

Testing for distributional features in varying coefficient panel data models

Alexandra Soberon¹, Winfried Stute² and Juan M. Rodriguez-Poo³

¹Department of Economics, University of Cantabria, Avda. Los Castros s/n

²Mathematical Institute, University of Giessen, Arndtstr. 2, D-35392 Giessen, Germany

³Department of Economics, University of Cantabria and University of Genève

October 7, 2016

Abstract

This paper provides several tests for skewness, kurtosis, and normality for the heterogeneity and the idiosyncratic error terms in a one-way component varying coefficient panel data model. However, the methodology is rather general and it can be applied to other models of interest. Using nonparametric residuals, the test statistics are derived in a moment condition framework. As in order to obtain these test statistics calculation of higher order moments are needed, we also propose an alternative technique to estimate them. Finally, and as a by-product, to obtain the nonparametric residuals we propose a local constant estimator that is based on a pairwise differencing transformation of the original model. It turns out that this estimator, under some standard assumptions, exhibits the same properties than others in the field, but its calculation is much simpler and it does not require iterative steps. The proposed estimators and test statistics are augmented by simulation studies and they are also illustrated in an empirical analysis about the technical efficiency of the European Union companies.^[1]

Keywords: asymptotic normality, moment estimator, pairwise difference, longitudinal data, skewness, kurtosis, normality.

^[1]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. Stute’s work was partly done while he was on leave at BCAM, the Basque Center of Applied Mathematics in Bilbao.

1 Introduction

This paper is concerned with the nonparametric estimation and testing of panel data varying coefficient models when standard distributional assumptions such as normality or symmetry cannot be justified. In the last decade, this type of models has been subjected to an intensive research for 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”. Third, they have been justified on the grounds of the economic theory, see Chamberlain (1992). However, although several authors have already investigated the problem of defining nonparametric panel data estimators, tests for normality are not exploited. See Su and Ullah (2011), Chen et al. (2013) and Rodriguez-Poo and Soberon (2016) for some surveys.

As it is well-known, the normality assumption plays a crucial role in the validity of several inferential techniques and therefore, testing for skewness and kurtosis is a relevant topic in many fields in economics such as in the financial economics literature, see for example Bai and Ng (2005), Dufour et al. (2003), Jarque and Bera (1981), Montes-Rojas and Sosa-Escudero (2011) and Premaratne and Bera (2005). In addition, in those cases in which normality is corroborated, more asymptotically efficient estimators can be proposed such as local maximum likelihood estimators. Nevertheless, although the tests above are able to detect departures away from normality in the form of skewness and kurtosis for cross-sectional or time series data, the results for panel data models are scarce. In fact, a natural complication in longitudinal data is that the lack of normality can arise from any of the two components of the composed error term and the previous tests are unable to identify which element departures from the Gaussian distribution. In a fully parametric context, in Horowitz and Markatou (1996) a technique to estimate nonparametrically the densities of the error components is proposed. Unfortunately, they do not provide a formal method to test for a pre-specified parametric density function. Also, in a fully parametric context, in Galvao et al. (2013) tests for skewness and kurtosis are derived. Nevertheless, the distribution of both coefficients is unknown and they need to approximate it through a bootstrap technique.

In this situation, the aim of this paper is twofold. On the one hand, to develop some skewness, kurtosis, and normality tests for both components of the error, jointly and separately. The test statistics are derived in a moment condition framework. Furthermore, to approximate the asymptotic distribution of these test statistics, in order to avoid bootstrap techniques, estimation of the higher order moments of the error components are needed. Then, by extending the contributions in Cox and Hall (2002) and Wu et al. (2012) we propose a technique to estimate these moments in a nonparametric panel data setting. To the best of our knowledge, this is a completely new proposal. On the other hand, since the estimators of higher order moments require the previous estimation of the varying coefficient functions, in this paper we also propose a new estimation technique based on a pairwise differencing transformation. The

interesting feature of the resulting nonparametric estimator is that it achieves nearly optimal rates of convergence without having to resort to iterative procedures such as those proposed in Wang (2003), Henderson et al. (2008), Qian and Wang (2012), and Rodriguez-Poo and Soberon (2015), among others.

As a by-product of the previous contributions, note that estimators of higher order moments are obtained in an easy-to-compute closed form expressions that can be applied to several fields such as non-linear, semi-parametric or nonparametric panel data models. Also, the asymptotic distribution of the different test statistics under the null hypothesis, the consistency of higher order moment estimators, and the nearly optimal asymptotic behavior of the varying coefficient estimators are all obtained without no further distributional assumptions on the error components. Furthermore, correlation between the unobserved individual effects and some/all the covariates of the model is allowed for. We would like to emphasize that our proposal to estimate higher order moments and the battery of tests could be also based on residuals obtained from a fully parametric model. However, for the sake of generality, we decided to work in the nonparametric setting. Needless to say that all results in this paper are straightforwardly applied to fully parametric panel data models.

Finally, to illustrate the usefulness of the results of this paper a Monte Carlo study is conducted to assess the finite sample performance of the proposed estimators and test statistics. Later, an empirical study about the production efficiency of the European Union (EU) companies is implemented. Note that the tests proposed in this paper can be also very useful for several fields such as the risk literature in which they are usually interested in the higher order moments.

The rest of the paper is organized as follows: In Section 2 we set up the model of interest and introduce the estimation of the moments of the error terms and the pairwise difference estimator. In Section 3 we study their asymptotic distributions. In Section 4 we derive some tests for skewness, kurtosis, and normality and study their asymptotic distributions. In Section 5 we apply our results to a production efficiency study and compare the estimators and test statistics considered via Monte Carlo experiments. Section 6 provides a summary of the paper. The proofs of the main results are collected in the Appendix.

2 Statistical model and estimation procedure

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

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

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 Y_{it} and X_{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} . Note that in this paper we are not willing to impose any condition on the statistical relationship between b_i and the covariates of the model.

As a first step to propose some test statistics for symmetry and kurtosis, root-N estimators for higher order moments of v_{it} and b_i are derived. 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

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

that can be approximated through the following Lemma.

Lemma 2.1 *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 $f_j^k(i)$ will vanish when we take expectations. Thus, it provides a set of estimating equations, each of them leading to consistent estimators. In order to obtain the efficient estimator for

$$\sigma_v^2 = E(v_{it}^2), \quad \sigma_b^2 = E(b_i^2), \quad \gamma_v^k = E(v_{it}^k) \quad \text{and} \quad \gamma_b^k = E(b_i^k),$$

finding a suitable combination of these polynomial functions of the residuals is essential when $k = 3, 4, \dots, 8$.

If we focus our attention on the second moments of the idiosyncratic error term, σ_v^2 , from Lemma 2.1, we obtain

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

and σ_v^2 can be represented in terms of the f 's, i.e.,

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

As the reader can notice, this equation does not incorporate b_i , so it may serve as a basis for the estimation of σ_v^2 . Then, averaging over $1 \leq i \leq N$ and replacing the unknown ϵ_{it} with the residuals $\hat{\epsilon}_{it}$, the estimator of σ_v^2 has the form

$$\hat{\sigma}_v^2 = \frac{1}{NT(T-1)} \sum_{i=1}^N \left[T \sum_{t=1}^T \hat{\epsilon}_{it}^2 - \left(\sum_{t=1}^T \hat{\epsilon}_{it} \right)^2 \right], \quad (2.2)$$

where $\hat{\epsilon}_{it} = Y_{it} - X_{it}^\top \hat{m}(Z_{it}; H)$ and $\hat{m}(Z_{it}; H)$ is a nonparametric estimator that will be defined in the next section.

Similarly, combining these expressions as

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

we can propose an estimator for σ_b^2

$$\hat{\sigma}_b^2 = \frac{1}{NT(T-1)} \sum_{i=1}^N \left[\left(\sum_{t=1}^T \hat{\epsilon}_{it} \right)^2 - \sum_{t=1}^T \hat{\epsilon}_{it}^2 \right]. \quad (2.3)$$

As pointed out in Wu et al. (2012), this lemma enables us to obtain estimators for the second order moment without having to impose distributional assumptions on b_i or v_{it} . However, for γ_b^k and γ_v^k things are more complex since there are several $f_j^k(i)$ to combine and some combinations can lead to inefficient estimators in terms of the variance. In the next section, we present the suitable combination that provides the estimators of the higher order moments with the minimal variance.

2.1 Estimation of the higher order moments

Focusing on the third and fourth order moments for γ_v^k and γ_b^k , from Lemma 2.1, we can write

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

from which the estimator for γ_v^3 has the form

$$\hat{\gamma}_v^3 = \frac{1}{NT(T-1)(T-2)} \sum_{i=1}^N \left[2 \left(\sum_{t=1}^T \hat{\epsilon}_{it} \right)^3 + T^2 \sum_{t=1}^T \hat{\epsilon}_{it}^3 - 3T \left(\sum_{t=1}^T \hat{\epsilon}_{it}^2 \right) \sum_{t=1}^T \hat{\epsilon}_{it} \right]. \quad (2.4)$$

Similarly, for γ_b^3 we obtain 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 for γ_b^3 is

$$\hat{\gamma}_b^3 = \frac{1}{NT(T-1)(T-2)} \sum_{i=1}^N \left[\left(\sum_{t=1}^T \hat{\epsilon}_{it} \right)^3 - 3 \left(\sum_{t=1}^T \hat{\epsilon}_{it}^2 \right) \sum_{t=1}^T \hat{\epsilon}_{it} + 2 \sum_{t=1}^T \hat{\epsilon}_{it}^3 \right]. \quad (2.5)$$

Finally, from Lemma 2.1 we obtain several polynomial functions to combine, i.e. $f_j^4(i)$, for $1 \leq j \leq 4$. Then, there are a great deal of possible combinations that lead to consistent estimators, but some of them can be inefficient. However, following Wu et al. (2012), the proper combination that provides the efficient estimator for γ_v^4 is

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

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

$$\begin{aligned} \hat{\gamma}_v^4 &= \frac{1}{NT(T-1)(T-2)(T-3)} \sum_{i=1}^N \left[(T^2 - 2T + 3) \left(T \sum_{t=1}^T \hat{\epsilon}_{it}^4 - 4 \sum_{t=1}^T \hat{\epsilon}_{it}^3 \sum_{t=1}^T \hat{\epsilon}_{it} \right) \right. \\ &\quad \left. + 6T \sum_{t=1}^T \hat{\epsilon}_{it}^2 \left(\sum_{t=1}^T \hat{\epsilon}_{it} \right)^2 - 3 \left(\sum_{t=1}^T \hat{\epsilon}_{it} \right)^4 - 3(2T - 3) \left(\sum_{t=1}^T \hat{\epsilon}_{it}^2 \right)^2 \right]. \end{aligned} \quad (2.6)$$

Meanwhile, for 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 resulting estimator for γ_b^4 is of the form

$$\begin{aligned} \hat{\gamma}_b^4 &= \frac{1}{NT(T-1)(T-2)(T-3)} \sum_{i=1}^N \left[\left(\sum_{t=1}^T \hat{\epsilon}_{it} \right)^4 - 6 \sum_{t=1}^T \hat{\epsilon}_{it}^2 \left(\sum_{t=1}^T \hat{\epsilon}_{it} \right)^2 + 8 \sum_{t=1}^T \hat{\epsilon}_{it}^3 \sum_{t=1}^T \hat{\epsilon}_{it} \right. \\ &\quad \left. - 6 \sum_{t=1}^T \hat{\epsilon}_{it}^4 + 3 \left(\sum_{t=1}^T \hat{\epsilon}_{it}^2 \right)^2 \right]. \end{aligned} \quad (2.7)$$

Note that all these estimators are based on the residuals of a varying coefficient nonparametric panel data model, i.e. $\hat{\epsilon}_{it} = Y_{it} - X_{it}^\top \hat{m}(Z_{it}; H)$. As an estimator for $m(\cdot)$, in the next section we propose a local constant estimator based on pairwise differences.

2.2 Pairwise difference estimation

As it is well-known, differencing techniques are usually used to remove the unobserved individual heterogeneity from the regression model to estimate. However, the transformed regression model appears as an additive function and iterative techniques such as marginal integration or backfitting are needed, see Wang (2003), Henderson et al. (2008), Qian and Wang (2012), and Rodriguez-Poo and Soberon (2015), among others. To avoid this situation, in this section a pairwise differencing transformation is proposed.

Note that this differencing technique is specially appealing because of two reasons. First, it removes the individual effects from the regression model to estimate. Second, it enables us to

obtain some efficiency gains concerning to the estimator because this transformation considers all time-dependences within the observations of each individual.

Inspired by Stromberg et al. (2000) and Honoré and Powell (2005), the pairwise differencing transformation implies subtracting from time t of (2.1) time s , for $s \neq t$, yielding

$$Y_{it} - Y_{is} = X_{it}^\top m(Z_{it}) - X_{is}^\top m(Z_{is}) + v_{it} - v_{is}, \quad i = 1, \dots, N; \quad t, s = 1, \dots, T, \quad s \neq t. \quad (2.8)$$

Direct estimation of $m(\cdot)$ in (2.8) through any standard nonparametric technique ends up in a non-negligible asymptotic bias, see for example Rodriguez-Poo and Soberon (2015) for further details. In order to overcome it, we define a kernel weight that controls the distance between any Z_{it}, Z_{is} .

To illustrate this technique we start by the simplest case ($d = q = 1$). In this situation, the proposed technique to estimate $m(\cdot)$ and its derivatives is 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 \mathbb{R} , obtaining

$$\begin{aligned} X_{it}m(Z_{it}) - X_{is}m(Z_{is}) &\approx (X_{it} - X_{is})m(z) + [X_{it}(Z_{it} - z) - X_{is}(Z_{is} - z)]m'(z) \\ &+ \frac{1}{2} [X_{it}(Z_{it} - z)^2 - X_{is}(Z_{is} - z)^2]m''(z) + \dots \\ &+ \frac{1}{p!} [X_{it}(Z_{it} - z)^p - X_{is}(Z_{is} - z)^p]m^{(p)}(z) \\ &\equiv \sum_{\lambda=0}^p \beta_\lambda [X_{it}(Z_{it} - z)^\lambda - X_{is}(Z_{is} - z)^\lambda]. \end{aligned} \quad (2.9)$$

That suggests that we estimate $m(z), m'(z), \dots, m^{(p)}(z)$ by regressing $Y_{it} - Y_{is}$ on the terms of the right-hand side of (2.9) with kernel weights.

Focusing on the particular case $p = 0$ and defining $\tilde{Y}_{its} = Y_{it} - Y_{is}$ and $\tilde{X}_{its} = X_{it} - X_{is}$, the unknown $\beta = m(z)$ can be estimated by minimizing the following criterion function,

$$\sum_{i=1}^N \sum_{t=1}^{T-1} \sum_{s=t+1}^T \left(\tilde{Y}_{its} - \tilde{X}_{its}^\top \beta \right)^2 K_H(Z_{it} - z) K_H(Z_{is} - z), \quad (2.10)$$

see Fan and Gijbels (1995) and Ruppert and Wand (1994), where H is a $q \times q$ symmetric positive-definite bandwidth matrix and, for each u , K a multivariate kernel such as

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

Let $\hat{\beta}$ be the minimizer of (2.10). It is equal to

$$\hat{m}(z; H) = S_{\tilde{X}\tilde{X}}^{-1}(z) S_{\tilde{X}\tilde{Y}}(z), \quad (2.11)$$

where

$$S_{\tilde{X}\tilde{X}}(z) = \left(\frac{T}{2} \right)^{-1} \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^{T-1} \sum_{s=t+1}^T K_H(Z_{it} - z) K_H(Z_{is} - z) \tilde{X}_{its} \tilde{X}_{its}^\top,$$

and

$$S_{\tilde{X}\tilde{Y}}(z) = \binom{T}{2}^{-1} \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^{T-1} \sum_{s=t+1}^T K_H(Z_{it} - z) K_H(Z_{is} - z) \tilde{X}_{its} \tilde{Y}_{its}.$$

Following this technique, it is straightforward to provide a local linear estimator for $m(\cdot)$. However, we believe that the local constant estimator is enough to obtain residuals with good properties for the estimation of the higher order moments.

3 Asymptotic properties

In order to show the asymptotic properties of the estimators proposed in the previous section, let us consider some assumptions.

Assumption 3.1 *Let $\{(Y_{it}, Z_{it}, X_{it})\}$ be a set of independent and identically distributed (i.i.d.) \mathbb{R}^{1+q+d} -random variables in the subscript i for each fixed t , and strictly stationary over t for fixed i .*

Assumption 3.2 *The random errors v_{it} are i.i.d. with zero mean and finite variance, $\sigma_v^2 < \infty$, and are independent of X_{it} and Z_{it} for all i and t . The individual effect b_i is i.i.d. with zero mean and finite variance, $\sigma_b^2 < \infty$. Also, b_i is independent of all v_{it} .*

Assumption 3.3 *Let $f_{Z_1}(\cdot)$ be the probability density function of Z_1 . All density functions are continuously differentiable in all their arguments and they are bounded away from zero in any point of their support.*

Assumption 3.4 *Let $\|A\| = \sqrt{\text{tr}(A^\top A)}$, $E \left[\|\tilde{X}_{its} \tilde{X}_{its}^\top\|^2 | Z_{it} = z_1, Z_{is} = z_2 \right]$ is bounded and uniformly continuous in its support. In addition, the matrix function $E \left[\tilde{X}_{its} \tilde{X}_{its}^\top | Z_{it} = z_1, Z_{is} = z_2 \right]$ is bounded and uniformly continuous in its support, \mathcal{Z} .*

Assumption 3.5 *The matrix $E \left[\tilde{X}_{its} \tilde{X}_{its}^\top | Z_{it} = z_1, Z_{is} = z_2 \right]$ is positive definite for any interior point of (z_1, z_2) in the support of $f_{Z_{it}, Z_{is}}(z_1, z_2)$.*

Assumption 3.6 *Let z be an interior point in the support of f_{Z_1} . All second-order derivatives of $m_1(\cdot), m_2(\cdot), \dots, m_d(\cdot)$ are bounded and uniformly continuous.*

Assumption 3.7 *The q -variate kernel functions K are compactly supported and bounded such that $\int uu^\top K(u)du = \mu_2(K)I_q$ 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 $\int u_1^{\iota_1} \dots u_q^{\iota_q} K(u)du = 0$, for all nonnegative integers ι_1, \dots, ι_q such that their sum is odd.*

Assumption 3.8 *The bandwidth matrix H is symmetric and strictly definite positive. Also, as $N \rightarrow \infty$ each entry of the matrix tends to zero in such a way that $N|H| \rightarrow \infty$.*

Assumption 3.9 *For some $\delta > 0$, the following function $E [|X_{it}v_{it}|^{2+\delta} | Z_{it} = z_1, Z_{is} = z_2]$ is bounded and uniformly continuous in any point of their support.*

Assumptions 3.1 and 3.2 are standard in panel data analysis although other time-dependence settings could be considered. However, since the asymptotic properties of the proposed estimator are analyzed for panels with large cross-section and fixed time-series dimension, it is enough to assume strict stationarity. Assumptions 3.3, 3.4 and 3.6 are basically smoothness and boundedness conditions on the density function and moments functionals. Assumption 3.5 is a generalization of the well-known rank condition of parametric models that guarantees that $m(\cdot)$ is identified. Also, Assumptions 3.7 and 3.8 are standard in the literature of the local linear regression for the kernel function and bandwidth matrix. Finally, Assumption 3.9 guarantees that a multivariate version of the Lindeberg-Lévy central limit theorem for $N \rightarrow \infty$ and fixed T can be used to establish the asymptotic normality of this estimator.

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

Theorem 3.1 *Under Assumptions 3.1-3.9, as N tends to infinity and T is fixed*

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

where

$$\begin{aligned} B(z; H) &= \mu_2(K) \left[\text{diag}_d(\text{tr}(HD_f(z)D_{m_r}(z))) \iota_d f_{Z_{it}, Z_{is}}^{-1}(z, z) + \frac{1}{2} \text{diag}_d(\text{tr}(H\mathcal{H}_{m_r}(z))) \iota_d \right] \\ V(z; H) &= \begin{pmatrix} T \\ 2 \end{pmatrix}^{-1} 2\sigma_v^2 R^2(K) \mathcal{B}_{\widetilde{X}\widetilde{X}}^{-1}(z, z), \end{aligned}$$

for $r = 1, \dots, d$, D_{m_r} is the first-order derivative vector of the r th 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

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

In addition, $\text{diag}_d(\text{tr}(H\mathcal{H}_{m_r}(z)))$ and $\text{diag}_d(\text{tr}(HD_f(z)D_{m_r}(z)))$ stand for a diagonal matrix of elements of $\text{tr}(H\mathcal{H}_{m_r}(z))$ and $\text{tr}(HD_f(z)D_{m_r}(z))$, respectively, being ι_d a $d \times 1$ unit vector.

The proof of this theorem is postponed to the Appendix. Furthermore, we also need the following result.

Theorem 3.2 *Under Assumptions 3.1-3.9, as N tends to infinity and T is fixed*

$$\sup_{z \in A} \|\widehat{m}(z; H) - m(z)\| = O_{\mathbb{P}} \left(\text{tr}(H) + \left(\frac{\log N}{N|H|} \right)^{1/2} \right).$$

The proof of this result follows the same lines as in Theorem 8 in Hansen (2008), so it is omitted.

The results shown in Theorem 3.1 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 is large and when there is sparser data near z . In addition, a useful feature of this estimation scheme is its computational simplicity. In one step it is possible to obtain a nonparametric estimator which almost achieves the optimal rate of convergence of this type of problems, i.e. $N|H|^{1/2}$.

Focusing now on the asymptotic properties of the estimators of the higher order moments of the error components, we also need the following assumption.

Assumption 3.10 *For fixed i , there is a $\delta > 0$ such that $E\|X_{it}^k \epsilon_{it}\|^{2+\delta} < \infty$, for $k = 2, 3, \dots, 8$.*

Assumption 3.10 is needed to bound high order moments relates to the residuals obtained in a nonparametric framework. Specifically, the main implication of these assumptions is that inference on disturbances can be performed using these residuals since the estimation of $m(\cdot)$ does not affect the limiting distributions of the skewness coefficient, kurtosis, and normality tests as $N \rightarrow \infty$ and T fixed.

The next theorem contains the main statistical properties of the variance estimators of both random error and unobserved individual heterogeneity.

Theorem 3.3 *Under Assumptions 3.1-3.10, when both $\gamma_v^4 = E(v_{it}^4)$ and $\gamma_b^4 = E(b_i^4)$ are finite, as $N \rightarrow \infty$ and T is fixed we have*

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

and

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

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

The proof of this result is shown in the Appendix.

Finally, the following theorem contains the main asymptotic properties of the estimators for the third and fourth order moments of both b_i and v_{it} .

Theorem 3.4 *Under Assumptions 3.1-3.10, when $\gamma_v^6 = E(v_{it}^6)$, $\gamma_b^6 = E(b_i^6)$, $\gamma_v^8 = E(v_{it}^8)$ and $\gamma_b^8 = E(b_i^8)$ are finite, as $N \rightarrow \infty$ and T is fixed,*

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

and

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

where

$$\begin{aligned}
\mu_b^3 &= \gamma_b^6 - (\gamma_b^3)^2 + \frac{9}{T}\gamma_b^4\sigma_v^2 + \frac{18\sigma_b^2\sigma_v^4}{T(T-1)} + \frac{6\sigma_v^6}{T(T-1)(T-2)}, \\
\mu_v^3 &= \gamma_v^6 - (\gamma_v^3)^2 + \left(\frac{9}{(T-1)} - 6\right)\gamma_v^4\sigma_v^2 + \frac{9(\gamma_v^3)^2}{(T-1)} + \left(\frac{24}{(T-1)(T-2)} + 9\right)\sigma_v^6, \\
\mu_b^4 &= \gamma_b^8 - (\gamma_b^4)^2 + \frac{16}{T}\gamma_b^6\sigma_v^2 + \frac{72\gamma_b^4\sigma_v^4}{T(T-1)} + \frac{96\sigma_b^2\sigma_v^6}{T(T-1)(T-2)} + \frac{24\sigma_v^8}{T(T-1)(T-2)(T-3)}, \\
\mu_v^4 &= \gamma_v^8 - (\gamma_v^4)^2 - 8\gamma_v^5\gamma_v^3 + \frac{16}{(T-1)}\gamma_v^6\sigma_v^2 + \frac{16}{(T-1)}(\gamma_v^4)^2 + \left(\frac{72}{(T-1)(T-2)} + 16(T-2)\right)(\gamma_v^3)^2\sigma_v^2 \\
&\quad + \left(\frac{72}{(T-1)(T-2)} - 96\right)\gamma_v^4\sigma_v^4 + \left(\frac{216}{(T-1)(T-2)(T-3)} + \frac{72(T-3)}{(T-1)(T-2)}\right)\sigma_v^8.
\end{aligned}$$

The proof of this result is shown in the Appendix.

Looking at the results of Theorem 3.3 and 3.4, it can be pointed out that the estimators proposed here for the higher order moments are asymptotically normal. In addition, given that μ_b^j and μ_v^j , for $j = 2, 3, 4$, denote the minimum variances when N goes to infinity and T is fixed, they can be achieved for empirical estimators based on the true b_i and v_{it} , respectively.

Finally, in order to estimate the asymptotic variances in Theorem 3.3 and 3.4 consistent estimators for the fifth, sixth and eighth order moments of v_{it} and b_i have to be computed. In particular, the following estimators for these terms can be proposed using Lemma 2.1.

$$\begin{aligned}
\hat{\gamma}_v^5 &= \frac{1}{NT(T-1)(T-2)} \sum_{i=1}^N \left[T^2 \sum_{t=1}^T \hat{\epsilon}_{it}^5 - 3T \sum_{t=1}^T \hat{\epsilon}_{it}^4 \left(\sum_{s=1}^T \hat{\epsilon}_{is} \right) + 2 \sum_{t=1}^T \hat{\epsilon}_{it}^3 \left(\sum_{t=1}^T \hat{\epsilon}_{it} \right)^2 \right] \\
&\quad - \frac{1}{(T-2)} (2\hat{\gamma}_v^3\hat{\sigma}_v^2 - 6\hat{\gamma}_v^3\hat{\sigma}_b^2 - 2T\hat{\gamma}_b^3\hat{\sigma}_v^2), \tag{3.1}
\end{aligned}$$

$$\begin{aligned}
\hat{\gamma}_b^5 &= \frac{1}{NT(T-1)(T-2)} \sum_{i=1}^N \left[\sum_{t=1}^T \hat{\epsilon}_{it}^3 \left(\sum_{t=1}^T \hat{\epsilon}_{it} \right)^2 - 3 \sum_{t=1}^T \hat{\epsilon}_{it}^4 \left(\sum_{s=1}^T \hat{\epsilon}_{is} \right) + 2 \sum_{t=1}^T \hat{\epsilon}_{it}^5 \right] \\
&\quad - \frac{1}{(T-2)} [\hat{\gamma}_v^3\hat{\sigma}_v^2 + (T-5)\hat{\sigma}_b^2\hat{\gamma}_v^3 + (3T-8)\hat{\gamma}_b^3\hat{\sigma}_v^2], \tag{3.2}
\end{aligned}$$

$$\begin{aligned}
\hat{\gamma}_v^6 &= \frac{1}{NT(T-1)(T-2)} \sum_{i=1}^N \left[T^2 \sum_{t=1}^T \hat{\epsilon}_{it}^6 - 3T \sum_{t=1}^T \hat{\epsilon}_{it}^5 \left(\sum_{t=1}^T \hat{\epsilon}_{it} \right) + 2 \sum_{t=1}^T \hat{\epsilon}_{it}^4 \left(\sum_{t=1}^T \hat{\epsilon}_{it} \right)^2 \right] \\
&\quad - \frac{1}{(T-2)} [2\hat{\gamma}_v^4\hat{\sigma}_v^2 + 12\hat{\sigma}_b^2(\hat{\sigma}_v^2)^2 + 2(T-6)\hat{\sigma}_b^2\hat{\gamma}_v^4 - 2(T+4)\hat{\gamma}_b^3\hat{\gamma}_v^3 - 3T\hat{\gamma}_b^4\hat{\sigma}_v^2], \tag{3.3}
\end{aligned}$$

$$\begin{aligned}
\hat{\gamma}_b^6 &= \frac{1}{NT(T-1)(T-2)} \sum_{i=1}^N \left[\sum_{t=1}^T \hat{\epsilon}_{it}^4 \left(\sum_{t=1}^T \hat{\epsilon}_{it} \right)^2 - 3 \sum_{t=1}^T \hat{\epsilon}_{it}^5 \left(\sum_{t=1}^T \hat{\epsilon}_{it} \right) + 2 \sum_{t=1}^T \hat{\epsilon}_{it}^6 \right] \\
&\quad - \frac{1}{(T-2)} [\hat{\gamma}_v^4\hat{\sigma}_v^2 + 6\hat{\sigma}_b^2(\hat{\sigma}_v^2)^2 + (T-6)\hat{\sigma}_b^2\hat{\gamma}_v^4 + 2(2T-7)\hat{\gamma}_v^3\hat{\gamma}_b^3 + 3(2T-5)\hat{\gamma}_b^4\hat{\sigma}_v^2], \tag{3.4}
\end{aligned}$$

$$\begin{aligned}
\widehat{\gamma}_v^8 &= \frac{1}{NT(T-1)(T-2)} \sum_{i=1}^N \left[T^2 \sum_{t=1}^T \widehat{\epsilon}_{it}^8 - 3T \sum_{t=1}^T \widehat{\epsilon}_{it}^7 \left(\sum_{s=1}^T \widehat{\epsilon}_{is} \right) + 2 \sum_{t=1}^T \widehat{\epsilon}_{it}^6 \left(\sum_{t=1}^T \widehat{\epsilon}_{it} \right)^2 \right] \\
&- \frac{1}{(T-2)} [2\widehat{\gamma}_v^6 \widehat{\sigma}_v^2 + 40\widehat{\gamma}_b^3 \widehat{\gamma}_v^3 \widehat{\sigma}_v^2 + (9T-30)\widehat{\sigma}_b^2 \widehat{\gamma}_v^6 + 30\widehat{\sigma}_b^2 \widehat{\sigma}_v^2 \widehat{\gamma}_v^4 + (5T-40)\widehat{\gamma}_b^3 \widehat{\gamma}_v^5 \\
&- (5T+30)\widehat{\gamma}_b^4 \widehat{\gamma}_v^4 - (9T+12)\widehat{\gamma}_b^5 \widehat{\gamma}_v^3 + 30\widehat{\gamma}_b^4 (\widehat{\sigma}_v^2)^2 - 5T\widehat{\gamma}_b^6 \widehat{\sigma}_v^2], \tag{3.5}
\end{aligned}$$

$$\begin{aligned}
\widehat{\gamma}_b^8 &= \frac{1}{NT(T-1)(T-2)} \sum_{i=1}^N \left[\sum_{t=1}^T \widehat{\epsilon}_{it}^6 \left(\sum_{t=1}^T \widehat{\epsilon}_{it} \right)^2 - 3 \sum_{t=1}^T \widehat{\epsilon}_{it}^7 \left(\sum_{t=1}^T \widehat{\epsilon}_{it} \right) + 2 \sum_{t=1}^T \widehat{\epsilon}_{it}^8 \right] \\
&- \frac{1}{(T-2)} [\widehat{\gamma}_v^6 \widehat{\sigma}_v^2 + 15\widehat{\sigma}_b^2 \widehat{\sigma}_v^2 \widehat{\gamma}_v^4 + (T-8)\widehat{\sigma}_b^2 \widehat{\gamma}_v^6 + 20\widehat{\gamma}_b^3 \widehat{\gamma}_v^3 \widehat{\sigma}_v^2 + (6T-27)\widehat{\gamma}_b^3 \widehat{\gamma}_v^5 \\
&+ (15T-50)\widehat{\gamma}_b^4 \widehat{\gamma}_v^4 + 15\widehat{\gamma}_b^4 (\widehat{\sigma}_v^2)^2 + (20T-55)\widehat{\gamma}_b^5 \widehat{\gamma}_v^3 + (15T-35)\widehat{\gamma}_b^6 \widehat{\sigma}_v^2]. \tag{3.6}
\end{aligned}$$

In addition, under the assumptions of Theorem 4.4 it is possible to show that $\widehat{\gamma}_v^5 \xrightarrow{p} \gamma_v^5$, $\widehat{\gamma}_v^6 \xrightarrow{p} \gamma_v^6$, and $\widehat{\gamma}_v^8 \xrightarrow{p} \gamma_v^8$ as $N \rightarrow \infty$ and T is fixed. Similar results are obtained for b_i .

4 Testing

In this section we provide tests for skewness, kurtosis, and normality in the individual and the idiosyncratic error components.

4.1 Testing for skewness

In order to test for skewness in both the individual heterogeneity and the remainder components, the quantities of interest for each component are $SK_v = \gamma_v^3/\sigma_v^3$ and $SK_b = \gamma_b^3/\sigma_b^3$, respectively. In this section, we derive the limiting distribution of the corresponding test statistics under arbitrary SK_v and SK_b . Note that the simplicity of this result enables us to propose a test for symmetry, that is, we can test the null hypothesis of $SK_v = 0$ and/or $SK_b = 0$.

In this case, the proposed statistics are

$$\widehat{SK}_v = \frac{\widehat{\gamma}_v^3}{\widehat{\sigma}_v^3} \quad \text{and} \quad \widehat{SK}_b = \frac{\widehat{\gamma}_b^3}{\widehat{\sigma}_b^3}, \tag{4.1}$$

where $\widehat{\sigma}_v^3 = (\widehat{\sigma}_v^2)^{3/2}$ and $\widehat{\sigma}_b^3 = (\widehat{\sigma}_b^2)^{3/2}$, while $\widehat{\gamma}_v^3$, $\widehat{\gamma}_b^3$, $\widehat{\sigma}_v^2$, and $\widehat{\sigma}_b^2$ are the consistent estimators for γ_v^3 , γ_b^3 , σ_v^2 , and σ_b^2 , respectively, obtained previously.

Theorem 4.1 *Suppose that $\gamma_v^6 = E(v_{it}^6)$ and $\gamma_b^6 = E(b_i^6)$ are finite. Assuming conditions 3.1-3.10 hold, as $N \rightarrow \infty$ and T is fixed, we have*

$$\begin{aligned}
\sqrt{N}(\widehat{SK}_v - SK_v) &\xrightarrow{d} \mathcal{N}\left(0, \frac{\alpha_v^\top \Gamma_v \alpha_v}{T\sigma_v^6}\right), \\
\sqrt{N}(\widehat{SK}_b - SK_b) &\xrightarrow{d} \mathcal{N}\left(0, \frac{\alpha_b^\top \Gamma_b \alpha_b}{\sigma_b^6}\right)
\end{aligned}$$

where $\alpha_v = [1, -\frac{3SK_v\sigma_v}{2}]^\top$ and $\alpha_b = [1, -\frac{3SK_b\sigma_b}{2}]^\top$ are vectors of dimension 2×1 , while Γ_v and Γ_b are 2×2 matrices of the form

$$\Gamma_v = \begin{bmatrix} \mu_v^3 & \mu_v^{32} \\ \mu_v^{32} & \mu_v^2 \end{bmatrix} \quad \text{and} \quad \Gamma_b = \begin{bmatrix} \mu_b^3 & \mu_b^{32} \\ \mu_b^{32} & \mu_b^2 \end{bmatrix},$$

where $\mu_v^{32} = \gamma_v^5 - (\frac{4T-10}{T-1})\sigma_v^2\gamma_v^3$ and $\mu_b^{32} = \gamma_b^5 - \gamma_b^3\sigma_b^2 + \frac{6}{T}\gamma_b^3\sigma_b^2$.

The proof of this result is shown in the Appendix.

In what follows, we obtain the distribution of these test statistics under the null hypothesis of symmetry by standardizing the previous results.

Theorem 4.2 *Suppose that $\gamma_v^6 = E(v_{it}^6)$ and $\gamma_b^6 = E(b_i^6)$ are finite. Assuming conditions 3.1-3.10 hold, under the null hypothesis of symmetry, i.e. $SK_v = 0$ and/or $SK_b = 0$, as $N \rightarrow \infty$ and T is fixed,*

$$\begin{aligned} \widehat{\pi}_{v3} &= \frac{\sqrt{N}\widehat{SK}_v}{sd(\widehat{SK}_v)} \xrightarrow{d} \mathcal{N}(0, 1), \\ \widehat{\pi}_{b3} &= \frac{\sqrt{N}\widehat{SK}_b}{sd(\widehat{SK}_b)} \xrightarrow{d} \mathcal{N}(0, 1), \end{aligned}$$

where $sd(\widehat{SK}_v) = \frac{1}{\sqrt{T}\widehat{\sigma}_v^6}\widehat{\Gamma}_{v0}^{1/2}$ and $sd(\widehat{SK}_b) = \frac{1}{\sqrt{\widehat{\sigma}_b^6}}\widehat{\Gamma}_{b0}^{1/2}$. Also, $\widehat{\sigma}_v^2$, $\widehat{\sigma}_b^2$, $\widehat{\Gamma}_{v0}$, and $\widehat{\Gamma}_{b0}$ are the consistent estimators for σ_v^2 , σ_b^2 , Γ_{v0} , and Γ_{b0} , respectively, whereas $\widehat{\Gamma}_{v0} = \widehat{\gamma}_v^6 - (\frac{6T-15}{T-1})\widehat{\gamma}_v^4\widehat{\sigma}_v^2 + (\frac{9(T-2)^2+24}{(T-1)(T-2)})\widehat{\sigma}_v^6$ and $\widehat{\Gamma}_{b0} = \widehat{\gamma}_b^6 + \frac{9}{T}\widehat{\gamma}_b^4\widehat{\sigma}_v^2 + \frac{18}{T(T-1)}\widehat{\sigma}_b^2\widehat{\sigma}_v^4 + \frac{6}{T(T-1)(T-2)}\widehat{\sigma}_v^6$.

The proof of this theorem follows the same lines as the corresponding for Theorem 4.1 so it is omitted.

This theorem indicates that a t-test for skewness can be implemented by standardizing \widehat{SK}_v and \widehat{SK}_b with the square root of a consistent estimator of their variances. Furthermore, \widehat{SK}_v is robust to the presence of skewness (or kurtosis) in b_i even in small samples. Similarly, this result holds for \widehat{SK}_b .

4.2 Testing for kurtosis

Focus now on testing for kurtosis, the quantities of interest are $KU_v = \gamma_v^4/\sigma_v^4$ and $KU_b = \gamma_b^4/\sigma_b^4$ for the random error and individual effects, respectively. As discussed previously, we first derive the limiting distribution of the proposed statistics under arbitrary KU_v and KU_b , and later we extend the results to the particular case in which $KU_v = 3$ and/or $KU_b = 3$.

The proposed statistics to construct tests for kurtosis are

$$\widehat{KU}_v = \frac{\widehat{\gamma}_v^4}{\widehat{\sigma}_v^4} \quad \text{and} \quad \widehat{KU}_b = \frac{\widehat{\gamma}_b^4}{\widehat{\sigma}_b^4}, \quad (4.2)$$

where $\widehat{\gamma}_v^4$, $\widehat{\gamma}_b^4$, $\widehat{\sigma}_v^4$ and $\widehat{\sigma}_b^4$ are the consistent estimators obtained previously for γ_v^4 , γ_b^4 , σ_v^4 and σ_b^4 , respectively.

Theorem 4.3 *Suppose that $\gamma_v^8 = E(v_{it}^8)$ and $\gamma_b^8 = E(b_i^8)$ are finite. Assuming conditions 3.1-3.10 hold, as $N \rightarrow \infty$ and T is fixed,*

$$\begin{aligned}\sqrt{N}(\widehat{KU}_v - KU_v) &\xrightarrow{d} \mathcal{N}\left(0, \frac{\beta_v^\top \Omega_v \beta_v}{T\sigma_v^8}\right), \\ \sqrt{N}(\widehat{KU}_b - KU_b) &\xrightarrow{d} \mathcal{N}\left(0, \frac{\beta_b^\top \Omega_b \beta_b}{\sigma_b^8}\right),\end{aligned}$$

where $\beta_v = [1, -2KU_v\sigma_v^2]^\top$ and $\beta_b = [1, -2KU_b\sigma_b^2]^\top$ are vectors of dimension 2×1 , while Ω_v and Ω_b are 2×2 matrices of the form

$$\Omega_v = \begin{bmatrix} \mu_v^4 & \mu_v^{42} \\ \mu_v^{42} & \mu_v^2 \end{bmatrix} \quad \text{and} \quad \Omega_b = \begin{bmatrix} \mu_b^4 & \mu_b^{42} \\ \mu_b^{42} & \mu_b^2 \end{bmatrix},$$

where $\mu_v^{42} = \gamma_v^6 - \frac{(T-9)}{(T-1)}\gamma_v^4\sigma_v^2 - 4(\gamma_v^3)^2 - \frac{12}{(T-1)}\sigma_v^6$ and $\mu_b^{42} = \gamma_b^6 - \gamma_b^4\sigma_b^2 + \frac{8}{T}\gamma_b^4\sigma_v^2 + \frac{12}{T(T-1)}\sigma_b^2\sigma_v^4$.

The proof of this result is shown in the Appendix.

In the particular case in which we are interested in testing whether the kurtosis coefficient is equal to 3, the null hypothesis is $KU_v = 3$ and $KU_b = 3$. Then, by standardizing the above results we obtain the following result.

Theorem 4.4 *Suppose that $\gamma_v^8 = E(v_{it}^8)$ and $\gamma_b^8 = E(b_i^8)$ are finite. Assuming conditions 3.1-3.10 hold, under the null hypothesis that the kurtosis coefficient is equal to 3, i.e. $KU_v = 3$ and/or $KU_b = 3$, as $N \rightarrow \infty$ and T is fixed,*

$$\begin{aligned}\widehat{\pi}_{v4} &= \frac{\sqrt{N}(\widehat{KU}_v - 3)}{sd(\widehat{KU}_v)} \xrightarrow{d} \mathcal{N}(0, 1), \\ \widehat{\pi}_{b4} &= \frac{\sqrt{N}(\widehat{KU}_b - 3)}{sd(\widehat{KU}_b)} \xrightarrow{d} \mathcal{N}(0, 1),\end{aligned}$$

where $sd(\widehat{KU}_v) = \frac{1}{\sqrt{T\widehat{\sigma}_v^8}} \left(\widehat{\beta}_{v3}^\top \widehat{\Omega}_{v3} \widehat{\beta}_{v3} \right)^{1/2}$ and $sd(\widehat{KU}_b) = \frac{1}{\sqrt{T\widehat{\sigma}_b^8}} \left(\widehat{\beta}_{b3}^\top \widehat{\Omega}_{b3} \widehat{\beta}_{b3} \right)^{1/2}$. Also, $\widehat{\beta}_{v3}$, $\widehat{\beta}_{b3}$, $\widehat{\Omega}_{v3}$, and $\widehat{\Omega}_{b3}$ are the consistent estimators for β_{v3} , β_{b3} , Ω_{v3} , and Ω_{b3} , respectively, whereas $\widehat{\beta}_{v3} = [1, -6\widehat{\sigma}_v^2]^\top$ and $\widehat{\beta}_{b3} = [1, -6\widehat{\sigma}_b^2]^\top$ are vectors of dimension 2×1 , and $\widehat{\Omega}_{v3}$ and $\widehat{\Omega}_{b3}$ are the corresponding expressions of $\widehat{\Omega}_v$ and $\widehat{\Omega}_b$ under the null, respectively.

The proof of this Theorem follows the same lines as the corresponding for Theorem 4.3, so it is therefore omitted.

As it is pointed out in Bai and Ng (2005) and Galvao et al. (2013), one of the main criticism when testing for kurtosis based on moment conditions is that skewness might influence kurtosis.

As the reader can appreciate in the results of Theorem 4.4, effectively \widehat{KU}_v and \widehat{KU}_b are affected by the skewness coefficient. Therefore, it is expected that SK_v and SK_b will likely be underestimated in practice given that the skewness coefficient can deviate these statistics substantially from its true value. Nevertheless, as we are going to show in the following, the skewness and kurtosis statistics proposed in this paper are very useful because they enable us to provide a joint test for each components of the error which assesses whether the data conform to any distribution of interest.

4.3 Testing for normality

The above skewness and kurtosis statistics provide a powerful scheme for assessing normality against a wide variety of alternative forms for each component of the error. In fact, a very appealing feature of the kurtosis statistic is that it enables us to identify how skewness affects kurtosis.

Let $\tilde{\pi}_{v3}$ and $\tilde{\pi}_{v4}$ be the test statistics for skewness and kurtosis, respectively, evaluated under the null hypothesis of normality of the random errors, i.e., $SK_v = 0$ and $KU_v = 3$. In this situation, it is easy to show that $\tilde{\pi}_{v3}$ and $\tilde{\pi}_{v4}$ are asymptotically independent. Thus, by extending the proposal in Jarque and Bera (1981) to panel data models, the proposed statistic to test for the normality of the random errors is of the form

$$\tilde{\pi}_{v34} = \tilde{\pi}_{v3}^2 + \tilde{\pi}_{v4}^2. \quad (4.3)$$

Similarly, under the null hypothesis of normality of the individual effects, $SK_b = 0$ and $KU_b = 3$, so the proposed statistic is

$$\tilde{\pi}_{b34} = \tilde{\pi}_{b3}^2 + \tilde{\pi}_{b4}^2, \quad (4.4)$$

where $\tilde{\pi}_{b3}$ and $\tilde{\pi}_{b4}$ are the corresponding test statistics for skewness and kurtosis evaluated under the null, respectively.

Theorem 4.5 *Suppose that $\gamma_v^8 = E(v_{it}^8)$ and $\gamma_b^8 = E(b_i^8)$ are finite and assuming conditions 3.1-3.10 hold. Under the null hypothesis of normality, as $N \rightarrow \infty$ and T is fixed,*

$$\begin{aligned} \tilde{\pi}_{v34} &\xrightarrow{d} \chi_2^2, \\ \tilde{\pi}_{b34} &\xrightarrow{d} \chi_2^2. \end{aligned}$$

The proof of this result follows the same lines as the corresponding for Theorems 4.1 and 4.3 and it is therefore omitted.

5 Monte Carlo simulations and application

To illustrate the feasibility and possible gains of the proposed method in this paper, we first carry out some simulation studies to demonstrate the finite sample performance of the estimators and tests proposed in the above sections. In addition, for the nonparametric component, the behavior of the pairwise difference estimator developed here is analyzed with respect to other estimators proposed for varying coefficient panel data models. Later, we apply the proposed method and tests to analyze a real data example.

5.1 Monte Carlo experiment

We consider the following data generating process,

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

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 variables in $[0, \pi/2]$ and Gaussian random variables $NID(0, 1)$, respectively. For the random errors v_{it} and the individual effects b_i , we consider the following cases:

$$\begin{aligned} DGP1. & \quad v_{it} \sim i.i.d. 0.5N(0, 1) & \quad \text{and} & \quad b_i \sim i.i.d. 0.5N(0, 1); \\ DGP2. & \quad v_{it} \sim i.i.d. 0.5N(0, 1) & \quad \text{and} & \quad b_i \sim i.i.d. 0.5t(9); \\ DGP3. & \quad v_{it} \sim i.i.d. 0.5t(9) & \quad \text{and} & \quad b_i \sim i.i.d. 0.5N(0, 1); \\ DGP4. & \quad v_{it} \sim i.i.d. 0.5t(9) & \quad \text{and} & \quad b_i \sim i.i.d. 0.5t(9); \\ DGP5. & \quad v_{it} \sim i.i.d. 0.5N(0, 1) & \quad \text{and} & \quad b_i \sim i.i.d. 0.5exp(N(0, 1)); \\ DGP6. & \quad v_{it} \sim i.i.d. 0.5exp(N(0, 1)) & \quad \text{and} & \quad b_i \sim i.i.d. 0.5N(0, 1). \end{aligned}$$

The simulation results are based on 1000 samples of data $\{(X_{it}, Z_{it}, Y_{it}) : i = 1, \dots, N, t = 1, \dots, T\}$. 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 for simplicity the bandwidth is chosen following Silverman's rule-of-thumb, i.e. $\hat{H} = \hat{h}I = \hat{\sigma}_z(NT)^{-1/5}$, where $\hat{\sigma}_z$ is the sample standard deviation of Z_{it} . Of course, a more specific bandwidth technique could be used, but this is beyond the scope of this paper.

The finite sample performance of the proposed estimators. Focusing on the effectiveness of the estimators proposed in Section 2, we consider cases *DGP1* – *DGP4*. In particular, for the higher order estimators we calculate the bias, standard deviations (SD) and the root mean squared error (RMSE), whereas the estimator for $m(\cdot)$ is assessed via the square root of the averaged squared errors (SRASE) defined as

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

Some of the representative results for the estimators of the second, third, and fourth order moments of the individual effects and random errors are listed in Table 1, whereas Table 2 presents the results of the proposed estimators. For the sake of comparison, we present the mean and SD of the SRASE of the nonparametric estimators that we consider in this study: the pairwise difference estimator (PDE) developed in this paper, the fixed effects (FEE) and the one-step backfitting (OBE) estimators proposed in Rodriguez-Poo and Soberon (2015) and the profile least square estimator (PLSE) presented in Sun et al. (2009). These results are computed together with the relative efficiencies (RE) of these estimators defined as the ratio of the SRASE of each estimator to that of the benchmark estimator (i.e., the FEE).

From Table 1 it can be pointed out that, for all T , as N increases, the bias and the standard deviation of the estimators for the second, third, and fourth order moments of the individual and error components are reduced in all the cases of study. Also, as it was expected from their asymptotic properties described in Section 4, the MSEs of these estimators are lower.

In Table 2, the comparison between the nonparametric estimators shows that our pairwise difference estimator is very competitive. In one step, it achieves results rather similar to the one step backfitting estimator in Rodriguez-Poo and Soberon (2015), corroborating the theoretical results in Section 4. In addition, these estimators are even closer as the sample size increases. Therefore, the gain of this new estimation method is corroborated.

The finite sample performance of the skewness coefficient, kurtosis and normality test. A well-known drawback of the tests that require high order moments is that they are not theoretically valid for numerous distributions, see Bai and Ng (2005) and Galvao et al. (2013) among others. With the aim of corroborating this fact, in this experiment we assess the finite sample performance of the proposed tests under different non-Gaussianity distributions. Thus, in Table 3 the size of the tests is evaluated using some symmetric distributions as those considered in $DGP1 - DGP3$. Later, in Table 4 the power of the tests is assessed by considering some asymmetric distributions as those in $DGP5 - DGP6$.

For practical implementation, the following Wald test statistics are used to test for skewness, (i) $\hat{\pi}_{v3}^2$ and (ii) $\hat{\pi}_{b3}^2$, for kurtosis we use (iii) $\hat{\pi}_{v4}^2$, (iv) $\hat{\pi}_{b4}^2$, whereas for normality the corresponding statistics are (v) $\tilde{\pi}_{v34}$ and (vi) $\tilde{\pi}_{b34}$. Thus, under the corresponding null hypotheses, the statistics (i), (ii), (iii), and (iv) have χ_1^2 asymptotic distribution, whereas (v) and (vi) have χ_2^2 asymptotic distribution.

In $DGP1$ all the tests should have empirical size close to the level of significance. Looking at the results in Table 3, the proposed tests exhibit a very good empirical size for the different sample size considered. Even for the smaller sample size, these tests achieve the correct size.

$DGP2$ considers that the individual effects follow a t_9 -Student distribution that is symmetric but presents excess kurtosis. Then, it is expected that the power of the tests for kurtosis in b_i increases as the sample size increases. In fact, in Table 3 the test for kurtosis in b_i exhibits non trivial empirical power, while the tests for skewness and kurtosis in v_{it} not. The opposite

Table 1.

Finite sample performance of the estimators for the higher order moments.

		Remainder component			Individual component		
		$\hat{\sigma}_v^2$	$\hat{\gamma}_v^3$	$\hat{\gamma}_v^4$	$\hat{\sigma}_b^2$	$\hat{\gamma}_b^3$	$\hat{\gamma}_b^4$
		$v_{it} \sim_{i.i.d.} 0.5\mathcal{N}(0, 1),$			$b_i \sim_{i.i.d} 0.5\mathcal{N}(0, 1)$		
N=50	Bias	0.320	-0.004	0.992	0.015	0.007	0.014
	SD	0.077	0.151	0.564	0.094	0.144	0.261
	MSE	0.328	0.151	1.140	0.074	0.113	0.228
N=100	Bias	0.291	-0.003	0.820	0.007	0.001	0.820
	SD	0.050	0.093	0.319	0.066	0.096	0.171
	MSE	0.295	0.095	0.879	0.049	0.075	0.151
N=150	Bias	0.274	-0.003	0.762	0.005	0.001	0.012
	SD	0.041	0.077	0.264	0.050	0.076	0.129
	MSE	0.277	0.078	0.805	0.038	0.064	0.112
		$v_{it} \sim_{i.i.d.} 0.5\mathcal{N}(0, 1),$			$b_i \sim_{i.i.d} 0.5t(9)$		
N=50	Bias	0.320	-0.009	0.959	0.011	0.017	0.040
	SD	0.079	0.148	0.542	0.123	0.271	0.684
	MSE	0.329	0.149	1.100	0.082	0.156	0.394
N=100	Bias	0.285	-0.004	0.799	0.010	0.007	0.021
	SD	0.047	0.093	0.299	0.083	0.170	0.386
	MSE	0.288	0.095	0.852	0.055	0.107	0.242
N=150	Bias	0.269	0.001	0.758	0.007	0.003	0.001
	SD	0.040	0.073	0.221	0.067	0.149	0.408
	MSE	0.272	0.075	0.789	0.039	0.071	0.188
		$v_{it} \sim_{i.i.d.} 0.5t(9),$			$b_i \sim_{i.i.d} 0.5\mathcal{N}(0, 1)$		
N=50	Bias	0.318	0.008	1.123	0.008	0.013	0.023
	SD	0.097	0.212	0.816	0.097	0.148	0.266
	MSE	0.329	0.202	1.339	0.074	0.121	0.232
N=100	Bias	0.282	0.007	0.896	0.002	0.002	0.010
	SD	0.058	0.133	0.524	0.067	0.101	0.162
	MSE	0.286	0.126	0.990	0.052	0.087	0.150
N=150	Bias	0.277	0.001	0.889	0.002	0.002	0.006
	SD	0.048	0.105	0.427	0.050	0.076	0.132
	MSE	0.280	0.099	0.973	0.039	0.057	0.116
		$v_{it} \sim_{i.i.d.} 0.5t(9),$			$b_i \sim_{i.i.d} 0.5t(9)$		
N=50	Bias	0.310	0.005	1.053	0.011	-0.008	0.013
	SD	0.089	0.189	0.757	0.126	0.265	0.766
	MSE	0.320	0.177	1.256	0.086	0.157	0.352
N=100	Bias	0.290	0.004	0.950	0.008	-0.006	0.017
	SD	0.057	0.128	0.464	0.084	0.166	0.347
	MSE	0.295	0.122	1.041	0.058	0.113	0.352
N=150	Bias	0.277	0.003	0.880	0.013	0.007	0.036
	SD	0.047	0.099	0.354	0.076	0.135	0.312
	MSE	0.276	0.089	0.934	0.047	0.090	0.213

Table 2.

Finite sample performance of the estimators for the nonparametric component.

		FEE	OBE	PLSE	PDE
		$v_{it} \sim_{i.i.d.} 0.5\mathcal{N}(0, 1),$		$b_i \sim_{i.i.d.} 0.5\mathcal{N}(0, 1)$	
N=50	Mean	0.500	0.555	0.769	0.513
	SD	0.086	0.029	0.049	0.040
	RE	1.000	0.337	0.465	0.465
N=100	Mean	0.448	0.528	0.763	0.479
	SD	0.061	0.028	0.046	0.030
	RE	1.000	0.466	0.766	0.500
N=150	Mean	0.426	0.510	0.769	0.461
	SD	0.051	0.018	0.045	0.025
	RE	1.000	0.352	0.882	0.490
		$v_{it} \sim_{i.i.d.} 0.5\mathcal{N}(0, 1),$		$b_i \sim_{i.i.d.} 0.5t(9)$	
N=50	Mean	0.503	0.555	0.756	0.511
	SD	0.087	0.027	0.051	0.039
	RE	1.000	0.310	0.586	0.448
N=100	Mean	0.447	0.529	0.751	0.479
	SD	0.058	0.026	0.043	0.029
	RE	1.000	0.448	0.741	0.500
N=150	Mean	0.425	0.511	0.750	0.462
	SD	0.055	0.018	0.050	0.025
	RE	1.000	0.327	0.909	0.454
		$v_{it} \sim_{i.i.d.} 0.5t(9),$		$b_i \sim_{i.i.d.} 0.5\mathcal{N}(0, 1)$	
N=50	Mean	0.498	0.554	0.772	0.511
	SD	0.086	0.031	0.054	0.045
	RE	1.000	0.360	0.627	0.511
N=100	Mean	0.445	0.526	0.766	0.476
	SD	0.066	0.026	0.046	0.034
	RE	1.000	0.393	0.696	0.515
N=150	Mean	0.433	0.514	0.764	0.465
	SD	0.050	0.019	0.043	0.025
	RE	1.000	0.380	0.860	0.500
		$v_{it} \sim_{i.i.d.} 0.5t(9),$		$b_i \sim_{i.i.d.} 0.5t(9)$	
N=50	Mean	0.498	0.555	0.767	0.510
	SD	0.088	0.030	0.047	0.045
	RE	1.000	0.341	0.534	0.511
N=100	Mean	0.450	0.529	0.765	0.483
	SD	0.058	0.029	0.046	0.031
	RE	1.000	0.500	0.793	0.534
N=150	Mean	0.431	0.514	0.758	0.465
	SD	0.053	0.017	0.041	0.023
	RE	1.000	0.320	0.773	0.433

Table 3.

Size and power of the symmetry, kurtosis, and normality tests.

		Remainder component			Individual component		
		Level of significance			Level of significance		
		0.10	0.05	0.01	0.10	0.05	0.01
		$v_{it} \sim_{i.i.d.} 0.5\mathcal{N}(0, 1),$			$b_i \sim_{i.i.d} 0.5\mathcal{N}(0, 1)$		
N=50	Skewness	0.087	0.047	0.017	0.132	0.065	0.006
	Kurtosis	0.093	0.052	0.016	0.079	0.037	0.009
	Normality	0.072	0.033	0.006	0.070	0.029	0.004
N=100	Skewness	0.108	0.055	0.020	0.098	0.036	0.003
	Kurtosis	0.083	0.054	0.017	0.104	0.061	0.017
	Normality	0.090	0.052	0.007	0.074	0.032	0.002
N=150	Skewness	0.109	0.051	0.012	0.114	0.054	0.005
	Kurtosis	0.105	0.057	0.020	0.107	0.054	0.029
	Normality	0.107	0.052	0.006	0.101	0.052	0.008
		$v_{it} \sim_{i.i.d.} 0.5\mathcal{N}(0, 1),$			$b_i \sim_{i.i.d} 0.5t(9)$		
N=50	Skewness	0.086	0.053	0.018	0.093	0.035	0.005
	Kurtosis	0.062	0.039	0.007	0.114	0.067	0.020
	Normality	0.063	0.035	0.002	0.089	0.040	0.003
N=100	Skewness	0.091	0.051	0.013	0.091	0.038	0.004
	Kurtosis	0.067	0.040	0.012	0.178	0.102	0.028
	Normality	0.078	0.041	0.003	0.133	0.078	0.006
N=150	Skewness	0.094	0.049	0.011	0.101	0.047	0.006
	Kurtosis	0.094	0.055	0.012	0.225	0.132	0.036
	Normality	0.099	0.045	0.002	0.162	0.093	0.008
		$v_{it} \sim_{i.i.d.} 0.5t(9),$			$b_i \sim_{i.i.d} 0.5\mathcal{N}(0, 1)$		
N=50	Skewness	0.087	0.050	0.013	0.113	0.058	0.011
	Kurtosis	0.164	0.085	0.025	0.075	0.039	0.010
	Normality	0.103	0.046	0.002	0.068	0.042	0.004
N=100	Skewness	0.090	0.045	0.008	0.122	0.062	0.014
	Kurtosis	0.264	0.153	0.037	0.103	0.056	0.025
	Normality	0.162	0.084	0.007	0.086	0.051	0.013
N=150	Skewness	0.116	0.050	0.010	0.105	0.051	0.008
	Kurtosis	0.348	0.227	0.064	0.100	0.055	0.019
	Normality	0.247	0.140	0.020	0.094	0.042	0.001

Table 4.

Size and power of the symmetry, kurtosis, and normality tests.

		Remainder component			Individual component		
		Level of significance			Level of significance		
		0.10	0.05	0.01	0.10	0.05	0.01
		<i>v_{it} ~ i.i.d. 0.5N(0, 1),</i>			<i>b_i ~ i.i.d 0.5exp(N(0, 1))</i>		
N=50	Skewness	0.035	0.027	0.015	0.624	0.442	0.120
	Kurtosis	0.117	0.090	0.039	0.253	0.145	0.041
	Normality	0.020	0.014	0.003	0.357	0.225	0.033
N=100	Skewness	0.053	0.029	0.016	0.646	0.475	0.202
	Kurtosis	0.114	0.102	0.047	0.391	0.223	0.072
	Normality	0.026	0.024	0.004	0.373	0.254	0.035
N=150	Skewness	0.098	0.045	0.019	0.623	0.506	0.184
	Kurtosis	0.111	0.106	0.057	0.422	0.235	0.088
	Normality	0.038	0.033	0.004	0.408	0.287	0.045
		<i>v_{it} ~ i.i.d. 0.5exp(N(0, 1)),</i>			<i>b_i ~ i.i.d 0.5N(0, 1)</i>		
N=50	Skewness	0.528	0.280	0.028	0.116	0.075	0.031
	Kurtosis	0.140	0.047	0.001	0.023	0.016	0.003
	Normality	0.397	0.200	0.015	0.087	0.060	0.015
N=100	Skewness	0.696	0.502	0.142	0.084	0.061	0.023
	Kurtosis	0.277	0.104	0.010	0.046	0.021	0.003
	Normality	0.505	0.349	0.043	0.056	0.035	0.003
N=150	Skewness	0.715	0.568	0.214	0.094	0.053	0.022
	Kurtosis	0.382	0.176	0.020	0.097	0.039	0.004
	Normality	0.529	0.367	0.063	0.081	0.049	0.002

is obtained for *DGP3*, where it is assumed that the random effects follow the t_9 Student distribution.

Table 4 reports results for *DGP5* – *DGP6*, where a log-normal distribution is considered. As it is well-known, this is an asymmetric function with a high level of kurtosis, so it is expected that the power of the tests for skewness and kurtosis of the composed error increases with the sample size. In fact, the tests for b_i in *DGP6* have non trivial empirical power, while the tests for the other component have empirical sizes close to the level of significance. The opposite is obtained for *DGP6*.

Finally, based on these results it can be pointed out that the proposed tests detect departures from the null hypothesis of skewness and/or kurtosis in each component. In addition, these tests are robust to the presence of skewness and/or kurtosis in the other component.

5.2 Application

Let us consider the estimation of technical efficiency in a panel of EU the companies. For nearly three decades, starting with Aigner et al. (1977) the stochastic frontier literature has brought forth models to estimate technical efficiency. Since then, many developments have taken place but some issues have remained unchanged. First, researchers mainly focus their attention on the mean of the inefficiency term. Second, production functions are assumed to belong to some pre-specified family of parametric functions usually recommended by economic theory and their parameters of interest remain constant along time.

It is well-known that higher order moments can often shed additional light on the distribution of a variable. Although most papers in this literature are only interested in the first moment there are some exceptions as in Kumbhakar (2002) where, in order to estimate risk preferences, the first three moments of the distribution of inefficiency are used. Furthermore, in the analysis of technical inefficiency, skewness plays an important role (see Almanidis and Sickles (2011)) but researchers usually have to deal with the so-called “wrong skewness” anomaly, see Simar and Wilson (2010). More precisely, in those cases in which the expected and the estimated sign of the skewness of the error terms are different it is necessary either to re-specify the model or to obtain a new sample. In fact, none of these solutions are very appealing for researchers and therefore a test for skewness appears in this issue of interest.

Regarding the choice of the functional form in fully parametric models, a standard assumption in this literature is indeed that capital and labor elasticities are constant along 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, see Li et al. (2002) for example. Then, if we are willing to assume that the parameters in the production function vary according to some set of explanatory variables it becomes much harder the choice of the functional form and therefore the risk of misspecification is high. On these

grounds, nonparametric varying coefficient models are a natural way to extend this assumption of constant elasticities.

In this situation, the varying coefficient panel data model that we propose to estimate is of the following form

$$\ln Y_{it} = \ln W_{it} \beta_1(Z_{it}) + \ln L_{it} \beta_2(Z_{it}) + \ln K_{it} \beta_3(Z_{it}) + b_i + v_{it}, \quad i = 1, \dots, N; \quad t = 1, \dots, T,$$

where Y represents the sales of the firm, W the liquid capital, L the labor input, K the fixed capital, and Z the firm's R&D expenses. Note that in this specification the R&D variable has a neutral effect on the production function by shifting the level of the production frontier but also affects the labor and/or capital marginal productivity. In addition, $v_{it} = \nu_{it} - u_{it}$ is a composed error term, where ν_{it} is the idiosyncratic error term and u_{it} represents the inefficiency of the firm i , that has an expected value equal to $E(v_{it}) = -E(u_{it})$ and a third central moment such as

$$E(v_{it} - E(v_{it}))^3 = E(\nu_{it} - u_{it} + E(u_{it}))^3 = -E(u_{it} - E(u_{it}))^3.$$

Thus, as it is noted in Carree (2002) among others, a positively skewed distribution of the inefficiencies u_{it} implies that the composed error term v_{it} has a negative skewness, so $\gamma_v^3 < 0$ is expected in the presence of inefficiencies.

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 million private and public European companies. After removing firms with missing values, we get a final sample of 1,120 observations, i.e., 120 companies and 7 time periods.

Varying coefficient estimates are plotted against the R&D expenses where the continuous line denotes the estimated varying coefficients and dotted lines represent 95% pointwise confidence intervals 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 $\beta_1(z) + \beta_2(z) + \beta_3(z)$.

Focusing on the results of Figure 1, we can realize that the marginal productivity of liquid capital and R&D expenses are slightly negatively related. However, for those firms with a low level of R&D expenses, we certainly observe a positive relationship. 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 firms with large expenses. In this way, those companies with larger expenses do have incentives to increase their R&D expenses since this will end up in higher marginal productivity of fixed capital.

Analyzing the results of the marginal productivity of labor, in Figure 3 we observe that it does not seem to be relationship between marginal productivity of labor and R&D expenses.

Figure 1.

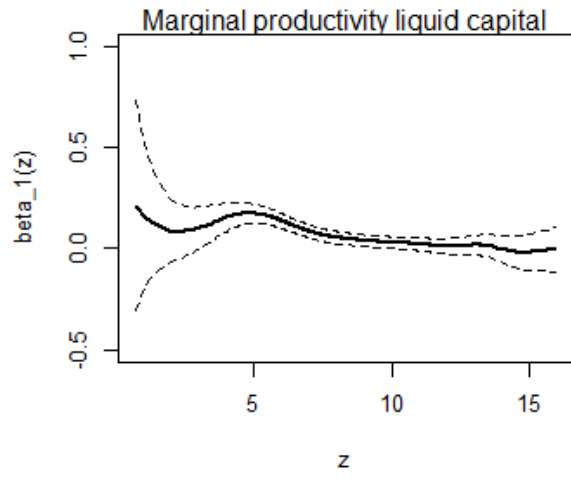


Figure 2.

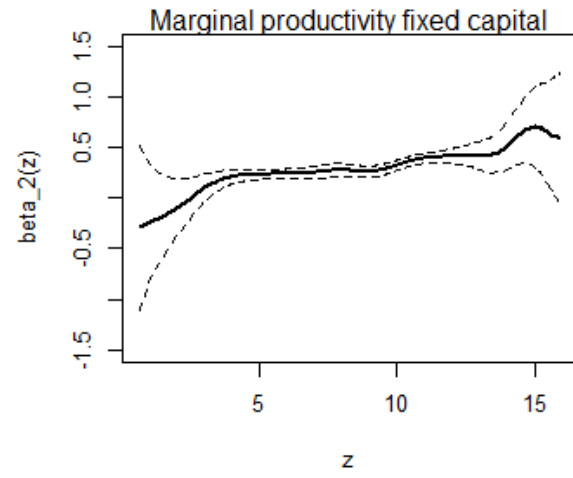


Figure 3.

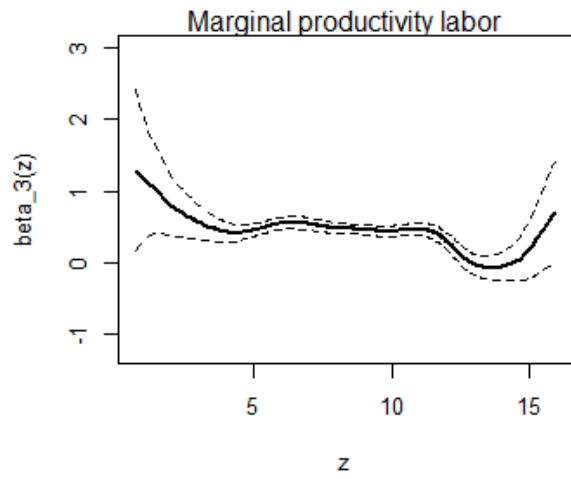
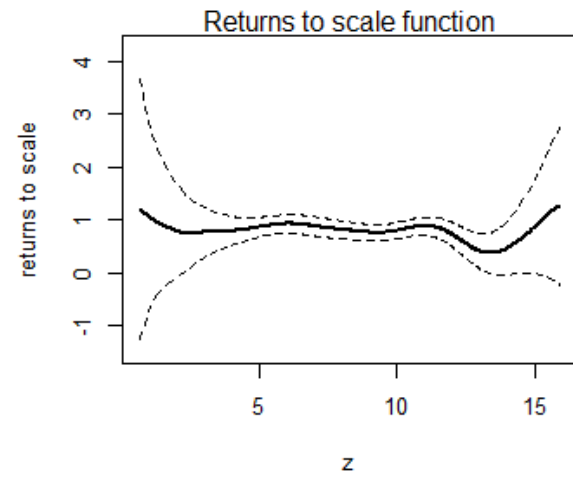


Figure 4.



On the boundaries, we can observe some non-linearities but they might be due to the so-called boundary effects. However, for firms with larger R&D expenses we might observe a drop followed by a sharp increase at the end. This inverted bell shape form suggests that lower 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 smaller than 1 in general. Specifically, firms with a low level of R&D expenses exhibit constant returns to scale, whereas companies with larger expenditures in R&D exhibit decreasing returns to scale.

Then, from all these results we can sum up that there exists a relationship between the marginal productivity of inputs and the R&D expenses. Also, this relationship appears as highly non-linear. In addition, companies with larger R&D expenses exhibit decreasing returns to scale and the marginal productivity of fixed and liquid capital is sensitive to this expenditures. Further, firms with moderate R&D expenses present constant returns to scale but the marginal productivity of the liquid capital is not very sensitive to these expenses.

Finally, Table 5 contains the estimated parameters for the higher order moments of b_i and v_{it} and the proposed tests statistics.

Table 5.
Estimated parameters stochastic frontier model.

	$\hat{\sigma}^2$	$\hat{\gamma}^3$	$\hat{\gamma}^4$	Skewness	Kurtosis	Normality
v_{it}	0.132	-0.069	0.332	-0.007	0.001	-0.005
b_i	31.576	202.672	1457.847	60.391	5.481	65.727

Analyzing the results of Table 5 we can highlight that this model presents inefficiencies since $\hat{\gamma}_v^3$ is greater than zero. In addition, when we test for skewness, kurtosis and normality of the firm level component, b_i , we get that it is largely asymmetric and exhibits excess kurtosis (rejecting the corresponding null hypothesis at the 1% significance level). Meanwhile, for the remainder component, v_{it} , it is probable that it follows a normal distribution since the joint test for the null hypothesis $H_0 : SK_v = 0$ and $KU_v = 3$ cannot be rejected.

In view of these results, it is clear that effectively this varying coefficient model enables to capture some relevant features of the covariates that were not possible with fully parametric or nonparametric models. In addition, the existence of inefficiencies in the error term with a positively skewed distribution has also been detected so it is possible to conclude that this model is not subject to the “wrong skewness” anomaly.

6 Conclusion

The normality assumption plays a crucial role in the validity of several specification tests and testing for skewness and kurtosis is a relevant topic in the finance or in the productive efficiency literature, for example. However, although there are plenty of studies for cross-sectional or time series data, the results for panel data models are scarce. In order to overcome this situation, this paper is concerned with the estimation and testing of varying coefficient panel data models when standard distributional assumptions cannot be justified.

In this context, we propose several tests for skewness, kurtosis, and normality for both components of the error, jointly and separately. Also, in order to avoid bootstrap techniques, estimators of higher order moments for both error components are proposed in a nonparametric framework. However, since the methodology proposed in this paper is rather general, its application to other models of interest is straightforward. In addition, in order to obtain the residuals for these tests, a local constant estimator based on a pairwise differencing transformation is presented. Unlike other estimators in the literature, this pairwise estimator is very appealing since, under some standard assumptions, it is shown that it exhibits the same properties than others in the field without having to resort to iterative procedures. Further, some simulations are used to examine the finite sample performance of the estimators and tests proposed in this paper and they are also illustrated in an empirical application about the production efficiency of the EU companies. ■

Appendix

Proof of Theorem 3.1. 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

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

Using the multivariate Taylor's theorem and by the regularity conditions of Theorem 3.1, we obtain

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

where $D_m(z)$ is a $dq \times 1$ vector such that $D_m(z) = \text{vec}(\partial m(z) / \partial z^\top)$ 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 $m(\cdot)$.

Replacing (6.1) in (2.11) and rearranging terms, $\widehat{m}(z; H)$ can be written as

$$\widehat{m}(z; H) - m(z) = \Psi_N^{-1} \left(B_N^{(1)} + B_N^{(2)} + U_N \right), \quad (6.2)$$

where

$$\begin{aligned} \Psi_N &= \binom{T}{2}^{-1} \frac{1}{N} \sum_{its} K_{it} K_{is} \widetilde{X}_{its} \widetilde{X}_{its}^\top \\ B_N^{(1)} &= \binom{T}{2}^{-1} \frac{1}{N} \sum_{its} K_{it} K_{is} \widetilde{X}_{its} \left(X_{it}^\top \otimes (Z_{it} - z)^\top - X_{is}^\top \otimes (Z_{is} - z)^\top \right) D_m(z), \\ B_N^{(2)} &= \binom{T}{2}^{-1} \frac{1}{2N} \sum_{its} K_{it} K_{is} \widetilde{X}_{its} \left(X_{it}^\top \otimes (Z_{it} - z)^\top \mathcal{H}_m(z) (Z_{it} - z) - X_{is}^\top \otimes (Z_{is} - z)^\top \mathcal{H}_m(z) (Z_{is} - z) \right), \\ U_N &= \binom{T}{2}^{-1} \frac{1}{N} \sum_{its} K_{it} K_{is} \widetilde{X}_{its} \widetilde{v}_{its}. \end{aligned}$$

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

$$\sqrt{N|H|} \left(\widehat{m}(z; H) - m(z) \right) - \sqrt{N|H|} \Psi_N^{-1} \left(B_N^{(1)} + B_N^{(2)} \right) = \sqrt{N|H|} \Psi_N^{-1} U_N, \quad (6.3)$$

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

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

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, by the law of iterated expectations and the strict stationarity condition

$$\begin{aligned} E(\Psi_N) &= \int E \left[\widetilde{X}_{its} \widetilde{X}_{its}^\top | Z_{it} = z + H^{1/2}u, Z_{is} = z + H^{1/2}v \right] f_{Z_{it}, Z_{is}}(Z_{it} = z + H^{1/2}u, Z_{is} = z + H^{1/2}v) \\ &\quad \times K(u)K(v) dudv, \end{aligned}$$

where by a Taylor expansion and Assumption 3.1 the expression (6.4) holds. Also, to complete the proof it is necessary to show that $\text{Var}(\Psi_N) \rightarrow 0$ as $N \rightarrow \infty$.

$$\begin{aligned} \text{Var}(\Psi_N) &= \frac{1}{N} \binom{T}{2}^{-1} \text{Var} \left(K_{it} K_{is} \widetilde{X}_{its} \widetilde{X}_{its}^\top \right) \\ &\quad + \frac{1}{N^2} \binom{T}{2}^{-2} \sum_{i=1}^N \sum_{t=1}^{T-1} \sum_{s=t+1}^T \sum_{t'=1}^{T-1} \sum_{s'=t'+1}^T \text{Cov} \left(K_{it} K_{is} \widetilde{X}_{its} \widetilde{X}_{its}^\top, K_{it'} K_{is'} \widetilde{X}_{it's'} \widetilde{X}_{it's'}^\top \right), \end{aligned}$$

where, under Assumptions 3.1 and 3.2, the first element is $O_{\mathbb{P}}\left(\frac{1}{N|H|}\right)$ whereas the second one is the sum of $o_{\mathbb{P}}\left(\frac{1}{N|H|}\right)$ terms. Then, as $N|H| \rightarrow \infty$ the variance term tends to zero and (6.4) holds.

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

$$\Psi_N^{-1} = \mathcal{B}_{\tilde{X}\tilde{X}}^{-1}(z, z) + o_{\mathbb{P}}(1). \quad (6.5)$$

Similarly, by the law of iterated expectations and the stationarity condition we can show

$$\begin{aligned} E\left(B_N^{(1)}\right) &= \binom{T}{2}^{-1} \sum_{ts} E \left[\left(E(\tilde{X}_{its}X_{it}^{\top} | Z_{it}, Z_{is}) \otimes (Z_{it} - z)^{\top} - E(\tilde{X}_{its}X_{is}^{\top} | Z_{it}, Z_{is}) \otimes (Z_{is} - z)^{\top} \right) D_m(z) \right. \\ &\quad \times \left. K_{it}K_{is} \right] \\ &= \binom{T}{2}^{-1} \sum_{ts} \int \left(E(\tilde{X}_{its}X_{it}^{\top} | Z_{it} = z, Z_{is} = z) D_f(z)(H^{1/2}u) \right) \otimes (H^{1/2}u)^{\top} D_m(z) K(u)K(v) dudv \\ &\quad - \binom{T}{2}^{-1} \sum_{ts} \int \left(E(\tilde{X}_{its}X_{is}^{\top} | Z_{it} = z, Z_{is} = z) D_f(z)(H^{1/2}v) \right) \otimes (H^{1/2}v)^{\top} D_m(z) K(u)K(v) dudv \\ &= \mu_2(K) \mathcal{B}_{\tilde{X}\tilde{X}}^{-1}(z, z) \text{diag}_d(\text{tr}(HD_f(z)D_{m_r}(z))) \iota_d f_{Z_{it}, Z_{is}}^{-1}(z, z) + o_{\mathbb{P}}(\text{tr}(H)), \end{aligned} \quad (6.6)$$

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

Following a similar procedure it is straightforward to show

$$\begin{aligned} E\left(B_N^{(2)}\right) &= \binom{T}{2}^{-1} \frac{1}{2} \sum_{ts} \int E(\tilde{X}_{its}X_{it}^{\top} | Z_{it} = z, Z_{is} = z) f_{Z_{it}, Z_{is}}(z, z) \otimes \text{tr}(H\mathcal{H}_{m_r}(z)) u^{\top} u K(u)K(v) dudv \\ &\quad - \binom{T}{2}^{-1} \frac{1}{2} \sum_{ts} \int E(\tilde{X}_{its}X_{is}^{\top} | Z_{it} = z, Z_{is} = z) f_{Z_{it}, Z_{is}}(z, z) \otimes \text{tr}(H\mathcal{H}_{m_r}(z)) v^{\top} v K(u)K(v) dudv \\ &= \mu_2(K) \mathcal{B}_{\tilde{X}\tilde{X}}^{-1}(z, z) \text{diag}_d(\text{tr}(H\mathcal{H}_{m_r}(z))) \iota_d + o_{\mathbb{P}}(\text{tr}(H)), \end{aligned} \quad (6.7)$$

where $\mathcal{H}_{m_r}(z)$ is the Hessian matrix of the r th 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 $H \rightarrow 0$ and $N|H| \rightarrow \infty$. Then, using (6.5)-(6.7) and applying the Cramér-Wold device it is proved that the asymptotic bias of $\hat{m}(z; H)$ is

$$\begin{aligned} \Psi_N^{-1} \left(B_N^{(1)} + B_N^{(2)} \right) &= \mu_2(K) \mathcal{B}_{\tilde{X}\tilde{X}}^{-1}(z, z) \mathcal{B}_{\tilde{X}\tilde{X}}^{-1}(z, z) \left(\text{diag}_d(\text{tr}(HD_f(z)D_{m_r}(z))) \iota_d f_{Z_{it}, Z_{is}}^{-1}(z, z) \right. \\ &\quad \left. + \frac{1}{2} \text{diag}_d(\text{tr}(H\mathcal{H}_{m_r}(z)) \iota_d) \right) + o_{\mathbb{P}}(\text{tr}(H)). \end{aligned} \quad (6.8)$$

So the first part of the proof is done.

In order 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 vector of observed covariates and denote $\tilde{v}_i =$

$(\tilde{v}_{i12}, \dots, \tilde{v}_{i1T}, \dots, \tilde{v}_{i(T-1)T})^\top$ as a $\frac{T(T-1)}{2} \times 1$ vector, $E(\tilde{v}_i \tilde{v}_{i'}^\top | \mathbb{X}) = 0$ for $\forall i \neq i'$ and

$$E(\tilde{v}_i \tilde{v}_i^\top | \mathbb{X}) = \begin{cases} 2\sigma_v^2 & \text{for } t = t', s = s', \\ \sigma_v^2 & \text{for } t = t', s \neq s', \\ -\sigma_v^2 & \text{for } t = s', t \neq t', s \neq s', \\ 0 & \text{for } t \neq t', s \neq s', t' \neq s, t \neq s'. \end{cases} \quad \begin{array}{l} \text{or } s = s', t \neq t' \\ \text{or } t' = s, t \neq t', s \neq s' \end{array}$$

When we analyze U_N we claim that by the law of iterated expectations and Assumptions 3.1, 3.5, and 3.6,

$$\begin{aligned} N|H|Var(U_N) &= |H| \frac{1}{N} \binom{T}{2}^{-2} \sum_{its} E \left[\tilde{X}_{its} E(\tilde{v}_{its} \tilde{v}_{its} | \mathbb{X}) \tilde{X}_{its}^\top K_{it}^2 K_{is}^2 \right] \\ &+ |H| \frac{2}{N} \binom{T}{2}^{-2} \sum_{its} \sum_{s' \neq s} E \left[\tilde{X}_{its} E(\tilde{v}_{its} \tilde{v}_{its'} | \mathbb{X}) \tilde{X}_{its'}^\top K_{it}^2 K_{is} K_{is'} \right] \\ &+ |H| \frac{2}{N} \binom{T}{2}^{-2} \sum_{its} \sum_{t' \neq t} E \left[\tilde{X}_{its} E(\tilde{v}_{its} \tilde{v}_{it't} | \mathbb{X}) \tilde{X}_{it't}^\top K_{it}^2 K_{is}^2 K_{it'} \right] \\ &+ |H| \frac{1}{N} \binom{T}{2}^{-2} \sum_{its} \sum_{t' \neq t} \sum_{s' \neq s} E \left[\tilde{X}_{its} E(\tilde{v}_{its} \tilde{v}_{it's'} | \mathbb{X}) \tilde{X}_{it's'}^\top K_{it} K_{it'} K_{is} K_{is'} \right] \\ &= \mathbf{I}_{1N} + \mathbf{I}_{2N} + \mathbf{I}_{3N} + \mathbf{I}_{4N}. \end{aligned}$$

Then, analyzing each of these terms separately we obtain

$$N|H|Var(U_N) = \binom{T}{2}^{-1} 2\sigma_v^2 R^2(K) \mathcal{B}_{\tilde{X}\tilde{X}}(z, z) (1 + o_{\mathbb{P}}(1)) \quad (6.9)$$

given that, under similar arguments as above, it can be proved that

$$\begin{aligned} \mathbf{I}_{1N} &= \binom{T}{2}^{-2} 2\sigma_v^2 \sum_{t=1}^{T-1} \sum_{s=t+1}^T \int E \left[\tilde{X}_{its} \tilde{X}_{its}^\top | Z_{it} = z, Z_{is} = z \right] f_{Z_{it}, Z_{is}}(z, z) K^2(u) K^2(v) dudv (1 + o_{\mathbb{P}}(1)), \\ \mathbf{I}_{2N} &= \binom{T}{2}^{-2} 2\sigma_v^2 |H|^{1/2} R(K) \sum_{ts} \sum_{s' \neq s} E \left[\tilde{X}_{its} \tilde{X}_{its'}^\top | Z_{it} = z, Z_{is} = z, Z_{is'=z} \right] f_{Z_{it}, Z_{is}, Z_{is'}}(z, z, z) (1 + o_{\mathbb{P}}(1)), \\ \mathbf{I}_{3N} &= -\binom{T}{2}^{-2} 2\sigma_v^2 |H|^{1/2} R(K) \sum_{ts} \sum_{t' \neq t} \sum_{s' = t'+1} E \left[\tilde{X}_{its} \tilde{X}_{iss'}^\top | Z_{it} = z, Z_{is} = z, Z_{is'=z} \right] f_{Z_{it}, Z_{is}, Z_{is'}}(z, z, z) \\ &\times (1 + o_{\mathbb{P}}(1)), \end{aligned}$$

and $\mathbf{I}_{4N} = o_{\mathbb{P}}(1)$ since $E(\tilde{v}_{its} \tilde{v}_{it's'} | \mathbb{X}) = 0$.

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

$$N|H|Var(\Psi_N^{-1} U_N) = \binom{T}{2}^{-1} 2\sigma_v^2 R^2(K) \mathcal{B}_{\tilde{X}\tilde{X}}^{-1}(z, z) \mathcal{B}_{\tilde{X}\tilde{X}}(z, z) \mathcal{B}_{\tilde{X}\tilde{X}}^{-1}(z, z) (1 + o_{\mathbb{P}}(1)). \quad (6.10)$$

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

Finally, to complete the proof of Theorem 3.1 it is necessary to show that, as N tends to infinity, for T fixed,

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

where

$$B(z, H) = \mu_2(K) \left(\text{diag}_d(\text{tr}(HD_f(z)D_{m_r}(z))) \iota_d f_{Z_{it}, Z_{is}}^{-1}(z, z) + \frac{1}{2} \text{diag}_d(\text{tr}(H\mathcal{H}_{m_r}(z))) \iota_d \right).$$

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

$$\sqrt{N|H|} \frac{1}{N} \binom{T}{2}^{-1} \sum_{its} K_{it} K_{is} \widetilde{X}_{its} \widetilde{v}_{its} = \frac{1}{\sqrt{N}} \binom{T}{2}^{-1} \sum_{its} \lambda_{its}, \quad (6.12)$$

where

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

By Theorem 3.1 and using the previous proofs,

$$\begin{aligned} \text{Var}(\lambda_{its}) &= \binom{T}{2}^{-1} 2\sigma_v^2 R^2(K) \mathcal{B}_{\widetilde{X}\widetilde{X}}^{-1}(z, z) (1 + o_{\mathbb{P}}(1)), \\ \text{Cov}(\lambda_{its}, \lambda_{it's'}) &= o_{\mathbb{P}}(1). \end{aligned}$$

Defining $\lambda_{N,i} = \binom{T}{2}^{-1} \sum_{its} \lambda_{its}$ and by the Minkowsky inequality we get

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

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

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

Hence, it is proved that

$$E|\lambda_{N,i}|^{2+\delta} = N^{-(2+\delta)/2} \sum_i E|\lambda_{N,i}|^{2+\delta} \leq CO_{\mathbb{P}}((N|H|)^{-\delta/2}).$$

Finally, since this term tends to zero as $N|H| \rightarrow \infty$, 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 3.1 is done. ■

Proof of Theorem 3.3. We first focus on the asymptotic properties of $\widehat{\sigma}_v^2$ and later on $\widehat{\sigma}_b^2$. Inserting (2.1) into $\widehat{\epsilon}_{it} = Y_{it} - X_{it}^\top \widehat{m}(Z_{it}; H)$, throughout the proof the residuals are written as $\widehat{\epsilon}_{it} = \epsilon_{it} - X_{it}^\top [\widehat{m}(Z_{it}; H) - m(Z_{it}; H)]$ and the expression to analyze is

$$\begin{aligned}\widehat{\sigma}_v^2 &= \frac{1}{NT(T-1)} \sum_{i=1}^N \left[T \sum_{t=1}^T \left(\epsilon_{it} - X_{it}^\top [\widehat{m}(Z_{it}; H) - m(Z_{it}; H)] \right)^2 - \left(\sum_{t=1}^T [\epsilon_{it} - X_{it}^\top [\widehat{m}(Z_{it}; H) - m(Z_{it}; H)]] \right)^2 \right] \\ &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T v_{it}^2 - \frac{1}{NT(T-1)} \sum_{i=1}^N \sum_{t=1}^T \sum_{s \neq t}^T v_{it} v_{is} + \mathbb{I}_v^{(1)} - \mathbb{I}_v^{(2)} + \mathbb{I}_v^{(3)} - \mathbb{I}_v^{(4)},\end{aligned}\tag{6.13}$$

where after rearranging terms

$$\begin{aligned}\mathbb{I}_v^{(1)} &= \frac{1}{NT} \sum_{it} (X_{it}^\top [\widehat{m}(Z_{it}; H) - m(Z_{it}; H)])^2, \\ \mathbb{I}_v^{(2)} &= \frac{1}{NT} \sum_{it} \epsilon_{it} X_{it}^\top [\widehat{m}(Z_{it}; H) - m(Z_{it}; H)], \\ \mathbb{I}_v^{(3)} &= \frac{1}{NT(T-1)} \sum_{its} X_{it}^\top [\widehat{m}(Z_{it}; H) - m(Z_{it}; H)] X_{is}^\top [\widehat{m}(Z_{is}; H) - m(Z_{is}; H)] \\ \mathbb{I}_v^{(4)} &= \frac{1}{NT(T-1)} \sum_{its} X_{it}^\top [\widehat{m}(Z_{it}; H) - m(Z_{it}; H)] \epsilon_{is}.\end{aligned}$$

As we are going to show, of these six terms of (6.13) only the first two will be the leading terms, whereas $\mathbb{I}_v^{(1)}$, $\mathbb{I}_v^{(2)}$, $\mathbb{I}_v^{(3)}$, and $\mathbb{I}_v^{(4)}$ are residual terms. Analyzing each term separately and using uniform convergence results as the ones established in Theorem 6 in Masry (1996), by Assumptions 3.1-3.3 and rearranging terms,

$$\begin{aligned}\mathbb{I}_v^{(1)} &\leq \frac{1}{NT} \sum_{it} |X_{it}^\top X_{it}| \sup_{\{Z_{it} \in A\}} |(\widehat{m}(Z_{it}; H) - m(Z_{it}; H))^\top| \sup_{\{Z_{it} \in A\}} |\widehat{m}(Z_{it}; H) - m(Z_{it}; H)| \\ &= O_{\mathbb{P}} \left(\text{tr}(H)^2 + \frac{\log N}{N|H|} \right),\end{aligned}\tag{6.14}$$

since $(NT)^{-1} \sum_{it} |X_{it}^\top X_{it}| = O_{\mathbb{P}}(1)$. A similar result holds for $\mathbb{I}_v^{(3)}$.

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

$$\mathbb{I}_v^{(2)} \leq \frac{2}{NT} \sum_{it} |\epsilon_{it} X_{it}^\top| \sup_{\{Z_{it} \in A\}} |\widehat{m}(Z_{it}; H) - m(Z_{it}; H)| = o_p \left(\frac{1}{\sqrt{N}} \right),\tag{6.15}$$

given that, under Assumption 3.8, it can be proved that the first term is $o_p(N^{-1/2})$ using the same argument as that in Lemma 2 in Gao (1995), whereas by Theorem 3.2 the second one is $o_p(1)$. Following this same procedure we get a similar result for $\mathbb{I}_v^{(4)}$.

Using these results in (6.13) and under Assumption 3.8,

$$\sqrt{N} (\widehat{\sigma}_v^2 - \sigma_v^2) = \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T (v_{it}^2 - \sigma_v^2) + o_{\mathbb{P}}(1)\tag{6.16}$$

and by the central limit theorem the first part of Theorem 3.3 is proved.

Focusing now on $\widehat{\sigma}_b^2$, (2.3) can be rewritten as

$$\widehat{\sigma}_b^2 = \frac{1}{NT(T-1)} \sum_{it} \sum_{s \neq t} \epsilon_{it} \epsilon_{is} - \mathbf{I}_b^{(1)} - \mathbf{I}_b^{(2)}, \quad (6.17)$$

where

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

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

$$\sqrt{N}(\widehat{\sigma}_b^2 - \sigma_b^2) = \frac{1}{\sqrt{NT}(T-1)} \sum_{i=1}^N \sum_{t=1}^T \sum_{s \neq t}^T (\epsilon_{it} \epsilon_{is} - \sigma_b^2) + o_{\mathbb{P}}(1) \quad (6.18)$$

from which it follows that the second part of Theorem 3.3 holds. ■

Proof of Theorem 3.4. Following the same argument as in the proof above, we first focus on the behavior of the estimators of the third order moments and later on the fourth order moments. As previously, using the same argument as that in proving Lemma 2 in Gao (1995), the result of Theorem 3.2 and Assumption 3.10, all contributions involving $X_{it}^\top (\widehat{m}(Z_{it}; H) - m(Z_{it}))$ may be neglected. Then, the final expressions to study are

$$\sqrt{N} \widehat{\gamma}_v^3 = \frac{1}{\sqrt{NT}} \sum_{it} \left[v_{it}^3 - \frac{3}{(T-1)} \sum_{s \neq t} v_{it}^2 v_{is} + \frac{2}{(T-1)(T-2)} \sum_{s \neq t} \sum_{r \neq s} v_{it} v_{is} v_{ir} \right] + o_{\mathbb{P}}(1) \quad (6.19)$$

and

$$\sqrt{N} \widehat{\gamma}_b^3 = \frac{1}{\sqrt{NT}(T-1)(T-2)} \sum_{it} \sum_{s \neq t} \sum_{r \neq s} \epsilon_{it} \epsilon_{is} \epsilon_{ir} + o_{\mathbb{P}}(1) \quad (6.20)$$

where, after centering v_{it}^3 and ϵ_{it}^3 , respectively, the first part of Theorem 3.4 is provided by the central limit theorem.

Similarly, if we again ignore the higher order terms of $X_{it}^\top (\widehat{m}(Z_{it}; H) - m(Z_{it}))$ for the estimators of the fourth order moment we get

$$\begin{aligned} \sqrt{N} \widehat{\gamma}_v^4 &= \frac{1}{\sqrt{NT}} \sum_{it} \left[v_{it}^4 - \frac{4}{(T-1)} \sum_{s \neq t} v_{it}^3 v_{is} + \frac{6}{(T-1)(T-2)} \sum_{s \neq t} \sum_{r \neq s} v_{it}^2 v_{is} v_{ir} \right. \\ &\quad \left. - \frac{3}{(T-1)(T-2)(T-3)} \sum_{s \neq t} \sum_{r \neq s} \sum_{h \neq r} v_{it} v_{is} v_{ir} v_{ih} \right] + o_{\mathbb{P}}(1) \end{aligned} \quad (6.21)$$

and

$$\sqrt{N}\widehat{\gamma}_b^4 = \frac{1}{\sqrt{NT}(T-1)(T-2)(T-3)} \sum_{it} \sum_{s \neq t} \sum_{r \neq s} \sum_{h \neq r} \epsilon_{it}\epsilon_{is}\epsilon_{ir}\epsilon_{ih} + o_{\mathbb{P}}(1). \quad (6.22)$$

Finally, as previously, after centering these expressions and using the central limit theorem, the second part of Theorem 3.4 holds. ■

Proof of Theorem 4.1. In order to prove the results of this theorem we follow the standard proofs scheme as in Bai and Ng (2005). First, we focus on the behavior of \widehat{SK}_v and later we analyze the properties of \widehat{SK}_b .

$$\widehat{SK}_v - SK_v = \left(\frac{\widehat{\gamma}_v^3 - \gamma_v^3}{\widehat{\sigma}_v^3} \right) - SK_v \left(\frac{(\widehat{\sigma}_v)^{3/2} - (\sigma_v^2)^{3/2}}{\widehat{\sigma}_v^3} \right).$$

For any estimator of the variance, by the delta method we get

$$\sqrt{N} \left((\widehat{\sigma}^2)^{k/2} - (\sigma^2)^{k/2} \right) = \frac{k}{2} (\sigma^2)^{k/2-1} \sqrt{N} (\widehat{\sigma}^2 - \sigma^2) + o_{\mathbb{P}}(1). \quad (6.23)$$

Then, for $k = 3$, we replace (6.16), (6.19) and (6.23) into the previous equation and rearranging terms the expression to analyze is

$$\sqrt{NT}(\widehat{SK}_v - SK_v) = \frac{\alpha_v^\top}{\widehat{\sigma}_v^3} \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T SK_{it}^v + o_{\mathbb{P}}(1), \quad (6.24)$$

where $\alpha_v = [1, \frac{3SK_v\sigma_v}{2}]^\top$ and $SK_{it}^v = [SK_{1it}^v, SK_{2it}^v]^\top$ are 2×1 vectors and

$$\begin{aligned} SK_{1it}^v &= (v_{it}^3 - \gamma_v^3) - \frac{3}{(T-1)} \sum_{s \neq t} v_{it}^2 v_{is} + \frac{2}{(T-1)(T-2)} \sum_{s \neq t} \sum_{r \neq s} v_{it} v_{is} v_{ir} \\ SK_{2it}^v &= (v_{it}^2 - \sigma_v^2) - \frac{1}{(T-1)} \sum_{s \neq t} v_{it} v_{is}. \end{aligned}$$

Under the assumptions of Theorem 4.1 it is easy to show that $E \left[\frac{1}{\sqrt{NT}} \sum_{it} SK_{it}^v \right] = 0$ and its variance-covariance matrix is

$$\Gamma_v = \begin{bmatrix} \Psi_{v3}^{(1)} & \Psi_{v3}^{(2)} \\ \Psi_{v3}^{(2)} & \Psi_{v3}^{(3)} \end{bmatrix}. \quad (6.25)$$

Specifically, under the assumptions of the theorem it is straightforward to show

$$\begin{aligned}
\Psi_{v3}^{(1)} &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \sum_{t'=1}^T E[SK_{1it}^v SK_{1it'}^v] \\
&= \gamma_v^6 - (\gamma_v^3)^2 - \left(\frac{6T-15}{T-1} \right) \gamma_v^4 \sigma_v^2 + \frac{9}{(T-1)} (\gamma_v^3)^2 + \left(\frac{24+9(T-2)^2}{(T-1)(T-2)} \right) \sigma_v^6, \\
\Psi_{v3}^{(2)} &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \sum_{t'=1}^T E[SK_{1it}^v SK_{2it'}^v] = \gamma_v^5 - \left(\frac{4T-10}{T-1} \right) \gamma_v^3 \sigma_v^2, \\
\Psi_{v3}^{(3)} &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \sum_{t'=1}^T E[SK_{2it}^v SK_{2it'}^v] = \gamma_v^4 - \sigma_v^4 + \frac{2\sigma_v^4}{(T-1)}.
\end{aligned}$$

By Theorem 3.3, $\widehat{\sigma}_v^2 \xrightarrow{p} \sigma_v^2$. Then, using this result and by the central limit theorem the first part of Theorem 4.1 is proved.

Similarly, focus on the properties of the skewness statistic for the individual effects and using (6.19), (6.20) and (6.23) for $k=3$, it can be written

$$\sqrt{N}(\widehat{SK}_b - SK_b) = \frac{\alpha_b^\top}{\widehat{\sigma}_b^3} \frac{1}{\sqrt{N}} \sum_{i=1}^N SK_i^b + o_{\mathbb{P}}(1), \quad (6.26)$$

where $\alpha_b = [1, \frac{3SK_b\sigma_b}{2}]^\top$ and $SK_{it}^b = [SK_{1i}^b, SK_{2i}^b]$ are 2×1 vectors and

$$\begin{aligned}
SK_{1i}^b &= (b_i^3 - \gamma_b^3) + \frac{3}{T} \sum_t b_i^2 v_{it} + \frac{3}{T(T-1)} \sum_t \sum_{s \neq t} b_i v_{it} v_{is} + \frac{1}{T(T-1)(T-2)} \sum_t \sum_{s \neq t} \sum_{r \neq s} v_{it} v_{is} v_{ir}, \\
SK_{2i}^b &= (b_i^2 - \sigma_b^2) + \frac{2}{T} \sum_t b_i v_{it} + \frac{1}{T(T-1)} \sum_t \sum_{s \neq t} v_{it} v_{is}.
\end{aligned}$$

Under the assumptions of this theorem, $E \left[\frac{1}{\sqrt{N}} \sum_{i=1}^N SK_i^b \right] = 0$ and its variance-covariance matrix is

$$\Gamma_b = \begin{bmatrix} \Psi_{b3}^{(1)} & \Psi_{b3}^{(2)} \\ \Psi_{b3}^{(2)} & \Psi_{b3}^{(3)} \end{bmatrix},$$

where it can be shown that

$$\begin{aligned}
\Psi_{b3}^{(1)} &= \frac{1}{N} \sum_i E[SK_{1i}^b SK_{1i}^b] = \gamma_b^6 - (\gamma_b^3)^2 + \frac{9}{T} \gamma_b^4 \sigma_b^2 + \frac{18}{T(T-1)} \sigma_b^2 \sigma_b^4 + \frac{6}{T(T-1)(T-2)} \sigma_b^6, \\
\Psi_{b3}^{(2)} &= \frac{1}{N} \sum_i E[SK_{1i}^b SK_{2i}^b] = \gamma_b^5 - \gamma_b^3 \sigma_b^2 + \frac{6}{T} \gamma_b^3 \sigma_b^2, \\
\Psi_{b3}^{(3)} &= \frac{1}{N} \sum_i E[SK_{2i}^b SK_{2i}^b] = \gamma_b^4 - \sigma_b^4 + \frac{4}{T} \sigma_b^2 \sigma_b^2 + \frac{2}{T(T-1)} \sigma_b^4.
\end{aligned}$$

Finally, using the results of Theorem 3.3, $\widehat{\sigma}_b^2 \xrightarrow{p} \sigma_b^2$, and by the central limit theorem, the second part of Theorem 4.1 holds.

■

Proof of Theorem 4.3. Focus on the properties of \widehat{KU}_v when $k = 4$, it can be written

$$\widehat{KU}_b - KU_b = \left(\frac{\widehat{\gamma}_v^4 - \gamma_v^4}{\widehat{\sigma}_v^4} \right) - KU_v \left(\frac{(\widehat{\sigma}_v^2)^2 - (\sigma_v^2)^2}{\widehat{\sigma}_v^4} \right)$$

and using (6.23), for $k = 4$, (6.16) and (6.21), the previous equation turns into

$$\sqrt{NT}(\widehat{KU}_v - KU_v) = \frac{\beta_v^\top}{\widehat{\sigma}_v^4} \frac{1}{\sqrt{NT}} \sum_{it} KU_{it}^v + o_{\mathbb{P}}(1), \quad (6.27)$$

where $\beta_v = [1, -2KU_v\sigma_v^2]^\top$ and $KU_{it}^v = [KU_{1it}^v, SK_{2it}^v]^\top$ are 2×1 vectors and

$$\begin{aligned} KU_{1it}^v &= (v_{it}^4 - \gamma_v^4) - \frac{4}{(T-1)} \sum_{s \neq t} v_{it}^3 v_{is} + \frac{6}{(T-1)(T-2)} \sum_{s \neq t} \sum_{r \neq s} v_{it}^2 v_{is} v_{ir} \\ &\quad - \frac{3}{(T-1)(T-2)(T-3)} \sum_{s \neq t} \sum_{r \neq s} \sum_{h \neq r} v_{it} v_{is} v_{ir} v_{ih}. \end{aligned}$$

Again, under the assumptions of Theorem 4.3 it can be proved that $E \left[\frac{1}{\sqrt{NT}} \sum_{it} KU_{it}^v \right] = 0$,

whereas $Var \left[\frac{1}{\sqrt{NT}} \sum_{it} KU_{it}^v \right]$ is of the form

$$\Omega_v = \begin{bmatrix} \Psi_{v4}^{(1)} & \Psi_{v4}^{(2)} \\ \Psi_{v4}^{(2)} & \Psi_{v4}^{(3)} \end{bmatrix},$$

where it is easy to show

$$\begin{aligned} \Psi_{v4}^{(1)} &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \sum_{t'=1}^T E[KU_{1it}^v KU_{1it'}^v] = \gamma_v^8 - (\gamma_v^4)^2 - 8\gamma_v^5 \gamma_v^3 + \frac{(16\gamma_v^6 \sigma_v^2 + 16(\gamma_v^4)^2)}{(T-1)} \\ &\quad + \frac{(16(T-2)^2 + 72)(\gamma_v^3)^2 \sigma_v^2}{(T-1)(T-2)} + \frac{(72 - 96(T-2))\gamma_v^4 \sigma_v^4}{(T-1)(T-2)} + \frac{72((T-3)^2 + 3)\sigma_v^8}{(T-1)(T-2)(T-3)}, \\ \Psi_{v4}^{(2)} &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \sum_{t'=1}^T E[KU_{1it}^v SK_{2it'}^v] = \gamma_v^6 - \left(\frac{T-9}{T-1} \right) \gamma_v^4 \sigma_v^4 - 4(\gamma_v^3)^2 - \frac{12}{(T-1)} \sigma_v^6, \\ \Psi_{v4}^{(3)} &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \sum_{t'=1}^T E[SK_{2it}^v SK_{2it'}^v] = \gamma_v^4 - \sigma_v^4 + \frac{2\sigma_v^4}{(T-1)}. \end{aligned}$$

Then, using Theorem 3.3 and the central limit theorem, the first part of the Theorem is proved.

Focus now on the behavior of \widehat{KU}_b , the expression to analyze is

$$\sqrt{N}(\widehat{KU}_b - KU_b) = \frac{\beta_b^\top}{\widehat{\sigma}_b^4} \frac{1}{\sqrt{N}} \sum_i KU_i^b + o_{\mathbb{P}}(1), \quad (6.28)$$

where $\beta_b = [1, -2KU_b\sigma_b^2]^\top$ and $KU_i = [KU_{i1}^b, SK_{i2}^b]^\top$ are 2×1 vectors of the form and

$$\begin{aligned} KU_{i1}^b &= (b_i^4 - \gamma_b^4) + \frac{4}{T} \sum_t b_i^3 v_{it} \frac{6}{T(T-1)} \sum_t \sum_{s \neq t} b_i^2 v_{it} v_{is} + \frac{4}{T(T-1)(T-2)} \sum_t \sum_{s \neq t} \sum_{r \neq s} b_i v_{it} v_{is} v_{ir} \\ &+ \frac{1}{T(T-1)(T-2)(T-3)} \sum_t \sum_{s \neq t} \sum_{r \neq s} \sum_{h \neq r} v_{it} v_{is} v_{ir} v_{ih}. \end{aligned}$$

Under the assumptions of this theorem, the mean of $\frac{1}{\sqrt{N}} \sum_i KU_i^b$ is zero and the variance-covariance matrix is

$$\Omega_b = \begin{bmatrix} \Psi_{b4}^{(1)} & \Psi_{b4}^{(2)} \\ \Psi_{b4}^{(2)} & \Psi_{b4}^{(3)} \end{bmatrix}, \quad (6.29)$$

where

$$\begin{aligned} \Psi_{b4}^{(1)} &= \frac{1}{N} \sum_i E[KU_{1i}^b KU_{1i}^b] = \gamma_b^8 - (\gamma_b^4)^2 + \frac{16}{T} \gamma_b^6 \sigma_v^2 + \frac{72}{T(T-1)} \gamma_b^4 \sigma_v^4 + \frac{96 \sigma_b^2 \sigma_v^6}{T(T-1)(T-2)} \\ &+ \frac{24 \sigma_v^8}{T(T-1)(T-2)(T-3)}, \\ \Psi_{b4}^{(2)} &= \frac{1}{N} \sum_i E[KU_{1i}^b SK_{2i}^b] = \gamma_b^6 - \gamma_b^4 \sigma_v^2 + \frac{8}{T} \gamma_b^4 \sigma_v^2 + \frac{12}{T(T-1)} \sigma_b^2 \sigma_v^4, \\ \Psi_{b4}^{(3)} &= \frac{1}{N} \sum_i E[SK_{2i}^b SK_{2i}^b] = \gamma_b^4 - (\sigma_b^2)^2 + \frac{4}{T} \sigma_b^2 \sigma_v^2 + \frac{2 \sigma_v^4}{T(T-1)}. \end{aligned}$$

Finally, using the results of Theorem 3.3, $\hat{\sigma}_b^2 \xrightarrow{p} \sigma_b^2$, and by the central limit theorem, the second part of Theorem 4.3 holds. ■

References

- Aigner, D., C. Lovell, and P. Schmidt (1977). “Formulation and estimation of stochastic frontier production function models”. *Journal of Econometrics* 6, 21–37.
- Almanidis, P. and R. Sickles (2011). The skewness issue in stochastic frontiers models: fact or fiction? In I. van Keilegom and P. Wilson (Eds.), *Exploring research frontiers in contemporary statistics and econometrics*, Chapter 10, pp. 201–227. Springer-Verlag Berlin Heidelberg.
- Bai, J. and S. Ng (2005). Tests for skewness, kurtosis, and normality for time series data. *Journal of Business & Economic Statistics* 23, 49–60.
- Carree, M. (2002). Technological inefficiency and the skewness of the error component in stochastic frontier analysis. *Empirical Letters* 77, 101–107.

- 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 effects model with nonnormal distributions. *Biometrika* 89, 831–840.
- Dufour, J., L. Khlaf, and M. Beaulieu (2003). Exact skewness-kurtosis tests for multivariate normality and goodness-of-fit in multivariate regressions with application to asset pricing models. *Oxford Bulletin of Economics and Statistics* 65, 891–906.
- Fan, J. and I. Gijbels (1995). *Local polynomial modelling and its applications*. Chapman & Hall.
- Galvao, A., G. Montes-Rojas, W. Sosa-Escudero, and L. Wang (2013). Tests for skewness and kurtosis in the one-way error component model. *Journal of Multivariate Analysis* 122, 35–52.
- Gao, J. (1995). The laws of the iterated logarithm of some estimates in a partly linear models. *Statistics & Probability Letters* 25, 153–162.
- Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* 24, 726–748.
- Henderson, D. J., R. J. Carroll, and Q. Li (2008). Nonparametric estimation and testing of fixed effects panel data model. *Journal of Econometrics* 144, 257–275.
- 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.
- Horowitz, J. L. and M. Markatou (1996). Semiparametric estimation of regression models for panel data. *The Review of Economic Studies* 63, 145–168.
- Jarque, C. and A. Bera (1981). Efficient tests for normality, homoscedasticity and serial independence of regression residuals: Monte Carlo evidence. *Economics Letters* 7, 313–318.
- Kumbhakar, S. C. (2002). Specification and estimation of production risk, risk preferences and technical efficiency. *American Journal of Agricultural Economics* 84, 8–22.
- Li, Q., J. H. Cliff, D. Li, and T. Fu (2002). Semiparametric smooth coefficient models. *Journal of Business & Economic Statistics* 20, 412–422.

- Masry, E. (1996). Multivariate local polynomial regression for time series: uniform strong consistency rate. *Journal of Time Series Analysis* 17, 571–599.
- Montes-Rojas, G. and W. Sosa-Escudero (2011). Robust tests for heteroskedasticity in the one-way error components model. *Journal of Econometrics* 160, 300–310.
- Premaratne, G. and A. Bera (2005). A test for symmetry with leptokurtic financial data. *Journal of Financial Econometrics* 3, 169–187.
- Qian, J. and L. Wang (2012). Estimating semiparametric panel data models by marginal integration. *Journal of Econometrics* 167, 483–493.
- 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.
- Rodriguez-Poo, J. M. and A. Soberon (2016). Nonparametric and Semiparametric Panel Data Models: Recent Developments. *Forthcoming in Journal of Economic Surveys*.
- Ruppert, D. and M. P. Wand (1994). Multivariate locally weighted least squares regression. *The Annals of Statistics* 22(3), 1346–1370.
- Simar, L. and P. Wilson (2010). Inference from cross-sectional stochastic frontier models. *Econometric Reviews* 29, 62–98.
- 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 (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.
- Sun, Y., R. J. Carroll, and D. Li (2009). Semiparametric estimation of fixed effects panel data varying coefficient models. *Advances in Econometrics* 25, 101–130.
- Wang, N. (2003). Marginal nonparametric kernel regression accounting for within-subject correlation. *Biometrika* 90, 43–52.
- Wu, P., W. Stute, and L. Zhu (2012). Efficient estimation of moments in linear mixed models. *Bernoulli* 18, 206–228.