed capital marginal productivity but not in the liquid capital marginal productivity. We decompose the capital input into two terms, i.e., liquid capital and fixed capital.

Note that in a standard parametric model the R&D variable usually only has a neutral effect on the production function by shifting the level of the production frontier. However, in this section we propose a varying coefficient model that allows this variable to affect also the labor and/or capital marginal productivity. Then, the varying coefficient panel data model that we propose to estimate is of the following form

$$lnY_{it} = lnW_{it}\beta_1(Z_{it}) + lnL_{it}\beta_2(Z_{it}) + lnK_{it}\beta_3(Z_{it}) + b_i + v_{it}, \quad i = 1, \cdots, N; \quad t = 1, \cdots, T,$$

where Y represents the sales of the company, W the liquid capital, L the labor input, K the fixed capital, and Z the firm's R&D expenses.

The estimated curves are plotted against the R&D where the continuous line denotes the nonparametric curve while dotted line is the 95% pointwise confidence interval obtained using the results of Section 4. Figures 1 and 2 show the results for the marginal productivity of liquid and fixed capital, respectively. Figure 3 exhibits the estimation results of the marginal productivity of labor. Finally Figure 4 graphs the returns to scale function defined as $\beta_1(z) + \beta_2(z) + \beta_3(z)$.

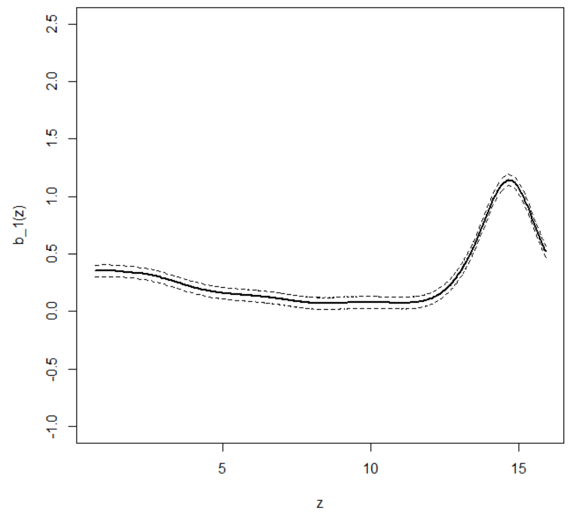

Figure 1. Marginal productivity liquid capital.

Focusing on the results of Figure 1 we can realize that the marginal productivity of liquid capital is more or less steady as long as the R&D expenses increases. However, for those companies with large R&D the marginal productivity first increase with z and later decreases sharply. On its part, analyzing the results of Figure 2 we can highlight that the marginal productivity of fixed capital is not a linear function of the variable R&D. Specifically, there is an upward trend in general, with a bell shape form for companies with large expenses. In this way, those companies with larger expenses in R&D exhibit higher marginal productivity of the fixed capital.

Analyzing the results of the marginal productivity of labor, from Figure 3 we observe that it first increases with the R&D expenses. Later, as this spending increases, this productivity is reduced followed by an increase at the end. This inverted bell shape form suggests that a low R&D expenses can improve the marginal productivity of labor, while larger expenses are related to lower levels of productivity. Note that this behavior is characteristic of companies that use the R&D expenses to improve the performance of their machines instead of allocating this investment to train workers. Finally, focusing on the results of Figure 4 we can see that the returns to scale are less than 1 in general. Specifically, companies with a low level of R&D expenses exhibit constant returns to scale, whereas companies with larger spending show decreasing returns to scale, passing from a value of 0.9 to 0.4.

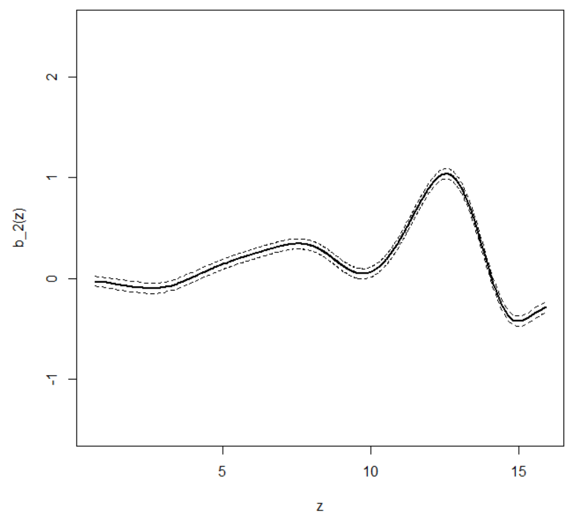

Figure 2. Marginal productivity fixed capital.

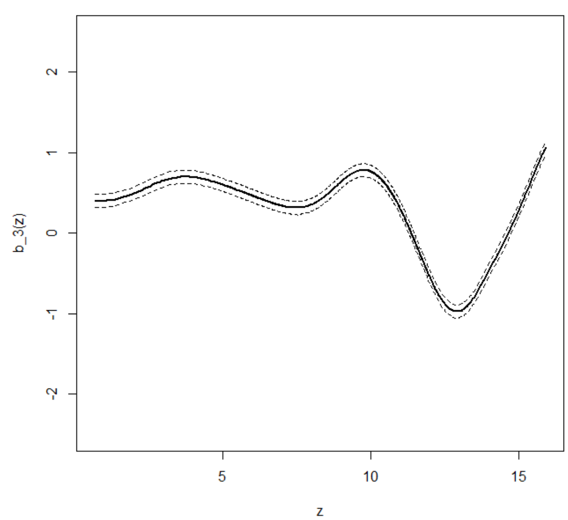

Figure 3. Marginal productivity labor.

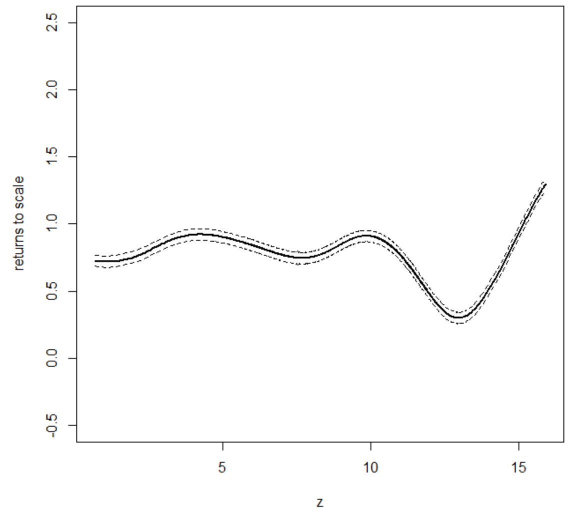

Figure 4. Returns to scale function.

Then, from all these results we can sum up that the R&D expenses have a different impact on the marginal productivity of the inputs depending on the magnitude of these expenses. On one hand, companies with larger R&D expenses exhibit decreasing returns to scale and the marginal productivity of fixed and liquid capital is sensitive to this investment. On the other hand, companies with moderate expenses present constant returns to scale but the marginal productivity of the liquid capital is not very sensitive to the R&D expenses. In addition, regarding to the marginal productivity of labor we get that a modest spending in R&D can increase the marginal productivity of labor, whereas larger R&D expenses are related to lower levels of productivity.

6 Conclusion

In the recent literature of panel data models there is a growing interest in the specification of nonparametric and semiparametric models by incorporating the variance components. However, most of these studies are based on standard distributional assumptions such as normality or symmetry of both the error and random variables that cannot be supported by data. In order to overcome it, this paper is concerned with the efficient and rate optimal nonparametric estimation and inference of varying coefficient panel data models. A two-step estimator that turns out to be asymptotically efficient exhibits an optimal nonparametric rate of convergence. In the first step, a pairwise differencing transformation is taken to remove the unobserved individual effects. In the second step, a backfitting algorithm is implemented to incorporate the information from the error covariance matrix. Note that, despite the results in this literature, it is shown that the proposed GLS-type estimator is asymptotically more efficient than the previous one without imposing any closeness property within groups.

In addition, by extending the fundamental lemma proposed in Wu et al. (2012) from linear mixed models to the nonparametric panel data models a root-n consistent estimator of the variance components is proposed. As a by-product of the previous results, a very simple and intuitive test is implemented to check the significance of individual effects. Finally, some simulations are used to examine the finite sample performance of the proposed estimators and test statistic. Also, in order to see how useful our results are, we illustrate an application about the production efficiency of the companies in the EU. From this empirical application, we obtain that R&D expenses have a different impact on the marginal productivity of the inputs depending on the magnitude of these expenses. On one hand, companies with larger R&D expenses exhibit decreasing returns to scale and the marginal productivity of fixed and liquid capital is sensitive to this investment. On the other hand, companies with moderate expenses present constant returns to scale but the marginal productivity of the liquid capital is not very sensitive to the R&D expenses.

∎

Appendix

Proof of Theorem 2.1. The proof of this theorem consists of three parts. First, the bias of the local constant estimator in (2.6) is obtained. Second, we give the variance term of this estimator and we conclude by obtaining the asymptotic distribution of our estimator.

For the sake of simplicity let us denote

$$K_{it} = |H_1|^{-1/2} K\left(H_1^{-1/2}(Z_{it} - z)\right) \quad \text{and} \quad K_{is} = |H_1|^{-1/2} K\left(H_1^{-1/2}(Z_{is} - z)\right).$$

Using the multivariate Taylor's theorem and since by the regularity conditions of Theorem 2.1 we obtain

$$
\begin{aligned}
\widetilde{Y}_{its} &= \widetilde{X}_{its}^{\top} m(z) + \left(X_{it}^{\top} \otimes (Z_{it}-z)^{\top} - X_{is}^{\top} \otimes (Z_{is}-z)^{\top} \right) D_m(z) \\
&+ \frac{1}{2} \left(X_{it}^{\top} \otimes (Z_{it}-z)^{\top} \mathcal{H}_m(z)(Z_{it}-z) - X_{is}^{\top} \otimes (Z_{is}-z)^{\top} \mathcal{H}_m(z)(Z_{is}-z) \right) + \widetilde{v}_{its} + o_{\mathbb{P}}(1),
\end{aligned} \tag{6.1}
$$

where $D_m(z)$ is a $dq \times 1$ vector such that $D_m(z) = vec(\partial m(z)/\partial z^{\top})$ is the first-order derivative vector of $m(\cdot)$ and $\mathcal{H}_m(z)$ is a $dq \times q$ matrix such that $\mathcal{H}_m(z) = \partial m(z)/\partial z \partial z^{\top}$ is the Hessian matrix of $m(\cdot)$.

Let $n = NT(T-1)$. If we replace (6.1) in (2.4) and rearrange terms we can write $\widehat{m}(z; H_1)$ as

$$
\widehat{m}(z; H_1) - m(z) = \Psi_n^{-1} \left(B_n^{(1)} + B_n^{(2)} + U_n \right), \tag{6.2}
$$

where

$$
\begin{aligned}
\Psi_n &= \frac{1}{n} \sum_{its} K_{it} K_{is} \widetilde{X}_{its} \widetilde{X}_{its}^{\top} \\
B_n^{(1)} &= \frac{1}{n} \sum_{its} K_{it} K_{is} \widetilde{X}_{its} \left(X_{it}^{\top} \otimes (Z_{it}-z)^{\top} - X_{is}^{\top} \otimes (Z_{is}-z)^{\top} \right) D_m(z), \\
B_n^{(2)} &= \frac{1}{2n} \sum_{its} K_{it} K_{is} \widetilde{X}_{its} \left(X_{it}^{\top} \otimes (Z_{it}-z)^{\top} \mathcal{H}_m(z)(Z_{it}-z) - X_{is}^{\top} \otimes (Z_{is}-z)^{\top} \mathcal{H}_m(z)(Z_{is}-z) \right), \\
U_n &= \frac{1}{n} \sum_{its} K_{it} K_{is} \widetilde{X}_{its} \widetilde{v}_{its}.
\end{aligned}
$$

Then, to analyze the asymptotic behavior of this estimator it is enough to show

$$
\sqrt{N|H_1|} \left(\widehat{m}(z; H_1) - m(z) \right) - \sqrt{N|H_1|} \Psi_n^{-1} \left(B_n^{(1)} + B_n^{(2)} \right) = \sqrt{N|H_1|} \Psi_n^{-1} U_n, \tag{6.3}
$$

where we will demonstrate that $\Psi_n^{-1} B_n^{(j)}$, for $j = 1, 2$, contributes to the asymptotic bias, whereas the right-hand side term of (6.3) is asymptotically normal.

Starting with the bias term of this estimator, we first focus on Ψ_n^{-1}. Under the assumption that X_{it} and v_{it} are *i.i.d.* across i for each fixed t and as N tends to infinity, we get

$$
\Psi_n = \mathcal{B}_{\widetilde{X}\widetilde{X}}(z, z)(1 + o_{\mathbb{P}}(1)), \tag{6.4}
$$

where $\mathcal{B}_{\widetilde{X}\widetilde{X}}(z, z)$ is a $q \times q$ matrix of the form

$$
\mathcal{B}_{\widetilde{X}\widetilde{X}}(z, z) = E\left[\widetilde{X}_{its} \widetilde{X}_{its}^{\top} | Z_{it} = z, Z_{is} = z \right] f_{Z_{it}, Z_{is}}(z, z).
$$

In order to show this result we obtain that by the law of iterated expectations and the strict stationarity condition

$$E\left(\Psi_n\right) = \int E\left[\widetilde{X}_{its}\widetilde{X}_{its}^{\top}|Z_{it} = z + H_1^{1/2}u, Z_{is} = z + H_1^{1/2}v\right] f_{Z_{it},Z_{is}}(Z_{it} = z + H_1^{1/2}u, Z_{is} = z + H_1^{1/2}v) \times K(u)K(v)dudv,$$

where by a Taylor expansion and Assumption 2.1 the expression (6.4) holds. Also, to complete the proof it is necessary to show that $Var(\Psi_n) \to 0$ as $N \to \infty$. For this,

$$Var\left(\Psi_n\right) = \frac{1}{n}Var\left(K_{it}K_{is}\widetilde{X}_{its}\widetilde{X}_{its}^{\top}\right) + \frac{1}{n^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s\neq t}^{T}\sum_{t'=1}^{T}\sum_{s'\neq t'}^{T} Cov\left(K_{it}K_{is}\widetilde{X}_{its}\widetilde{X}_{its}^{\top}, K_{it'}K_{is'}\widetilde{X}_{it's'}\widetilde{X}_{it's'}^{\top}\right),$$

where under Assumptions 2.1 and 2.2 we can show that the first element is $O_{\mathbb{P}}\left(\frac{1}{N|H_1|}\right)$ whereas the second one is $o_{\mathbb{P}}\left(\frac{1}{N|H_1|}\right)$. Then, if $N|H_1| \to \infty$ the variance term tends to zero and (6.4) holds.

Using this result and the inverse matrix of S_n it is proved that by the Slutsky theorem,

$$\Psi_n^{-1} = \mathcal{B}_{\widetilde{X}\widetilde{X}}^{-1}(z,z) + o_{\mathbb{P}}(1). \tag{6.5}$$

Similarly, for the standard case $\mu_2(K_u) = \mu_2(K_v) = \mu_2(K)$ we can show that by the law of iterated expectations and the stationarity condition,

$$\begin{aligned} E\left(B_n^{(1)}\right) &= \frac{1}{T(T-1)}\sum_{ts} E\left[K_{it}K_{is}\left(E(\widetilde{X}_{its}X_{it}^{\top}|Z_{it}, Z_{is}) \otimes (Z_{it} - z)^{\top} - E(\widetilde{X}_{its}X_{is}^{\top}|Z_{it}, Z_{is}) \otimes (Z_{is} - z)^{\top}\right) D_m(z)\right] \\ &= \frac{1}{T(T-1)}\sum_{ts}\int \left(E(\widetilde{X}_{its}X_{it}^{\top}|Z_{it} = z, Z_{is} = z)D_f(z)(H_1^{1/2}u)\right) \otimes (H_1^{1/2}u)^{\top} D_m(z)K(u)K(v)dudv \\ &- \frac{1}{T(T-1)}\sum_{ts}\int \left(E(\widetilde{X}_{its}X_{is}^{\top}|Z_{it} = z, Z_{is} = z)D_f(z)(H_1^{1/2}v)\right) \otimes (H_1^{1/2}v)^{\top} D_m(z)K(u)K(v)dudv \\ &= \mu_2(K)\mathcal{B}_{\widetilde{X}\widetilde{X}}(z,z)diag_d\left(tr(H_1 D_f(z) D_{m_r}(z))\right)\imath_d f_{Z_{it},Z_{is}}^{-1}(z,z) + o_{\mathbb{P}}(tr(H_1)), \end{aligned} \tag{6.6}$$

where for $r = 1, \cdots, d$, D_{m_r} is the first-order derivative vector of the rth component of $m(\cdot)$ and $\imath_d$ is a $d \times 1$ unit vector.

Following a similar procedure it is straightforward to show

$$\begin{aligned} E\left(B_n^{(2)}\right) &= \frac{1}{2T(T-1)}\sum_{ts}\int E(\widetilde{X}_{its}X_{it}^{\top}|Z_{it} = z, Z_{is} = z)f(z,z) \otimes tr(H_1\mathcal{H}_{m_r}(z))u^{\top}uK(u)K(v)dudv \\ &- \frac{1}{2T(T-1)}\sum_{ts}\int E(\widetilde{X}_{its}X_{is}^{\top}|Z_{it} = z, Z_{is} = z)f(z,z) \otimes tr(H_1\mathcal{H}_{m_r}(z))v^{\top}vK(u)K(v)dudv \\ &= \frac{\mu_2(K)}{2}\mathcal{B}_{\widetilde{X}\widetilde{X}}(z,z)diag_d\left(tr(H_1\mathcal{H}_{m_r}(z))\right)\imath_d + o_{\mathbb{P}}(tr(H_1)), \end{aligned} \tag{6.7}$$

where $\mathcal{H}_{m_r}(z)$ is the Hessian matrix of the rth component of $m(\cdot)$.

Using similar arguments as above we can show that any component of the variance of $B_n^{(1)}$ and $B_n^{(2)}$ converges to zero as $H_1 \to 0$ and $N|H_1| \to \infty$. Then, using (6.5)-(6.7) and applying the Cramér-Wold device it is proved that the asymptotic bias of $\widehat{m}(z; H_1)$ is

$$\begin{aligned}\Psi_n^{-1}(B_n^{(1)} + B_n^{(2)}) &= \mu_2(K)\mathcal{B}_{\widetilde{X}\widetilde{X}}^{-1}(z,z)\mathcal{B}_{\widetilde{X}\widetilde{X}}(z,z)\left(diag_d\left(tr(H_1 D_f(z) D_{m_r}(z))\right)\imath_d f_{Z_{it},Z_{is}}^{-1}(z,z)\right. \\ &+ \left.\frac{1}{2} diag_d\left(tr(H_1\mathcal{H}_{m_r}(z))\imath_d\right)\right) + o_{\mathbb{P}}(tr(H_1)). \qquad (6.8)\end{aligned}$$

So the first part of the proof is done.

Now, to obtain the asymptotic variance of the right-hand side of (6.3) we have to analyze the behavior of U_n. Let $\mathbb{X}$ be the observed covariates vector and denote with $\widetilde{v} = (\widetilde{v}_1, \cdots, \widetilde{v}_N)^\top$ the $NT(T-1)$ vector with $\widetilde{v}_i = (\widetilde{v}_{i12}, \cdots, \widetilde{v}_{i1T}, \cdots, \widetilde{v}_{iT1}, \cdots, \widetilde{v}_{iT(T-1)})^\top$, $E(\widetilde{v}_i \widetilde{v}_i'^\top|\mathbb{X}) = 0$ for $\forall i \neq i'$ and

$$E\left(\widetilde{v}_i\widetilde{v}_i^\top|\mathbb{X}\right) = \begin{cases} 2\sigma_v^2 & \text{for} \quad t = t', s = s', \\ \sigma_v^2 & \text{for} \quad t = t', s \neq s', \\ 0 & \text{for} \quad t \neq t', s \neq s'. \end{cases}$$

When we analyze U_n we claim that by the law of iterated expectations and Assumptions 2.1, 2.5, and 2.6,

$$\begin{aligned} n|H_1|Var(U_n) &= |H_1|n^{-1}\sum_{its} E\left[\widetilde{X}_{its}E\left(\widetilde{v}_{its}\widetilde{v}_{its}|\mathbb{X}\right)\widetilde{X}_{its}^\top K_{it}^2 K_{is}^2\right] \\ &+ |H_1|n^{-1}\sum_{its}\sum_{s'\neq s} E\left[\widetilde{X}_{its}E\left(\widetilde{v}_{its}\widetilde{v}_{its'}|\mathbb{X}\right)\widetilde{X}_{its'}^\top K_{it}K_{is}^2 K_{is'}\right] \\ &+ |H_1|n^{-1}\sum_{its}\sum_{t'\neq t}\sum_{s'\neq s} E\left[\widetilde{X}_{its}E\left(\widetilde{v}_{its}\widetilde{v}_{it's'}|\mathbb{X}\right)\widetilde{X}_{it's'}^\top K_{it}K_{it'}K_{is}K_{is'}\right] \\ &= \mathbf{I}_{1n} + \mathbf{I}_{2n} + \mathbf{I}_{3n}. \end{aligned}$$

Then, analyzing each of these terms separately we obtain

$$n|H_1|Var(U_n) = 2\sigma_v^2 R^2(K)\mathcal{B}_{\widetilde{X}\widetilde{X}}(z,z)(1 + o_{\mathbb{P}}(1)). \qquad (6.9)$$

With similar arguments as above we can prove that

$$\mathbf{I}_{1n} = \frac{2\sigma_v^2}{T(T-1)}\sum_{t=1}^{T}\sum_{s\neq t}^{T}\int E\left[\widetilde{X}_{its}\widetilde{X}_{its}^\top|Z_{it} = z, Z_{is} = z\right] f(z,z)K^2(u)K^2(v)dudv(1 + o_{\mathbb{P}}(1)),$$

$$\mathbf{I}_{2n} = \sigma_v^2|H_1|^{1/2}(T-2)E\left[\widetilde{X}_{its}\widetilde{X}_{its'}^\top|Z_{it} = z, Z_{is} = z, Z_{is'=z}\right] f_{Z_{it},Z_{is},Z_{is'}}(z,z,z)(1 + o_{\mathbb{P}}(1)),$$

and $\mathbf{I}_{3n} = o_{\mathbb{P}}(1)$ since $E(\widetilde{v}_{its}\widetilde{v}_{it's'}|\mathbb{X}) = 0$.

Then, using (6.4) and (6.9), and by the Cramér-Wold device, as $N|H_1| \to \infty$,

$$n|H_1|Var\left(\Psi_n^{-1}U_n\right) = 2\sigma_v^2 R^2(K)\mathcal{B}_{\widetilde{X}\widetilde{X}}^{-1}(z,z)\mathcal{B}_{\widetilde{X}\widetilde{X}}(z,z)\mathcal{B}_{\widetilde{X}\widetilde{X}}^{-1}(z,z)(1+o_{\mathbb{P}}(1)). \tag{6.10}$$

Note that the conditions established for H_1 are enough to show that the other terms of the variance are $o_{\mathbb{P}}(1)$.

Finally, to complete the proof of Theorem 2.1 it is necessary to show that, as N tends to infinity, for T fixed,

$$\sqrt{N|H_1|}\left(\widehat{m}(z,H_1) - m(z) - B(z,H_1)\right) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, 2\sigma_v^2 R^2(K)\mathcal{B}_{\widetilde{X}\widetilde{X}}^{-1}(z,z)\right), \tag{6.11}$$

where

$$B(z,H_1) = \mu_2(K)diag_d\left(tr(H_1 D_f(z) D_{m_r}(z))\right)\imath_d f_{Z_{it},Z_{is}}^{-1}(z,z) + \frac{\mu_2(K)}{2}diag_d\left(tr(H_1\mathcal{H}_{m_r}(z))\right)\imath_d.$$

In order to show this result, we check the Lindeberg condition for which we can write

$$\sqrt{N|H_1|}n^{-1}\sum_{its} K_{it}K_{is}\widetilde{X}_{its}\widetilde{v}_{its} = \frac{1}{\sqrt{N}}\frac{1}{T(T-1)}\sum_{its}\lambda_{its}, \tag{6.12}$$

where

$$\lambda_{its} = |H_1|^{1/2}K_{it}K_{is}\widetilde{X}_{its}\widetilde{v}_{its}.$$

By Theorem 2.1 and using the previous proofs we see that as $H_1 \to 0$,

$$\begin{aligned} Var(\lambda_{its}) &= 2\sigma_v^2 R^2(K)\mathcal{B}_{\widetilde{X}\widetilde{X}}^{-1}(z,z)(1+o_{\mathbb{P}}(1)), \\ Cov(\lambda_{i1s},\lambda_{i1s'}) &= o_{\mathbb{P}}(1). \end{aligned}$$

Defining $\lambda_{n,i} = \frac{1}{T(T-1)}\sum_{ts}\lambda_{its}$ and by the Minkowsky inequality we get

$$E|\lambda_{n,i}|^{2+\lambda} \leq C(T(T-1))^{(2+\delta)/2}E|\lambda_{its}|.$$

For analyzing the behavior of λ_{its}, by the law of iterated expectations, we get

$$\begin{aligned} E|\lambda_{it}|^{2+\delta} &\leq |H_1|^{(2+\delta)/2}E|\widetilde{X}_{its}\widetilde{v}_{its}K_{it}^{2+\delta}K_{is}^{2+\delta}| \\ &= |H_1|^{\delta/2}E\left(|\widetilde{X}_{its}\widetilde{v}_{its}|^{2+\delta}|Z_{it}=z, Z_{is}=z\right)f_{Z_{it},Z_{is}}(z,z)\int K^{2+\delta}(u)K^{2+\delta}(v)dudv \\ &+ o_{\mathbb{P}}\left(|H_1|^{-\delta/2}\right). \end{aligned}$$

Hence, it is proved that

$$E|\lambda_{n,i}|^{2+\delta} = N^{-(2+\delta)/2}\sum_{i} E\left|\lambda_{n,i}\right|^{2+\delta} \leq CO_{\mathbb{P}}\left((N|H_1|)^{-\delta/2}\right).$$

Finally, since this term tends to zero as $N|H_1| \to \infty$, the Lindeberg condition is verified. So we can conclude that the Lyapunov Central Limit Theorem can be used to verify (6.11), and the proof of Theorem 2.1 is done.

∎

Proof of Theorem 3.1. In order to obtain the desired results of Theorem 3.1 we follow the same lines of the proof above. To this end, let us define

$$\widetilde{m}(z;H_2) = S_n^{-1}(z)\frac{\kappa}{n}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s\neq t}^{T} K_{H_2}(Z_{it}-z)X_{it}\left(\widetilde{Y}^*_{its} - \tau\sum_{r\neq t}^{T}\widetilde{Y}^*_{itr}\right). \tag{6.13}$$

Using this definition, the GLS nonparametric estimator (3.6) can be rewritten as

$$\widehat{m}(z;H_2) = (\widehat{m}(z;H_2) - \widetilde{m}(z;H_2)) + \widetilde{m}(z;H_2). \tag{6.14}$$

According to this expression, the asymptotic properties of this estimator can be obtained showing that, under the conditions established in Theorem 3.1, the first element of the right-hand side of (6.13) is asymptotically negligible whereas the second one provides the convergence in distribution of this estimator, as N tends to infinity. These results are proved in Lemmas 6.1 and 6.2.

Lemma 6.1 *Under the conditions of Theorem 3.1, as $N \to \infty$ and T is fixed,*

$$\sqrt{N|H_2|}\left(\widehat{m}(z;H_2) - \widetilde{m}(z;H_2)\right) = o_{\mathbb{P}}(1) \quad \textit{uniformly in } z.$$

Proof of Lemma 6.1 Focus on the first element of (6.13). Combining (6.13) and (6.14), we see that

$$\widehat{m}(z;H_2) - \widetilde{m}(z;H_2) = S_n^{-1}(z)\frac{\kappa}{n}\sum_{its} K_{H_2}(Z_{it}-z)X_{it}\left(\left(\widehat{Y}_{its} - \tau\sum_{r\neq t}^{T}\widehat{Y}_{itr}\right) - \left(\widetilde{Y}^*_{its} - \tau\sum_{r\neq t}^{T}\widetilde{Y}^*_{itr}\right)\right).$$

Analyzing the behavior of $S_n(z)$ and following a similar reasoning as in (6.4) we see that

$$S_n^{-1}(z) = \mathcal{B}_{XX}^{-1}(z) + o_{\mathbb{P}}(1), \tag{6.15}$$

where $\mathcal{B}_{XX}(z)$ is a $d \times d$ matrix of the form

$$\mathcal{B}_{XX}(z) = E\left[X_{it}X_{it}^{\top}|Z_{it} = z\right] f_{Z_{it}}(z).$$

Then, by the Slutsky theorem it is proved

$$S_n^{-1}(z) = \mathcal{B}_{XX}^{-1}(z)(1 + o_{\mathbb{P}}(1)). \tag{6.16}$$

Focus now on the behavior of the numerator of (6.15) and replacing $\widehat{Y}_{its} - \tau \sum_{r\neq t}^{T} \widehat{Y}_{itr}$ and $\widetilde{Y}^*_{its} - \tau \sum_{r\neq t}^{T} \widetilde{Y}^*_{itr}$ by (3.6) and (3.7), respectively, we can show that under the assumptions of Theorem 3.1, as N tends to infinity

$$\begin{aligned}
& \frac{\kappa}{n} \sum_{its} K_{H_2}(Z_{it} - z) X_{it} \left(\left(\widehat{Y}_{its} - \tau \sum_{r\neq t}^{T} \widehat{Y}_{itr} \right) - \left(\widetilde{Y}^*_{its} - \tau \sum_{r\neq t}^{T} \widetilde{Y}^*_{itr} \right) \right) \\
= \ & \frac{\kappa}{n} \sum_{its} K_{H_2}(Z_{it} - z) X_{it} \left[X_{is}^{\top} \left(\widehat{m}(Z_{is}; H_1) - m(Z_{is}) \right) - \tau \sum_{r\neq t}^{T} X_{ir}^{\top} \left(\widehat{m}(Z_{ir}; H_1) - m(Z_{ir}) \right) \right] \\
& \leq \frac{\kappa}{n} \sum_{its} |K_{H_2}(Z_{it} - z) X_{it} X_{is}^{\top}| \sup_{\{Z_{is} \in \mathcal{Z}\}} |\widehat{m}(Z_{is}; H_2) - m(Z_{is})| \\
& + \frac{\kappa\tau}{n} \sum_{itsr} |K_{H_2}(Z_{it} - z) X_{it} X_{ir}^{\top}| \sup_{\{Z_{ir} \in \mathcal{Z}\}} |\widehat{m}(Z_{ir}; H_2) - m(Z_{ir})| = o_{\mathbb{P}}(1),
\end{aligned} \tag{6.17}$$

where we use the uniform convergence results of Theorem 6 in Masry (1996). It is straightforward to show that $n^{-1} \sum_{its} |K_{H_2}(Z_{it} - z) X_{it} X_{is}^{\top}| = O_{\mathbb{P}}(1)$ and $n^{-1} \sum_{itsr} |K_{H_2}(Z_{it} - z) X_{it} X_{ir}^{\top}| = O_{\mathbb{P}}(1)$.

Then, using (6.15) and (6.17) and the Cramér-Wold device we obtain

$$\widehat{m}(z; H_2) - \widetilde{m}(z; H_2) = o_{\mathbb{P}}(1).$$

So Lemma 6.1 is proved.

∎

On the other hand, in order to prove the asymptotic distribution of the GLS estimator (3.7) we can write (6.13) as

$$\sqrt{N|H_2|^{1/2}} \left(\widehat{m}(z; H_2) - m(z) \right) = \sqrt{N|H_2|^{1/2}} \left(\widehat{m}(z; H_2) - \widetilde{m}(z; H_2) \right) + \sqrt{N|H_2|} \left(\widetilde{m}(z; H_2) - m(z) \right),$$

where using Lemma 6.1

$$\sqrt{N|H_2|^{1/2}} \left(\widehat{m}(z; H_2) - m(z) \right) = \sqrt{N|H_2|^{1/2}} \left(\widetilde{m}(z; H_2) - m(z) \right) + o_{\mathbb{P}}(1) \tag{6.18}$$

which follows from Lemma 6.2 below.

Lemma 6.2

$$\sqrt{N|H_2|^{1/2}} \left(\widehat{m}(z; H_2) - m(z) - B(z; H_2) \right) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, V(z; H_2)\right)$$

where

$$B(z;H_2) = \mu_2(K)\left[diag_d(tr(H_2 D_f(z) D_{m_r}(z))) f_{Z_{it}}^{-1}(z) + diag_d(tr(H_2 \mathcal{H}_{m_r}(z)))\right]\imath_d,$$
$$V(z;H_2) = \sigma_v^2\kappa^2 R(K)\mathcal{B}_{XX}^{-1}(z).$$

Proof of Lemma 6.2. By the assumptions of Theorem 3.1 a Taylor expansion implies

$$\begin{aligned}\widetilde{Y}_{its}^* - \tau\sum_{r\neq t}^{T}\widetilde{Y}_{itr}^* &= X_{it}^\top m(Z_{it}) + X_{it}^\top \otimes (Z_{it}-z)^\top D_m(z) + \frac{1}{2}X_{it}^\top \otimes (Z_{it}-z)^\top \mathcal{H}_m(z)(Z_{it}-z) \\ &+ \widetilde{v}_{its} - \tau\sum_{r\neq t}^{T}\widetilde{v}_{itr} + o_{\mathbb{P}}(1). \end{aligned} \tag{6.19}$$

Replacing (6.19) in (6.13) and rearranging terms $\widetilde{m}(z;H_2)$ can be written as

$$\widetilde{m}(z;H_2) - m(z) = S_n^{-1}\left(\widetilde{B}_n^{(1)} + \widetilde{B}_n^{(2)} + \widetilde{U}_n\right), \tag{6.20}$$

where

$$\widetilde{B}_n^{(1)} = \frac{\kappa}{n}\sum_{its} K_{H_2}(Z_{it}-z)X_{it}X_{it}^\top \otimes (Z_{it}-z)^\top D_m(z),$$
$$\widetilde{B}_n^{(2)} = \frac{\kappa}{2n}\sum_{its} K_{H_2}(Z_{it}-z)X_{it}X_{it}^\top \otimes (Z_{it}-z)^\top \mathcal{H}_m(z)(Z_{it}-z),$$
$$\widetilde{U}_n = \frac{\kappa}{n}\sum_{its} K_{H_2}(Z_{it}-z)X_{it}\left(\widetilde{v}_{its} - \tau\sum_{r\neq t}\widetilde{v}_{itr}\right).$$

Then, to complete the proof of Lemma 6.2 it is enough to show that

$$\sqrt{N|H_2|^{1/2}}\left(\widehat{m}(z;H_2) - m(z)\right) - \sqrt{N|H_2|^{1/2}}S_n^{-1}\left(\widetilde{B}_n^{(1)} + \widetilde{B}_n^{(2)}\right) = \sqrt{N|H_2|^{1/2}S_n^{-1}\widetilde{U}_n}, \tag{6.21}$$

where $S_n^{-1}\widetilde{B}_n^j$, for $j = 1, 2$, will contribute to the asymptotic bias while $S_n^{-1}\widetilde{U}_n$ will be asymptotically normal.

On the one hand, for analyzing the behavior of the bias term, we can show that as in the proof of Lemma 6.1 as $N \to \infty$,

$$E\left(\widetilde{B}_n^{(1)}\right) = \kappa\mu_2(K)\mathcal{B}_{XX}(z) diag_d(tr(H_2)D_f(z)D_{m_r}(z)) f_{Z_{it}}^{-1}(z)\imath_d + o_{\mathbb{P}}(tr(H_2)) \tag{6.22}$$

and

$$E\left(\widetilde{B}_n^{(2)}\right) = \frac{\kappa\mu_2(K)}{2}\mathcal{B}_{XX}(z) diag_d(tr(H_2\mathcal{H}_{m_r}(z)))\imath_d + o_{\mathbb{P}}(tr(H_2)), \tag{6.23}$$

where $diag_d(tr(H_2)D_f(z)D_{m_r}(z))$ and $diag_d(tr(H_2\mathcal{H}_m(z)))$ stands for a diagonal matrix of elements $tr(H_2)D_f(z)D_{m_r}(z)$ and $tr(H_2\mathcal{H}_m(z))$, respectively.

By this procedure it is easy to show that any component of the variance of $\widetilde{B}_n^{(j)}$, for $j = 1, 2$, converges to zero as $H_2 \to 0$ and $N|H_2| \to \infty$. Then, inserting (6.22) and (6.23) into the second element of the left-hand side of (6.20) and using the Cramér-Wold device we can conclude that the bias term has the form

$$
\begin{aligned}
S_n^{-1}(z)\left(\widetilde{B}_n^{(1)} + \widetilde{B}_n^{(2)}\right) &= \kappa\mu_2(K)\mathcal{B}_{XX}^{-1}(x)\mathcal{B}_{XX}(z)\left[diag_d(tr(H_2D_f(z)D_{m_r}(z))f_{Z_{it}}^{-1}(z)+\right. \\
&+ \left.\frac{1}{2}diag_d(tr(H_2\mathcal{H}_{m_r}(z)))\right]\imath_d + o_{\mathbb{P}}(tr(H_2)). \qquad (6.24)
\end{aligned}
$$

Now we consider the asymptotic variance of $\widetilde{m}(z; H_2)$. To this end, remember that in Section 3 we premultiply the model of interest by V_i^{-1}. So the disturbances of the resulting regression model are spherical. By this reason and following a similar arguments as in (6.9) the variance term of $\widetilde{U}_n$ can be written as

$$
\begin{aligned}
n|H_2|^{1/2}Var\left(\widetilde{U}_n\right) &= |H_2|^{1/2}\sigma_v^2\kappa^2(NT)^{-1}\sum_{it}\sum_{t'}E\left[K_{H_2}(Z_{it}-z)K_{H_2}(Z_{it'}-z)X_{it}X_{it'}^\top\right] \\
&= \sigma_v^2\kappa^2R(K)\mathcal{B}_{XX}(z)(1+o_{\mathbb{P}}(1)). \qquad (6.25)
\end{aligned}
$$

Then, resorting to the Cramér-Wold device and using the above result together with (6.15) we obtain,

$$
n|H_2|^{1/2}Var(S_n^{-1}(z)\widetilde{U}_n) = \sigma_v^2\kappa^2R(K)\mathcal{B}_{XX}^{-1}(z)\mathcal{B}_{XX}(z)\mathcal{B}_{XX}^{-1}(z)(1+o_{\mathbb{P}}(1)). \qquad (6.26)
$$

Finally, to complete the proof of the Lemma the Lyapunov condition has to be checked. For this, we can write

$$
\sqrt{N|H_2|^{1/2}}\frac{1}{n}\sum_{its}K_{H_2}(Z_{it}-z)X_{it}\left(\widetilde{v}_{its} - \tau\sum_{r\neq t}^{T}\widetilde{v}_{itr}\right) = \frac{1}{\sqrt{N}}\left(\frac{1}{T(T-1)}\right)\sum_{its}\widetilde{\lambda}_{its}, \qquad (6.27)
$$

where $\widetilde{\lambda}_{its} = |H_2|^{1/2}K_{H_2}(Z_{it}-z)X_{it}\left(\widetilde{v}_{its} - \tau\sum_{r\neq t}^{T}\widetilde{v}_{itr}\right)$. Following the same lines as in the proof of Theorem 2.1 we can prove

$$
E\left|\widetilde{\lambda}_{its}\right|^{2+\delta} \leq CO_{\mathbb{P}}\left((N|H_2|^{1/2})^{-\delta/2}\right).
$$

Then, using this result and replacing (6.24) and (6.26) in (6.21) we can conclude

$$
\sqrt{N|H_2|^{1/2}}\left(\widehat{m}(z;H_2) - m(z) - B(z;H_2)\right) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, \sigma_v^2\kappa^2R(K)\mathcal{B}_{XX}^{-1}(z)\mathcal{B}_{XX}(z)\mathcal{B}_{XX}^{-1}(z)\right)
$$

and the proof is completed.

∎

Proof of Theorem 4.1. We first focus on the asymptotic properties of $\widehat{\sigma}_v^2$ and later on $\widehat{\sigma}_b^2$. From (4.1) and $\widehat{\epsilon}_{it} = \epsilon_{it} - \left(X_{it}^\top[\widehat{m}(Z_{it}; H_1) - m(Z_{it})]\right)$ we can write

$$\begin{aligned}\widehat{\sigma}_v^2 &= \frac{1}{NT(T-1)}\sum_{i=1}^{N}\left[T\sum_{t=1}^{T}\left(\epsilon_{it} - X_{it}^\top[\widehat{m}(Z_{it}; H_1) - m(Z_{it})]\right)^2 - \left(\sum_{t=1}^{T}\left[\epsilon_{it} - X_{it}^T[\widehat{m}(Z_{it}; H_1) - m(Z_{it})]\right]\right)^2\right] \\ &= I\!I_v^{(1)} + I\!I_v^{(2)} - I\!I_v^{(3)} + I\!I_v^{(4)} + I\!I_v^{(5)} - I\!I_v^{(6)}, \qquad (6.28)\end{aligned}$$

where after rearranging terms

$$\begin{aligned}I\!I_v^{(2)} &= \frac{1}{NT}\sum_{it} X_{it}^\top[\widehat{m}(Z_{it}; H_1) - m(Z_{it})]X_{it}^\top[\widehat{m}(Z_{it}; H_1) - m(Z_{it})], \\ I\!I_v^{(3)} &= \frac{2}{NT}\sum_{it}\epsilon_{it}X_{it}^\top[\widehat{m}(Z_{it}; H_1) - m(Z_{it})], \\ I\!I_v^{(4)} &= \frac{1}{n}\sum_{its}\epsilon_{it}X_{is}^\top[\widehat{m}(Z_{is}; H_1) - m(Z_{is})] \\ I\!I_v^{(5)} &= \frac{1}{n}\sum_{its}X_{it}^\top[\widehat{m}(Z_{it}; H_1) - m(Z_{it})]\epsilon_{is} \\ I\!I_v^{(6)} &= \frac{1}{n}\sum_{its}X_{it}^\top[\widehat{m}(Z_{it}; H_1) - m(Z_{it})]X_{is}^\top[\widehat{m}(Z_{is}; H_1) - m(Z_{is})].\end{aligned}$$

As we are going to show, of these six terms only the first one will be a leading term. Analyzing each term separately and using uniform convergence results as the ones established in Theorem 6 of Masry (1996), by Assumptions 2.1-2.3 and rearranging terms,

$$\begin{aligned}I\!I_v^{(2)} &\leq \frac{1}{NT}\sum_{it}|X_{it}^\top X_{it}|\sup_{\{Z_{it}\}}|\left(\widehat{m}(Z_{it}; H_1) - m(Z_{it})\right)^\top||\sup_{\{Z_{it}\in\mathcal{Z}\}}|\widehat{m}(Z_{it}; H_1) - m(Z_{it})| \\ &= O_{\mathbb{P}}\left(\frac{lnN}{N|H_1|}\right) + O_{\mathbb{P}}(tr(H_1)^2) \qquad (6.29)\end{aligned}$$

since we can show $(NT)^{-1}\sum_{it}|X_{it}X_{it}^\top| = O_{\mathbb{P}}(1)$. A similar result holds for $I\!I_v^{(6)}$.

Also, and under the same reasoning as the above, it is straightforward to show that

$$\begin{aligned}I\!I_v^{(3)} &\leq \frac{2}{NT}\sum_{it}|\epsilon_{it}X_{it}^\top|\sup_{\{Z_{it}\in\mathcal{Z}\}}|\widehat{m}(Z_{it}; H_1) - m(Z_{it})| \\ &= o_{\mathbb{P}}\left(\frac{\sqrt{lnN}}{\sqrt{N|H_1|}}\right) + o_{\mathbb{P}}(tr(H_1)), \qquad (6.30)\end{aligned}$$

given that $(NT)^{-1}\sum_{it}|\epsilon_{it}X_{it}^\top| = O_{\mathbb{P}}(1)$. Following this same procedure we get a similar result for $I\!I_v^{(4)}$ and $I\!I_v^{(5)}$.

Using these results in (6.28),

$$\widehat{\sigma}_v^2 - \sigma_v^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(v_{it}^2 - \sigma_v^2\right) + O_{\mathbb{P}}\left(\frac{\sqrt{lnN}}{\sqrt{N|H_1|}}\right) + O_{\mathbb{P}}(tr(H_1))$$

and under Assumption 2.6 this previous equation can be written as

$$\sqrt{NT}\left(\widehat{\sigma}_v^2 - \sigma_v^2\right) = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(v_{it}^2 - \sigma_v^2\right) + o_{\mathbb{P}}(1). \tag{6.31}$$

Finally, by the central limit theorem it can be proved that, as N goes to infinity and T is fixed,

$$\sqrt{N}\left(\widehat{\sigma}_v^2 - \sigma_v^2\right) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, \gamma_v^4 - (\sigma_v^2)^2\right) \tag{6.32}$$

and the first part of Theorem 4.1 is done.

Focusing now on $\widehat{\sigma}_b^2$, (4.2) can be rewritten as

$$\widehat{\sigma}_b^2 \;=\; \frac{1}{NT(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s\neq t}\epsilon_{it}\epsilon_{is} - \mathbf{I}_b^{(1)} - \mathbf{I}_b^{(2)} + \mathbf{I}_b^{(3)}, \tag{6.33}$$

where

$$\begin{aligned}
\mathbf{I}_b^{(1)} &= \frac{1}{n}\sum_{it}\sum_{s\neq t}\epsilon_{it}X_{is}^{\top}[\widehat{m}(Z_{is};H_1) - m(Z_{is})],\\
\mathbf{I}_b^{(2)} &= \frac{1}{n}\sum_{it}\sum_{s\neq t}\epsilon_{is}X_{it}^{\top}[\widehat{m}(Z_{it};H_1) - m(Z_{it)}],\\
\mathbf{I}_b^{(3)} &= \frac{1}{n}\sum_{it}\sum_{s\neq t}X_{it}^{\top}[\widehat{m}(Z_{it};H_1) - m(Z_{it})]X_{is}^{\top}[\widehat{m}(Z_{is};H_1) - m(Z_{is})].
\end{aligned}$$

Using similar arguments as those above, we can show that as N goes to infinity and T is fixed,

$$\widehat{\sigma}_b^2 - \sigma_b^2 = \frac{1}{NT(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s\neq t}^{T}\left(\epsilon_{it}\epsilon_{is} - \sigma_b^2\right) + O_{\mathbb{P}}\left(\frac{\sqrt{lnN}}{\sqrt{N|H_1|}}\right) + O_{\mathbb{P}}(tr(H_1)).$$

Finally, under Assumption 2.6 and by the central limit theorem it is straightforward to show

$$\sqrt{N}\left(\widehat{\sigma}_b^2 - \sigma_b^2\right) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, \gamma_b^4 - (\sigma_b^2)^2 + \left(\frac{4T-1}{T(T-1)}\right)\sigma_b^2\sigma_v^2 + \frac{2}{T(T-1)}(\sigma_v^2)^2\right). \tag{6.34}$$

∎

References

Ahmad, I., S. Leelahanon, and Q. Li (2005). Efficient estimation of a semiparametric partially linear varying coefficient model. *The Annals of Statistics 33*, 258–283.

Breusch, T. S. and A. . R. Pagan (1980). The Lagrange multiplier test and its application to model specification in econometrics. *Review of Economics Studies 47*, 239–253.

Chamberlain, G. (1992). Efficiency bounds for semiparametric regression. *Econometrica 60*, 567–596.

Chen, J., D. Li, and J. Gao (2013). Non- and semi-parametric panel data models: a selective review. Working Paper 18/13 Monash University.

Cox, D. R. and P. Hall (2002). Estimation in a simple random effecs model with nonnormal distributions. *Biometrika 89*, 831–840.

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

Henderson, D. J. and A. Ullah (2005). A nonparametric random effects estimator. *Economics Letters 88*, 403–407.

Honda, Y. (1985). Testing the error components model with nonnormal disturbances. *Review of Economic Studies 52*, 71–79.

Honoré, B. E. and J. L. Powell (2005). Pairwise difference estimation of nonlinear models. In D. W. K. Andrews and J. H. Stock (Eds.), *Identification and Inference for Econometric Models. Essays in Honor of Thomas Rothenberg*, pp. 520–553. Cambridge University Press, Cambridge.

Lin, X. and R. J. Carroll (2000). Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of the American Statistical Association 95*, 520–534.

Masry, E. (1996). Multivariate local polynomial regression for time series: uniform strong consistency rate. *Journal of Time Series Analysis 17*, 571–599.

Moulton, B. R. and W. C. Randolph (1989). Alternative tests of the error components model. *Econometrica 57*, 685–693.

Rodriguez-Poo, J. M. and A. Soberon (2015). Nonparametric estimation of fixed effects panel data varying coefficient models. *Journal of Multivariate Analysis 133*, 95–122.

Ruckstuhl, A. F., A. H. Welsh, and R. J. Carroll (2000). Nonparametric function estimation of the relationship between two repeatedly measured variables. *Statistica Sinica 10*, 51–71.

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

Stromberg, A. J., O. Hossjer, and D. M. Hawkins (2000). The least trimmed differences regression estimator and alternatives. *Journal of the American Statistical Association 95*(451), 853–864.

Su, L. and A. Ullah (2007). More efficient estimation of nonparametric panel data models with random effects. *Economics Letters 96*, 375–380.

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.

Wooldridge, J. M. (2003). *Econometric analysis of cross section and panel data models*. The MIT Press, Cambridge, Massachusets.

Wu, P., W. Stute, and L. Zhu (2012). Efficient estimation of moments in linear mixed models. *Bernoulli 18*, 206–228.

Zhou, B., J. You, Q. Xu, and G. Chen (2010). Weighted profile least squares estimation for a panel data varying-coefficient partially linear model. *Chinese Annals of Mathematics 31B*(2), 247–272.