Testing for distributional features in varying coefficient panel data models

Alexandra Soberon[1], Winfried Stute[2] and Juan M. Rodriguez-Poo[1]

[1]Department of Economics, University of Cantabria, E-39005 Santander, Spain

[2]Mathematical Institute, University of Giessen, D-35392 Giessen, Germany

January 30, 2019

Abstract

This paper provides several tests for skewness and kurtosis for the error terms in a one-way fixed-effects varying coefficient panel data model. To obtain these tests, estimators of higher-order moments of both error components are obtained as solutions of estimating equations. Additionally, to obtain the nonparametric residuals, a local constant estimator based on a pairwise differencing transformation is proposed. The asymptotic properties of these estimators and tests are established. The proposed estimators and test statistics are augmented by simulation studies, and they are also illustrated in an empirical analysis regarding the technical efficiency of European Union companies.[1]

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

1 Introduction

Testing for skewness and kurtosis is a relevant topic in many fields in economics, among others in finance and productivity analysis. In finance, commonly used financial models, such as the capital asset pricing model (see Sharpe (1964) and Lintner (1965)) and the options pricing model (see Black and Sholes (1973)) are developed based on the assumption of symmetry. Furthermore, in Mandelbrot (1963) has been observed the presence of leptokurtosis in the empirical distribution of price changes. This fact has motivated the development of financial data models based on nonnormal distributions. In productivity analysis, one main stream to introduce inefficiency in production function models has been the use of stochastic frontier models (see Aigner et al. (1977), Meeusen and van den Broeck (1977) and Battese and Coelli (1988) for panel data). These models decompose the error structure in the econometric model into two terms, an idiosyncratic symmetric error component and an asymmetric error term that accounts for the inefficiency. In this framework, a test for skewness in this second error component is of great interest because it is equivalent to a test for inefficiency at firm levels. Given the importance of skewness and kurtosis in these fields of economics, it is useful to have tests and, in general, estimators of higher-order moments that can correctly identify these features.

In the context of cross-sectional and time series data, there is an extensive literature that studies the issues raised above. See Bai and Ng (2005), Dufour et al. (2003), Jarque and Bera (1981), Montes-Rojas and Sosa-Escudero (2011) and Premaratne and Bera (2005), among others. In the panel data framework, techniques to estimate higher-order moments and tests of skewness and kurtosis for the different random components are much more scarce. A natural complication with detecting departures away from normality or skewness is the identification of which component is causing the departure. Assuming that the regression function is fully parametric, in Horowitz and Markatou (1996) the densities of the error components are estimated nonparametrically, but they avoid testing the distributional features of these terms. In this same context, in Wu et al. (2012) and Galvao et al. (2013)

estimators for higher-order moments and tests for skewness and kurtosis are derived. The asymptotic properties of these estimators are not robust either to the presence of correlation between the individual heterogeneity term and the explanatory variables (i.e., fixed-effects) or to misspecification of the (parametric) form of the regression function.

This paper extends the contributions in Cox and Hall (2002) and Wu et al. (2012); we propose estimators of higher-order moments and tests of skewness and kurtosis for the different components of a fixed-effects panel data model where the regression function has the form of a nonparametric varying coefficient model. It turns out that by a proper combination of polynomial functions of the residuals, we can obtain higher-order moment estimators, which are asymptotically normal and have the same limit variance as if the unknown errors were known. 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 estimators 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. Varying coefficient models are currently very common in the specification of econometric models, which is due to several reasons: First, varying coefficient models encompass a great variety of econometric models, such as partially linear models. Second, they mitigate the "curse of dimensionality". Third, they have been justified on the grounds of economic theory; see Chamberlain (1992). To the best of our knowledge, this is a completely new proposal, and the easy-to-compute closed-form expressions of these estimators can be used in many fields such as nonlinear, semi-parametric or nonparametric panel data models.

We would like to emphasize that our proposal to estimate higher-order moments and the battery of tests could be also based on root-N consistent residuals obtained from a fully parametric model. To our knowledge, this simpler specification has not been studied yet. However, we have chosen to use nonparametric residuals to generalize our results to

estimators that are more robust to misspecification in the conditional mean. To assess the finite sample performance of the proposed estimators and test statistics of this paper, a Monte Carlo study is conducted. Finally, an empirical study on the production efficiency of European Union (EU) companies is implemented.

The rest of the paper is organized as follows: In Section 2, we set up the model of interest, we introduce the pairwise difference estimator and we give its asymptotic properties. In Section 3, we present the estimators of higher-order moments and we provide their asymptotic properties. 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 detailed mathematical proofs of the main results and additional Monte Carlo results are collected in the supplement, Appendix C.

2 Econometric model and estimation procedure

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

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

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 heterogeneity effect b_i, and an idiosyncratic error term v_{it}. Assumptions about all components of the model will be formally stated below.

As is well known, differencing techniques are usually used to remove the unobserved individual heterogeneity from the regression model to be estimated. 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),

4

Qian and Wang (2012), and Rodriguez-Poo and Soberon (2015), among others.

To overcome these difficulties, a very appealing alternative pairwise differencing transformation is proposed in this section. First, this approach removes the individual effects from the regression model to be estimated. Second, it enables us to obtain some gains in efficiency because this transformation considers all time-dependencies within the observations of each individual. Third, the resulting estimator almost achieves optimality in only one step.

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

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

Nevertheless, the application of any standard nonparametric technique in (2.2) ends up with a non-negligible asymptotic bias. See Rodriguez-Poo and Soberon (2015) for a more detailed explanation about this issue. To solve it, in this paper we propose to estimate the quantities of interest by defining a kernel weight, which controls the distance between any (Z_{it}, Z_{is}). Thus, for a given point $z \in \mathbb{R}^q$ and for Z_{it} and Z_{is} in a neighborhood of z, the unknown $\beta = m(z)$ can be estimated by minimizing the following local constant criterion function:

$$\sum_{i=1}^N \sum_{t=1}^{T-1} \sum_{s=t+1}^T \left(\widetilde{Y}_{its} - \widetilde{X}_{its}^\top \beta \right)^2 K_H(Z_{it} - z) K_H(Z_{is} - z), \quad (2.3)$$

with respect to β, where $\widetilde{Y}_{its} = Y_{it} - Y_{is}$ and $\widetilde{X}_{its} = X_{it} - X_{is}$. See Fan and Gijbels (1995) and Ruppert and Wand (1994) for a detailed description of this technique. Note that H is a $q \times q$ symmetric positive-definite bandwidth matrix that needs to be selected empirically 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 $\widehat{\beta}$ be the minimizer of (2.3). It is equal to

$$\widehat{m}(z; H) = S_{\widetilde{X}\widetilde{X}}^{-1}(z)S_{\widetilde{X}\widetilde{Y}}(z), \tag{2.4}$$

where

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

and

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

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

To investigate the asymptotic properties of the nonparametric estimator of the varying coefficient function, we consider the following assumptions. Some additional notation is needed as well. More precisely, let $v_{\cdot t} = (v_{1t}, \ldots, v_{Nt})^{\top}$ and $\eta_{\cdot t} = (\eta_{1t}, \ldots, \eta_{Nt})^{\top}$ be vectors of $N \times 1$ dimension. Additionally, for a matrix A we have $\|A\| = \sqrt{tr(A^{\top}A)}$.

Assumption 2.1 *Let $\{(Z_{it}, X_{it})\}$ be a set of independent and identically distributed (i.i.d.) $\mathbb{R}^{q+d}$-random variables in the subscript i for each fixed t, and strictly stationary over t for fixed i.*

Assumption 2.2 *Let $v_{\cdot t} = \Omega\eta_{\cdot t}$ be a $N \times 1$ dimensional vector, where Ω is a $N \times N$ nonstochastic definite positive weighting matrix that does not contain unknown parameters and whose elements are known by the researcher. The random error η_{it} is i.i.d. for i and t with zero mean and finite variance σ_η^2. For some $\delta > 0$, $E|\eta_{it}|^{4+\delta} < \infty$ and η_{it} is independent of X_{it} and Z_{it} for all i and t. Let ω_{ij} be the (i,j) element of Ω, and assuming $\overline{\lim}_{N\to\infty} N^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} \omega_{ij}^2 < \infty$.*

Assumption 2.3 Z_{it} has a bounded density function $f_{Z_{i1}}(\cdot)$ and it is continuously differentiable in all its arguments, at any point of its support. Furthermore, the joint density of distinct elements of (Z_{it}, Z_{is}), for $s > t$, is bounded and continuously differentiable in all its arguments, at any point of its support.

Assumption 2.4 For $\kappa = |t - s|$, where $\kappa \in \{1, \ldots, (T - 1)\}$, the matrix functions $E\left[\|\widetilde{X}_{i11}\widetilde{X}_{i11}^\top\|^2 | Z_{i1} = z_1, Z_{i(1+\kappa)} = z_2\right]$, $E\left[\widetilde{X}_{i1}\widetilde{X}_{i1(1+\kappa)}^\top | Z_{i1} = z_1, Z_{i(1+\kappa)} = z_2\right]$, and $E\left[\widetilde{X}_{i1(1+\kappa)}X_{i1}^\top | Z_{i1} = z_1, Z_{i(1+\kappa)} = z_2\right]$ are bounded and uniformly continuous at any interior point, (z_1, z_2), in the support of $f_{Z_{i1}, Z_{i\kappa}}(z_1, z_2)$.

Assumption 2.5 For $\kappa = |t - s|$, where $\kappa \in \{1, \ldots, (T - 1)\}$, the matrix function $E\left[X_{i1}\widetilde{X}_{i1(1+\kappa)}^\top | Z_{it} = z_1, Z_{is} = z_2\right]$ is positive definite at any interior point, (z_1, z_2), in the support of $f_{Z_{i1}, Z_{i(1+\kappa)}}(z_1, z_2)$.

Assumption 2.6 All second-order derivatives of $m_1(\cdot), m_2(\cdot), \ldots, m_d(\cdot)$ are bounded and uniformly continuous at any interior point in the support of $f_{Z_{i1}}(\cdot)$.

Assumption 2.7 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, $\int u_1^{\iota_1} \ldots u_q^{\iota_q} K(u)du = 0$, for all non-negative integers $\iota_1, \ldots, \iota_q$ such that their sum is odd.

Assumption 2.8 The bandwidth matrix H is symmetric and strictly definite positive. Additionally, as $N \to \infty$ each entry of the matrix tends to zero in such a way that $N|H| \to \infty$.

Assumption 2.9 For some $\delta > 0$, the function $E\left[|X_{i1}X_{i1}^\top|^{2+\delta} | Z_{i1} = z_1, Z_{i(1+\kappa)} = z_2\right]$ is bounded and uniformly continuous at any interior point, (z_1, z_2), in the support of $f_{Z_{i1}, Z_{i(1+\kappa)}}(z_1, z_2)$.

Assumption 2.1 is rather standard in panel data analysis although other time-dependence settings could be considered, such as strong mixing conditions or nonstationary time series. However, since the asymptotic properties of the proposed estimator are analyzed for panels with large cross-sections and fixed time-series dimensions, it is sufficient to assume strict stationarity. Assumption 2.2 determines the behavior of the cross-sectional dependence, and it is slightly weaker than condition C.3 in Bai and Ng (2002). It implies that the largest eigenvalue (and hence all of the eigenvalues) of $\Omega\Omega^\top$ are bounded by $\max_i \sum_{j=1}^N \omega_{ij}^2$. Assumptions 2.3, 2.4 and 2.6 are basically smoothness and boundedness conditions on the density function and moments functionals. Assumption 2.5 is a generalization of the rank condition of parametric models that guarantees that $m(\cdot)$ is identified. Further, since t and s are integers, this assumption provides us with a lower positive bound for all t, s, when $t < s$. Assumptions 2.7 and 2.8 are standard in the literature of the local linear regression for the kernel function and bandwidth matrix. Finally, Assumptions 2.2 and 2.9 guarantee that a multivariate version of the Lindeberg-Lévy central limit theorem for $N \to \infty$ 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 2.1 *Under Assumptions 2.1-2.9, as N tends to infinity and T is fixed,*

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

where

$$
\begin{aligned}
B(z; H) &= \mu_2(K)\mathcal{B}_{X\widetilde{X}}^{-1}(z, z)\mathcal{B}_{X\widetilde{X}}(z, z)\left[diag_d\left(tr(HD_f(z)D_{m_r}(z))\right)\imath_d f_{Z_{i1}, Z_{i(1+\kappa)}}^{-1}(z, z)\right. \\
&\quad + \left.\frac{1}{2}diag_d\left(tr(H\mathcal{H}_{m_r}(z))\right)\imath_d\right], \\
V(z; H) &= \sigma_\eta^2 \omega_N R^2(K)\mathcal{B}_{X\widetilde{X}}^{-1}(z, z)\mathcal{B}_{X\widetilde{X}}^V(z, z)\mathcal{B}_{\widetilde{X}X}^{-1}(z, z),
\end{aligned}
$$

and $\omega_N = N^{-1}\sum_{i=1}^N \sum_{j=1}^N \omega_{ij}^2$. For $r = 1, \ldots, d$, D_{m_r} is the first-order derivative vector

of the rth component of $m(\cdot)$, $\mathcal{H}_{m_r}(z)$ the Hessian matrix, $D_f(z)$ the first-order derivative vector of the density function, and

$$\mathcal{B}_{X\widetilde{X}}(z,z) = \sum_{\kappa=1}^{T-1} \left(1 - \frac{\kappa}{T}\right) E\left[X_{i1}\widetilde{X}_{i1(1+\kappa)}^{\top} | Z_{i1} = z, Z_{i(1+\kappa)} = z\right] f_{Z_{i1}, Z_{i(1+\kappa)}}(z,z),$$

$$\mathcal{B}_{\widetilde{X}X}(z,z) = \sum_{\kappa=1}^{T-1} \left(1 - \frac{\kappa}{T}\right) E\left[\widetilde{X}_{i1(1+\kappa)}X_{i1}^{\top} | Z_{i1} = z, Z_{i(1+\kappa)} = z\right] f_{Z_{i1}, Z_{i(1+\kappa)}}(z,z),$$

$$\mathcal{B}_{X\widetilde{X}}^{V}(z,z) = \sum_{\kappa=1}^{T-1} \left(1 - \frac{\kappa}{T}\right)^2 E\left[X_{i1}\widetilde{X}_{i1(1+\kappa)}^{\top} | Z_{i1} = z, Z_{i(1+\kappa)} = z\right] f_{Z_{i1}, Z_{i(1+\kappa)}}(z,z).$$

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

The proof of this theorem is postponed to the supplement, Appendix A. Furthermore, in order to obtain the convergence results for higher-order moments we need the following result:

Theorem 2.2 *Under Assumptions 2.1-2.9, as N tends to infinity and T is fixed,*

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

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 2.1 are rather standard. However, there are some differences that need to be pointed out. More precisely, as far as we have more curvature in $m(\cdot)$, the bias is enlarged. On its part, the variance will be penalized when H is large and 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 that almost achieves the optimal rate of convergence of this type of problems, i.e., $N|H|^{1/2}$.

9

3   Estimation of moments

To achieve one of the main aims of this paper, that is, to propose and develop some tests for symmetry and kurtosis for both the individual heterogeneity effects and the idiosyncratic error term, higher-order moments are needed. We will assume the following:

Assumption 3.1 *The η_{it}'s are random variables with finite 8-th order moments, i.e. $\gamma_\eta^8 < \infty$ and the b_i's are i.i.d. zero-mean random variables also with finite 8-th order moments, i.e., $\gamma_b^8 < \infty$. Moreover, b_i is independent of all η_{it}. Additionally, for fixed i, there is a $\delta > 0$ such that $E\|X_{it}^k v_{it}\|^{2+\delta} < \infty$ and $E\|X_{it}^k b_i\|^{2+\delta} < \infty$, for $k = 2, \ldots, 8$.*

Assumption 3.1 is needed to bound the higher-order moments related to the residuals that were obtained in a nonparametric framework. Furthermore, we have included an assumption about the behavior of b_i. We could have included it before, but we want to point out that indeed, no assumption about b_i is needed to obtain the results in Theorem 2.1. This type of assumption is only needed to show the asymptotic results that follow.

Let us now introduce a new technique to obtain $\sqrt{N}$-consistent estimators for these moments. To do so following the same ideas as in Wu et al. (2012), we present special nonlinear functions of the composite error terms that will be important tools to derive estimating equations for

$$\gamma_b^k = E\left(b_i^k\right) \quad \text{and} \quad \gamma_v^k = E\left(v_{it}^k\right) \quad i = 1, \ldots, N; \quad t = 1, \ldots, T,$$

where $k \in \{2, \ldots, 8\}$ is the k-th order moment in which we are interested. Using Assumption 3.1, the following set of nonlinear functions is introduced:

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

Moreover, to obtain the suitable combination that provides the estimators for the higher-order moments, the following lemma is crucial:

Lemma 3.1 *Let $a \wedge b$ and $a \vee b$ be the minimum and maximum, respectively, of two real numbers a and b, we have*

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

This lemma is based on simple arithmetic. When we take expectations, usually many of the terms in the expansion of $f_j^k(i)$ will be set equal to zero, mainly because the v_{it}'s and b_i's are centered and independent. Moreover, by taking proper linear combinations of the functions, $f_j^k(i)$, we should be able to represent the moments γ_b^k and γ_v^k in terms of the functions obtaining the estimating equations that will lead to the associated estimators. For example, if we are interested in the estimation of second order moments, i.e. $\sigma_v^2 \equiv \gamma_v^2 = E(v_{it}^2)$ and $\sigma_b^2 \equiv \gamma_b^2 = E(b_i^2)$, using Lemma 3.1, we obtain

$$f_2^2(i) = \sum_{t=1}^{T} v_{it}^2 + T b_i^2 + 2 b_i \sum_{t=1}^{T} v_{it},$$

$$f_1^2(i) = T^2 b_i^2 + 2 T b_i \sum_{t=1}^{T} v_{it} + \left(\sum_{t=1}^{T} v_{it} \right)^2.$$

Hence, in terms of the f's, σ_v^2 can be represented as

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

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 $1 \le i \le N$ and replacing the unknown ϵ_{it} with the residuals $\widehat{\epsilon}_{it}$, the estimator of σ_v^2 has the form

$$\widehat{\sigma}_v^2 = \frac{1}{NT(T-1)} \sum_{i=1}^{N} \left[T \sum_{t=1}^{T} \widehat{\epsilon}_{it}^2 - \left(\sum_{t=1}^{T} \widehat{\epsilon}_{it} \right)^2 \right], \tag{3.1}$$

where $\widehat{\epsilon}_{it} = Y_{it} - X_{it}^{\top} \widehat{m}(Z_{it}; H)$ and $\widehat{m}(Z_{it}; H)$ is the pairwise differencing estimator.

11

Similarly, combining these expressions as

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

the following estimator for σ_b^2 can be proposed

$$\widehat{\sigma}_b^2 = \frac{1}{NT(T-1)} \sum_{i=1}^{N} \left[\left(\sum_{t=1}^{T} \widehat{\epsilon}_{it} \right)^2 - \sum_{t=1}^{T} \widehat{\epsilon}_{it}^2 \right]. \tag{3.2}$$

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}. Focusing on the estimation of higher-order moments, i.e., γ_b^k and γ_v^k for $k = 3, 4, \ldots$, things are somewhat more complex than before. In these particular cases, there are several $f_j^k(i)$ to combine, and some of them can lead to inefficient estimators. In the following, we present the suitable combinations that provides efficient estimators.

Considering the estimation of the third-order moments. From Lemma 3.1, and proceeding in the same way as before to obtain (3.1) and (3.2) we now obtain

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

from which the estimator for γ_v^3 has the form

$$\widehat{\gamma}_v^3 = \frac{1}{NT(T-1)(T-2)} \sum_{i=1}^{N} \left[2 \left(\sum_{t=1}^{T} \widehat{\epsilon}_{it} \right)^3 + T^2 \sum_{t=1}^{T} \widehat{\epsilon}_{it}^3 - 3T \left(\sum_{t=1}^{T} \widehat{\epsilon}_{it}^2 \right) \sum_{t=1}^{T} \widehat{\epsilon}_{it} \right]. \tag{3.3}$$

Similarly, for γ_b^3, we obtain the following estimating equation:

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

Thus, the resulting estimator for γ_b^3 is

$$\widehat{\gamma}_b^3 = \frac{1}{NT(T-1)(T-2)} \sum_{i=1}^{N} \left[\left(\sum_{t=1}^{T} \widehat{\epsilon}_{it} \right)^3 - 3 \left(\sum_{t=1}^{T} \widehat{\epsilon}_{it}^2 \right) \sum_{t=1}^{T} \widehat{\epsilon}_{it} + 2 \sum_{t=1}^{T} \widehat{\epsilon}_{it}^3 \right]. \qquad (3.4)$$

For the fourth-order moment, in addition to the $f_j^k(i)$ with $j \leq k$, we need

$$f_5^4(i) = \left(\sum_{t=1}^{T} \epsilon_{it}^2 \right)^2.$$

The proper combination that provides the efficient estimator for γ_v^4 is

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

Therefore, the resulting estimator for γ_v^4 is

$$\begin{aligned}
\widehat{\gamma}_v^4 &= \frac{1}{NT(T-1)(T-2)(T-3)} \sum_{i=1}^{N} \left[(T^2 - 2T + 3) \left(T \sum_{t=1}^{T} \widehat{\epsilon}_{it}^4 - 4 \sum_{t=1}^{T} \widehat{\epsilon}_{it}^3 \sum_{t=1}^{T} \widehat{\epsilon}_{it} \right) \right. \\
&\quad + \left. 6T \sum_{t=1}^{T} \widehat{\epsilon}_{it}^2 \left(\sum_{t=1}^{T} \widehat{\epsilon}_{it} \right)^2 - 3 \left(\sum_{t=1}^{T} \widehat{\epsilon}_{it} \right)^4 - 3(2T-3) \left(\sum_{t=1}^{T} \widehat{\epsilon}_{it}^2 \right)^2 \right]. \qquad (3.5)
\end{aligned}$$

Meanwhile, for γ_b^4 we propose the following estimating equation

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

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

$$\begin{aligned}
\widehat{\gamma}_b^4 &= \frac{1}{NT(T-1)(T-2)(T-3)} \sum_{i=1}^{N} \left[\left(\sum_{t=1}^{T} \widehat{\epsilon}_{it} \right)^4 - 6 \sum_{t=1}^{T} \widehat{\epsilon}_{it}^2 \left(\sum_{t=1}^{T} \widehat{\epsilon}_{it} \right)^2 + 8 \sum_{t=1}^{T} \widehat{\epsilon}_{it}^3 \sum_{t=1}^{T} \widehat{\epsilon}_{it} \right. \\
&\quad - \left. 6 \sum_{t=1}^{T} \widehat{\epsilon}_{it}^4 + 3 \left(\sum_{t=1}^{T} \widehat{\epsilon}_{it}^2 \right)^2 \right]. \qquad (3.6)
\end{aligned}$$

The estimators given in (3.5) and (3.6) have been obtained also using Lemma 3.1 and proceeding in the same way as before to obtain (3.1) and (3.2).

The next theorem contains the main statistical properties of the variance estimators of both random error and unobserved individual heterogeneity. For the sake of simplicity, we assume that Ω is an identity matrix. Of course, these results can be extended to specific error structures, such as in Assumption 2.2, but at the price of considerably enlarging the complexity of the proofs.

Theorem 3.1 *Under Assumptions 2.1-2.9 and 3.1, as $N \to \infty$ and T is fixed, we have*

$$\sqrt{N}(\widehat{\sigma}_v^2 - \sigma_v^2) \xrightarrow{d} \mathcal{N}(0, \mu_{2,v}), \tag{3.7}$$

and

$$\sqrt{N}(\widehat{\sigma}_b^2 - \sigma_b^2) \xrightarrow{d} \mathcal{N}(0, \mu_{2,b}), \tag{3.8}$$

where $\mu_{2,v} = \frac{\gamma_v^4}{T} - \frac{\sigma_v^4}{T} + \frac{\sigma_v^4}{T(T-1)}$ *and* $\mu_{2,b} = \gamma_b^4 - \sigma_b^4 + \frac{4}{T}\sigma_b^2\sigma_v^2 + \frac{\sigma_v^4}{T(T-1)}$.

The proof of this result is shown in the supplement, Appendix A. It is interesting to note that (3.7) and (3.8) will be shown by verifying

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

and

$$\sqrt{N}(\widehat{\sigma}_b^2 - \sigma_b^2) = \frac{1}{\sqrt{N}}\sum_i\left(b_i^2 - \sigma_b^2\right) + \frac{2}{\sqrt{NT}}\sum_{it}b_i v_{it} + \frac{1}{\sqrt{NT(T-1)}}\sum_{its}v_{it}v_{is} + o_{\mathbb{P}}(1).$$

In other words, $\widehat{\sigma}_v^2$ and $\widehat{\sigma}_b^2$ are as efficient as the moment estimators based on the true (but unknown) v_{it} and b_i.

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 3.2 *Under Assumptions 2.1-2.9 and 3.1, as $N \to \infty$ and T is fixed,*

$$\sqrt{N}(\widehat{\gamma}_v^3 - \gamma_v^3) \quad \overset{d}{\longrightarrow} \quad \mathcal{N}(0, \mu_{3,v}) \quad , \quad \sqrt{N}(\widehat{\gamma}_b^3 - \gamma_b^3) \quad \overset{d}{\longrightarrow} \quad \mathcal{N}(0, \mu_{3,b}),$$

and

$$\sqrt{N}(\widehat{\gamma}_v^4 - \gamma_v^4) \quad \overset{d}{\longrightarrow} \quad \mathcal{N}(0, \mu_{4,v}) \quad , \quad \sqrt{N}(\widehat{\gamma}_b^4 - \gamma_b^4) \quad \overset{d}{\longrightarrow} \quad \mathcal{N}(0, \mu_{4,b}),$$

where $\mu_{3,b} = \gamma_b^6 - (\gamma_b^3)^2 + \frac{9}{T}\gamma_b^4\sigma_v^2 + \frac{9\sigma_b^2\sigma_v^4}{T(T-1)} + \frac{\sigma_v^6}{T(T-1)(T-2)}$, $\mu_{3,v} = \frac{\gamma_v^6}{T} - \frac{(\gamma_v^3)^2}{T} - \left(\frac{6T-15}{2T(T-1)}\right)\gamma_v^4\sigma_v^2 + \frac{9(\gamma_v^3)^2}{2T(T-1)} + \left(\frac{3(T-2)^2+8}{2T(T-1)(T-2)}\right)\sigma_v^6$, $\mu_{4,b} = \gamma_b^8 - (\gamma_b^4)^2 + \frac{16}{T}\gamma_b^6\sigma_v^2 + \frac{36\gamma_b^4\sigma_v^4}{T(T-1)} + \frac{16\sigma_b^2\sigma_v^6}{T(T-1)(T-2)} + \frac{\sigma_v^8}{T(T-1)(T-2)(T-3)}$, $\mu_{4,v} = \frac{\gamma_v^8}{T} - \frac{(\gamma_v^4)^2}{T} - \frac{4}{T}\gamma_v^5\gamma_v^3 + \frac{8}{T(T-1)}\gamma_v^6\sigma_v^2 + \frac{8}{T(T-1)}(\gamma_v^4)^2 + \left(\frac{8(T-2)^2+36}{3T(T-1)(T-2)}\right)(\gamma_v^3)^2\sigma_v^2 - \left(\frac{16T-20}{T(T-1)(T-2)}\right)\gamma_v^4\sigma_v^4 + \left(\frac{3(T-3)^2+9}{T(T-1)(T-2)(T-3)}\right)\sigma_v^8$.

The proof of this result is shown in the supplement, Appendix A. For second moments, these quantities denote the minimum variances, which may be achieved for empirical estimators based on the true v_{it} and b_i, respectively. The main interest of Theorems 3.1 and 3.2 is that it is possible to make inference on several moments of both b_i and v_{it} without having to assume a prespecified distribution. Only the existence of higher-order moments is needed. Furthermore, an-easy-to-compute expression for the estimators of the asymptotic variances is provided just by plugging consistent estimators of the fifth-, sixth- and eighth-order moments of v_{it} and b_i into the expressions for $\mu_{2,v}$, $\mu_{2,b}$, $\mu_{3,v}$, $\mu_{3,b}$, $\mu_{4,v}$ and $\mu_{4,b}$ given in the theorems. Indeed, they can be easily obtained using Lemma 3.1. These consistent estimators are given in the supplement, Appendix B. Under the assumptions of Theorem 4.4 it is possible to show that $\widehat{\gamma}_v^5 \overset{p}{\to} \gamma_v^5$, $\widehat{\gamma}_v^6 \overset{p}{\to} \gamma_v^6$, and $\widehat{\gamma}_v^8 \overset{p}{\to} \gamma_v^8$, as $N \to \infty$ and T is fixed. Similar results are obtained for b_i. The asymptotic variances of the test statistics of next section are estimated in the same way. Finally, note that the properties of these estimators would hold also for root-N residuals. That is, for a econometric model where the regression function is fully parametric.

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

To test for skewness in both the individual heterogeneity and the remainder components, the quantities of interest for v_{it} and b_i 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 any value of SK_v and SK_b. Further, the simplicity of this result enables us to propose a test for symmetry, that is, testing the null hypothesis of $SK_v = 0$ and/or $SK_b = 0$.

In this case, the proposed test statistics would be

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

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 estimators for γ_v^3, γ_b^3, σ_v^2, and σ_b^2, respectively.

Theorem 4.1 *Under Assumptions 2.1-2.9 and 3.1, as $N \to \infty$ and T is fixed, we have*

$$\sqrt{N}(\widehat{SK}_v - SK_v) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, \frac{\alpha_v^\top \Gamma_v \alpha_v}{T\sigma_v^6}\right),$$

$$\sqrt{N}(\widehat{SK}_b - SK_b) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, \frac{\alpha_b^\top \Gamma_b \alpha_b}{\sigma_b^6}\right)$$

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

$$\Gamma_v = \begin{pmatrix} \mu_{3,v} & \mu_{32,v} \\ \mu_{32,v} & \mu_{2,v} \end{pmatrix} \quad \text{and} \quad \Gamma_b = \begin{pmatrix} \mu_{3,b} & \mu_{32,b} \\ \mu_{32,b} & \mu_{2,b} \end{pmatrix},$$

where $\mu_{32,v} = \frac{\gamma_v^5}{T} - \left(\frac{5T-11}{2T(T-1)}\right)\sigma_v^2\gamma_v^3$ and $\mu_{32,b} = \gamma_b^5 - \gamma_b^3\sigma_b^2 + \frac{6}{T}\gamma_b^3\sigma_v^2$.

The proof of this result is shown in the supplement, Appendix A.

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

Theorem 4.2 *Let Assumptions 2.1-2.9 and 3.1 hold, under the null hypothesis of symmetry, i.e. $SK_v = 0$ and/or $SK_b = 0$, as $N \to \infty$ and T is fixed,*

$$\widehat{\pi}_{v3} = \frac{\sqrt{N}\widehat{SK}_v}{sd(\widehat{SK}_v)} \xrightarrow{d} \mathcal{N}(0,1),$$

$$\widehat{\pi}_{b3} = \frac{\sqrt{N}\widehat{SK}_b}{sd(\widehat{SK}_b)} \xrightarrow{d} \mathcal{N}(0,1),$$

where $sd(\widehat{SK}_v) = (\widehat{\mu}_{3,v0}/(T\widehat{\sigma}_v^6))^{1/2}$ and $sd(\widehat{SK}_b) = (\widehat{\mu}_{3,b0}/\widehat{\sigma}_b^6)^{1/2}$. We remark that $\widehat{\sigma}_v^2$, $\widehat{\sigma}_b^2$, $\widehat{\mu}_{3,v0}$ and $\widehat{\mu}_{3,b0}$ are consistent estimators for σ_v^2, σ_b^2, $\mu_{3,v0}$ and $\mu_{3,b0}$, respectively. In addition, $\mu_{3,v0}$ and $\mu_{3,b0}$ are the values of $\mu_{3,v}$ and $\mu_{3,b}$, respectively, when the null hypothesis of symmetry is imposed.

The proof of this theorem follows the same lines as the corresponding proof 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 $\widehat{SK}_b$. It is important to notice that, under $H_0 : SK_v = 0$, the statistic $\widehat{SK}_v$ is robust to the presence of skewness (or kurtosis) in b_i; however, it is not robust to kurtosis on its own. Similarly, under $H_0 : SK_b = 0$, the statistic $\widehat{SK}_b$ is not affected by skewness (or kurtosis) in the random error (v_{it}), but it is not robust to kurtosis on its own. These facts are confirmed by simulations in the next section.

4.2   Testing for kurtosis

For the test of kurtosis, the quantities of interest for the random error and the individual effects are $KU_v = \gamma_v^4/\sigma_v^4$ and $KU_b = \gamma_b^4/\sigma_b^4$, respectively. As in the previous subsection,

we will proceed first by deriving the limiting distribution of the proposed statistics under any value of KU_v and KU_b, and we will apply these results to the particular case in which $KU_v = 3$ and/or $KU_b = 3$.

To construct the test statistics for kurtosis, we replace the population expressions with sample moments, obtaining the following statistics:

$$\widehat{KU}_v = \frac{\widehat{\gamma}_v^4}{\widehat{\sigma}_v^4} \qquad \text{and} \qquad \widehat{KU}_b = \frac{\widehat{\gamma}_b^4}{\widehat{\sigma}_b^4}, \tag{4.2}$$

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

Theorem 4.3 *Under Assumptions 2.1-2.9 and 3.1, as $N \to \infty$ and T is fixed,*

$$\sqrt{N}(\widehat{KU}_v - KU_v) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, \frac{\beta_v^\top \Phi_v \beta_v}{T\sigma_v^8}\right),$$

$$\sqrt{N}(\widehat{KU}_b - KU_b) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, \frac{\beta_b^\top \Phi_b \beta_b}{\sigma_b^8}\right),$$

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 Φ_v and Φ_b are 2×2 matrices of the form

$$\Phi_v = \begin{pmatrix} \mu_{4,v} & \mu_{42,v} \\ \mu_{42,v} & \mu_{2,v} \end{pmatrix} \quad \text{and} \quad \Phi_b = \begin{pmatrix} \mu_{4,b} & \mu_{42,b} \\ \mu_{42,b} & \mu_{2,b} \end{pmatrix},$$

where $\mu_{42,v} = \frac{\gamma_v^6}{T} - \left(\frac{T-5}{T(T-1)}\right)\gamma_v^4\sigma_v^2 - \frac{2}{T}(\gamma_v^3)^2 - \frac{2}{T(T-1)}\sigma_v^6$ and $\mu_{42,b} = \gamma_b^6 - \gamma_b^4\sigma_b^2 + \frac{8}{T}\gamma_b^4\sigma_v^2 + \frac{6}{T(T-1)}\sigma_b^2\sigma_v^4$.

The proof of this result is shown in the supplement, Appendix A.

In the particular case where we are interested in testing whether the kurtosis coefficient is equal to 3, the null hypothesis will be $KU_v = 3$ for the idiosyncratic error term or $KU_b = 3$ for the individual heterogeneity. Then, by standardizing the results above, we obtain the following.

Theorem 4.4 *Suppose Assumptions 2.1-2.9 and 3.1 hold, under the null hypothesis that the kurtosis coefficient is equal to 3, i.e. $KU_v = 3$ and/or $KU_b = 3$, as $N \to \infty$ and T is fixed,*

$$\widehat{\pi}_{v4} = \frac{\sqrt{N}(\widehat{KU}_v - 3)}{sd(\widehat{KU}_v)} \quad \xrightarrow{d} \quad \mathcal{N}(0,1),$$

$$\widehat{\pi}_{b4} = \frac{\sqrt{N}(\widehat{KU}_b - 3)}{sd(\widehat{KU}_b)} \quad \xrightarrow{d} \quad \mathcal{N}(0,1),$$

where $sd(\widehat{KU}_v) = \left(\widehat{\beta}_{v3}^{\top}\widehat{\Phi}_{v3}\widehat{\beta}_{v3}/(T\widehat{\sigma}_v^8)\right)^{1/2}$ and $sd(\widehat{KU}_b) = \left(\widehat{\beta}_{b3}^{\top}\widehat{\Phi}_{b3}\widehat{\beta}_{b3}/\widehat{\sigma}_b^8\right)^{1/2}$. We remark that $\widehat{\beta}_{v3}$, $\widehat{\beta}_{b3}$, $\widehat{\Phi}_{v3}$, and $\widehat{\Phi}_{b3}$ are consistent estimators for β_{v3}, β_{b3}, Φ_{v3}, and Φ_{b3}, respectively. Additionally, β_{v3}, β_{b3}, Φ_{v3} and Φ_{b3} are the values of β_v, β_b, Φ_v and Φ_b, respectively, when the null hypothesis of kurtosis is imposed.

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

From this theorem, it can be pointed out that under $H_0 : KU_v = 3$, the statistic $\widehat{KU}_v$ is not affected by kurtosis (or skewness) in b_i. Similarly, under $H_0 : KU_b = 3$, $\widehat{KU}_b$ is not affected by kurtosis (or skewness) in v_{it}. However, one of the main criticisms when testing for kurtosis based on moment conditions is that skewness might influence kurtosis. See Bai and Ng (2005) and Galvao et al. (2013) for further details. As the reader can appreciate in Theorem 4.4, both statistics $\widehat{KU}_v$ and $\widehat{KU}_b$ are affected by its own skewness coefficient. Therefore, underestimated $\widehat{KU}_v$ and $\widehat{KU}_b$ are expected in practice given that the skewness coefficient can cause these statistics to deviate substantially from their true values. That result will be confirmed later by simulations.

Finally, note that these statistics of skewness and kurtosis are very useful because they enable us to provide a joint test for each component of the error, which assesses whether the data conform to any distribution of interest. Let $\widetilde{\pi}_{v3}$ and $\widetilde{\pi}_{v4}$ 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 shown in the supplement,

Appendix A, that $\widetilde{\pi}_{v3}$ and $\widetilde{\pi}_{v4}$ 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

$$\widetilde{\pi}_{v34} = \widetilde{\pi}_{v3}^2 + \widetilde{\pi}_{v4}^2. \tag{4.3}$$

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

$$\widetilde{\pi}_{b34} = \widetilde{\pi}_{b3}^2 + \widetilde{\pi}_{b4}^2, \tag{4.4}$$

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

Theorem 4.5 *Suppose Assumptions 2.1-2.9 and 3.1 hold. Under the null hypothesis of normality, as $N \to \infty$ and T is fixed,*

$$\widetilde{\pi}_{v34} \xrightarrow{d} \chi_2^2,$$
$$\widetilde{\pi}_{b34} \xrightarrow{d} \chi_2^2.$$

The proof of this results is straightforward if we are able to show that both statistics are asymptotically independent. This is shown in the supplement, Appendix A. Although the need of testing for normality in a nonparametric regression context does not look very appealing, in this setting, it can be justified on the grounds of efficiency. That is, in a varying coefficient parameter setting, if the normality assumption of the error terms is not rejected, then through the use of local maximum likelihood techniques, it is possible to obtain more efficient estimators than those obtained here in the previous section.

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. Later, we apply the proposed estimator and test statistics to analyze a real data example.

5.1 Monte Carlo experiment

To evaluate whether the nonparametric estimators and test statistics proposed in this article are valid for several distributions and to check their sensibility to the presence of technical inefficiency, three sets of simulations are conducted: one set with the random error exhibiting technical inefficiency and the other sets under no technical inefficiency.

In both sets of experiments the data generating process (DGP) is as such:

$$
\begin{aligned}
Y_{it} &= X_{it}^\top m(Z_{it}) + b_i + v_{it}, \quad i = 1, \ldots, N; \quad t = 1, \ldots, T, \\
v_{it} &= \nu_{it} - u_{it},
\end{aligned}
$$

where v_{it} is a composed error term that has two components, a non-negative error term to account for technical inefficiency, u_{it}, and a symmetric error term to account of other random errors, ν_{it}. Following Aigner et al. (1977), it is assumed that $\nu_{it} \sim i.i.d. 0.5\mathcal{N}(0, \sigma_\nu^2)$ and it is independent of the u_{it}, that is assumed to be $u_{it} \sim i.i.d.\mathcal{N}^+(0, \sigma_u^2)$. The notation "+" indicates that the underlying distribution is truncated from below at zero so that $u_{it} \geq 0$. Additionally, to generate data for the simulation, 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 $Z_{it} = \varpi_{it} + \varpi_{i(t-1)}$, where $\varpi_{it} \sim i.i.d.U[0, \pi/2]$ and $\xi_{it} \sim i.i.d.\mathcal{N}(0, 1)$.

For the above sets of experiments, we consider different processes for both the composed error term, v_{it}, and the individual effects, b_i. Further, the variance ratio $\lambda = \sigma_u/\sigma_\nu$ may affect model estimation. Therefore, in the following, we hold σ_ν^2 fixed at 0.1 and consider alternative values of λ.

Case A: no technical inefficiency and symmetric distributions (i.e., $v_{it} = \nu_{it}$):

$$DGP.A1. \quad \nu_{it} \sim i.i.d. \ 0.5\mathcal{N}(0,1) \quad ; \quad b_i \sim i.i.d. \ 0.5\mathcal{N}(0,1);$$

$$DGP.A2. \quad \nu_{it} \sim i.i.d. \ 0.5\mathcal{N}(0,1) \quad ; \quad b_i \sim i.i.d. \ 0.5t(9);$$

$$DGP.A3. \quad \nu_{it} \sim i.i.d. \ 0.5t(9) \quad \quad ; \quad b_i \sim i.i.d. \ 0.5\mathcal{N}(0,1);$$

$$DGP.A4. \quad \nu_{it} \sim i.i.d. \ 0.5t(9) \quad \quad ; \quad b_i \sim i.i.d. \ 0.5t(9).$$

Case B: no technical inefficiency and asymmetric distributions (i.e., $v_{it} = \nu_{it}$):

$$DGP.B1. \quad \nu_{it} \sim i.i.d. \ 0.5\mathcal{N}(0,1) \quad \quad ; \quad b_i \sim i.i.d. \ 0.5exp(\mathcal{N}(0,1)));$$

$$DGP.B2. \quad \nu_{it} \sim i.i.d. \ 0.5exp(\mathcal{N}(0,1)) \quad ; \quad b_i \sim i.i.d. \ 0.5\mathcal{N}(0,1);$$

$$DGP.B3. \quad \nu_{it} \sim i.i.d. \ 0.5exp(\mathcal{N}(0,1)) \quad ; \quad b_i \sim i.i.d. \ 0.5exp(\mathcal{N}(0,1)).$$

Case C: technical inefficiency (i.e., $v_{it} = \nu_{it} - u_{it}$), where $\nu_{it} \sim i.i.d. 0.5\mathcal{N}(0, \sigma_\nu^2)$:

$$DGP.C1. \quad u_{it} \sim i.i.d. \ \mathcal{N}^+(0, \sigma_\nu^2\lambda^2) \quad ; \quad b_i \sim i.i.d. \ 0.5\mathcal{N}(0,1) \quad ; \quad \lambda = 0.5;$$

$$DGP.C2. \quad u_{it} \sim i.i.d. \ \mathcal{N}^+(0, \sigma_\nu^2\lambda^2) \quad ; \quad b_i \sim i.i.d. \ 0.5\mathcal{N}(0,1) \quad ; \quad \lambda = 0.75;$$

$$DGP.C3. \quad u_{it} \sim i.i.d. \ \mathcal{N}^+(0, \sigma_\nu^2\lambda^2) \quad ; \quad b_i \sim i.i.d. \ 0.5\mathcal{N}(0,1) \quad ; \quad \lambda = 1.$$

The simulation results are based on 1000 samples of data $\{(X_{it}, Z_{it}, Y_{it}) : i = 1, \ldots, N, t = 1, \ldots, T\}$. 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 Epanechnikov kernel has been used, and for simplicity the bandwidth is chosen following Silverman's rule-of-thumb, i.e., $\widehat{H} = \widehat{h}I = \widehat{\sigma}_z(NT)^{-1/5}$, where $\widehat{\sigma}_z$ 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. To save space, in Tables 1-3, we report only the simulations results for DGP.A1-DGP.A2, DGP.B1-DGP.B2, and DGP.C2. The complete set of simulation results is available in the supplement, Appendix C, of this paper.

Finite sample performance of the proposed estimators. The behavior of the estimator for

$m(\cdot)$ is assessed via the following root mean squared error (RMSE)

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

We are also interested in the performance of the nonparametric estimator proposed in this paper with respect to other estimators proposed in the literature for this type of models. Then, in Table 1 the mean and standard deviation (Sd) of the above expression for the pairwise difference estimator (PDE) are compared to those obtained for the fixed-effects (FEE) and 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).

Table 1. Mean and Sd of the RMSE of the estimators for the nonparametric component.

	N	Results	FEE	OBE	PLSE	PDE		N	Results	FEE	OBE	PLSE	PDE
DGP.A1	50	Mean	0.500	0.555	0.769	0.513	DGP.B1	50	Mean	0.500	0.556	0.759	0.513
		Sd	0.086	0.029	0.049	0.040			Sd	0.084	0.028	0.049	0.040
	100	Mean	0.448	0.528	0.763	0.479		100	Mean	0.449	0.529	0.764	0.481
		Sd	0.061	0.028	0.046	0.030			Sd	0.060	0.053	0.048	0.030
	150	Mean	0.426	0.510	0.769	0.461		150	Mean	0.430	0.511	0.767	0.464
		Sd	0.051	0.018	0.045	0.025			Sd	0.049	0.018	0.048	0.026
DGP.A2	50	Mean	0.503	0.555	0.756	0.511	DGP.B2	50	Mean	0.494	0.555	0.760	0.511
		Sd	0.087	0.027	0.051	0.039			Sd	0.080	0.028	0.051	0.040
	100	Mean	0.447	0.529	0.751	0.479		100	Mean	0.448	0.529	0.765	0.481
		Sd	0.058	0.026	0.043	0.029			Sd	0.060	0.026	0.046	0.030
	150	Mean	0.425	0.511	0.750	0.462		150	Mean	0.431	0.514	0.768	0.465
		Sd	0.055	0.018	0.050	0.025			Sd	0.046	0.016	0.051	0.020
DGP.A3	50	Mean	0.498	0.554	0.772	0.511	DGP.C2	50	Mean	0.490	0.552	0.755	0.508
		Sd	0.086	0.031	0.054	0.045			Sd	0.081	0.026	0.047	0.037
	100	Mean	0.445	0.526	0.766	0.476		100	Mean	0.456	0.527	0.756	0.481
		Sd	0.066	0.026	0.046	0.034			Sd	0.061	0.025	0.042	0.027
	150	Mean	0.433	0.514	0.764	0.465		150	Mean	0.411	0.509	0.772	0.455
		Sd	0.050	0.019	0.043	0.025			Sd	0.052	0.019	0.041	0.023

In Tables 1-2, it is proved that the finite sample performance of the estimators proposed in this paper is robust to alternative distributional processes. As it is expected from their theoretical properties, the RMSEs of all these estimators are lower in all cases of study. Further, the results in Table 1 show that the PDE is very competitive; in only one step does it achieve rather similar results to the OEB, corroborating our theoretical findings. In addition, these estimators are even closer as the sample size increases. Therefore, the gain of this new estimation method is corroborated.

Table 2. Bias, Sd, and RMSE of the estimators for higher-order moments.

	N	Results	Remainder component			Individual component		
			$\widehat{\sigma}_v^2$	$\widehat{\gamma}_v^3$	$\widehat{\gamma}_v^4$	$\widehat{\sigma}_b^2$	$\widehat{\gamma}_b^3$	$\widehat{\gamma}_b^4$
DGP.A1		True	0.250	0.000	0.187	0.250	0.000	0.187
	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
		RMSE	0.328	0.151	1.140	0.074	0.113	0.228
	100	Bias	0.291	0.001	0.820	0.007	0.011	0.013
		Sd	0.050	0.093	0.319	0.066	0.096	0.171
		RMSE	0.295	0.095	0.879	0.049	0.075	0.151
	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
		RMSE	0.277	0.078	0.805	0.038	0.064	0.112
DGP.A2		True	0.250	0.000	0.187	0.321	0.000	0.444
	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
		RMSE	0.329	0.149	1.100	0.082	0.156	0.394
	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
		RMSE	0.288	0.095	0.852	0.055	0.107	0.242
	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
		RMSE	0.272	0.075	0.789	0.039	0.071	0.188
DGP.A3		True	0.321	0.000	0.434	0.250	0.002	0.187
	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
		RMSE	0.329	0.202	1.339	0.074	0.121	0.232
	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
		RMSE	0.286	0.126	0.990	0.052	0.087	0.150
	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
		RMSE	0.280	0.099	0.973	0.039	0.057	0.116
DGP.B1		True	0.250	0.000	0.187	1.857	12.887	273.23
	50	Bias	0.319	0.009	0.967	0.009	0.019	0.138
		Sd	0.075	0.144	0.510	0.279	1.947	18.977
		RMSE	0.327	0.145	1.091	0.087	0.346	3.488
	100	Bias	0.292	-0.003	0.854	0.007	0.024	0.854
		Sd	0.053	0.095	0.339	0.202	1.875	12.790
		RMSE	0.296	0.094	0.981	0.062	0.319	2.441
	150	Bias	0.275	-0.002	0.776	0.004	0.002	0.019
		Sd	0.040	0.086	0.260	0.130	0.758	5.903
		RMSE	0.277	0.085	0.817	0.041	0.131	1.053
DGP.B2		True	1.847	11.280	188.9	0.250	0.000	0.187
	50	Bias	0.215	-0.226	0.229	0.264	0.421	0.507
		Sd	0.128	0.910	9.151	0.124	0.219	0.408
		RMSE	0.228	0.363	2.443	0.280	0.454	0.610
	100	Bias	0.185	-0.228	-0.128	0.271	0.429	0.532
		Sd	0.100	0.883	7.517	0.083	0.150	0.294
		RMSE	0.191	0.303	2.261	0.279	0.446	0.590
	150	Bias	0.167	-0.201	0.073	0.273	0.437	0.534
		Sd	0.073	0.596	7.133	0.068	0.109	0.199
		RMSE	0.171	0.248	0.444	0.278	0.444	0.559
DGP.C2		True	0.100	-0.049	0.030	0.250	0.000	0.187
	50	Bias	0.270	0.095	0.671	-0.049	-0.179	-0.070
		Sd	0.066	0.113	0.393	0.076	0.111	0.196
		RMSE	0.278	0.149	0.779	0.079	0.202	0.193
	100	Bias	0.254	0.093	0.641	-0.049	-0.179	-0.076
		Sd	0.046	0.077	0.263	0.054	0.077	0.124
		RMSE	0.258	0.120	0.693	0.067	0.191	0.140
	150	Bias	0.224	0.093	0.508	-0.051	-0.176	-0.075
		Sd	0.032	0.055	0.144	0.044	0.039	0.075
		RMSE	0.226	0.106	0.527	0.060	0.179	0.099

	N	α	Remainder component					Individual component				
			SK_u	KU_u	$\widehat{\pi}_{v3}$	$\widehat{\pi}_{v4}$	$\widetilde{\pi}_{v34}$	SK_b	KU_b	$\widehat{\pi}_{b3}$	$\widehat{\pi}_{b4}$	$\widetilde{\pi}_{b34}$
DGP.A1	50	0.05	0	3	0.047	0.052	0.033	0	3	0.065	0.037	0.029
	100	0.05	0	3	0.055	0.054	0.052	0	3	0.036	0.061	0.032
	150	0.05	0	3	0.051	0.057	0.052	0	3	0.054	0.054	0.052
DGP.A2	50	0.05	0	3	0.053	0.039	0.035	0	4.309	0.035	0.067	0.040
	100	0.05	0	3	0.051	0.040	0.041	0	4.309	0.038	0.102	0.078
	150	0.05	0	3	0.049	0.055	0.045	0	4.309	0.047	0.132	0.093
DGP.A3	50	0.05	0	4.309	0.050	0.085	0.046	0	3	0.058	0.039	0.042
	100	0.05	0	4.309	0.045	0.153	0.084	0	3	0.062	0.056	0.051
	150	0.05	0	4.309	0.050	0.227	0.140	0	3	0.051	0.055	0.042
DGP.B1	50	0.05	0	3	0.027	0.090	0.014	5.092	79.233	0.442	0.145	0.225
	100	0.05	0	3	0.029	0.102	0.024	5.092	79.233	0.475	0.223	0.254
	150	0.05	0	3	0.045	0.106	0.033	5.092	79.233	0.506	0.235	0.287
DGP.B2	50	0.05	5.092	79.233	0.280	0.047	0.200	0	3	0.075	0.016	0.060
	100	0.05	5.092	79.233	0.502	0.104	0.349	0	3	0.061	0.021	0.035
	150	0.05	5.092	79.233	0.586	0.176	0.367	0	3	0.053	0.039	0.049
DGP.C2	50	0.05	-1.549	3	0.617	0.005	0.598	0	3	0.048	0.029	0.044
	100	0.05	-1.549	3	0.657	0.002	0.655	0	3	0.026	0.037	0.041
	150	0.05	-1.549	3	0.659	0.004	0.659	0	3	0.020	0.025	0.035

Finite sample performance of the skewness coefficient, kurtosis and normality test. For practical implementation, the following Wald test statistics are used to test for skewness, (i) $\widehat{\pi}_{v3}^2$ and (ii) $\widehat{\pi}_{b3}^2$. For kurtosis we use (iii) $\widehat{\pi}_{v4}^2$, (iv) $\widehat{\pi}_{b4}^2$, whereas for normality, the corresponding statistics are (v) $\widetilde{\pi}_{v34}$ and (vi) $\widetilde{\pi}_{b34}$. 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 Table 3, the size and power of these tests are displayed.

In $DGP.A1$, $\nu_{it} \sim 0.5\mathcal{N}(0,1)$ and $b_i \sim 0.5\mathcal{N}(0,1)$ is considered. All of the proposed tests are expected to have empirical size close to 0.05. The results in Table 3 corroborate the expectation. Even for smaller sample sizes, all of the tests achieve the correct empirical size.

In $DGP.A2$-$DGP.A3$ a t_9-Student distribution is considered to analyze the effect of kurtosis in one element but not in the other. In the first case, it is expected that the test for kurtosis in b_i has relevant power, while the tests for skewness in b_i and skewness and kurtosis in the random error do not. The results for $DGP.A2$ corroborate that expected behavior. More precisely, the kurtosis test for b_i has power increasing with N (for $N = 50$ the rejection rate is 0.067 and it increases to 0.132 for $N = 150$) and the remaining tests have size close to the nominal size. Additionally, the test for skewness starts with undersized values

for $N = 50$ but increases to 0.047 when $N = 150$. All of those observations corroborate that the skewness test is robust to the presence of kurtosis in the other element, but not to its own kurtosis. Further, although the kurtosis test for b_i has power to reject kurtosis, the power is not very large. Because of this fact, the normality test has low power. This result is expected when a distribution is symmetric and a larger sample size should be considered. See Bai and Ng (2002) for a deeper discussion. Opposite results are obtained for $DGP.A3$, corroborating the theoretical findings in Section 4.

In $DGP.B1$-$DGP.B2$ the results for a log-normal distribution that exhibits an asymmetric behavior with a high level of kurtosis are reported. In this situation, it is expected that the power of the tests for skewness and kurtosis of the composed error increases with the sample size. Looking at the results in $DGP.B1$ we see that the tests for b_i effectively 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 $DGP.B2$.

To corroborate the fact that skewness affects kurtosis, in $DGP.C2$ we consider a distribution with both skewness and kurtosis. The results in Table 3 reveal that the proposed test for the composed error is effectively able to detect departures from symmetry, but the power of the kurtosis test has been negatively affected (the largest rejection rate is 0.037 for $N = 100$). In contrast, the skewness, kurtosis and normality test for the individual heterogeneity remain unchanged to the presence of skewness in the other component. As can be seen in the supplement, Appendix C, all of those results hold for the different values of λ considered in this experiment.

In summary, it can be remarked 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

To illustrate our proposed method, in this section, we perform an empirical analysis of the technical efficiency in a panel of EU companies. Traditionally, studies of this type are based on the parametric stochastic frontier analysis (SFA), proposed originally in Aigner et al. (1977) and Meeusen and van den Broeck (1977)). Despite its limited computational complexity, they exhibit the following important drawbacks.

First, the stochastic frontier production function is characterized by an error term that has two components, a non-negative error term to account for technical inefficiency and a symmetric error term to account for other random effects. In this situation, the estimation procedure in the panel data context is considerably more complicated, especially when the unobserved individual heterogeneity is correlated with the covariates of the model. See Greene (2005) and Wang and Ho (2010) for further details. Second, testing symmetry of the composed error is of great interest in this literature. The reason is that researchers who estimate this type of models typically reject models, samples, or both when residuals have skewness in the "wrong" direction, i.e., when the expected and the estimated sign of the skewness of the composer error are different. See the proposals in Coelli (1995), Simar and Wilson (2010) and Wang et al. (2011), among others. In addition, tests on the distributional features of the error components are harder to derive without further identification conditions. Third, these parametric models are characterized by a lack of flexibility. Thus, the risk of misspecification of the production function is high and it can lead to misleading conclusions. To overcome this latter limitation, several semi-parametric and nonparametric techniques have been introduced in the literature, see Fan et al. (1996) or Kumbhakar et al. (2007) for example.

Recently, Li et al. (2002) or Ahmad et al. (2005), among others, have found that capital and labor elasticities vary according to other features of the companies such as the R&D expenses, and varying coefficient models can be understood as a natural way to solve it. Further, a standard belief in this literature is that the level of R&D has some impact on

the marginal productivity of the fixed capital but not in the liquid capital. To corroborate this fact and trying to overcome most of the previous difficulties, we propose to estimate a varying coefficient panel data model of the following form:

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

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 SFA models the R&D variable usually has a neutral effect on the production function, but in this specification, it affects the marginal productivity of both labor and capital. In addition, $v_{it} = \nu_{it} - u_{it}$ is a composite error term where ν_{it} is the idiosyncratic error term with zero mean, and u_{it} represents the technical inefficiency of the firm i. Independence between ν_{it} and u_{it} is assumed. Then, the composite error term v_{it} has an expected value equal to $E(v_{it}) = -E(u_{it})$ and a third central moment such as

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

Therefore, 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, see Greene (2000) and Carree (2002), among others.

The data used for this study are drawn from the Analyze 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 obtain a final sample of $1,120$ observations, i.e., 160 companies and 7 time periods. The variables used in this study are defined in Table 4, including some basic statistics.

In the following figures, varying coefficient estimates are plotted against the R&D expenses, where the continuous line denotes the estimated varying coefficients and the dotted

Table 4. Statistics of the variables: mean and standard deviation in brackets.

| Variable	Description	Mean	Std
ln Y	Logarithm of the firm' sales of the firm	12.393	(2.8040)
ln W	Logarithm of the firm's liquid capital	10.930	(2.8060)
ln L	Logarithm of the firm's labor input	6.9735	(2.7991)
ln K	Logarithm of the firm's fixed capital	10.757	(2.8978)
Z	Logarithm of the firm's R&D expenses	8.8431	(2.7850)

lines represent pointwise 95% 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 does not appear to be affected by the level of R&D, corroborating the standard beliefs in this literature. On the boundaries, some nonlinearities can be observed, but they might be due to the boundary effects. Figure 2 exhibits a general upward trend with R&D, indicating that firms with large R&D expenses yield relatively higher marginal productivity of fixed capital. In this way, those companies with larger expenses do have incentives to increase their R&D expenses, since that will end up in higher marginal productivity of fixed capital.

Figure 3 shows that the marginal productivity of labor is a nonlinear function of R&D expenses. In general, the marginal productivity of labor first increases with R&D and then decreases as these expenses increase further. This bell shape of the curve corroborates what is obtained in Ahmad et al. (2005), for example. This behavior suggests that, while modest R&D tends to improve labor productivity, higher R&D expenses are related to lower labor productivity. Note that a different behavior is obtained for the extreme cases of this kind of expense. Finally, Figure 4 shows that the returns to scale are well below one (i.e., the constant return to scale level) in general. Firms with lower levels of R&D expenses exhibit decreasing returns to scale, whereas companies with larger expenditures in R&D exhibit
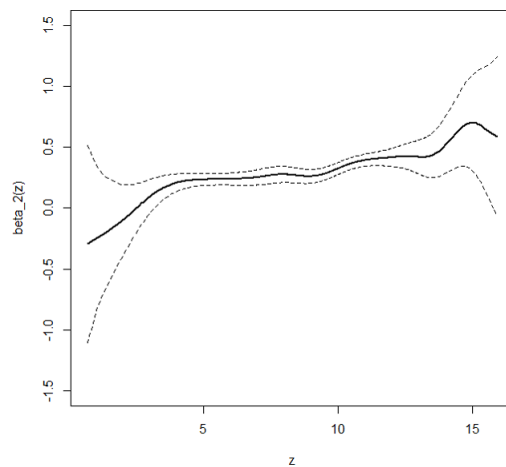
Figure 1.
Marginal productivity liquid capital
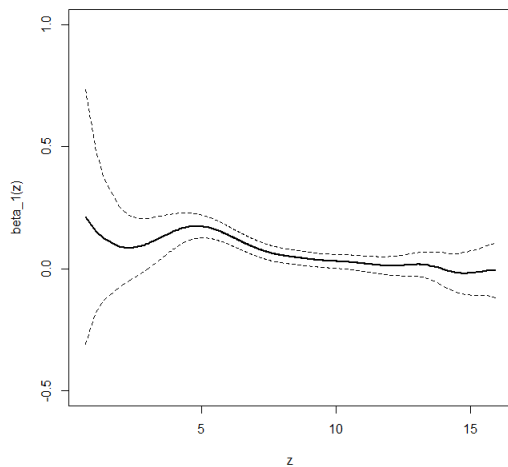
Figure 2.
Marginal productivity fixed capital

Figure 3.
Marginal productivity labor
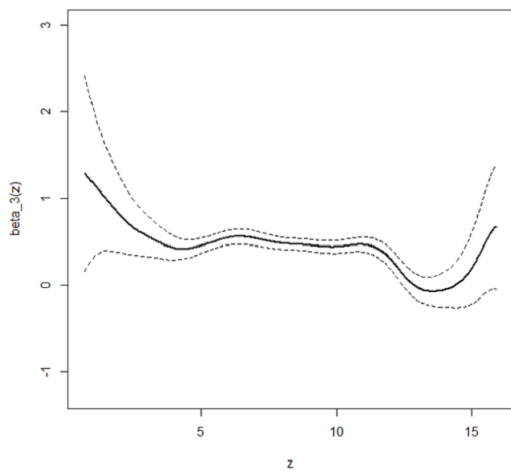
Figure 4.
Returns to scale function
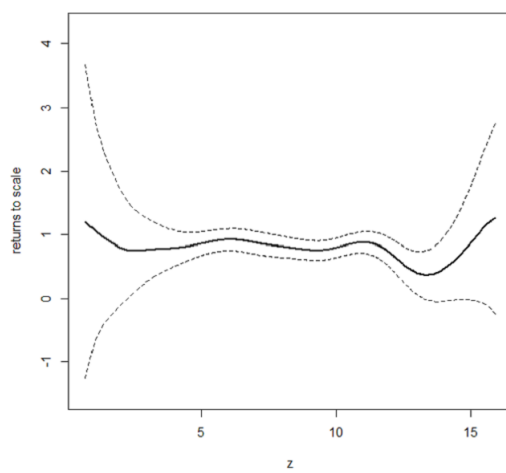
increasing returns to scale.

From all these results, it can be summed up that there exists a highly nonlinear relationship between the marginal productivity of both fixed capital and labor input and the R&D expenses. However, the marginal productivity of liquid capital does not appear to be sensitive to these kind of expenses. More precisely, companies with larger R&D expenses exhibit increasing returns to scale and higher marginal productivity of both fixed and labor input.

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

Table 5.
Estimated parameters for the stochastic frontier model.

	$\widehat{\sigma}^2$	$\widehat{\gamma}^3$	$\widehat{\gamma}^4$	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

The empirical results denote that v_{it} exhibits a nearly Gaussian distribution with a slight asymmetry. In fact the null hypothesis $H_0 : SK_v = 0$ and $KU_v = 3$ cannot be rejected. Therefore, no technical inefficiencies are present in this data set at the firm level. On the other side, the behavior of the heterogeneity term, b_i, is largely asymmetric and exhibits excess kurtosis (rejecting the corresponding null hypothesis at the 1% significance level). In fact, as we have pointed out both in Section 4 and the simulations section, the symmetry (kurtosis) test of b_i might be affected by its kurtosis (asymmetry); therefore, the results are not very reliable.

In view of these results, it is clear that this varying coefficient model effectively enables us to capture some relevant features of the covariates that were not possible with fully parametric or nonparametric models.

6   Conclusion

In this paper, a new technique to test for skewness and kurtosis in varying coefficient panel data models is proposed. This new proposal is robust to fixed-effects and distributional assumptions. In this paper, a pairwise difference estimator to compute these varying coefficients parameters is also introduced. This nonparametric estimator is almost asymptotically efficient, and it exhibits nice computational properties. 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 regarding the production efficiency of EU companies.

References

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

Aigner, D., C. Lovell, and P. Schmidt (1977). "Formulation and estimation of stochastic frontier production function models". *Journal of Econometrics 6*, 21–37.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*, 191–221.

Bai, J. and S. Ng (2005). Tests for skewness, kurtosis, and normality for time series data. *Journal of Business & Economic Statistics 23*, 49–60.

Battese, G. and T. Coelli (1988). Prediction of firm-level technical efficiencies with a generalised frontier production function and panel data. *Journal of Econometrics 38*, 387–399.

Black, F. and M. Sholes (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy 81*, 637–659.

Carree, M. (2002). Technological inefficiency and the skewness of the error component in stochastic frontier analysis. *Empirical Economics Letters 77*, 101–107.

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

Coelli, T. J. (1995). Estimators and hypotesis tests for a stochastic frontier frunction: a monte carlo analysis. *Journal of Productivity Analysis 6*, 247–268.

Cox, D. R. and P. Hall (2002). Estimation in a simple random effecs 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.

Fan, Y., Q. Li, and A. Weersink (1996). Estimation of stochastic production frontier models. *Journal of Business & Economic Statistics 14*, 460–468.

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.

Greene, W. (2000). A gamma distributed stochastic frontier model. *Journal of Econometrics 46*, 141–164.

Greene, W. (2005). Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of Econometrics 126*, 269–303.

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., B. Park, L. Simar, and E. Tsionas (2007). Nonparametric stochastic fontiers: a local maximum likelihood approach. *Journal of Econometrics 137*, 1–27.

Li, Q., J. H. Cliff, D. Li, and T. Fu (2002). Semiparametric smooth coefficient models. *Journal of Business & Economic Statistics 20*, 412–422.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics 47*, 13–37.

Mandelbrot, B. (1963). The Variation of Certain Speculative Prices. *Journal of Business 36*, 394–419.

Meeusen, W. and J. van den Broeck (1977). Efficient estimation from cobb-douglas production functions with composed error. *International Economic Review 18*, 435–444.

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.

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

Sharpe, W. (1964). Capital asset prices: a theory of market equilibrium under conditions of risk. *Journal of Finance 19*, 425–442.

Simar, L. and P. Wilson (2010). Inference from cross-sectional stochastic frontier models. *Econometric Reviews 29*, 62–98.

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

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, H. and C. Ho (2010). Estimating fixed-effect panel stochastic frontier models by model transformation. *Journal of Econometrics 157*, 286–296.

Wang, N. (2003). Marginal nonparametric kernel regression accounting for within-subject correlation. *Biometrika 90*, 43–52.

Wang, W., C. Amsler, and P. Schmidt (2011). Goodness of fit tests in stochastic frontier models. *Journal of Productivity Analysis 35*, 95–118.

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

Supplement to "Testing for distributional features in varying coefficient panel data models"
by A. Soberon, University of Cantabria,
W. Stute, Mathematical Institute, University of Giessen,
J.M. Rodriguez-Poo, University of Cantabria
January 2019

This supplement contains three sections. Appendix A provides the proofs of the main results presented in the paper. Appendix B collects the proposed estimators for the higher-order moments up to the eight moment for both individual heterogeneity and random error. Appendix C includes additional simulations to that discussed in the paper.

Appendix A

Proof of Theorem 2.1. The proof of this theorem consists of three parts. First, the bias of the local constant estimator in (3.5) in the paper 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

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

Let $Y_{\cdot t} = (Y_{1t}, \ldots, Y_{Nt})^\top$, $v_{\cdot t} = (v_{1t}, \ldots, v_{Nt})^\top$, and $b = (b_1, \ldots, b_N)^\top$ be vectors of $N \times 1$ dimension, it is possible to rewrite (2.1) in the paper in vectorial form as

$$Y_{\cdot t} = \Upsilon_{\cdot t}(X, m(Z)) + b + v_{\cdot t}, \qquad t = 1, \ldots, T, \tag{A.1}$$

where $\Upsilon_{\cdot t}(X, m(Z))$ stacks all the individual observations of $X_{it}^\top m(Z_{it})$ in a N-dimensional vector such as $\Upsilon_{\cdot t}(X, m(Z)) = (X_{1t}^\top m(Z_{1t}), \ldots, X_{Nt}^\top m(Z_{Nt}))^\top$.

Let $\widetilde{Y}_{\cdot ts}$ and $\widetilde{v}_{\cdot ts}$ be N-dimensional vectors such that $\widetilde{Y}_{ts} = [\widetilde{Y}_{1ts}, \ldots, \widetilde{Y}_{Nts}]^\top$ and $\widetilde{v}_{ts} = [\widetilde{v}_{1ts}, \ldots, \widetilde{v}_{Nts}]^\top$, respectively. Using the multivariate Taylor's theorem in (A.1) we obtain

$$\widetilde{Y}_{\cdot ts} = \widetilde{X}_{\cdot ts} m(z) + (D_{\cdot t}(z) - D_{\cdot s}(z)) + \frac{1}{2} (H_{\cdot t}(z) - H_{\cdot s}(z)) + (R_{\cdot t}(z) - R_{\cdot s}(z)) + \widetilde{v}_{\cdot ts},$$

where $R_{\cdot t}(z)$ and $R_{\cdot s}(z)$ are the residual terms of the Taylor expansion. Additionally, $\widetilde{X}_{\cdot ts} = [\widetilde{X}_{1ts}^\top, \ldots, \widetilde{X}_{Nts}^\top]^\top$ is a matrix of $N \times d$ dimension, and $D_{\cdot t}(z) = [D_{1t}^\top, \ldots, D_{Nt}^\top]^\top$, $H_{\cdot t}(z) = [H_{1t}^\top, \ldots, H_{Nt}^\top]^\top$ and $R_{\cdot t}(z) = [R_{1t}^\top, \ldots, R_{Nt}^\top]^\top$ are N-dimensional vectors whose corresponding entries are

$$\begin{aligned} D_{it}(z) &= X_{it}^\top \otimes (Z_{it} - z)^\top D_m(z), & H_{it}(z) &= X_{it}^\top \otimes (Z_{it} - z)^\top \mathcal{H}_m(z)(Z_{it} - z), \\ R_{it}(z) &= X_{it}^\top \otimes (Z_{it} - z)^\top \mathcal{R}_m(Z_{it}, z)(Z_{it} - z). \end{aligned}$$

Similar definitions for $D_{\cdot s}(z)$, $H_{\cdot s}(z)$, and $R_{\cdot s}(z)$. Additionally, $D_m(z)$ is a $dq \times 1$ vector such that $D_m(z) = vec(\partial m(z)/\partial z^\top)$ is the first-order derivative vector of $m(\cdot)$, $\mathcal{H}_m(z)$ 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 $m(\cdot)$, and $\mathcal{R}_m(Z_{it}, z) = (\mathcal{R}_{m_1}^\top(Z_{it}, z), \ldots, \mathcal{R}_{m_d}^\top(Z_{it}; z))^\top$ is a d-dimensional vector where

$$\mathcal{R}_{m_r}(Z_{it}, z) = \int_0^1 \left[\frac{\partial^2 m_r}{\partial z \partial z^\top}(z + \varphi(Z_{it} - z)) - \frac{\partial^2 m_r(z)}{\partial z \partial z^\top} \right] (1 - \varphi) d\varphi, \tag{A.2}$$

φ is a weight function and $r = 1, \ldots, d$. There is a similar definition for $\mathcal{R}_m(Z_{is}, z)$.

Replacing (A.2) in (2.4) in the paper and rearranging terms, $\widehat{m}(z; H)$ can be written as

$$\widehat{m}(z; H) - m(z) = \Psi_N^{-1} \left(A_N^{(1)} + A_N^{(2)} + U_N + R_N \right), \tag{A.3}$$

where

$$\begin{array}{ll}
\Psi_N = \binom{T}{2}^{-1} \frac{1}{N} \sum_{ts} \widetilde{X}_{\cdot ts}^\top W_{\cdot ts}(z) \widetilde{X}_{\cdot ts}, & A_N^{(1)} = \binom{T}{2}^{-1} \frac{1}{N} \sum_{ts} \widetilde{X}_{\cdot ts}^\top W_{\cdot ts}(z)(D_{\cdot t}(z) - D_{\cdot s}(z)), \\
U_N = \binom{T}{2}^{-1} \frac{1}{N} \sum_{ts} \widetilde{X}_{\cdot ts}^\top W_{\cdot ts}(z) \widetilde{v}_{\cdot ts}, & A_N^{(2)} = \binom{T}{2}^{-1} \frac{1}{2N} \sum_{ts} \widetilde{X}_{\cdot ts}^\top W_{\cdot ts}(z)(H_{\cdot t}(z) - H_{\cdot s}(z)), \\
R_N = \binom{T}{2}^{-1} \frac{1}{N} \sum_{ts} \widetilde{X}_{\cdot ts}^\top W_{\cdot ts}(z)(R_{\cdot t}(z) - R_{\cdot s}(z)), &
\end{array}$$

and $W_{\cdot ts}(z) = diag\left(K_H(Z_{1t} - z) K_H(Z_{1s} - z), \ldots, K_H(Z_{Nt} - z) K_H(Z_{Ns} - z) \right)$ is a $N \times N$ dimensional matrix.

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

$$\sqrt{N|H|} \left(\widehat{m}(z; H) - m(z) \right) - \sqrt{N|H|} \Psi_N^{-1} \left(A_N^{(1)} + A_N^{(2)} + R_N(z) \right) = \sqrt{N|H|} \Psi_N^{-1} U_N, \tag{A.4}$$

where we will demonstrate that $\Psi_N^{-1} A_N^{(j)}$, for $j = 1, 2$, contributes to the asymptotic bias, whereas the right-hand side term of (A.4) is asymptotically normal.

Starting with the bias term of this estimator, we first focus on Ψ_N^{-1}. Under Assumption 2.1 X_{it} are $i.i.d.$ across i, for each fixed t. Then, as N tends to infinity and T is fixed, we obtain

$$\Psi_N = \frac{2}{(T-1)} \mathcal{B}_{X\widetilde{X}}(z, z)(1 + o_\mathbb{P}(1)), \tag{A.5}$$

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

$$\mathcal{B}_{X\widetilde{X}}(z, z) = \sum_{\kappa=1}^{T-1} \left(1 - \frac{\kappa}{T} \right) E\left[X_{i1} \widetilde{X}_{i1(1+\kappa)}^\top | Z_{i1} = z, Z_{i(1+\kappa)} = z \right] f_{Z_{i1}, Z_{i(1+\kappa)}}(z, z)$$

and $\widetilde{X}_{i1(1+\kappa)} = X_{i1} - X_{i(1+\kappa)}$. To show this result, by taking expectation and under Assumption 2.1,

$$\begin{aligned}
E\left(\Psi_N \right) &= \frac{2}{NT(T-1)} \sum_{its} E[(X_{it} - X_{is})(X_{it} - X_{is})^\top K_{it} K_{is}] \\
&= \frac{4}{N(T-1)} \sum_{i\kappa} \left(1 - \frac{\kappa}{T} \right) E[X_{i1}(X_{i1} - X_{i(1+\kappa)})^\top K_{i1} K_{i(1+\kappa)}] \\
&= \frac{4}{(T-1)} \sum_{\kappa} \int E[X_{i1} \widetilde{X}_{i(1+\kappa)}^\top | Z_{i1} = z + H^{1/2} u, Z_{i(1+\kappa)} = z + H^{1/2} v] \\
&\quad \times f_{Z_{i1}, Z_{i(1+\kappa)}}(Z_{i1} = z + H^{1/2} u, Z_{i(1+\kappa)} = z + H^{1/2} v) K(u) K(v) du dv,
\end{aligned}$$

2

and by a Taylor expansion the expression (A.5) holds. Additionally, to complete the proof, it is necessary to show that $Var(\Psi_N) \to 0$ as $N \to \infty$.

$$
\begin{aligned}
Var\left(\Psi_N\right) &= \frac{4}{N}\binom{T}{2}^{-2}\sum_{\kappa}(1-\kappa)^2\, Var[X_{i1}\widetilde{X}_{i1(1+\kappa)}^{\top}K_{i1}K_{i(1+\kappa)}] \\
&+ \frac{4}{N}\binom{T}{2}^{-2}\sum_{\kappa}\sum_{\kappa<\kappa'}(1-\kappa)\left(1-\kappa'\right)Cov[X_{i1}\widetilde{X}_{i1(1+\kappa)}^{\top}K_{i1}K_{i(1+\kappa)}, X_{i1}\widetilde{X}_{i1(1+\kappa')}^{\top}K_{i1}K_{i(1+\kappa')}]
\end{aligned}
$$

where, under Assumptions 2.1 and 2.8, the first element is $O_{\mathbb{P}}\left(\frac{1}{N|H|}\right)$, whereas the second one is $o_{\mathbb{P}}\left(\frac{1}{N|H|}\right)$. Then, as $N|H| \to \infty$ the variance term tends to zero and (A.5) holds.

Using this result and the inverse matrix of Ψ_N, by the Cramér-Wold device it is proved that

$$
\Psi_N^{-1} = \frac{(T-1)}{4}\mathcal{B}_{X\widetilde{X}}^{-1}(z,z) + o_{\mathbb{P}}(1). \tag{A.6}
$$

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

$$
\begin{aligned}
E\left(A_N^{(1)}\right) &= \frac{2}{(T-1)}\sum_{\kappa}\left(1-\frac{\kappa}{T}\right)\int\left(E(\widetilde{X}_{i1(1+\kappa)}X_{i1}^{\top}|Z_{i1}=z,Z_{i(1+\kappa)}=z)D_f(z)(H^{1/2}u)\right) \\
&\otimes\ (H^{1/2}u)^{\top}D_m(z)K(u)K(v)dudv \\
&-\frac{2}{(T-1)}\sum_{\kappa}\left(1-\frac{\kappa}{T}\right)\int\left(E(\widetilde{X}_{i1(1+\kappa)}X_{i(1+\kappa)}^{\top}|Z_{i1}=z,Z_{i(1+\kappa)}=z)D_f(z)(H^{1/2}v)\right) \\
&\otimes\ (H^{1/2}v)^{\top}D_m(z)K(u)K(v)dudv \\
&= \frac{4\mu_2(K)}{(T-1)}\mathcal{B}_{X\widetilde{X}}(z,z)diag_d\left(tr(HD_{f_r}(z)D_{m_r}(z))\right)\imath_d f_{Z_{i1},Z_{i(1+\kappa)}}^{-1}(z,z) + o_{\mathbb{P}}(tr(H)), \tag{A.7}
\end{aligned}
$$

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

Following a similar procedure, it is straightforward to show that

$$
\begin{aligned}
E\left(A_N^{(2)}\right) &= \frac{1}{(T-1)}\sum_{\kappa}E(\widetilde{X}_{i1(1+\kappa)}X_{i1}^{\top}|Z_{i1}=z,Z_{i(1+\kappa)}=z)f_{Z_{i1},Z_{i(1\kappa)}}(z,z) \\
&\otimes\ tr(H\mathcal{H}_{m_r}(z))\int u^{\top}uK(u)K(v)dudv \\
&-\frac{1}{(T-1)}\sum_{\kappa}E(\widetilde{X}_{i1(1+\kappa)}X_{i(1+\kappa)}^{\top}|Z_{i1}=z,Z_{i(1+\kappa)}=z)f_{Z_{i1},Z_{i(1+\kappa)}}(z,z) \\
&\otimes\ tr(H\mathcal{H}_{m_r}(z))\int v^{\top}vK(u)K(v)dudv \\
&= \frac{2\mu_2(K)}{(T-1)}\mathcal{B}_{X\widetilde{X}}(z,z)diag_d\left(tr(H\mathcal{H}_{m_r}(z))\right)\imath_d + o_{\mathbb{P}}(tr(H)), \tag{A.8}
\end{aligned}
$$

where $\mathcal{H}_{m_r}(z)$ is the Hessian matrix of the rth component of $m(\cdot)$. Additionally, using similar arguments as above, we can show that any component of the variance of $A_N^{(1)}$ and $A_N^{(2)}$ converges to zero as $H \to 0$ and $N|H| \to \infty$.

3

Finally, to show that the local constant estimator $\widehat{m}(z; H)$ is asymptotically unbiased, it is necessary to show that the residual terms of the Taylor expansion are negligible as N goes to infinity and T is fixed. To analyze the asymptotic behavior of R_N, we add and subtract $\widetilde{X}_{its} X_{is}^\top (Z_{is} - z)^\top \mathcal{R}_m(Z_{it}; z)(Z_{is} - z)$. Then, rearranging terms, it is possible to decompose R_N into the following two terms:

$$
\begin{aligned}
R_N &= \frac{1}{N} \binom{T}{2}^{-1} \sum_{its} \widetilde{X}_{its} \left(X_{it}^\top \otimes (Z_{it} - z)^\top \mathcal{R}_m(Z_{it}; z)(Z_{it} - z) - X_{is}^\top \otimes (Z_{is} - z)^\top \mathcal{R}_m(Z_{is}; z)(Z_{is} - z) \right) \\
&\quad \times K_{it} K_{is} \\
&= \Lambda_1(z) - \Lambda_2(z),
\end{aligned}
$$

where

$$
\begin{aligned}
\Lambda_1(z) &= \frac{1}{N} \binom{T}{2}^{-1} \sum_{its} \widetilde{X}_{its} \left(X_{it}^\top \otimes (Z_{it} - z)^\top \mathcal{R}_m(Z_{it}; z)(Z_{it} - z) - X_{is}^\top \otimes (Z_{is} - z)^\top \mathcal{R}_m(Z_{it}; z)(Z_{is} - z) \right) \\
&\quad \times K_{it} K_{is} \\
\Lambda_2(z) &= \frac{1}{N} \binom{T}{2}^{-1} \sum_{its} \widetilde{X}_{its} X_{is}^\top \otimes (Z_{is} - z)^\top (\mathcal{R}_m(Z_{it}; z) - \mathcal{R}_m(Z_{is}; z))(Z_{is} - z) K_{it} K_{is}.
\end{aligned}
$$

Thus, what we need to show is that as $N \to \infty$ and T is fixed,

$$
E(R_N) = o_\mathbb{P}(tr(H)). \tag{A.9}
$$

Analysing $\Lambda_1(z)$ and $\Lambda_2(z)$ separately, under Assumption 2.1 and following a similar procedure as above,

$$
\begin{aligned}
E(\Lambda_1(z)) &= \frac{2}{(T-1)} \sum_\kappa \left(1 - \frac{\kappa}{T}\right) \int E[\widetilde{X}_{i1(1+\kappa)} X_{i1}^\top | Z_{i1} = z + H^{1/2}u, Z_{i(1+\kappa)} = z + H^{1/2}v] \otimes (H^{1/2}u)^\top \\
&\quad \times \mathcal{R}_m(z + H^{1/2}u, z)(H^{1/2}u) f_{Z_{i1}, Z_{i(1+\kappa)}}(z + H^{1/2}u, z + H^{1/2}v) K(u)K(v)dudv, \\
&\quad - \frac{2}{(T-1)} \sum_\kappa \left(1 - \frac{\kappa}{T}\right) \int E[\widetilde{X}_{i1(1+\kappa)} X_{i(1+\kappa)}^\top | Z_{i1} = z + H^{1/2}u, Z_{i(1+\kappa)} = z + H^{1/2}v] \otimes (H^{1/2}v)^\top \\
&\quad \times \mathcal{R}_m(z + H^{1/2}u, z)(H^{1/2}v) f_{Z_{i1}, Z_{i(1+\kappa)}}(z + H^{1/2}u, z + H^{1/2}v) K(u)K(v)dudv.
\end{aligned}
$$

By definition (A.2) and Assumption 2.3, for $r = 1, \ldots, d$,

$$
|\mathcal{R}_{m_r}(z + H^{1/2}u; z)| \leq \int_0^1 \varsigma(\varphi \|H^{1/2}u\|)(1 - \varphi)d\varphi,
$$

where $\varsigma(\cdot)$ is the modulus of continuity of $\frac{\partial^2 m_r(z)}{\partial z \partial z^\top}$. Hence, under Assumptions 2.3-2.4, and given that we impose $\varsigma(\varphi \|H^{1/2}u\|) \to 0$ as $N \to \infty$,

$$
\begin{aligned}
E|\Lambda_1(z)| &\leq C_1 \int \int_0^1 |(H^{1/2}u)^\top| \|\varsigma(\varphi\|H^{1/2}u\|)\| \|H^{1/2}u\| |K(u)||K(v)|(1 - \varphi)dudvd\varphi \\
&\quad - C_2 \int \int_0^1 |(H^{1/2}v)^\top| \|\varsigma(\varphi\|H^{1/2}u\|)\| \|H^{1/2}v\| |K(u)||K(v)|(1 - \varphi)dudvd\varphi + o_\mathbb{P}(tr(H)) \\
&= o_\mathbb{P}(tr(H)).
\end{aligned}
$$

4

Focus now on $\varepsilon_2(z)$. By the law of iterated expectations and Assumption 2.1,

$$
\begin{aligned}
E(\Lambda_2(z)) &= \frac{2}{(T-1)} \sum_\kappa \left(1 - \frac{\kappa}{T}\right) \int E[\widetilde{X}_{i1(1+\kappa)} X_{i(1+\kappa)}^\top | Z_{i1} = z + H^{1/2}u, Z_{i(1+\kappa)} = z + H^{1/2}v] \otimes (H^{1/2}v)^\top \\
&\times \left(\mathcal{R}_m(z + H^{1/2}u) - \mathcal{R}_m(z + H^{1/2}v)\right) (H^{1/2}v) f_{Z_{i1},Z_{i(1+\kappa)}}(z + H^{1/2}u, z + H^{1/2}v) K(u)K(v) du dv
\end{aligned}
$$

and following a similar procedure as previously

$$
E|\Lambda_2(z)| \leq C_2 \int \int_0^1 |(H^{1/2}v)^\top || \varsigma(\varphi \| H^{1/2}u \| - \varphi \| H^{1/2}v \|) || H^{1/2}v \| |K(u)||K(v)|(1-\varphi) du dv d\varphi + o_{\mathbb{P}}(tr(H)).
$$

Under the dominated convergence, $\varsigma(\varphi \| H^{1/2}u \|) \to 0$ and $\varsigma(\varphi \| H^{1/2}v \|) \to 0$ as $N \to \infty$, so $E(\Lambda_2(z)) = o_{\mathbb{P}}(tr(H))$ is proved and (A.9) holds.

Then, using (A.6)-(A.9) and applying the Cramér-Wold device it is proved that the asymptotic bias of $\widehat{m}(z; H)$ is

$$
\begin{aligned}
\Psi_N^{-1}\left(A_N^{(1)} + A_N^{(2)} + R_N\right) &= \mu_2(K)\mathcal{B}_{X\widetilde{X}}^{-1}(z, z)\mathcal{B}_{X\widetilde{X}}(z, z)\left(diag_d\left(tr(HD_{f_r}(z)D_{m_r}(z))\right) \imath_d f_{Z_{i1},Z_{i(1+\kappa)}}^{-1}(z, z)\right. \\
&+ \left. \frac{1}{2}diag_d\left(tr(H\mathcal{H}_{m_r}(z))\imath_d\right)\right) + o_{\mathbb{P}}(tr(H)) + o_{\mathbb{P}}\left(\frac{1}{\sqrt{N|H|}}\right). \quad (A.10)
\end{aligned}
$$

Thus, the first part of the proof is done.

To obtain the asymptotic variance of the right-hand side of (A.4), we have to analyze the behavior of U_N. Let $\mathbb{X}$ be the vector of observed covariates. By the law of iterated expectations, Assumption 2.1 and strict stationarity,

$$
Var(U_N) = \frac{1}{N^2}\binom{T}{2}^{-2} \sum_{\kappa=1}^{T-1} \sum_{\kappa'=1}^{T-1} (T-\kappa)(T-\kappa') E[\widetilde{X}_{\cdot 1(1+\kappa)}^\top W_{\cdot 1(1+\kappa)}(z) E(\widetilde{v}_{\cdot 1(1+\kappa)} \widetilde{v}_{\cdot 1(1+\kappa')}^\top | \mathbb{X}) W_{\cdot 1(1+\kappa')}(z) \widetilde{X}_{\cdot 1(1+\kappa')}]
$$

and by (A.1) we know $E(\widetilde{v}_{i1(1+\kappa)} \widetilde{v}_{j1(1+\kappa)} | \mathbb{X}) = 0$, for $\forall i \neq j$, and

$$
E\left(\widetilde{v}_{\cdot 1(1+\kappa)} \widetilde{v}_{\cdot 1(1+\kappa')}^\top | \mathbb{X}\right) = \begin{cases} 2\sigma_\eta^2 \Omega\Omega^\top & \text{for } \kappa = \kappa', \\ \sigma_\eta^2 \Omega\Omega^\top & \text{for } \kappa \neq \kappa'. \end{cases}
$$

In addition, remember that by Assumption 2.2, η_{it} is $i.i.d$ in i. Then, by the law of iterated expectations and Assumptions 2.1, 2.3-2.4, and 2.7,

$$
\begin{aligned}
N|H|Var(U_N) &= \frac{1}{N} \sum_{ij} \omega_{ij}^2 \frac{8\sigma_\eta^2 |H|}{(T-1)^2} \sum_\kappa \left(1 - \frac{\kappa}{T}\right)^2 E\left[\widetilde{X}_{i1(1+\kappa)} \widetilde{X}_{i1(1+\kappa)}^\top K_{i1}^2 K_{i(1+\kappa)}^2\right] \\
&+ \frac{1}{N} \sum_{ij} \omega_{ij}^2 \frac{4\sigma_\eta^2 |H|}{(T-1)^2} \sum_\kappa \sum_{\kappa \neq \kappa'} \left(1 - \frac{\kappa}{T}\right)^2 E\left[\widetilde{X}_{i1(1+\kappa)} \widetilde{X}_{i1(1+\kappa')}^\top K_{it}^2 K_{i(1+\kappa)} K_{i(1+\kappa')}\right] \\
&= \mathbf{I}_{1N} + \mathbf{I}_{2N}.
\end{aligned}
$$

Analysing each of these terms separately, we obtain

$$
N|H|Var(U_N) = \frac{16\sigma_\eta^2 R^2(K)}{(T-1)^2} \varpi_N \mathcal{B}_{X\widetilde{X}}^V(z, z)(1 + o_{\mathbb{P}}(1))
$$

5

where

$$
\mathcal{B}^V_{X\widetilde{X}}(z,z) = \sum_{\kappa=1}^{T-1} \left(1 - \frac{\kappa}{T}\right)^2 E[X_{i1}\widetilde{X}^\top_{i1(1+\kappa)}|Z_{i1}=z, Z_{i(1+\kappa)}=z]f_{Z_{i1},Z_{i(1+\kappa)}}(z,z).
$$

With the aim of showing this result, under similar arguments as above, it can be proved that

$$
\mathbf{I}_{1N} = \frac{16\sigma_\eta^2 R^2(K)}{(T-1)^2}\varpi_N \sum_\kappa \left(1 - \frac{\kappa}{T}\right)^2 E[X_{i1}\widetilde{X}^\top_{i1(1+\kappa)}|Z_{i1}=z, Z_{i(1+\kappa)}=z]f_{Z_{i1},Z_{i(1+\kappa)}}(z,z)(1+o_{\mathbb{P}}(1))
$$

and $\mathbf{I}_{2N} = O_{\mathbb{P}}(H^{1/2})$, so the expression (A.11) holds.

Then, using (A.5) and (A.11), by the Cramér-Wold device, as $N|H| \to \infty$,

$$
N|H|Var\left(\Psi_N^{-1}U_N\right) = \sigma_\eta^2 R^2(K)\varpi_N \mathcal{B}^{-1}_{X\widetilde{X}}(z,z)\mathcal{B}^V_{X\widetilde{X}}(z,z)\mathcal{B}^{-1}_{\widetilde{X}X}(z,z)(1+o_{\mathbb{P}}(1)). \quad \text{(A.11)}
$$

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

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

$$
\sqrt{N|H|}\left(\widehat{m}(z,H) - m(z) - B(z,H)\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma_\eta^2 R^2(K)\varpi_N \mathcal{B}^{-1}_{X\widetilde{X}}(z,z)\mathcal{B}^V_{X\widetilde{X}}(z,z)\mathcal{B}^{-1}_{\widetilde{X}X}(z,z)\right), \text{(A.12)}
$$

where

$$
\begin{aligned}
B(z,H) &= \mu_2(K)\mathcal{B}^{-1}_{X\widetilde{X}}(z,z)\mathcal{B}_{X\widetilde{X}}(z,z)\left(diag_d(tr(HD_{f_r}(z)D_{m_r}(z)))\imath_d f^{-1}_{Z_{i1},Z_{i(1+\kappa)}}(z,z)\right. \\
&\quad + \left.\frac{1}{2}diag_d(tr(H\mathcal{H}_{m_r}(z)))\imath_d\right) + o_{\mathbb{P}}(tr(H)).
\end{aligned}
$$

To verify the Lindeberg condition, it suffices to check the Lyapounov condition in the following expression:

$$
\frac{\sqrt{N|H|}}{N}\binom{T}{2}^{-1}\sum_{ts}\widetilde{X}^\top_{\cdot ts}W_{\cdot ts}(z)\widetilde{v}_{\cdot ts} = \frac{1}{\sqrt{N}}\binom{T}{2}^{-1}\sum_i \theta_i, \quad \text{(A.13)}
$$

where, under strict stationarity,

$$
\theta_i = \frac{2|H|^{1/2}}{(T-1)}\sum_\kappa \left(1 - \frac{\kappa}{T}\right)K_{i1}K_{i(1+\kappa)}\widetilde{X}_{i1(1+\kappa)}\widetilde{v}_{i1(1+\kappa)}.
$$

Using the results of Theorem 2.1, $Var\left(\sum_i \theta_i\right) = N^{(2+\delta)/2}$. Additionally, by the Minkowsky inequality, the Cauchy-Schwarz inequality, and the assumptions of this theorem, it is straightforward to show $E|\theta_i|^{(2+\delta)} = O_{\mathbb{P}}(|H|^{-\delta/2})$. Hence,

$$
N^{-(2+\delta)/2}\sum_i E|\theta_i|^{2+\delta} \leq CO_{\mathbb{P}}\left((N|H|)^{-\delta/2}\right).
$$

Then, it is proved that this term tends to zero as $N|H| \to \infty$, so the Lindeberg condition is verified. Therefore, the Lyapunov Central Limit Theorem can be used to verify (A.12), and the proof of Theorem 2.1 is done.

Proof of Theorem 3.1. We first focus on the asymptotic properties of $\widehat{\sigma}_v^2$ and later on $\widehat{\sigma}_b^2$. Inserting (2.1) of the paper 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

$$
\begin{aligned}
\widehat{\sigma}_v^2 &= \frac{1}{NT(T-1)} \sum_{i=1}^{N} \left[T \sum_t \left(\epsilon_{it} - X_{it}^\top [\widehat{m}(Z_{it}; H) - m(Z_{it})] \right)^2 - \left(\sum_t [\epsilon_{it} - X_{it}^T [\widehat{m}(Z_{it}; H) - m(Z_{it})]] \right)^2 \right] \\
&= \frac{1}{NT} \sum_{it} v_{it}^2 - \frac{1}{NT(T-1)} \sum_{its} v_{it} v_{is} + I\!\!I_v^{(1)} - I\!\!I_v^{(2)} + I\!\!I_v^{(3)} - I\!\!I_v^{(4)},
\end{aligned}
\tag{A.14}
$$

where after rearranging terms

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

As we will show, of these six terms of (A.14), only the first two will be the leading terms, whereas $I\!\!I_v^{(1)}$, $I\!\!I_v^{(2)}$, $I\!\!I_v^{(3)}$, and $I\!\!I_v^{(4)}$ are residual terms. Analysing each term separately and using uniform convergence results as the ones established in Theorem 6 in Masry (1996), by Assumptions 2.1-2.3 and rearranging terms,

$$
\begin{aligned}
I\!\!I_v^{(1)} &\leq \frac{1}{NT} \sum_{it} |X_{it}^\top X_{it}| \sup_{\{Z_{it} \in A\}} |(\widehat{m}(Z_{it}; H) - m(Z_{it}))^\top| \sup_{\{Z_{it} \in A\}} |\widehat{m}(Z_{it}; H) - m(Z_{it})| \\
&= O_{\mathbb{P}} \left(tr(H)^2 + \frac{logN}{N|H|} \right),
\end{aligned}
\tag{A.15}
$$

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

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

$$
I\!\!I_v^{(2)} \leq \frac{2}{NT} \sum_{it} |\epsilon_{it} X_{it}^\top| \sup_{\{Z_{it} \in A\}} |\widehat{m}(Z_{it}; H) - m(Z_{it})| = o_p \left(\frac{1}{\sqrt{N}} \right).
\tag{A.16}
$$

Given that, under Assumption 2.8, 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 2.2 the second one is $o_p(1)$. Following this same procedure, we obtain a similar result for $I\!\!I_v^{(4)}$.

Using these results in (A.14) and under Assumption 2.8,

$$
\sqrt{N} \left(\widehat{\sigma}_v^2 - \sigma_v^2 \right) = \frac{1}{\sqrt{NT}} \sum_{it} \left(v_{it}^2 - \sigma_v^2 \right) + o_{\mathbb{P}}(1)
\tag{A.17}
$$

7

and by the central limit theorem the first part of Theorem 3.1 is proved.

Focusing now on $\widehat{\sigma}_b^2$, equation (3.2) of the paper can be rewritten as

$$\widehat{\sigma}_b^2 \;=\; \frac{1}{N}\sum_i b_i^2 + \frac{2}{NT}\sum_{it} b_i v_{it} + \frac{1}{NT(T-1)}\sum_{its} v_{it}v_{is} + \mathbf{I}_b^{(1)} - 2\mathbf{I}_b^{(2)}, \qquad \text{(A.18)}$$

where

$$\mathbf{I}_b^{(1)} \;=\; \frac{1}{NT(T-1)}\sum_{its} X_{it}^\top[\widehat{m}(Z_{it};H)-m(Z_{it})]X_{is}^\top[\widehat{m}(Z_{is};H)-m(Z_{is})],$$

$$\mathbf{I}_b^{(2)} \;=\; \frac{1}{NT(T-1)}\sum_{its} \epsilon_{it} X_{is}^\top[\widehat{m}(Z_{is};H)-m(Z_{is})].$$

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

$$\sqrt{N}(\widehat{\sigma}_b^2-\sigma_b^2) = \frac{1}{\sqrt{N}}\sum_i \left(b_i^2-\sigma_b^2\right) + \frac{2}{\sqrt{N}T}\sum_{it} b_i v_{it} + \frac{1}{\sqrt{N}T(T-1)}\sum_{its} v_{it}v_{is} + o_\mathbb{P}(1) \quad \text{(A.19)}$$

from which it follows that the second part of Theorem 3.1 holds.

∎

Proof of Theorem 3.2. 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 2.2 and Assumption 3.1, all contributions involving $X_{it}^\top(\widehat{m}(Z_{it};H)-m(Z_{it}))$ may be neglected. Then, the final expressions to study are

$$\sqrt{N}\widehat{\gamma}_v^3 = \frac{1}{\sqrt{N}T}\sum_{it}\left[v_{it}^3 - \frac{3}{(T-1)}\sum_{s\neq t} v_{it}^2 v_{is} + \frac{2}{(T-1)(T-2)}\sum_{s\neq t}\sum_{r\neq s} v_{it}v_{is}v_{ir}\right] + o_\mathbb{P}(1) \qquad \text{(A.20)}$$

and

$$\sqrt{N}\widehat{\gamma}_b^3 = \frac{1}{\sqrt{N}}\sum_i\left[b_i^3 + \frac{3}{T}\sum_t b_i^2 v_{it} + \frac{3}{T(T-1)}\sum_{ts} b_i v_{it}v_{is} + \frac{1}{T(T-1)(T-2)}\sum_{tsr} v_{it}v_{is}v_{ir}\right] + o_\mathbb{P}(1) \tag{A.21}$$

where, after centering v_{it}^3 and ϵ_{it}^3, respectively, the first part of Theorem 3.2 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 obtain

$$\sqrt{N}\widehat{\gamma}_v^4 = \frac{1}{\sqrt{N}T}\sum_{it}\left[v_{it}^4 - \frac{4}{(T-1)}\sum_{s\neq t} v_{it}^3 v_{is} + \frac{6}{(T-1)(T-2)}\sum_{s\neq t}\sum_{r\neq s} v_{it}^2 v_{is}v_{ir}\right.$$
$$\left. - \frac{3}{(T-1)(T-2)(T-3)}\sum_{s\neq t}\sum_{r\neq s}\sum_{h\neq r}^T v_{it}v_{is}v_{ir}v_{ih}\right] + o_\mathbb{P}(1) \tag{A.22}$$

8

and

$$\sqrt{N}\widehat{\gamma}_b^4 = \frac{1}{\sqrt{N}} \sum_i \left[b_i^4 + \frac{4}{T} \sum_t b_i^3 v_{it} + \frac{6}{T(T-1)} \sum_{ts} b_i^2 v_{it} v_{is} + \frac{4}{T(T-1)(T-2)} \sum_{tsr} b_i v_{it} v_{is} v_{ir} \right.$$
$$\left. + \frac{1}{T(T-1)(T-2)(T-3)} \sum_{tsrh} v_{it} v_{is} v_{ir} v_{ih} \right] + o_{\mathbb{P}}(1). \tag{A.23}$$

Finally, as previously, after centering these expressions and using the central limit theorem, the second part of Theorem 3.2 holds.

∎

Proof of Theorem 4.1. To prove the results of this theorem, we follow the standard proof 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, i.e., $\widehat{\sigma}^2$, by the delta method, we obtain

$$\sqrt{N} \left((\widehat{\sigma}^2)^{k/2} - (\sigma^2)^{k/2} \right) = \frac{k}{2} (\sigma^2)^{k/2-1} \sqrt{N} \left(\widehat{\sigma}^2 - \sigma^2 \right) + o_{\mathbb{P}}(1). \tag{A.24}$$

Then, for $k = 3$, we replace (A.17), (A.20) and (A.24) in the previous equation and rearrange terms, so the expression to analyze is

$$\sqrt{NT}(\widehat{SK}_v - SK_v) = \frac{\alpha_v^\top}{\widehat{\sigma}_v^3} \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T SK_{v,it} + o_{\mathbb{P}}(1), \tag{A.25}$$

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

$$SK_{v,1it} = (v_{it}^3 - \gamma_v^3) - \frac{3}{(T-1)} \sum_{s \neq t} v_{it}^2 v_{is} + \frac{2}{(T-1)(T-2)} \sum_{s \neq t} \sum_{r \neq s} v_{it} v_{is} v_{ir}$$

$$SK_{v,2it} = (v_{it}^2 - \sigma_v^2) - \frac{1}{(T-1)} \sum_{s \neq t} v_{it} v_{is}.$$

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

$$\Gamma_v = \begin{bmatrix} \Gamma_{v,1} & \Gamma_{v,2} \\ \Gamma_{v,2} & \Gamma_{v,3} \end{bmatrix}. \tag{A.26}$$

9

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

$$\Gamma_{v,1} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{t'=1}^{T}E[SK_{v,1it}SK_{v,1it'}]$$

$$= \frac{\gamma_v^6}{T} - \frac{(\gamma_v^3)^2}{T} - \left(\frac{6T-15}{2T(T-1)}\right)\gamma_v^4\sigma_v^2 + \frac{9}{2T(T-1)}(\gamma_v^3)^2 + \left(\frac{3(T-2)^2+8}{2T(T-1)(T-2)}\right)\sigma_v^6,$$

$$\Gamma_{v,2} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{t'=1}^{T}E[SK_{v,1it}SK_{v,2it'}] = \frac{\gamma_v^5}{T} - \left(\frac{5T-11}{2T(T-1)}\right)\gamma_v^3\sigma_v^2,$$

$$\Gamma_{v,3} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{t'=1}^{T}E[SK_{v,2it}SK_{v,2it'}] = \frac{\gamma_v^4}{T} - \frac{\sigma_v^4}{T} + \frac{\sigma_v^4}{T(T-1)}.$$

By Theorem 3.1, $\widehat{\sigma}_v^2 \xrightarrow{p} \sigma_v^2$. Then, using this result and by the central limit theorem the first part of Theorem 4.1 is proved.

Similarly, focusing on the properties of the skewness statistic for the individual effects and using (A.20), (A.21) and (A.24) for $k = 3$, it can be written

$$\sqrt{N}(\widehat{SK}_b - SK_b) = \frac{\alpha_b^\top}{\widehat{\sigma}_b^3}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}SK_{b,i} + o_{\mathbb{P}}(1), \tag{A.27}$$

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

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

$$SK_{b,2i} = (b_i^2 - \sigma_b^2) + \frac{2}{T}\sum_t b_i v_{it} + \frac{1}{T(T-1)}\sum_t\sum_{s\neq t}v_{it}v_{is}.$$

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

$$\Gamma_b = \begin{bmatrix} \Gamma_{b,1} & \Gamma_{b,2} \\ \Gamma_{b,2} & \Gamma_{b,3} \end{bmatrix},$$

where it can be shown that

$$\Gamma_{b,1} = \frac{1}{N}\sum_i E[SK_{b,1i}^2] = \gamma_b^6 - (\gamma_b^3)^2 + \frac{9}{T}\gamma_b^4\sigma_v^2 + \frac{9}{T(T-1)}\sigma_b^2\sigma_v^4 + \frac{\sigma_v^6}{T(T-1)(T-2)},$$

$$\Gamma_{b,2} = \frac{1}{N}\sum_i E[SK_{b,1i}SK_{b,2i}] = \gamma_b^5 - \gamma_b^3\sigma_b^2 + \frac{6}{T}\gamma_b^3\sigma_v^2,$$

$$\Gamma_{b,3} = \frac{1}{N}\sum_i E[SK_{b,2i}^2] = \gamma_b^4 - \sigma_b^4 + \frac{4}{T}\sigma_b^2\sigma_v^2 + \frac{\sigma_v^4}{T(T-1)}.$$

Finally, using the results of Theorem 3.1, $\widehat{\sigma}_b^2 \xrightarrow{p} \sigma_b^2$, 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 (A.24), for $k = 4$, (A.17) and (A.22), the previous equation turns into

$$\sqrt{NT}(\widehat{KU}_v - KU_v) = \frac{\beta_v^\top}{\widehat{\sigma}_v^4} \frac{1}{\sqrt{NT}} \sum_{it} KU_{v,it} + o_{\mathbb{P}}(1), \tag{A.28}$$

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

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

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

$$\Phi_v = \left[\begin{array}{cc} \Phi_{v,1} & \Phi_{v,2} \\ \Phi_{v,2} & \Phi_{v,4} \end{array} \right],$$

where it is easy to show

$$\begin{aligned}
\Phi_{v,1} = & \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \sum_{t'=1}^T E[KU_{v,1it} KU_{v,1it'}] = \frac{\gamma_v^8}{T} - \frac{(\gamma_v^4)^2}{T} - \frac{4}{T} \gamma_v^5 \gamma_v^3 + \frac{(8\gamma_v^6 \sigma_v^2 + 8(\gamma_v^4)^2)}{T(T-1)} \\
& + \frac{(8(T-2)^2 + 36)(\gamma_v^3)^2 \sigma_v^2}{T(T-1)(T-2)} - \frac{(16T - 20)\gamma_v^4 \sigma_v^4}{T(T-1)(T-2)} + \frac{3(T-3)^2 + 9)\sigma_v^8}{T(T-1)(T-2)(T-3)}, \\
\Phi_{v,2} = & \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \sum_{t'=1}^T E[KU_{v,1it} SK_{v,2it'}] = \frac{\gamma_v^6}{T} - \left(\frac{T-5}{T(T-1)} \right) \gamma_v^4 \sigma_v^4 - \frac{2}{T}(\gamma_v^3)^2 - \frac{2}{T(T-1)} \sigma_v^6, \\
\Phi_{v,3} = & \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \sum_{t'=1}^T E[SK_{v,2it} SK_{v,2it'}] = \frac{\gamma_v^4}{T} - \frac{\sigma_v^4}{T} + \frac{\sigma_v^4}{T(T-1)}.
\end{aligned}$$

Then, using Theorem 3.1 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

$$\sqrt{N}(\widehat{KU}_b - KU_b) = \frac{\beta_b^\top}{\widehat{\sigma}_b^4} \frac{1}{\sqrt{N}} \sum_i KU_{b,i} + o_{\mathbb{P}}(1), \tag{A.29}$$

11

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

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

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

$$
\Phi_b = \begin{bmatrix} \Phi_{b,1} & \Phi_{b,2} \\ \Phi_{b,2} & \Phi_{b,3} \end{bmatrix}, \tag{A.30}
$$

where

$$
\begin{aligned}
\Phi_{b,1} &= \frac{1}{N}\sum_i E[KU_{b,1i}^2] = \gamma_b^8 - (\gamma_b^4)^2 + \frac{16}{T}\gamma_b^6\sigma_v^2 + \frac{36}{T(T-1)}\gamma_b^4\sigma_v^4 + \frac{16\sigma_b^2\sigma_v^6}{T(T-1)(T-2)} \\
&+ \frac{\sigma_v^8}{T(T-1)(T-2)(T-3)}, \\
\Phi_{b,2} &= \frac{1}{N}\sum_i E[KU_{b,1i}SK_{b,2i}] = \gamma_b^6 - \gamma_b^4\sigma_v^2 + \frac{8}{T}\gamma_b^4\sigma_v^2 + \frac{6}{T(T-1)}\sigma_b^2\sigma_v^4, \\
\Phi_{b,3} &= \frac{1}{N}\sum_i E[SK_{b,2i}^2] = \gamma_b^4 - (\sigma_b^2)^2 + \frac{4}{T}\sigma_b^2\sigma_v^2 + \frac{\sigma_v^4}{T(T-1)}.
\end{aligned}
$$

Finally, using the results of Theorem 3.1, $\widehat{\sigma}_b^2 \xrightarrow{p} \sigma_b^2$, and by the central limit theorem, the second part of Theorem 4.3 holds.

∎

Proof of Theorem 4.5 As it is pointed out in Bai and Ng (2005), in order to show that $\widetilde{\pi}_{v34} \xrightarrow{d} \chi_2^2$ it suffices to prove that $\widehat{\pi}_{v3}$ and $\widehat{\pi}_{v4}$ are asymptotically independent so that the squared value of these terms are also asymptotically independent.

In this situation, the limit of $\widehat{\pi}_{v3}$ is determined only by the behavior of $SK_{v,it} = SK_{v,1it}$ because the second element of α_v is zero under normality. Meanwhile, the limit of $\widehat{\pi}_{v4}$ is determined by $KU_{v,it}$, with $\gamma_v^4 = 3\sigma_v^4$. Following a similar procedure as in the proofs of Theorems 4.1 and 4.3, it is straightforward to show that $E\left[\frac{1}{\sqrt{NT}}\sum_{it} SK_{v,it}KU_{v,it}\right] = 0$ under normality. Further, this reasoning is also valid for $\widetilde{\pi}_{b34}$.

∎

Appendix B

As in the paper, using Lemma 2.1 and rearranging terms, the proposed estimators for the higher-order moments up to the eight moment of both individual heterogeneity and random error are of the form

$$\widehat{\gamma}_v^5 = \frac{1}{NT(T-1)\tau_5} \sum_{i=1}^{N} \left[T^3(2T^2 - 3T + 15) \sum_t \widehat{\epsilon}_{it}^5 - 5T^2(T^2 - 2T + 11) \sum_t \widehat{\epsilon}_{it}^4 \sum_t \widehat{\epsilon}_{it} \right.$$

$$- \left. 10T(T-7) \sum_t \widehat{\epsilon}_{it}^3 \left(\sum_t \widehat{\epsilon}_{it} \right)^2 + 10(T^2-3) \sum_t \widehat{\epsilon}_{it}^2 \left(\sum_t \widehat{\epsilon}_{it} \right)^3 - (7T-3) \left(\sum_t \widehat{\epsilon}_{it} \right)^5 \right], \qquad \text{(A.31)}$$

$$\widehat{\gamma}_b^5 = \frac{1}{NT(T-1)(T-2)\tau_5} \sum_{i=1}^{N} \left[(2T^2 + 8T - 9) \left(\sum_t \widehat{\epsilon}_{it} \right)^5 - 20T(T-2) \sum_t \widehat{\epsilon}_{it}^2 (\widehat{\epsilon}_{it})^3 \right.$$

$$- \left. 10(2T^2 + 6T - 5) \sum_t \widehat{\epsilon}_{it}^3 (\widehat{\epsilon}_{it})^2 + 5(2T^3 + 8T^2 - 12T - 27) \sum_t \widehat{\epsilon}_{it}^4 \widehat{\epsilon}_{it} - (10T^3 + 7T^2 + 18T - 54) \sum_t \widehat{\epsilon}_{it}^5 \right],$$
$$\text{(A.32)}$$

where $\tau_5 = 2T^4 - 6T^3 + 19T^2 - 36T + 27$. In addition,

$$\widehat{\gamma}_v^6 = \frac{1}{NT(T-1)(T-2)\tau_6} \sum_{i=1}^{N} \left[2T(T^6 - 15T^5 + 40T^4 - 237T^3 + 613T^2 - 264T - 60) \sum_t \widehat{\epsilon}_{it}^6 \right.$$

$$- 3(2T^6 + 30T^5 - 203T^4 + 75T^3 + 471T^2 + 48T - 666) \sum_t \widehat{\epsilon}_{it}^5 \left(\sum_t \widehat{\epsilon}_{it} \right)$$

$$- 15T(16T^2 - 72T - 25) \sum_t \widehat{\epsilon}_{it}^4 \left(\sum_t \widehat{\epsilon}_{it} \right)^2 - 30(3T^3 - 16T^2 - 5) \sum_t \widehat{\epsilon}_{it}^3 \left(\sum_t \widehat{\epsilon}_{it} \right)^3$$

$$- 30(T-1)(T^2 + T - 9) \sum_t \widehat{\epsilon}_{it}^2 \left(\sum_t \widehat{\epsilon}_{it} \right)^4 + 3(8T^2 - 15T + 16) \left(\sum_t \widehat{\epsilon}_{it} \right)^6$$

$$+ \left. 5(2T^6 - 15T^5 + 77T^4 - 213T^3 + 284T^2 - 132T - 30) \left(\sum_t \widehat{\epsilon}_{it}^3 \right)^2 \right]$$

$$+ \frac{45}{\tau_6} \left[(T-2)(2T^3 + 8T^2 - 35T + 34)(\widehat{\sigma}_v^2)^3 - (2T^6 - 19T^5 + 17T^4 - 50T^3 + 444T^2 - 120T + 50)\widehat{\sigma}_b^2(\widehat{\sigma}_v^2)^2 \right],$$
$$\text{(A.33)}$$

$$\widehat{\gamma}_b^6 = \frac{1}{NT(T-1)(T-2)\tau_6} \sum_{i=1}^{N} \left[2(T-1)(T^2 + 6T - 10) \left(\sum_t \widehat{\epsilon}_{it} \right)^6 - 30(T-2)(T^2 - T - 3) \sum_t \widehat{\epsilon}_{it}^2 (\widehat{\epsilon}_{it})^4 \right.$$

$$+ 10(5T^3 - 39T^2 + 103T - 66) \sum_t \widehat{\epsilon}_{it}^3 (\widehat{\epsilon}_{it})^3 - 30(5T^3 - 33T^2 + 76T - 45) \sum_t \widehat{\epsilon}_{it}^4 (\widehat{\epsilon}_{it})^2$$

$$+ 6(5T^4 + 24T^3 - 194T^2 + 315T - 54) \sum_t \widehat{\epsilon}_{it}^5 \sum_t \widehat{\epsilon}_{it} - 3(T-2)(5T^4 + 65T^3 - 113T^2 - 182T + 84) \sum_t \widehat{\epsilon}_{it}^6$$

$$+ \left. 5(3T^5 + 27T^4 - 149T^3 + 120T^2 + 146T - 132) \left(\sum_t \widehat{\epsilon}_{it}^3 \right)^2 \right] + \frac{30}{\tau_6}(T-2)(2T^3 - 14T^2 + 13T - 8)(\widehat{\sigma}_v^2)^3$$

$$+ \frac{45}{\tau_6}(T^5 - 33T^4 + 167T^3 - 354T^2 + 192T + 96)\widehat{\sigma}_b^2(\widehat{\sigma}_v^2)^2, \qquad \text{(A.34)}$$

where $\tau_6 = 2T^6 - 18T^5 + 41T^4 - 141T^3 + 569T^2 - 786T + 252$. Finally,

$$
\begin{aligned}
\widehat{\gamma}_v^8 &= \frac{1}{NT(T-1)(T-2)} \sum_{i=1}^N \left[T^2 \sum_{t=1}^T \widehat{\epsilon}_{it}^8 - 3T \sum_{t=1}^T \widehat{\epsilon}_{it}^7 \left(\sum_{s=1}^T \widehat{\epsilon}_{is} \right) + 2 \sum_{t=1}^T \widehat{\epsilon}_{it}^6 \left(\sum_{t=1}^T \widehat{\epsilon}_{it} \right)^2 \right] \\
&\quad - \frac{1}{(T-2)} \left[2\widehat{\gamma}_v^6 \widehat{\sigma}_v^2 + 40\widehat{\gamma}_b^3 \widehat{\gamma}_v^3 \widehat{\sigma}_v^2 + (9T-30)\widehat{\sigma}_b^2 \widehat{\gamma}_v^6 + 30\widehat{\sigma}_b^2 \widehat{\sigma}_v^2 \widehat{\gamma}_v^4 + (5T-40)\widehat{\gamma}_b^3 \widehat{\gamma}_v^5 \right. \\
&\quad\left. - (5T+30)\widehat{\gamma}_b^4 \widehat{\gamma}_v^4 - (9T+12)\widehat{\gamma}_b^5 \widehat{\gamma}_v^3 + 30\widehat{\gamma}_b^4 (\widehat{\sigma}_v^2)^2 - 5T\widehat{\gamma}_b^6 \widehat{\sigma}_v^2 \right], \quad\quad (A.35)\\
\widehat{\gamma}_b^8 &= \frac{1}{NT(T-1)(T-2)} \sum_{i=1}^N \left[\sum_{t=1}^T \widehat{\epsilon}_{it}^6 \left(\sum_{t=1}^T \widehat{\epsilon}_{it} \right)^2 - 3 \sum_{t=1}^T \widehat{\epsilon}_{it}^7 \left(\sum_{t=1}^T \widehat{\epsilon}_{it} \right) + 2 \sum_{t=1}^T \widehat{\epsilon}_{it}^8 \right] \\
&\quad - \frac{1}{(T-2)} \left[\widehat{\gamma}_v^6 \widehat{\sigma}_v^2 + 15\widehat{\sigma}_b^2 \widehat{\sigma}_v^2 \widehat{\gamma}_v^4 + (T-8)\widehat{\sigma}_b^2 \widehat{\gamma}_v^6 + 20\widehat{\gamma}_b^3 \widehat{\gamma}_v^3 \widehat{\sigma}_v^2 + (6T-27)\widehat{\gamma}_b^3 \widehat{\gamma}_v^5 \right. \\
&\quad\left. + (15T-50)\widehat{\gamma}_b^4 \widehat{\gamma}_v^4 + 15\widehat{\gamma}_b^4 (\widehat{\sigma}_v^2)^2 + (20T-55)\widehat{\gamma}_b^5 \widehat{\gamma}_v^3 + (15T-35)\widehat{\gamma}_b^6 \widehat{\sigma}_v^2 \right]. \quad\quad (A.36)
\end{aligned}
$$

∎

Appendix C

0.1   Monte Carlo experiment

In the paper we consider the following data generating process (DGP),

$$
Y_{it} = X_{it}^\top m(Z_{it}) + b_i + v_{it}, \quad i = 1, \dots, N; \quad t = 1, \dots, T, \quad\quad (A.37)
$$

where v_{it} is a composed error term, which has two components: a non-negative error term to account for technical inefficiency, u_{it}, and a symmetric error term to account of other random errors, ν_{it}. Following Aigner et al. (1977), it is assumed that $\nu_{it} \sim i.i.d.0.5\mathcal{N}(0, \sigma_\nu^2)$ and it is independent of u_{it}, that is assumed to be $u_{it} \sim i.i.d.\mathcal{N}^+(0, \sigma_u^2)$. The notation "+" indicates that the underlying distribution is truncated from below at zero so that $u_{it} \geq 0$. Additionally, to generate data for simulation, 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 $Z_{it} = \varpi_{it} + \varpi_{i(t-1)}$, where $\varpi_{it} \sim i.i.d.U[0, \pi/2]$ and $\xi_{it} \sim i.i.d.\mathcal{N}(0, 1)$.

For the above sets of experiments, we consider different processes for both the composed error term, v_{it}, and the individual effects, b_i. Further, the variance ratio $\lambda = \sigma_u/\sigma_\nu$ may affect model estimation. Therefore, in the following we hold σ_ν^2 fixed at 0.1 and consider alternative values of λ.

Case A: no technical inefficiency and symmetric distributions (i.e., $v_{it} = \nu_{it}$):

$$
\begin{aligned}
&DGP.A1. \quad \nu_{it} \sim i.i.d.\ 0.5\mathcal{N}(0,1) \quad ; \quad b_i \sim i.i.d.\ 0.5\mathcal{N}(0,1); \\
&DGP.A2. \quad \nu_{it} \sim i.i.d.\ 0.5\mathcal{N}(0,1) \quad ; \quad b_i \sim i.i.d.\ 0.5t(9); \\
&DGP.A3. \quad \nu_{it} \sim i.i.d.\ 0.5t(9) \quad\quad\ ; \quad b_i \sim i.i.d.\ 0.5\mathcal{N}(0,1); \\
&DGP.A4. \quad \nu_{it} \sim i.i.d.\ 0.5t(9) \quad\quad\ ; \quad b_i \sim i.i.d.\ 0.5t(9).
\end{aligned}
$$

Case B: no technical inefficiency and asymmetric distributions (i.e., $v_{it} = \nu_{it}$):

$$DGP.B1. \quad \nu_{it} \sim i.i.d.\ 0.5\mathcal{N}(0,1) \qquad\quad ; \quad b_i \sim i.i.d.\ 0.5exp(\mathcal{N}(0,1)));$$
$$DGP.B2. \quad \nu_{it} \sim i.i.d.\ 0.5exp(\mathcal{N}(0,1)) \quad ; \quad b_i \sim i.i.d.\ 0.5\mathcal{N}(0,1);$$
$$DGP.B3. \quad \nu_{it} \sim i.i.d.\ 0.5exp(\mathcal{N}(0,1)) \quad ; \quad b_i \sim i.i.d.\ 0.5exp(\mathcal{N}(0,1)).$$

Case C: technical inefficiency (i.e., $v_{it} = \nu_{it} - u_{it}$), where $\nu_{it} \sim i.i.d. 0.5\mathcal{N}(0,\sigma_\nu^2)$:

$$DGP.C1. \quad u_{it} \sim i.i.d.\ \mathcal{N}^+(0,\sigma_\nu^2\lambda^2) \quad ; \quad b_i \sim i.i.d.\ 0.5\mathcal{N}(0,1) \quad ; \quad \lambda = 0.5;$$
$$DGP.C2. \quad u_{it} \sim i.i.d.\ \mathcal{N}^+(0,\sigma_\nu^2\lambda^2) \quad ; \quad b_i \sim i.i.d.\ 0.5\mathcal{N}(0,1) \quad ; \quad \lambda = 0.75;$$
$$DGP.C3. \quad u_{it} \sim i.i.d.\ \mathcal{N}^+(0,\sigma_\nu^2\lambda^2) \quad ; \quad b_i \sim i.i.d.\ 0.5\mathcal{N}(0,1) \quad ; \quad \lambda = 1.$$

The simulation results are based on 1000 samples of data $\{(X_{it}, Z_{it}, Y_{it}) : i = 1, \ldots, N, t = 1, \ldots, T\}$. The number of time observations T is set at 4, while the number of cross-sections N is 50, 100, or 150. A complete set of simulation results is given in Tables 1-5.

References

Aigner, D., C. Lovell, and P. Schmidt (1977). "Formulation and estimation of stochastic frontier production function models". *Journal of Econometrics 6*, 21–37.

Bai, J. and S. Ng (2005). Tests for skewness, kurtosis, and normality for time series data. *Journal of Business & Economic Statistics 23*, 49–60.

Gao, J. (1995). The laws of the iterated logarithm of some estimates in a partly linear models. *Statistics & Probability Letters 25*, 153–162.

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

Table 1. Mean and standard deviation (Sd) of the RMSE of the estimators for the nonparametric component.

	N	Results	FEE	OBE	PLSE	PDE
DGP.A1	50	Mean	0.500	0.555	0.769	0.513
		Sd	0.086	0.029	0.049	0.040
	100	Mean	0.448	0.528	0.763	0.479
		Sd	0.061	0.028	0.046	0.030
	150	Mean	0.426	0.510	0.769	0.461
		Sd	0.051	0.018	0.045	0.025
DGP.A2	50	Mean	0.503	0.555	0.756	0.511
		Sd	0.087	0.027	0.051	0.039
	100	Mean	0.447	0.529	0.751	0.479
		Sd	0.058	0.026	0.043	0.029
	150	Mean	0.425	0.511	0.750	0.462
		Sd	0.055	0.018	0.050	0.025
DGP.A3	50	Mean	0.498	0.554	0.772	0.511
		Sd	0.086	0.031	0.054	0.045
	100	Mean	0.445	0.526	0.766	0.476
		Sd	0.066	0.026	0.046	0.034
	150	Mean	0.433	0.514	0.764	0.465
		Sd	0.050	0.019	0.043	0.025
DGP.A4	50	Mean	0.498	0.555	0.767	0.510
		Sd	0.088	0.030	0.047	0.045
	100	Mean	0.450	0.529	0.765	0.483
		Sd	0.058	0.029	0.046	0.031
	150	Mean	0.431	0.514	0.758	0.465
		Sd	0.053	0.017	0.041	0.023
DGP.B1	50	Mean	0.500	0.556	0.759	0.513
		Sd	0.084	0.028	0.049	0.040
	100	Mean	0.449	0.529	0.764	0.481
		Sd	0.060	0.053	0.048	0.030
	150	Mean	0.430	0.511	0.767	0.464
		Sd	0.049	0.018	0.048	0.026
DGP.B2	50	Mean	0.494	0.555	0.760	0.511
		Sd	0.080	0.028	0.051	0.040
	100	Mean	0.448	0.529	0.765	0.481
		Sd	0.060	0.026	0.046	0.030
	150	Mean	0.431	0.514	0.768	0.465
		Sd	0.046	0.016	0.051	0.020
DGP.B3	50	Mean	0.495	0.553	0.758	0.510
		Sd	0.081	0.028	0.051	0.039
	100	Mean	0.442	0.525	0.771	0.478
		Sd	0.062	0.025	0.051	0.028
	150	Mean	0.430	0.515	0.765	0.465
		Sd	0.058	0.021	0.044	0.021
DGP.C1	50	Mean	0.491	0.552	0.756	0.508
		Sd	0.077	0.025	0.049	0.037
	100	Mean	0.447	0.528	0.766	0.479
		Sd	0.059	0.024	0.048	0.029
	150	Mean	0.421	0.511	0.759	0.462
		Sd	0.053	0.016	0.041	0.024
DGP.C2	50	Mean	0.490	0.552	0.755	0.508
		Sd	0.081	0.026	0.047	0.037
	100	Mean	0.456	0.527	0.756	0.481
		Sd	0.061	0.025	0.042	0.027
	150	Mean	0.411	0.509	0.772	0.455
		Sd	0.052	0.019	0.041	0.023
DGP.C3	50	Mean	0.490	0.552	0.755	0.508
		Sd	0.081	0.026	0.047	0.037
	100	Mean	0.456	0.528	0.756	0.481
		Sd	0.061	0.025	0.042	0.027
	150	Mean	0.429	0.512	0.759	0.461
		Sd	0.061	0.016	0.039	0.021

Table 2. Bias, standard deviation (Sd), and RMSE of the estimators for higher-order moments.

	N	Results	Remainder component $\widehat{\sigma}_v^2$	$\widehat{\gamma}_v^3$	$\widehat{\gamma}_v^4$	Individual component $\widehat{\sigma}_b^2$	$\widehat{\gamma}_b^3$	$\widehat{\gamma}_b^4$
DGP.A1		True	0.250	0.000	0.187	0.250	0.000	0.187
	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
		RMSE	0.328	0.151	1.140	0.074	0.113	0.228
	100	Bias	0.291	0.001	0.820	0.007	0.011	0.013
		Sd	0.050	0.093	0.319	0.066	0.096	0.171
		RMSE	0.295	0.095	0.879	0.049	0.075	0.151
	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
		RMSE	0.277	0.078	0.805	0.038	0.064	0.112
DGP.A2		True	0.250	0.000	0.187	0.321	0.000	0.444
	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
		RMSE	0.329	0.149	1.100	0.082	0.156	0.394
	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
		RMSE	0.288	0.095	0.852	0.055	0.107	0.242
	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
		RMSE	0.272	0.075	0.789	0.039	0.071	0.188
DGP.A3		True	0.321	0.000	0.434	0.250	0.002	0.187
	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
		RMSE	0.329	0.202	1.339	0.074	0.121	0.232
	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
		RMSE	0.286	0.126	0.990	0.052	0.087	0.150
	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
		RMSE	0.280	0.099	0.973	0.039	0.057	0.116
DGP.A4		True	0.321	0.000	0.434	0.321	0.002	0.444
	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
		RMSE	0.320	0.177	1.256	0.086	0.157	0.352
	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
		RMSE	0.295	0.122	1.041	0.058	0.113	0.352
	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
		RMSE	0.276	0.089	0.934	0.047	0.090	0.213
DGP.B1		True	0.250	0.000	0.187	1.857	12.887	273.23
	50	Bias	0.319	0.009	0.967	0.009	0.019	0.138
		Sd	0.075	0.144	0.510	0.279	1.947	18.977
		RMSE	0.327	0.145	1.091	0.087	0.346	3.488
	100	Bias	0.292	-0.003	0.854	0.007	0.024	0.854
		Sd	0.053	0.095	0.339	0.202	1.875	12.790
		RMSE	0.296	0.094	0.981	0.062	0.319	2.441
	150	Bias	0.275	-0.002	0.776	0.004	0.002	0.019
		Sd	0.040	0.086	0.260	0.130	0.758	5.903
		RMSE	0.277	0.085	0.817	0.041	0.131	1.053

Table 3. Bias, standard deviation (Sd), and RMSE of the estimators for higher-order moments.

	N	Results	Remainder component			Individual component		
			$\widehat{\sigma}_v^2$	$\widehat{\gamma}_v^3$	$\widehat{\gamma}_v^4$	$\widehat{\sigma}_b^2$	$\widehat{\gamma}_b^3$	$\widehat{\gamma}_b^4$
DGP.B2		True	1.847	11.280	188.9	0.250	0.000	0.187
	50	Bias	0.215	-0.226	0.229	0.264	0.421	0.507
		Sd	0.128	0.910	9.151	0.124	0.219	0.408
		RMSE	0.228	0.363	2.443	0.280	0.454	0.610
	100	Bias	0.185	-0.228	-0.128	0.271	0.429	0.532
		Sd	0.100	0.883	7.517	0.083	0.150	0.294
		RMSE	0.191	0.303	2.261	0.279	0.446	0.590
	150	Bias	0.167	-0.201	0.073	0.273	0.437	0.534
		Sd	0.073	0.596	7.133	0.068	0.109	0.199
		RMSE	0.171	0.248	0.444	0.278	0.444	0.559
DGP.B3		True	1.847	11.280	188.9	1.847	12.887	273.23
	50	Bias	0.207	-0.274	-0.929	0.809	2.629	15.980
		Sd	0.195	3.375	77.863	1.360	18.778	325.851
		RMSE	0.222	0.805	17.766	0.832	3.268	38.696
	100	Bias	0.152	-0.219	-0.210	0.940	3.298	28.262
		Sd	0.139	1.354	19.478	1.250	12.744	296.265
		RMSE	0.173	0.318	5.097	0.781	3.216	28.8319
	150	Bias	0.165	-0.244	-0.250	0.822	2.536	12.689
		Sd	0.102	1.209	11.932	0.878	10.715	169.130
		RMSE	0.172	0.349	2.752	0.829	2.727	18.808
DGP.C1		True	0.075	-0.031	0.016	0.250	0.000	0.187
	50	Bias	0.292	0.074	0.681	-0.053	-0.146	-0.081
		Sd	0.072	0.110	0.384	0.077	0.108	0.384
		RMSE	0.301	0.134	0.781	0.083	0.168	0.181
	100	Bias	0.264	0.070	0.594	-0.045	-0.148	-0.071
		Sd	0.045	0.070	0.237	0.055	0.073	0.123
		RMSE	0.268	0.099	0.639	0.062	0.159	0.124
	150	Bias	0.240	0.075	0.521	-0.044	-0.145	-0.087
		Sd	0.037	0.056	0.226	0.039	0.056	0.105
		RMSE	0.243	0.094	0.568	0.053	0.151	0.103
DGP.C2		True	0.100	-0.049	0.030	0.250	0.000	0.187
	50	Bias	0.270	0.095	0.671	-0.049	-0.179	-0.070
		Sd	0.066	0.113	0.393	0.076	0.111	0.196
		RMSE	0.278	0.149	0.779	0.079	0.202	0.193
	100	Bias	0.254	0.093	0.641	-0.049	-0.179	-0.076
		Sd	0.046	0.077	0.263	0.054	0.077	0.124
		RMSE	0.258	0.120	0.693	0.067	0.191	0.140
	150	Bias	0.224	0.093	0.508	-0.051	-0.176	-0.075
		Sd	0.032	0.055	0.144	0.044	0.039	0.075
		RMSE	0.226	0.106	0.527	0.060	0.179	0.099
DGP.C3		True	0.047	-0.069	1.696	0.250	0.000	0.187
	50	Bias	0.255	0.121	0.668	-0.049	-0.205	-0.071
		Sd	0.067	0.116	0.405	0.077	0.112	0.197
		RMSE	0.264	0.169	0.782	0.082	0.227	0.197
	100	Bias	0.238	0.118	0.639	-0.049	-0.206	-0.077
		Sd	0.047	0.078	0.271	0.054	0.077	0.125
		RMSE	0.243	0.142	0.693	0.068	0.217	0.143
	150	Bias	0.215	0.113	0.510	-0.052	-0.206	-0.084
		Sd	0.039	0.056	0.223	0.043	0.048	0.078
		RMSE	0.219	0.125	0.556	0.064	0.211	0.106

Table 4. Size and power of the symmetry, kurtosis, and normality tests.

	N	α	SK_u	KU_u	$\widehat{\pi}_{v3}$	$\widehat{\pi}_{v4}$	$\widetilde{\pi}_{v34}$	SK_b	KU_b	$\widehat{\pi}_{b3}$	$\widehat{\pi}_{b4}$	$\widetilde{\pi}_{b34}$
					Remainder component					Individual component		
DGP.A1	50	0.10	0	3	0.087	0.093	0.072	0	3	0.132	0.079	0.070
		0.05	0	3	0.047	0.052	0.033	0	3	0.065	0.037	0.029
		0.01	0	3	0.017	0.016	0.006	0	3	0.006	0.009	0.004
	100	0.10	0	3	0.108	0.083	0.090	0	3	0.098	0.104	0.074
		0.05	0	3	0.055	0.054	0.052	0	3	0.036	0.061	0.032
		0.01	0	3	0.020	0.017	0.007	0	3	0.003	0.017	0.002
	150	0.10	0	3	0.109	0.105	0.107	0	3	0.114	0.107	0.101
		0.05	0	3	0.051	0.057	0.052	0	3	0.054	0.054	0.052
		0.01	0	3	0.012	0.020	0.006	0	3	0.005	0.029	0.008
DGP.A2	50	0.10	0	3	0.086	0.062	0.063	0	4.309	0.093	0.114	0.089
		0.05	0	3	0.053	0.039	0.035	0	4.309	0.035	0.067	0.040
		0.01	0	3	0.018	0.007	0.002	0	4.309	0.005	0.020	0.003
	100	0.10	0	3	0.091	0.067	0.078	0	4.309	0.091	0.178	0.133
		0.05	0	3	0.051	0.040	0.041	0	4.309	0.038	0.102	0.078
		0.01	0	3	0.013	0.012	0.003	0	4.309	0.004	0.028	0.006
	150	0.10	0	3	0.094	0.094	0.099	0	4.309	0.101	0.225	0.162
		0.05	0	3	0.049	0.055	0.045	0	4.309	0.047	0.132	0.093
		0.01	0	3	0.011	0.012	0.002	0	4.309	0.006	0.036	0.008
DGP.A3	50	0.10	0	4.309	0.087	0.164	0.103	0	3	0.113	0.075	0.068
		0.05	0	4.309	0.050	0.085	0.046	0	3	0.058	0.039	0.042
		0.01	0	4.309	0.013	0.025	0.002	0	3	0.011	0.010	0.004
	100	0.10	0	4.309	0.090	0.264	0.162	0	3	0.122	0.103	0.086
		0.05	0	4.309	0.045	0.153	0.084	0	3	0.062	0.056	0.051
		0.01	0	4.309	0.008	0.037	0.007	0	3	0.014	0.025	0.013
	150	0.10	0	4.309	0.116	0.348	0.247	0	3	0.105	0.100	0.094
		0.05	0	4.309	0.050	0.227	0.140	0	3	0.051	0.055	0.042
		0.01	0	4.309	0.010	0.064	0.020	0	3	0.008	0.019	0.001
DGP.A4	50	0.10	0	4.309	0.109	0.115	0.120	0	4.309	0.094	0.095	0.103
		0.05	0	4.309	0.065	0.062	0.070	0	4.309	0.040	0.067	0.060
		0.01	0	4.309	0.028	0.026	0.039	0	4.309	0.007	0.036	0.030
	100	0.10	0	4.309	0.109	0.234	0.190	0	4.309	0.096	0.159	0.152
		0.05	0	4.309	0.074	0.144	0.125	0	4.309	0.039	0.111	0.106
		0.01	0	4.309	0.027	0.050	0.049	0	4.309	0.005	0.060	0.049
	150	0.10	0	4.309	0.119	0.280	0.225	0	4.309	0.095	0.209	0.196
		0.05	0	4.309	0.067	0.182	0.146	0	4.309	0.041	0.143	0.126
		0.01	0	4.309	0.025	0.070	0.055	0	4.309	0.008	0.075	0.064
DGP.B1	50	0.10	0	3	0.035	0.117	0.020	5.092	79.233	0.624	0.253	0.357
		0.05	0	3	0.027	0.090	0.014	5.092	79.233	0.442	0.145	0.225
		0.01	0	3	0.015	0.039	0.003	5.092	79.233	0.120	0.041	0.033
	100	0.10	0	3	0.053	0.114	0.026	5.092	79.233	0.646	0.391	0.373
		0.05	0	3	0.029	0.102	0.024	5.092	79.233	0.475	0.223	0.254
		0.01	0	3	0.016	0.047	0.004	5.092	79.233	0.202	0.072	0.035
	150	0.10	0	3	0.098	0.111	0.038	5.092	79.233	0.623	0.422	0.408
		0.05	0	3	0.045	0.106	0.033	5.092	79.233	0.506	0.235	0.287
		0.01	0	3	0.019	0.057	0.004	5.092	79.233	0.184	0.088	0.045

Table 5. Size and power of the symmetry, kurtosis, and normality tests.

	N	α	Remainder component					Individual component				
			SK_u	KU_u	$\widehat{\pi}_{v3}$	$\widehat{\pi}_{v4}$	$\widetilde{\pi}_{v34}$	SK_b	KU_b	$\widehat{\pi}_{b3}$	$\widehat{\pi}_{b4}$	$\widetilde{\pi}_{b34}$
DGP.B2	50	0.10	5.092	79.233	0.528	0.140	0.397	0	3	0.116	0.023	0.087
		0.05	5.092	79.233	0.280	0.047	0.200	0	3	0.075	0.016	0.060
		0.01	5.092	79.233	0.028	0.001	0.015	0	3	0.031	0.003	0.015
	100	0.10	5.092	79.233	0.696	0.277	0.505	0	3	0.084	0.046	0.056
		0.05	5.092	79.233	0.502	0.104	0.349	0	3	0.061	0.021	0.035
		0.01	5.092	79.233	0.142	0.010	0.043	0	3	0.023	0.003	0.003
	150	0.10	5.092	79.233	0.715	0.382	0.529	0	3	0.094	0.097	0.081
		0.05	5.092	79.233	0.586	0.176	0.367	0	3	0.053	0.039	0.049
		0.01	5.092	79.233	0.214	0.020	0.063	0	3	0.022	0.004	0.002
DGP.B3	50	0.10	5.092	79.233	0.201	0.024	0.092	5.092	79.233	0.699	0.077	0.643
		0.05	5.092	79.233	0.419	0.007	0.047	5.092	79.233	0.609	0.043	0.547
		0.01	5.092	79.233	0.599	0.004	0.017	5.092	79.233	0.428	0.027	0.359
	100	0.10	5.092	79.233	0.419	0.024	0.237	5.092	79.233	0.723	0.090	0.675
		0.05	5.092	79.233	0.184	0.007	0.110	5.092	79.233	0.647	0.062	0.591
		0.01	5.092	79.233	0.046	0.004	0.035	5.092	79.233	0.496	0.027	0.437
	150	0.10	5.092	79.233	0.599	0.069	0.442	5.092	79.233	0.785	0.134	0.744
		0.05	5.092	79.233	0.404	0.023	0.265	5.092	79.233	0.726	0.097	0.673
		0.01	5.092	79.233	0.102	0.013	0.073	5.092	79.233	0.560	0.033	0.525
DGP.C1	50	0.10	-1.509	2.844	0.504	0.002	0.565	0	3	0.069	0.051	0.082
		0.05	-1.509	2.844	0.580	0.002	0.526	0	3	0.045	0.040	0.059
		0.01	-1.509	2.844	0.508	0.001	0.445	0	3	0.015	0.025	0.026
	100	0.10	-1.509	2.844	0.604	0.002	0.592	0	3	0.058	0.052	0.061
		0.05	0.088	2.844	0.590	0.001	0.591	0	3	0.032	0.038	0.040
		0.01	-1.509	2.844	0.508	0.001	0.584	0	3	0.007	0.022	0.024
	150	0.10	-1.509	2.844	0.604	0.001	0.601	0	3	0.054	0.050	0.068
		0.05	-1.509	2.844	0.603	0.001	0.601	0	3	0.029	0.033	0.036
		0.01	-1.509	2.844	0.601	0.001	0.600	0	3	0.004	0.017	0.013
DGP.C2	50	0.10	-1.549	3	0.618	0.005	0.612	0	3	0.069	0.037	0.065
		0.05	-1.549	3	0.617	0.005	0.598	0	3	0.048	0.029	0.044
		0.01	-1.549	3	0.588	0.001	0.553	0	3	0.017	0.014	0.016
	100	0.10	-1.549	3	0.657	0.002	0.656	0	3	0.044	0.053	0.063
		0.05	-1.549	3	0.657	0.002	0.655	0	3	0.026	0.037	0.041
		0.01	-1.549	3	0.656	0.002	0.655	0	3	0.011	0.023	0.021
	150	0.10	-1.549	3	0.679	0.004	0.669	0	3	0.035	0.042	0.048
		0.05	-1.549	3	0.659	0.004	0.659	0	3	0.020	0.025	0.035
		0.01	-1.549	3	0.659	0.001	0.658	0	3	0.011	0.011	0.015
DGP.C3	50	0.10	-6.772	767.77	0.643	0.002	0.643	0	3	0.078	0.028	0.063
		0.05	-6.772	767.77	0.642	0.001	0.641	0	3	0.049	0.022	0.037
		0.01	-6.772	767.77	0.640	0.001	0.631	0	3	0.015	0.015	0.018
	100	0.10	-6.772	767.77	0.653	0.002	0.656	0	3	0.040	0.049	0.057
		0.05	-6.772	767.77	0.652	0.000	0.656	0	3	0.022	0.034	0.035
		0.01	-6.772	767.77	0.650	0.000	0.655	0	3	0.009	0.019	0.021
	150	0.10	-6.772	767.77	0.678	0.005	0.679	0	3	0.036	0.027	0.038
		0.05	-6.772	767.77	0.676	0.005	0.678	0	3	0.019	0.019	0.022
		0.01	-6.772	767.77	0.670	0.002	0.677	0	3	0.003	0.011	0.007