



# Assessing the consistency of the fixed-effects estimator: a regression-based Wald test

Laura Spierdijk<sup>1</sup> 

Received: 9 June 2021 / Accepted: 17 August 2022 / Published online: 9 September 2022  
© The Author(s) 2022

## Abstract

Under large- $n$  and fixed- $T$  panel data asymptotics, we develop a method to test a sufficient condition for the FE estimator's consistency using a stacked regression framework. The resulting test exploits a previously unnoted relation between the fixed-effects estimator and the short- and long-differences estimators. It takes the familiar form of a panel-robust Wald test, but is also shown to be asymptotically equivalent to a GMM test. We provide a theoretical comparison between our test and two existing ones from the literature, which are shown to focus on generic strict exogeneity conditions instead of being specifically related to the FE estimator's moment conditions. We investigate our test's finite-sample properties in a simulation study, where we continue the comparison with the other tests. We show that our test has good finite-sample properties, especially if the estimator of the covariance matrix is based on a panel bootstrap. The practical use of our test is illustrated in two applications to existing data from the literature.

**Keywords** Linear panel regression · Fixed-effects estimator · Differences estimators · Wald test · GMM overidentifying test

**JEL Classification** C23 · C52

## 1 Introduction

Since the early days of econometrics, the fixed-effects (or within) estimator has been widely used to estimate the linear panel regression model in the presence of individual effects correlated with the regressors (Mundlak 1961; Mundlak and Hoch 1965). Because the fixed-effects (FE) estimator exploits a single moment condition for each

---

✉ Laura Spierdijk  
L.Spierdijk@utwente.nl

<sup>1</sup> Section Financial Engineering, Department of High-tech Business and Entrepreneurship, Faculty of Behavioural, Management and Social Sciences, University of Twente, Enschede, The Netherlands

covariate, it is just identified. Viewed from a GMM perspective, it is therefore not possible to test the validity of these moment conditions by means of the  $J$ -test for overidentifying moment conditions.<sup>1</sup> In other situations, we may have certain suspicions that some covariates are unlikely to satisfy the moment conditions imposed by the FE estimator. If instrumental variables (IVs) are available for these covariates, we could test the FE estimator against the FE-IV estimator using a Hausman test (Hausman 1978).

To our best knowledge, consistency tests for the FE estimator that do not require IVs are rare. The present study seeks to fill the gap in the literature by developing a test to validate a sufficient condition for the consistency of the FE estimator that does not use such side information. We derive the test by exploiting a previously unnoted relation between the FE estimator and the short- and long-differences estimators. The resulting test can be performed in the familiar form of a panel-robust Wald test for certain parameter restrictions in a stacked regression framework.

Because our test turns out asymptotically equivalent to a GMM test, our approach also fits in the familiar setting of GMM estimation and specification testing. Consequently, the asymptotic properties of our Wald test are standard and well documented in the literature (Newey 1985; Cameron and Trivedi 2005; Hall 2005). The theoretical part of our study uses the link with GMM testing to draw a formal comparison between our test and the ones proposed by Wooldridge (2010) and Su et al. (2016), where the latter is an extension of the former. Both tests make use of auxiliary regressions to assess the validity of certain strict exogeneity conditions. We show that our test validates sufficient conditions for the consistency of the FE estimator, while the other two focus on more generic strict exogeneity conditions that are neither sufficient nor necessary for the consistency of the FE estimator.

We investigate our test's finite-sample properties in a simulation study, where we continue the comparison between our Wald test and the strict exogeneity test of Su et al. (2016). We use a simulation design in which the two tests are either both consistent or both inconsistent. Our Wald test generally exhibits good finite-sample properties, especially if the estimator of the covariance matrix is based on a panel bootstrap.

The empirical behavior of our Wald test is illustrated in two empirical applications that elaborate on existing studies from the literature. Both McKinnish (2008) and Erickson and Whited (2000) apply the linear panel regression model to data sets containing an explanatory variable that is suspected to be subject to measurement error, which would render the FE estimator inconsistent. In the context of our theoretical results, these data sets provide a particularly relevant empirical case for our Wald test. For the linear panel regression model applied to the data of McKinnish (2008), our Wald test rejects the sufficient condition for the FE estimator's consistency. Applied to the data of Erickson and Whited (2000), however, our Wald test finds no such evidence. We also run the test of Su et al. (2016) and draw the comparison with the outcomes of our test.

Because our test validates a sufficient condition for the consistency of the FE estimator, it is possible that the FE estimator is consistent even though this condition does not hold. Furthermore, we will show that there is also a possibility that the Wald test

<sup>1</sup> The  $J$ -test is also known as the Hansen–Sargan, GMM or overidentifying test.

has low power in certain cases. We will provide recommendations on how to remedy such type I and type II errors using additional analysis. Hence, although our test does not require instrumental variables, it should be combined with further investigations.

Our approach connects to different strands of literature. From the time series literature, we take the idea of a test that exploits taking differences (e.g., Plosser et al., 1982, Davidson et al., 1985, Breusch and Godfrey, 1986, Thursby, 1989). We combine this idea with the insight of Griliches and Hausman (1986, p. 114) that the linear panel regression model is misspecified if short- and long-differences estimators differ significantly. The resulting stacked regression framework facilitates researchers to routinely run our Wald test. In this way, we extend the panel data literature about large- $n$  and fixed- $T$  specification testing, which includes but is not limited to tests for overidentifying restrictions (Hayakawa 2019), random effects vs. fixed effects and FE vs. FE-IV (Hausman 1978; Baltagi et al. 2003; Amini et al. 2012; Joshi and Wooldridge 2019), unit roots (Harris and Tzavalis 1999), selectivity bias (Verbeek and Nijman 1992; Wooldridge 1995), cross-sectional dependence (Sarafidis and Wansbeek 2012) and GMM-based test for autocorrelation in error terms (Arellano and Bond 1991).

The setup of the remainder of this study is as follows. Section 2 describes the regression framework that we propose to estimate and test the FE estimator. Section 3 introduces the test statistic and discusses its statistical properties based on the literature (asymptotic behavior), followed by a simulation study (finite-sample behavior) in Sect. 4. Both sections draw the comparison with the test proposed by Su et al. (2016). Our approach is illustrated in Sect. 5, where we provide two applications to existing data from the literature. Lastly, Sect. 6 concludes. An appendix with supplementary material is available.

## 2 Regression framework

We consider the situation that we are interested in estimating the static linear panel regression model with  $T \geq 3$  time observations, given by

$$\mathbf{y}_i = \gamma_i \boldsymbol{\iota}_T + \mathbf{X}_i \boldsymbol{\beta}_0 + \boldsymbol{\varepsilon}_i \quad [i = 1, \dots, n], \quad (1)$$

where  $\mathbf{y}_i$  ( $T \times 1$ ) is the dependent variable,  $\gamma_i$  the individual-specific intercept,  $\boldsymbol{\iota}_T$  ( $T \times 1$ ) a vector of ones,  $\mathbf{X}_i$  ( $T \times k$ ) the matrix of observed covariates,  $\boldsymbol{\beta}_0$  ( $k \times 1$ ) the unknown coefficient vector and  $\boldsymbol{\varepsilon}_i$  ( $T \times 1$ ) the error term.

### 2.1 FE and differences estimators

Let  $\mathbf{D}_j$  be the  $(T-j) \times T$  matrix that takes differences over time span  $j = 1, \dots, T-1$  and write  $\boldsymbol{\Delta}_j = \mathbf{D}_j' \mathbf{D}_j$ . We define  $\hat{\boldsymbol{\beta}}_j$  as the OLS estimator of  $\boldsymbol{\beta}_0$  in (1) after taking differences over time span  $j$ , yielding

$$\hat{\boldsymbol{\beta}}_j = \left( \sum_{i=1}^n \mathbf{X}_i' \boldsymbol{\Delta}_j \mathbf{X}_i \right)^{-1} \sum_{i=1}^n \mathbf{X}_i' \boldsymbol{\Delta}_j \mathbf{y}_i. \quad (2)$$

We denote the centering matrix of order  $T$  by  $\mathbf{A}_T = \mathbf{I}_T - \mathbf{t}_T \mathbf{t}'_T / T$ . This matrix is symmetric, idempotent of rank  $T - 1$  and orthogonal to  $\mathbf{t}_T$ . We observe that  $\Delta_j$  has the  $j$ th pseudo-diagonal equal to  $-1$ , with all other pseudo-diagonals are zero. Moreover,  $\sum_j \Delta_j$  has diagonal elements equal to  $T - 1$  since all rows add to zero. As a result,  $\mathbf{A}_T = (\Delta_1 + \dots + \Delta_{T-1})/T$ .

We use the relation between  $\mathbf{A}_T$  and the  $\Delta_j$ s to rewrite the FE estimator of  $\beta_0$  in terms of the differences estimators  $\hat{\beta}_j$ , resulting in

$$\hat{\beta}_{FE} = \sum_{j=1}^{T-1} \mathbf{W}_j \left( \sum_{i=1}^n \mathbf{X}'_i \Delta_j \mathbf{X}_i \right)^{-1} \sum_{i=1}^n \mathbf{X}'_i \Delta_j \mathbf{y}_i = \sum_{j=1}^{T-1} \mathbf{W}_j \hat{\beta}_j, \quad (3)$$

where

$$\mathbf{W}_j = \left( \sum_{i=1}^n \sum_{t=1}^{T-1} \mathbf{X}'_i \Delta_t \mathbf{X}_i \right)^{-1} \sum_{i=1}^n \mathbf{X}'_i \Delta_j \mathbf{X}_i \quad (4)$$

$$\sum_{j=1}^{T-1} \mathbf{W}_j = \mathbf{I}_k, \quad \mathbf{W}_j \text{ symmetric and positive definite.} \quad (5)$$

The complete derivation of this result is given in Appendix A. This leads to Result 1.

**Result 1** *The FE estimator is the weighted matrix average of differences estimators, i.e.,*

$$\hat{\beta}_{FE} = \sum_{j=1}^{T-1} \mathbf{W}_j \hat{\beta}_j, \quad (6)$$

with  $\mathbf{W}_j$  as in (4).

Now let  $\beta_j \equiv \text{plim}_{n \rightarrow \infty} \hat{\beta}_j$  and assume that standard regularity conditions for large- $n$  and fixed- $T$  panel data hold. We note that

$$\text{plim}_{n \rightarrow \infty} \mathbf{W}_j = \left( \sum_{t=1}^{T-1} \mathbb{E}(\mathbf{X}'_t \Delta_t \mathbf{X}_t) \right)^{-1} \mathbb{E}(\mathbf{X}'_j \Delta_j \mathbf{X}_j), \quad (7)$$

and observe that also these weight matrices sum to the identity matrix. Consequently, from (6) it follows that  $\text{plim}_{n \rightarrow \infty} \hat{\beta}_{FE} = \beta_0$  under  $H_0 : \beta_1 = \dots = \beta_{T-1} = \beta_0$ . Stated differently, a sufficient condition for the consistency of the FE estimator is that all of the differences estimators are consistent. This leads to Corollary 1.

**Corollary 1** *Assume that the large- $n$  and fixed- $T$  panel data regularity conditions for GMM estimators as listed in Su et al. (2016) hold. Then the FE estimator is consistent if each of the differences estimators is consistent; i.e., if  $\text{plim}_{n \rightarrow \infty} \hat{\beta}_j = \beta_0$  ( $j = 1, \dots, T - 1$ ), then  $\text{plim}_{n \rightarrow \infty} \hat{\beta}_{FE} = \beta_0$ .*

We note that  $H_0$  is not a necessary condition for the consistency of the FE estimator. Under  $H_1 : \beta_j \neq \beta_{j+1}$  (at least one  $j = 1, \dots, T - 2$ ), the FE estimator can still be consistent. This result follows directly from Result 1, but we will come back to it in Sect. 3.4.4.

## 2.2 Motivating examples

To illustrate the link between the (in-)consistency of the FE- and differences estimators, Table 1 provides four motivating examples. We consider the linear panel regression model with (i) classical measurement error, (ii) non-classical measurement error, (iii) omitted variables and (iv) simultaneity. The precise model specifications are described in the first column of Table 1 and given in more detail in Section B of the appendix with supplementary material. To ensure stationarity, we assume that all autoregressive parameters fall in the interval  $(-1, 1)$ . In each of the four cases, the FE estimator and differences estimators are inconsistent for non-trivial parameter values. The second and third column in Table 1 report the inconsistencies.<sup>2</sup>

## 2.3 Stacked regression

In order to use Corollary 1 for the construction of a statistical test for a sufficient condition for the FE estimator's consistency, it turns out useful to estimate the  $\beta_j$ s jointly. Let

$$\begin{aligned} \mathbf{y}_{i1} &= \mathbf{D}_1 \mathbf{y}_i = (y_{i2} - y_{i1}, y_{i3} - y_{i2}, \dots, y_{iT} - y_{i,T-1})' \\ \mathbf{y}_{i2} &= \mathbf{D}_2 \mathbf{y}_i = (y_{i3} - y_{i1}, y_{i4} - y_{i2}, \dots, y_{iT} - y_{i,T-2})' \\ &\vdots \\ \mathbf{y}_{i,T-1} &= \mathbf{D}_{T-1} \mathbf{y}_i = y_{iT} - y_{i1}, \end{aligned}$$

and define  $\mathbf{X}_{i1}, \dots, \mathbf{X}_{i,T-1}$  and  $\boldsymbol{\varepsilon}_{i1}, \dots, \boldsymbol{\varepsilon}_{i,T-1}$  analogously. Next, let

$$\tilde{\mathbf{y}}_i = \begin{pmatrix} \mathbf{y}_{i1} \\ \mathbf{y}_{i2} \\ \vdots \\ \mathbf{y}_{i,T-1} \end{pmatrix}, \quad \tilde{\mathbf{X}}_i = \begin{pmatrix} \mathbf{X}_{i1} & 0 & \dots & 0 \\ 0 & \mathbf{X}_{i2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{X}_{i,T-1} \end{pmatrix}, \quad \tilde{\boldsymbol{\varepsilon}}_i = \begin{pmatrix} \boldsymbol{\varepsilon}_{i1} \\ \boldsymbol{\varepsilon}_{i2} \\ \vdots \\ \boldsymbol{\varepsilon}_{i,T-1} \end{pmatrix}.$$

<sup>2</sup> The inconsistencies of the FE estimator are denoted in matrix notation to save space. According to Corollary 1, they equal the weighted average of the inconsistencies of the differences estimators. This result gives some intuition to the matrix expressions.

**Table 1** Motivating examples: inconsistency of difference and FE estimators

Model	Inconsistency $\widehat{\beta}_j$	Inconsistency $\widehat{\beta}_{FE}$
<b>CLASSICAL ME</b>		
$y_{it} = \alpha_i + \beta \xi_{it} + \varepsilon_{it}$ $x_{it} = \xi_{it} + v_{it}$ $\xi_{it} = \rho \xi_{i,t-1} + \theta_{it}$ $v_{it} = \delta v_{i,t-1} + \eta_{it}$ $\sigma_{\theta\eta} = 0$	$-\beta \frac{\sigma_\eta^2 \frac{1-\delta^j}{1-\delta^2}}{\sigma_\theta^2 \frac{1-\rho^j}{1-\rho^2} + \sigma_\eta^2 \frac{1-\delta^j}{1-\delta^2}}$	$\frac{-\text{tr}[\mathbf{A}\mathbf{\Sigma}_v]}{\text{tr}[\mathbf{A}(\mathbf{\Sigma}_\xi + \mathbf{\Sigma}_v)]} \beta$
<b>NON-CLASSICAL ME</b>		
as classical, but $\sigma_{\theta\eta} \neq 0$ $W_j \equiv \frac{1 - (\delta^j + \rho^j)/2}{1 - \delta\rho}$	$-\beta \frac{\sigma_\eta^2 \frac{1-\delta^j}{1-\delta^2} + \sigma_{\theta\eta} W_j}{\sigma_\theta^2 \frac{1-\rho^j}{1-\rho^2} + \sigma_\eta^2 \frac{1-\delta^j}{1-\delta^2} + 2\sigma_{\theta\eta} W_j}$	$\frac{\text{tr}[\mathbf{A}(\mathbf{\Sigma}_{\xi v} - \mathbf{\Sigma}_v)]}{\text{tr}[\mathbf{A}(\mathbf{\Sigma}_\xi + \mathbf{\Sigma}_v)]} \beta$
<b>OMITTED VARIABLES</b>		
$y_{it} = \alpha_i + \beta x_{it} + \gamma z_{it} + \varepsilon_{it}$ $x_{it} = \rho x_{i,t-1} + \theta_{it}$ $z_{it} = \delta z_{i,t-1} + \eta_{it}$	$\gamma \frac{\sigma_{\theta\eta} \frac{1 - (\delta^j + \rho^j)/2}{1 - \delta\rho}}{\sigma_\theta^2 \frac{1 - \rho^j}{1 - \rho^2}}$	$\frac{\text{tr}(\mathbf{A}\mathbf{\Sigma}_{zx})}{\text{tr}(\mathbf{A}\mathbf{\Sigma}_x)} \gamma$
<b>SIMULTANEITY</b>		
$y_{it} = \beta_i + \beta x_{it} + \varepsilon_{it}$ $x_{it} = \alpha_i + \alpha y_{it} + u_{it}$ $u_{it} = \rho u_{i,t-1} + \theta_{it}$	$\frac{(1 - \alpha\beta)\alpha\sigma_\varepsilon^2}{\alpha^2\sigma_\varepsilon^2 + \sigma_\theta^2 \frac{1 - \rho^j}{1 - \rho^2}}$	$\frac{\text{tr}(\mathbf{A}\mathbf{\Sigma}_\varepsilon)\alpha(1 - \alpha\beta)}{\text{tr}(\mathbf{A}\mathbf{\Sigma}_u) + \text{tr}(\mathbf{A}\mathbf{\Sigma}_\varepsilon)\alpha^2}$

*Notation:*  $\sigma_w^2$  denotes the variance of  $w_{it}$ , with  $w \in \{\eta, \theta, \varepsilon\}$ ;  $\sigma_{\theta\eta}$  stands for the covariance between  $\theta_{it}$  and  $\eta_{it}$ ;  $\mathbf{A}$  denotes the  $T \times T$  centering matrix;  $\mathbf{\Sigma}_w$  is the matrix containing the covariances  $\text{Cov}(w_{ns}, w_{nt})$  for  $w \in \{\xi, v, x, \varepsilon, u\}$ ; the matrix  $\mathbf{\Sigma}_{\xi v}$  contains the covariances  $\text{Cov}(\xi_{ns}, v_{nt})$ ;  $\mathbf{\Sigma}_{zx}$  is defined analogously

We then write

$$\tilde{\mathbf{y}}_i = \tilde{\mathbf{X}}_i \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{T-1} \end{pmatrix} + \tilde{\boldsymbol{\varepsilon}}_i. \quad (8)$$

The stacked regression model in (8) allows us to estimate the  $\beta_j$ s jointly by means of OLS. We observe that the FE estimator arises as the constrained OLS estimator of  $\beta_D = (\beta'_1, \beta'_2, \dots, \beta'_{T-1})'$  in (8), under the parameter restriction  $\beta_1 = \dots = \beta_{T-1}$ .

### 3 Test procedure

Corollary 1 states that, if each differences estimator is consistent, also the FE estimator must be consistent. This observation is the starting point of our Wald test. In brief, we first estimate  $\widehat{\beta}_D$  in (8) using (unconstrained) OLS. Subsequently, we use a Wald test to test  $H_0$  against  $H_1$ .

### 3.1 Wald test

To calculate the Wald test statistic, we need a cluster-robust estimator of the asymptotic covariance matrix in addition to  $\hat{\beta}_D$ . This estimator is given by

$$\hat{\Omega}_D = \left( \sum_{i=1}^n \tilde{\mathbf{X}}_i' \tilde{\mathbf{X}}_i \right)^{-1} \sum_{i=1}^n \tilde{\mathbf{X}}_i' \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i' \tilde{\mathbf{X}}_i \left( \sum_{i=1}^n \tilde{\mathbf{X}}_i' \tilde{\mathbf{X}}_i \right)^{-1}, \quad (9)$$

where  $\hat{\mathbf{u}}_i = \tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i' \hat{\beta}_D$ . The Wald test statistic for the parameter restrictions  $\beta_j = \beta_{j+1}$  ( $j = 1, \dots, T-2$ ) is given by

$$q_W = \hat{\beta}_D' \mathbf{R}' \left( \sum_{i=1}^n \mathbf{R} \hat{\Omega}_D \mathbf{R}' \right)^{-1} \mathbf{R} \hat{\beta}_D, \quad (10)$$

where  $\mathbf{R} = \mathbf{B} \otimes \mathbf{I}_k$  and  $\mathbf{B}$  is the  $k(T-2) \times (T-1)$  matrix taking first differences, given by

$$\mathbf{B} = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & -1 & 1 \end{pmatrix}. \quad (11)$$

Under  $H_0$ , the asymptotic distribution of the Wald test statistic is Chi-square with  $k(T-2)$  degrees of freedom, while under fixed alternatives the test statistic converges in probability to infinity (Cameron and Trivedi 2005, Section 7.6.2). We therefore reject  $H_0$  if  $q_W$  exceeds the  $(1-\alpha)\%$  critical value of the Chi-square distribution with  $k(T-2)$  degrees of freedom, with  $\alpha$  the chosen significance level. The usual asymptotic properties of the Wald test hold under standard large- $n$  and fixed- $T$  panel data regularity conditions, as summarized in the following result.

**Result 2** *Under the regularity assumptions as listed in Su et al. (2016), the asymptotic distribution of the Wald test statistic under  $H_0$  is Chi-square with  $k(T-2)$  degrees of freedom. The Wald test statistic converges in probability to infinity under  $H_1$ . The Wald test has nominal asymptotic size under  $H_0$  and unit asymptotic power under  $H_1$ .*

If  $H_0$  is rejected, the pattern in the  $\hat{\beta}_j$ s can help to assess the economic relevance of the rejection. This becomes particularly relevant if  $n$  is large, since large samples incur the risk of detecting economically minor violations of the null (Griliches and Hausman 1986, p. 110). If all  $\hat{\beta}_j$ s are close in value to the FE estimator, the economic importance of the rejection is considered limited. An informal visualization of the Wald test is obtained by plotting each element of  $\hat{\beta}_j$  as a function of  $j$ , with the value of the FE estimator of each covariate's coefficient added as a horizontal line. We will refer to these plots as the 'differences curves.' These curves will be illustrated in the section with empirical applications.

### 3.2 Relation to GMM tests

To analyze the properties of our Wald test in more detail, it turns out useful to draw the parallel with overidentifying tests in a GMM framework. From Newey and West (1987) and Newey and McFadden (1994) and the linearity of the moment conditions, we infer that  $q_W$  is numerically identical to a GMM test statistic for a stacked regression model. This is formalized in Result 3.

**Result 3** *Under the regularity assumptions as listed in Su et al. (2016), the Wald test statistic  $q_W$  is numerically identical to the overidentifying test statistic based on the two-step GMM estimator  $\hat{\beta}_{GMM}$  of  $\beta_0$  in the stacked regression model*

$$\tilde{y}_i = \begin{pmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{i,T-1} \end{pmatrix} \beta_0 + \tilde{\varepsilon}_i, \quad (12)$$

using the instrument matrix  $\tilde{Z}_i = \tilde{X}_i$  provided that both test statistics use the same estimator for the covariance matrix, with the requirement that this estimator is consistent under  $H_0$ .

Result 3 requires both test statistics to use the same consistent estimator for the covariance matrix. In practice, the GMM test statistic uses  $\hat{\beta}_{GMM}$  to obtain a panel-robust estimator of the covariance matrix, while the Wald test statistic uses  $\hat{\beta}_D$  to do so. Because both estimators of the covariance matrix are consistent under the null, this difference in covariance matrices does not matter for the asymptotic properties of the test statistics (Cameron and Trivedi 2005, ). We thus conclude that our Wald test statistic (with the panel-robust estimator of the covariance matrix based on  $\hat{\beta}_D$ ) is asymptotically equivalent to the GMM test statistic (with the panel-robust estimator of the covariance matrix based on  $\hat{\beta}_{GMM}$ ). The two tests have the same asymptotic power and size, under both the null and any (fixed or local) alternative hypothesis (Newey and West 1987; Newey and McFadden 1994).<sup>3</sup>

**Corollary 2** *Under the regularity assumptions as listed in Su et al. (2016), the Wald test statistic  $q_{tinyW}$  is asymptotically equivalent to the overidentifying J-statistic corresponding to the two-step estimator of  $\beta_0$  in (12) with instruments  $\tilde{Z}_i$ . The two tests have the same asymptotic power and size, under both the null and any (fixed or local) alternative hypothesis.*

With  $\tilde{Z}_i$  as the instrument matrix in the equivalent GMM test, we thus see that the overidentifying moment conditions are the  $k(T - 1)$  moment conditions imposed

<sup>3</sup> Based on Newey and West (1987) and Newey and McFadden (1994), we infer that the Wald test statistic is also identical to three other well-known test statistics to test  $H_0$ : the LM test statistic, the distance-difference test statistic (where the distance-difference test is the GMM equivalent of the likelihood-ratio test) and the minimum Chi-square test statistic. Again the equality only holds if the same consistent estimator for the covariance matrix is used for these test statistics.

by the differences estimators. Hence, our test boils down to a GMM test for the overidentifying moment conditions

$$\mathbb{E}(\mathbf{X}_i' \Delta_j \varepsilon_i) = \mathbf{0} \quad [j = 1, \dots, T - 1]. \quad (13)$$

These moment conditions arise by ‘unfolding’ the moment condition imposed by the FE estimator

$$\sum_{j=1}^{T-1} \mathbb{E}(\mathbf{X}_i' \Delta_j \varepsilon_i) = \mathbf{0}. \quad (14)$$

Corollary 1 already established the link between (the probability limits of)  $\widehat{\beta}_{FE}$  and the  $\widehat{\beta}_j$ s. By means of (13) and (14), we have now also shown how ‘unfolding’ connects the moment conditions of the FE and differences estimators.

### 3.3 Trivial power

Result 2 makes clear that the power of the Wald test arises from the differences in the  $\beta_j$ s for different values of  $j$  under  $H_1$ . In certain cases where the FE estimator is inconsistent, such differences may not exist though. Despite the FE estimator’s inconsistency, we will then find  $\beta_1 = \beta_2 = \dots = \beta_{T-1} \neq \beta_0$ . Consequently, the asymptotic rejection rate of the Wald test will be equal to the chosen significance level, yielding ‘trivial’ asymptotic power. As shown by Newey (1985) and (Hall 2005, Ch. 5), the issue of trivial power is inherent with overidentifying tests. These authors also provide a more technical discussion of the region where GMM tests have trivial power.

The practical implication of the existence of a parameter region with trivial asymptotic power is that our Wald test may have low empirical power in certain situations. In the motivating examples of omitted variables and measurement error shown in Table 1, trivial power arises for  $\rho = \delta$ . This can be inferred from the expressions for the  $\beta_j$ s in the table, which do not vary with  $j$  for  $\rho = \delta$ . A test to assess whether two panel variables have the same degree of persistence could therefore prove useful in this scenario. In practice, however, misspecification is likely to be much more complex than for the motivating examples of Table 1. Consequently, we typically do not know when trivial power will arise and to what extent it is related to the persistence in the observed variables. It is therefore hard to think of a statistical test that could be used to recognize a case of trivial power. In fact, to our best knowledge, no remedy against trivial power exists other than cautiously interpreting the outcomes of GMM tests (Parente and Santos Silva 2012). It therefore remains important to look for other evidence against the FE estimator if the test does not reject the null hypothesis, such as coefficient signs and magnitudes that are implausible from an economic perspective.

### 3.4 Comparison with existing tests

As mentioned in Introduction, tests for the consistency of the FE estimator that do not require IVs are rare. The two tests that come closest are the ones of (Wooldridge 2010, p. 324–325) and Su et al. (2016). These approaches also take the linear panel regression model in (1) as the starting point.

#### 3.4.1 Wald test of Wooldridge (2010)

The test of Wooldridge (2010) is based on the auxiliary OLS regression

$$\mathbf{L}_1 \mathbf{y}_i = \gamma_i \mathbf{t}_{T-1} + \mathbf{L}_1 \mathbf{X}_i \boldsymbol{\beta}_0 + \mathbf{F}_1 \mathbf{X}_i \boldsymbol{\zeta}_1 + \mathbf{L}_1 \boldsymbol{\varepsilon}_i, \quad (15)$$

after taking the within transformation. With  $\mathbf{I}_{T-1}$  the identity matrix of order  $T - 1$ , the matrices  $\mathbf{L}_1$  and  $\mathbf{F}_1$  are defined as the  $(T - 1) \times T$  block matrices

$$\mathbf{L}_1 = (\mathbf{I}_{T-1} \mid \mathbf{0}_{T-1}) \text{ and } \mathbf{F}_1 = (\mathbf{0}_{T-1} \mid \mathbf{I}_{T-1}), \quad (16)$$

with  $\mathbf{0}_{T-1}$  a  $(T - 1)$ -dimensional column vector of zeros.

The term  $\mathbf{F}_1 \mathbf{X}_i \boldsymbol{\zeta}_1$  in (15) ensures that the regression model contains the one-period ahead lead values of the covariates as regressors. Because this leads to the loss of the last time period, we need the matrix  $\mathbf{L}_1$  to ensure that the other vectors also contain the right time observations. In more familiar notation, we would write (15) as  $y_{it} = \gamma_i + \mathbf{x}'_{it} \boldsymbol{\beta}_0 + \mathbf{x}'_{i,t+1} \boldsymbol{\zeta}_1 + \varepsilon_{it}$ . The reason that we use the above alternative notation is to facilitate the comparison with our own ‘differences’ approach, as will become clear below.

The test takes the form of a standard panel-robust Wald test for the null hypothesis  $\bar{H}_0 : \boldsymbol{\zeta}_1 = \mathbf{0}$  against the alternative hypothesis  $\bar{H}_1 : \boldsymbol{\zeta}_1 \neq \mathbf{0}$ . It is motivated by the fact that, under strict exogeneity, the FE estimator of  $\boldsymbol{\zeta}_1$  will have a zero probability limit, while the FE estimator of  $\boldsymbol{\beta}_0$  will converge in probability to  $\boldsymbol{\beta}_0$ . Under the usual regularity conditions for panel data, the resulting test has nominal asymptotic size under  $\bar{H}_0$  and unit asymptotic power under  $\bar{H}_1$ .

#### 3.4.2 Sup-Wald test of Wooldridge (2010)

The extension proposed by Su et al. (2016) is based on the idea that Wooldridge’s approach of adding one-period ahead lead values to the regression model is rather arbitrary. They overcome this by allowing for a wider range of leads and lags. More specifically, Su et al. (2016) consider auxiliary regressions of the form

$$\begin{aligned} \mathbf{L}_\ell \mathbf{y}_i &= \gamma_i \mathbf{t}_{T-1} + \mathbf{L}_\ell \mathbf{X}_i \boldsymbol{\beta}_0 + \mathbf{F}_\ell \mathbf{X}_i \boldsymbol{\zeta}_\ell + \mathbf{L}_\ell \boldsymbol{\varepsilon}_i \\ \ell &\in \mathcal{S}_T = \{-T_2, -T_3, \dots, -1, 1, \dots, T_3, T_2\} \\ T_a &= T - a \quad \text{for } a \leq T - 1 \text{ integer,} \end{aligned} \quad (17)$$

where  $\mathbf{L}_\ell$  and  $\mathbf{F}_\ell$  are defined in analogy with (16).

Su et al. (2016) estimate the regression equation in (17) for all  $\ell \in \mathcal{S}_T$  by means of OLS after applying the within transformation.<sup>4</sup> Subsequently, they test the null hypothesis  $\tilde{H}_0 : \boldsymbol{\zeta}_\ell = \mathbf{0}$  for all  $\ell \in \mathcal{S}_T$  against the alternative hypothesis  $\tilde{H}_1 : \boldsymbol{\zeta}_\ell \neq \mathbf{0}$  for some  $\ell \in \mathcal{S}_T$  using a sup-Wald test. The sup-Wald test statistic is obtained as follows. For each individual null hypothesis  $\tilde{H}_0^\ell : \boldsymbol{\zeta}_\ell = \mathbf{0}$ , they calculate the corresponding individual Wald test statistic. Subsequently, the supremum is taken over the individual Wald test statistics, yielding the sup-Wald test statistic. The underlying idea is that the supremum of a range of test statistics will behave more like the most powerful among them. The critical values of the sup-Wald test statistic are determined by means of a panel bootstrap. Under standard panel data regularity conditions, the resulting sup-Wald test has nominal asymptotic size under  $\tilde{H}_0$  and unit asymptotic power under  $\tilde{H}_1$ .

### 3.4.3 GMM framework

To facilitate the comparison with our own approach, it is convenient to draw the parallel with GMM overidentifying tests one more time. From Newey and West (1987) and Newey and McFadden (1994), we infer that the Wald test of Wooldridge (2010) is asymptotically equivalent to a GMM test based on the two-step GMM estimator of  $\boldsymbol{\beta}_0$  in

$$\mathbf{L}_1 \mathbf{y}_i = \gamma_i \mathbf{t}_{T-1} + \mathbf{L}_1 \mathbf{X}_i \boldsymbol{\beta}_0 + \mathbf{L}_1 \boldsymbol{\varepsilon}_i, \quad (18)$$

with instruments  $\mathbf{L}_1 \mathbf{X}_i$  and  $\mathbf{F}_1 \mathbf{X}_i$ , after applying the within transformation to both the regression equation and the instruments; i.e., after pre-multiplying both with the centering matrix of order  $T - 1$ . Hence, Wooldridge's approach tests  $2k$  moment conditions, namely

$$\mathbb{E}(\mathbf{X}_i' \mathbf{A}_{T-1}^* \boldsymbol{\varepsilon}_i) = \mathbf{0} \quad \text{and} \quad \mathbb{E}(\mathbf{X}_i' \mathbf{A}_{T-1}^\dagger \boldsymbol{\varepsilon}_i) = \mathbf{0}, \quad (19)$$

where

$$\begin{aligned} \mathbf{A}_{T-1}^* &= \mathbf{L}_1' \mathbf{A}_{T-1} \mathbf{L}_1 = \left( \begin{array}{c|c} \mathbf{A}_{T-1} & \mathbf{0}_{T-1} \\ \hline \mathbf{0}_{T-1}' & 0 \end{array} \right) \quad \text{and} \\ \mathbf{A}_{T-1}^\dagger &= \mathbf{F}_1' \mathbf{A}_{T-1} \mathbf{L}_1 = \left( \begin{array}{c|c} \mathbf{0}_{T-1}' & 0 \\ \hline \mathbf{A}_{T-1} & \mathbf{0}_{T-1} \end{array} \right). \end{aligned} \quad (20)$$

We apply Newey and West (1987) and Newey and McFadden (1994) one more time to infer that each individual Wald test of  $\tilde{H}_0^\ell$  is equivalent to a GMM test of the  $2k$  overidentifying moment conditions

$$\mathbb{E}(\mathbf{X}_i' \mathbf{A}_{T-|\ell|}^* \boldsymbol{\varepsilon}_i) = \mathbf{0} \quad \text{and} \quad \mathbb{E}(\mathbf{X}_i' \mathbf{A}_{T-|\ell|}^\dagger \boldsymbol{\varepsilon}_i) = \mathbf{0}, \quad (21)$$

<sup>4</sup> Su et al. (2016) also discuss two varieties of their main test: one in which (17) is estimated using the first differences estimator and one in which a subset of  $\mathcal{S}_T$  is considered in  $\tilde{H}_0$ . For the sake of brevity, our discussion here is confined to the more efficient FE estimator that exploits the full set  $\mathcal{S}_T$ .

where  $\mathbf{A}_{T-|\ell|}^*$  and  $\mathbf{A}_{T-|\ell|}^\dagger$  are defined in analogy with (20). For fixed  $\ell$ , the first moment condition in (21) comes close to the moment condition of the within estimator—being  $\mathbb{E}(\mathbf{X}_i' \mathbf{A}_T \varepsilon_i) = \mathbf{0}$ —but applies to the subsample that excludes  $\ell$  years from the full sample. The second moment condition is similar to the first, but is formulated in terms of lead or lags of the regressors. The moment conditions in (21) are necessary—but not sufficient—for strict exogeneity.

We thus see that the Wald test of Wooldridge (2010) and the individual Wald tests constituting the sup-Wald test of Su et al. (2016) reduce to overidentifying tests, like our own Wald test. They both tests conditions that are necessary for strict exogeneity. Because strict exogeneity is a sufficient condition for the consistency of the FE estimator, the conditions validated by the sup-Wald test are neither necessary nor sufficient for the consistency of the FE estimator.

### 3.4.4 Comparison Wald and sup-Wald tests

*Similarities* Because the Wald test of Wooldridge (2010) and the individual Wald tests constituting the sup-Wald test of Su et al. (2016) also reduce to overidentifying tests, like our own Wald test, they will also have a parameter region with trivial power (Newey 1985; Hall 2005). This property will be illustrated in detail in Sect. 4.

*Differences* The main difference among the three tests is that each of them tests a different null hypothesis, which we have labeled as  $H_0$ ,  $\tilde{H}_0$  and  $\tilde{\tilde{H}}_0$ , respectively. Because the three tests involve different parameter restrictions in different auxiliary regressions, it is not straightforward how to compare them. The translation of each test to a GMM framework has turned out to facilitate the comparison of the three null hypotheses and has made clear that each of the three tests focuses on different moment conditions.

According to Result 3, our Wald test looks at the moment conditions of the differences estimators, which arise as the ‘unfolded’ moment condition of the FE estimator. The underlying relation between the (probability limits of the) FE and differences estimators is specified in Corollary 1 and holds under both  $H_0$  and  $H_1$ .

The test of Wooldridge (2010) considers the moment conditions in (19), while the sup-Wald test focuses on those in (21). Both tests focus on generic strict exogeneity conditions instead of being specifically related to the FE estimator’s orthogonality conditions. For fixed  $\ell$ , the FE estimator of  $\zeta_\ell$  in (17) has a zero probability limit under strict exogeneity, while the FE estimator of  $\beta_0$  then converges in probability to  $\beta_0$  (Su et al. 2016). In other cases, however, it is not known how the inconsistency of the FE estimator of  $\beta_0$  (our parameter of interest) relates to the probability limit of the FE estimator of  $\zeta_\ell$  (the auxiliary parameter to run the test).

To formalize the above considerations, let  $\beta_{FE} = \text{plim}_{n \rightarrow \infty} \hat{\beta}_{FE}$  and define  $H_0^{FE} : \beta_{FE} = \beta_0$  and  $H_1^{FE} : \beta_{FE} \neq \beta_0$ . We will now view the Wald and sup-Wald tests as tests of  $H_0^{FE}$  instead of  $H_0$  and  $\tilde{H}_0$ , which means that we will derive expressions for the rejection probabilities under  $H_0^{FE}$  and  $H_1^{FE}$ . This will allow us to compare both tests’ type I and type II errors if viewed as tests of  $H_0^{FE}$ .

We start with the Wald test and assume that  $H_0^{FE}$  is true. In this scenario, there are two cases: (i)  $H_0$  is true or (ii)  $H_1$  holds true. In case (i), the rejection probability under  $H_0^{FE}$  is also a rejection probability under  $H_0$ . The asymptotic value of this probability equals

the nominal size of the Wald test according to Result 2. Case (ii) can occur because consistency of the differences estimators is a sufficient but not a necessary assumption for the consistency of the FE estimator. In case (ii), the rejection probability has a unit asymptotic value, unless there is trivial power. In the latter case, the asymptotic value of the rejection probability is equal to the nominal size of the Wald test. In sum, the asymptotic size of the Wald test is at least nominal if we view the test as a test of  $H_0^{\text{FE}}$ .

At this point, we would ideally provide a specific example of case (ii), where some of the differences estimators are inconsistent while the FE estimator is still consistent. However, it is not straightforward to illustrate the Wald test's oversizedness in this way, because the process of constructing such an example quickly goes beyond what is still analytically tractable.<sup>5</sup> Fortunately, we can characterize the conditions under which case (ii) will occur in general terms using the GMM analogy. If (14) holds true but (13) does not, then the FE estimator is consistent while some of the differences estimators are inconsistent.

Next we assume that  $H_1^{\text{FE}}$  is true. Because  $H_0$  is a sufficient condition for the consistency of the FE estimator,  $H_1$  must also hold true. Hence, the rejection probability under  $H_1^{\text{FE}}$  is also a rejection probability under  $H_1$ , with a unit asymptotic value according to Result 2. In sum, the Wald test viewed as a test of  $H_0^{\text{FE}}$  has unit asymptotic power under  $H_1^{\text{FE}}$ , unless we encounter a scenario of trivial power.

For the sup-Wald test, the derivation of the rejection probability under  $H_0^{\text{FE}}$  is similar as above and leads to the insight that the asymptotic value of this probability is larger than or equal to the nominal size of the sup-Wald test. Under  $H_1^{\text{FE}}$ , the situation is more complex than above, since either  $\tilde{H}_0$  or  $\tilde{H}_1$  may hold true. The reason for this is that the sup-Wald test validates conditions that are neither necessary nor sufficient for the consistency of the FE estimator. As a result, there are two possibilities for the asymptotic power of the sup-Wald test if seen as a test of  $H_0^{\text{FE}}$ . In case (i), it has a unit value under  $H_1^{\text{FE}}$ , which occurs if  $\tilde{H}_1$  holds true and no trivial power arises. In case (ii), it has a value equal to the nominal size of the sup-Wald test. This occurs if  $\tilde{H}_0$  holds true, or if  $\tilde{H}_1$  holds true in combination with trivial power. Also at this point it is hard to construct analytically tractable examples that illustrate both cases.

In sum, if we view the Wald and sup-Wald tests as tests of the consistency of the FE estimator, we conclude that both of them are asymptotically oversized. Furthermore, the Wald test will have unit asymptotic power, unless we encounter a case of trivial power that is inherent with GMM testing. The sup-Wald test has one additional possibility for trivial asymptotic power, which arises because the specific moment conditions validated by this test are neither sufficient nor necessary for the consistency of the FE estimator.

## 4 Simulation study

Our simulation study focus on the motivating examples listed in Table 1, for which the three tests' finite-sample performance is an empirical matter. Because the sup-Wald test of Su et al. (2016) generally turns out superior to the test of Wooldridge (2010),

<sup>5</sup> We note that the differences and FE estimators are all inconsistent in the motivating examples of Table 1.

we confine the comparison to our Wald test and the sup-Wald test. For a detailed simulation study into the finite-sample differences between the tests of Wooldridge (2010) and Su et al. (2016), we refer to the latter study.

#### 4.1 Simulation setup

We start with some explanation about the design of our simulation study. As said, our simulations focus on the motivating examples in Table 1. With exception of two cases of trivial power, both our Wald test and the sup-Wald test turn out consistent in each of these examples. The consistency of our Wald test follows directly from the results in Table 1.

To determine whether the sup-Wald test is consistent in the motivating examples, we have to assess whether  $\tilde{H}_1$  holds true.<sup>6</sup> However, for fixed  $\ell$ , it already turns out infeasible to obtain an analytical expression for the probability limit of the FE estimator of  $\xi_\ell$  under  $\tilde{H}_1^\ell : \xi_\ell = 0$ . Calculations quickly grow complex because the constituent regressions of the sup-Wald test already contain two covariates in the simplest possible case. We therefore take a more practical view to determine whether the sup-Wald test is consistent in each of our motivating examples. Whenever our simulations yield unit empirical power for sufficiently large values of  $n$ , we view this as a reliable indicator that the sup-Wald test is consistent.

We set  $\rho = \delta$  in two of our simulation experiments. Both our Wald test and the sup-Wald test turn out to have trivial power in these cases. For our Wald test, the trivial power follows immediately from the results in Table 1, which show that the  $\beta_j$ s do not vary with  $j$ . For the sup-Wald test, the inconsistency follows from the low empirical power that persists across large values of  $n$  in our simulations.

The more technical details of our simulation design are as follows. We run simulations for three of the motivating examples considered in Table 1: classical measurement error ('ME'), omitted variables ('OV') and simultaneity ('S'). Throughout, we run 10,000 simulation runs for each simulation experiment to obtain the empirical rejection rates for our Wald test. We consider values  $n = 100$ ,  $n = 500$  and  $n = 1000$ , each with  $T = 5$  and  $T = 10$ . We set the significance level for each test equal to 5%. For classical measurement error and omitted variables, we simulate two scenarios in terms of the persistence parameters:  $\delta = 0.3$  and  $\rho = 0.9$  and  $\delta = 0.3$  and  $\rho = 0.6$ . For simultaneity, we consider the cases  $\rho = 0.6$  and  $\rho = 0.9$ . For each simulation experiment, we report the empirical size and power for both tests. The reported empirical power has not been size adjusted.

#### 4.2 Empirical power and size

The right-hand side of Table 3 reports the empirical power for our Wald test, with the full set of parameters for each of the three models as specified in the table notes. These notes also report the probability limits of the underlying models'  $R^2$ , as well as the reliability and noise-to-signal ratios for the models with classical measurement

<sup>6</sup> This follows from the result proved in Su et al. (2016), showing that the sup-Wald test has unit asymptotic power under  $\tilde{H}_1$ .

**Table 2** Simulation results: empirical size

	$n = 100$ sup-Wald	$n = 500$	$n = 1000$	$n = 100$ Wald	$n = 500$	$n = 1000$
$T = 5$						
$\rho = 0.9$						
ME	0.052	0.051	0.057	0.057	0.052	0.056
OV	0.062	0.055	0.052	0.070	0.054	0.050
S	0.048	0.044	0.061	0.068	0.048	0.056
$\rho = 0.6$						
ME	0.061	0.040	0.053	0.059	0.056	0.050
OV	0.047	0.049	0.055	0.063	0.049	0.049
S	0.047	0.054	0.058	0.066	0.054	0.050
$T = 10$						
$\rho = 0.9$						
ME	0.058	0.056	0.048	0.099 (0.050)	0.058	0.059
OV	0.052	0.061	0.051	0.102 (0.052)	0.059	0.060
S	0.064	0.053	0.045	0.117 (0.049)	0.051	0.052
$\rho = 0.6$						
ME	0.052	0.058	0.053	0.099 (0.049)	0.059	0.058
OV	0.058	0.060	0.048	0.110 (0.063)	0.064	0.059
S	0.050	0.055	0.044	0.099 (0.046)	0.050	0.047

*Notes:* All simulation results are based on 10,000 simulation runs. The empirical power is obtained as the fraction of the number of simulation runs in which the test rejects the null hypothesis, while the empirical size is obtained as the fraction of the number of simulation runs in which the test rejects the null hypothesis, while the null hypothesis is true. For the Wald test, the rejection rates for the bootstrap-based version of the test are in parentheses (1,000 bootstrap runs). Throughout, the significance level is 5%. The simulated models correspond to three of the illustrative cases listed in Table 1 of the main text. Parameters for classical measurement error ('ME'):  $\beta = 1$ ,  $\sigma_\varepsilon^2 = 1$ ,  $\sigma_\theta^2 = 1.44$ ,  $\sigma_\eta^2 = 0.64$ ,  $\delta = 0.3$ . This yields reliabilities of 0.92 ( $\rho = 0.9$ ) and 0.76 ( $\rho = 0.6$ ), noise-to-signal ratios of 0.09 ( $\rho = 0.9$ ) and 0.312 ( $\rho = 0.6$ ) and probability limits for the model's  $R^2$  of 0.88 ( $\rho = 0.9$ ) and 0.69 ( $\rho = 0.6$ ). Parameters for omitted variables ('OV'):  $\beta = 1$ ,  $\gamma = 1$ ,  $\sigma_\varepsilon^2 = 0.25$ ,  $\sigma_\theta^2 = 0.36$ ,  $\sigma_\eta^2 = 0.36$ ,  $\rho_{\theta\eta} = -0.6$ ,  $\delta = 0.3$ . This yields probability limits for the model's  $R^2$  of 0.89 ( $\rho = 0.9$ ) and 0.74 ( $\rho = 0.6$ ). Parameters for simultaneity ('S'):  $\beta = 1$ ,  $\alpha = 2$ ,  $\sigma_\varepsilon^2 = 4$ ,  $\sigma_\theta^2 = 1$ . This yields probability limits for the model's  $R^2$  of 0.84 ( $\rho = 0.9$ ) and 0.81 ( $\rho = 0.6$ ).

error. The simulation results show that the Wald test's finite-sample power can turn out relatively low for smaller values of  $n$  and  $T$ . This is especially the case if the distance between  $\rho$  and  $\delta$  is relatively small.

We also run simulations for these three models in the absence of any measurement error, omitted variables or simultaneity, yielding the empirical size of the Wald test. The results in the right-hand side of Table 2 point out that the size of the Wald test is substantially above nominal for  $n = 100$  and  $T = 10$ . In the other cases, the rejection rates are fairly close to nominal. For the large- $n$  and fixed- $T$  asymptotics to apply, we must have  $n \gg T$ . For  $n = 100$  and  $T = 10$ , the ratio  $n/T$  might be too small. We therefore try to improve the finite-sample results by resorting to a bootstrap-based test statistic. Indeed the above nominal rejection rates for  $n = 100$  and  $T = 10$  can

**Table 3** Simulation results: empirical power

	$n = 100$ sup-Wald	$n = 500$	$n = 1000$	$n = 100$ Wald	$n = 500$	$n = 1000$
$T = 5$						
$\rho = 0.9$						
ME	0.358	0.990	1.000	0.602	1.000	1.000
OV	0.995	1.000	1.000	0.866	1.000	1.000
S	0.502	1.000	1.000	0.769	1.000	1.000
$\rho = 0.6$						
ME	0.110	0.453	0.752	0.216	0.704	0.960
OV	0.514	0.999	1.000	0.289	0.850	0.991
S	0.239	0.952	1.000	0.526	0.998	1.000
$T = 10$						
$\rho = 0.9$						
ME	0.964	1.000	1.000	0.998 (0.990)	1.000	1.000
OV	1.000	1.000	1.000	1.000 (1.000)	1.000	1.000
S	0.999	1.000	1.000	1.000 (1.000)	1.000	1.000
$\rho = 0.6$						
ME	0.336	0.996	1.000	0.610 (0.435)	0.999	1.000
OV	0.971	1.000	1.000	0.726 (0.562)	1.000	1.000
S	0.880	1.000	1.000	0.961 (0.910)	1.000	1.000

*Notes:* All simulation results are based on 10,000 simulation runs. The empirical power is obtained as the fraction of the number of simulation runs in which the test rejects the null hypothesis, while the empirical size is obtained as the fraction of the number of simulation runs in which the test rejects the null hypothesis, while the null hypothesis is true. For the Wald test, the rejection rates for the bootstrap-based version of the test are in parentheses (1,000 bootstrap runs). Throughout, the significance level is 5%. The simulated models correspond to three of the illustrative cases listed in Table 1 of the main text. Parameters for classical measurement error ('ME'):  $\beta = 1$ ,  $\sigma_\varepsilon^2 = 1$ ,  $\sigma_\theta^2 = 1.44$ ,  $\sigma_\eta^2 = 0.64$ ,  $\delta = 0.3$ . This yields reliabilities of 0.92 ( $\rho = 0.9$ ) and 0.76 ( $\rho = 0.6$ ), noise-to-signal ratios of 0.09 ( $\rho = 0.9$ ) and 0.312 ( $\rho = 0.6$ ) and probability limits for the model's  $R^2$  of 0.88 ( $\rho = 0.9$ ) and 0.69 ( $\rho = 0.6$ ). Parameters for omitted variables ('OV'):  $\beta = 1$ ,  $\gamma = 1$ ,  $\sigma_\varepsilon^2 = 0.25$ ,  $\sigma_\theta^2 = 0.36$ ,  $\sigma_\eta^2 = 0.36$ ,  $\rho_{\theta\eta} = -0.6$ ,  $\delta = 0.3$ . This yields probability limits for the model's  $R^2$  of 0.89 ( $\rho = 0.9$ ) and 0.74 ( $\rho = 0.6$ ). Parameters for simultaneity ('S'):  $\beta = 1$ ,  $\alpha = 2$ ,  $\sigma_\varepsilon^2 = 4$ ,  $\sigma_\theta^2 = 1$ . This yields probability limits for the model's  $R^2$  of 0.84 ( $\rho = 0.9$ ) and 0.81 ( $\rho = 0.6$ ).

be circumvented by using a panel wild bootstrap to estimate the covariance matrix. The bootstrap-based covariance matrix then replaces the formula-based covariance matrix in the Wald test statistic.<sup>7</sup> Table 2 reports the bootstrap-based rejection rates in parentheses, which are close to nominal. The use of the bootstrap only turned out necessary for  $n = 100$  and  $T = 10$ , but from a robustness perspective one may consider using the bootstrap regardless of the panel dimensions.

We consider the same simulation experiments for the sup-Wald strict exogeneity test of Su et al. (2016) and report the outcomes in the left-hand sides of Tables 2 and

<sup>7</sup> Another possibility is to use the panel wild bootstrap to obtain bootstrap-based critical values for the Wald test statistic. Because this approach turned out to yield poorer empirical power and size, we do not consider it in our analysis.

3. Also this test's empirical size is close to nominal. In terms of empirical power, the simulation results make clear that there are two situations: one in which the sup-Wald test's empirical power is relatively low in comparison to our Wald test and one in which both tests perform similarly. Especially for  $n = 100$ , our Wald test tends to have relatively high empirical power. For the ME model with  $\rho = 0.6$ , this pattern also persists for larger values of  $n$ . However, there is one exception and that is the model with an omitted variable. Also here we identify the above two situations, but with the roles of the two tests reversed: one in which the sup-Wald test's empirical power is relatively high in comparison to our Wald test and one in which both tests perform similarly. We explain the good performance of the sup-Wald test in the presence of an omitted variable from the fact that the sup-Wald test's auxiliary regressions are close to the true panel regression model, because the observed covariate's lags and leads are correlated with the omitted variable. Hence, the sup-Wald test performs very well in this specific case of misspecification. As we will see below, however, this result does not hold true for all parameter values.

### 4.3 Additional simulations

We consider five additional sets of simulations. First, we simulate the three motivating examples (measurement error, omitted variable and simultaneity) with  $n$  and  $T$  equal to the values that we will later encounter in our empirical applications. This gives rise to simulations (i) and (ii). In (i), we take  $n = 51$  and  $T = 20$ , leading to a relatively small value of  $n/T$  instead of  $n \gg T$ . In (ii), we set  $n = 737$  and  $T = 4$ . In (iii), we extend the basic ME model with  $n = 51$  and  $n = 100$  as to include an additional error-free regressor. In (iv) and (v), we reconsider the basic ME and omitted variable model, respectively, with a parameter setting that results in trivial power for our Wald test according to Table 1 ( $\delta = \rho = 0.6$ ). Table 4 reports the empirical power and size for both tests in each of the five cases. For (i) and (iii), the table reports bootstrap-based rejection rates for our Wald test.

In (i), we observe that the sup-Wald test has higher empirical power in the omitted variables case, while the two tests perform comparably in terms of empirical power in the presence of simultaneity. In case of ME, our Wald test's empirical power is slightly better. In (ii), the sup-Wald only outperforms in terms of empirical power in the omitted variable case. In the other two cases, our test outperforms. In (iii), the sup-Wald test outperforms in terms of empirical power. In (iv) and (v), we observe that both tests have trivial power. For our test, this outcome was expected on the basis of Table 1. Apparently, also the sup-Wald test needs a difference in persistence between the unobserved regressor and the measurement error in (iv) and between the observed and omitted regressor in (v) in order to have non-trivial power.<sup>8</sup>

<sup>8</sup> We note that the power comparison between the two tests may not always be entirely fair, due to small differences in the tests' empirical sizes. Because the simulations for the sup-Wald test are computationally very intensive, it was not feasible to size adjust the test as discussed by, e.g., Lloyd (2005).

**Table 4** Additional simulation results: empirical size and power

	Size sup-Wald	Wald	Power sup-Wald	Wald
<i>(i) Alternative panel dimensions</i>				
$n = 51, T = 20, \text{ME}$	0.072	0.052	0.501	0.555
$n = 51, T = 20, \text{OV}$	0.071	0.049	0.978	0.682
$n = 51, T = 20, \text{S}$	0.073	0.047	0.979	0.964
<i>(ii) Alternative panel dimensions</i>				
$n = 737, T = 4, \text{ME}$	0.047	0.053	0.346	0.636
$n = 737, T = 4, \text{OV}$	0.059	0.051	0.997	0.795
$n = 737, T = 4, \text{S}$	0.047	0.055	0.905	0.985
<i>(iii) Additional error-free covariate (ME model)</i>				
$n = 51, T = 20$	0.059	0.055	0.979	0.900
$n = 100, T = 10$	0.057	0.044	0.947	0.818
<i>(iv) Trivial power (ME model)</i>				
$n = 100, T = 5$	0.061	0.059	0.062	0.077
$n = 500, T = 5$	0.040	0.056	0.049	0.045
$n = 1000, T = 5$	0.053	0.050	0.043	0.045
<i>(v) Trivial power (OV model)</i>				
$n = 100, T = 5$	0.047	0.063	0.060	0.072
$n = 500, T = 5$	0.049	0.049	0.045	0.053
$n = 1000, T = 5$	0.055	0.049	0.056	0.049

*Notes:* All simulation results are based on 10,000 simulation runs. The empirical power is obtained as the fraction of the number of simulation runs in which the test rejects the null hypothesis, while the empirical size is obtained as the fraction of the number of simulation runs in which the test rejects the null hypothesis, while the null hypothesis is true. Throughout, the significance level is 5%. The simulated models are the same as for Tables 2 and 3, with  $\rho = 0.6$  and  $\delta = 0.3$  in (i), (ii) and (iii) and with  $\rho = \delta = 0.6$  in (iv) and (v). Furthermore, in (iii) there is an additional i.i.d. normally distributed error-free covariate included in the ME model, with coefficient 1, mean 0 and variance 4. Its correlation with  $\theta$  is 0.75. The additional covariate is uncorrelated with the remaining model variables

#### 4.4 Cautionary remark

Because it is infeasible to do analytical calculations for the sup-Wald test and the constituent individual Wald tests, we do not provide an example that illustrates that the sup-Wald test indeed test a different null hypothesis than our Wald test. For the same reason, our simulation study considers misspecification that renders the two tests both consistent or both inconsistent. Although these simulations prove insightful, we emphasize that in more general cases of misspecification the two tests will typically test different null hypotheses. The null hypothesis of our Wald test is a sufficient condition for the FE estimator's consistency, while the null hypothesis of the sup-Wald test is neither sufficient nor necessary for the consistency of the FE estimator. The finding our simulation study that our Wald test does not always have higher empirical power than the sup-Wald test must be viewed in this context.

## 5 Empirical applications

This section considers two existing panel data sets from the literature, which each contain an explanatory variable that was suspected to be subject to measurement error in the studies that introduced these data. In the context of our theoretical results, these data sets provide a particularly relevant empirical case for our test. Our goal is to investigate the consistency of the FE estimator using our Wald test. For the sake of comparison, we will also report results for the sup-Wald test.

### 5.1 Birth rates and welfare

Economic theory suggests that a government transfer program that reduces the cost of supporting a child should lead to a rise in birth rates. As pointed out by McKinnish (2008), childbearing is a commitment to current and future consumption. We may therefore expect fertility decisions to be relatively unresponsive to transitory fluctuations in welfare benefits. This would imply that welfare benefits are erroneous relative to the conceptual variable of interest, even though these benefits are generally reported without error in administrative records. As explained by Griliches and Hausman (1986), this kind of ‘conceptual’ measurement error is isomorphic to the errors-in-variables model with measurement error that is less persistent than the unobserved regressor and would render the FE estimator inconsistent due to endogeneity of the observed regressor.

McKinnish (2008) aims to provide an empirical investigation of the presence of such conceptual measurement error in welfare benefits. She uses a panel data set consisting of US state-level birth rates by white women in the age group 20–24.5 years and AFDC benefit levels for a family of four with no additional income. The panel data set with  $n = 51$  and  $T = 20$  covers the 1973–1992 period. The data set also contains a measure of the earnings per capita in each state. Both welfare benefits and earnings per capita are deflated and expressed in prices of the base year 1982–84.

We consider the linear panel regression model specified as

$$\log(y_{it}) = \alpha_i + \delta_t + \beta_w \log(w_{it}) + \beta_e \log(e_{it}) + \varepsilon_{it}, \quad (22)$$

where  $y_{it}$  denotes the birth rate in state  $i$  in year  $t$ ,  $\alpha_i$  a state fixed effect,  $\delta_t$  a year fixed effect,  $w_{it}$  the welfare benefit (i.e., the allegedly error-ridden regressor), and  $e_{it}$  the earnings per capita.

McKinnish (2008) estimates the linear panel regression model in (22) using data that is differenced over a time span of  $j = 1, 3, 5, 7$  years. We denote the resulting coefficient estimates of the welfare benefit by  $\hat{\beta}_{w,j}$ . McKinnish (2008) compares the  $\hat{\beta}_{w,j}$ s for different values of  $j$ . In this way, she proceeds in a similar fashion as Goolsbee (2000). McKinnish (2008) establishes a monotonically increasing pattern in the  $\hat{\beta}_{w,j}$ s, which she contributes to the presence of conceptual measurement error.

We estimate the linear panel regression model in (22) using data that is differenced over a time span of  $j = 1, \dots, 10$  years. The estimation results are summarized in the upper panel of Table 5. This table also reports the results based on the FE estimator.

**Table 5** Estimation results

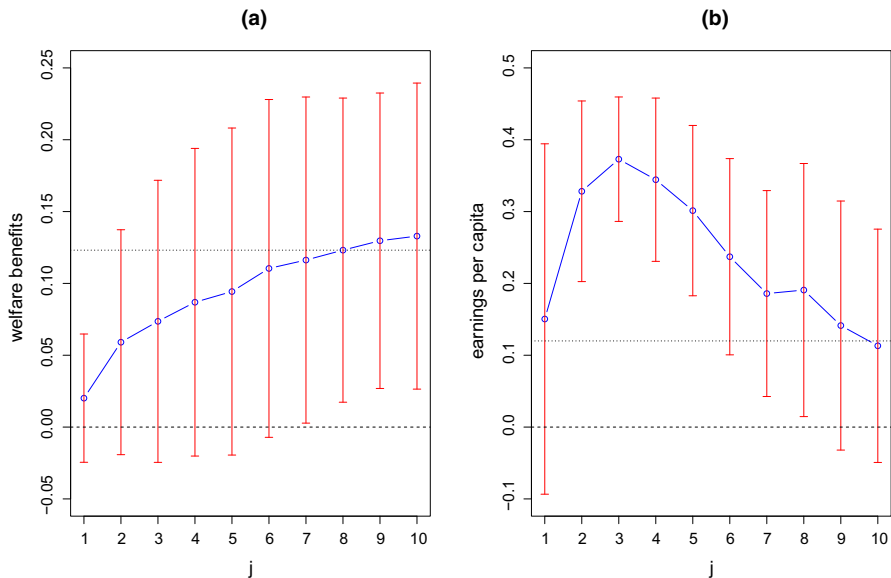
	FE	$D_1$	$D_2$	$D_3$	$D_4$	$D_5$	$D_6$	$D_7$	$D_8$	$D_9$	$D_{10}$
<b>McKinnish (2008)</b>											
$\log(w)$	0.123 (0.021)	0.020 (0.377)	0.059 (0.139)	0.074 (0.142)	0.087 (0.112)	0.094 (0.104)	0.110 (0.066)	0.116 (0.045)	0.123 (0.023)	0.130 (0.013)	0.133 (0.014)
$\log(e)$	0.120 (0.194)	0.150 (0.227)	0.328 (0.000)	0.373 (0.000)	0.344 (0.000)	0.301 (0.000)	0.237 (0.001)	0.186 (0.011)	0.191 (0.034)	0.141 (0.110)	0.113 (0.172)
adj. $R^2$	0.044	0.010	0.079	0.107	0.097	0.085	0.070	0.056	0.061	0.051	0.049
<b>Erickson and Whited (2000)</b>											
$q$	0.018 (0.000)	0.016 (0.000)	0.019 (0.000)	0.019 (0.000)							
$c$	0.113 (0.011)	0.095 (0.020)	0.125 (0.007)	0.120 (0.043)							
$cf$	-0.099 (0.030)	-0.084 (0.050)	-0.104 (0.029)	-0.113 (0.065)							
adj. $R^2$	0.087	0.049	0.104	0.120							

*Notes:* This table shows the estimation results for the model in (22) applied to the data of McKinnish (2008) (upper panel) and for the model in (23) applied to the data of Erickson and Whited (2000) (lower panel). The column captioned 'FE' reports the estimation results based on the FE estimator, while the columns captioned ' $D_j$ ' contain the estimates based on differences over time span  $j$ . The  $p$ -values for the estimated coefficients are in parentheses. The adjusted  $R^2$ 's correspond to the transformed models whenever a panel data transformation has been used. Abbreviations used:  $w$  = welfare benefit;  $e$  = earnings per capita;  $c$  = cash flow per unit of capital;  $f$  = 0–1 variable for financially constraint firms (based on the availability of a bond rating). The time fixed effects are not reported, but have been taken out by using the dependent variable and the regressors in deviations from their means per time period

**Table 6** Test outcomes

Data set	Test statistic	df	c.v.	<i>p</i> value
McKinnish (2008) (Wald)	139.6	18	28.87	0.00
McKinnish (2008) (Wald, bootstrap)	29.9	18	28.87	0.04
McKinnish (2008) (sup-Wald)	23.1		17.6	0.009
Erickson and Whited (2000) (Wald)	4.8	6	12.59	0.57
Erickson and Whited (2000) (Wald, bootstrap)	5.1	6	12.59	0.53
Erickson and Whited (2000) (sup-Wald)	19.0		10.8	0.002

*Notes:* This table shows the results of the Wald and sup-Wald tests for the model in (22) applied to the data of McKinnish (2008) (upper panel) and the model in (23) applied to the data of Erickson and Whited (2000) (lower panel). For each test, we report the critical value ('c.v.') and *p*-value. For the Wald test, we also report the degrees of freedom of the Chi-square distribution (d.f.), while the results for the bootstrap-based version of the test are provided in a separate row. The bootstrap-based Wald test statistic is based on 1,000 bootstrap runs and uses an estimator of the covariance matrix based on a panel wild bootstrap



**Fig. 1** Difference curves (McKinnish (2008) data). *Notes:* This figure shows the difference curves for the estimated coefficients of the natural log of (a) the welfare benefit and (b) the earnings per capita. They correspond to the model in (22), applied to the data of McKinnish (2008). The intervals in red show the 95% asymptotic confidence interval for each point estimate. The two dashed lines indicate the zero line and the value of the FE estimator

At the 5% significance level, our bootstrap-based Wald test rejects the null hypothesis  $H_0 : \beta_{e,j} = \beta_{e,j+1}; \beta_{w,j} = \beta_{w,j+1}$  for  $j = 1, \dots, 9$  at the 5% level; see Table 6. Although we do not have a case of large  $n$  here, we show the differences curves in

Fig. 1a and b for the welfare benefit and the earnings variables anyhow, for the sake of illustration. The differences curves confirm the economic relevance of the rejection.<sup>9</sup>

In sum, our test outcome substantiates the doubts of McKinnish (2008) about the consistency of the FE estimator of (22). As explained in Sect. 3.4.4, if the Wald test rejects it is still possible that the FE is consistent. Because of the additional evidence against the FE estimator's consistency provided by McKinnish (2008), we consider that possibility unlikely here. Although our test results are consistent with the presence of conceptual measurement error in the welfare benefit variable, the source of the inconsistency—measurement error or something else—remains an open question. For example, the data used by McKinnish (2008) are aggregated across different cohorts and states that may respond differently to changes in welfare over time, which may also render the FE estimator inconsistent.

We note that the sup-Wald test rejects  $\tilde{H}_0$  at the 5% level; see Table 6. As explained in Sect. 3.4.4, this test focuses on generic strict exogeneity conditions instead of being specifically related to the FE estimator's orthogonality conditions. We therefore view the rejection as a sign that other moment conditions related to strict exogeneity do not hold either. We refer to McKinnish (2008) for additional estimations that exploit less stringent moment conditions.

## 5.2 Investments and Tobin's $q$

Erickson and Whited (2000) analyze the impact of Tobin's  $q$  on the investment rate, with Tobin's  $q$  the ratio of the market valuation of a firm's capital stock to its replacement value. The theoretical motivation for studying this relation is the standard model of a perfectly competitive firm. This model is based on the maximization of net shareholder wealth, in the presence of convex adjustment costs following changes in the capital stock (e.g., Blundell et al. 1992). According to this model, Tobin's  $q$  has a positive effect on the investment rate. An empirical complication is the measurement error problem associated with Tobin's  $q$ . This problem arises due to the difference between marginal  $q$ , the conceptual variable of interest, and measured  $q$  as defined above. Erickson and Whited (2000) discuss the possible sources of measurement error in measured  $q$  and propose an estimator that controls for such error by exploring higher-order moments. Their empirical analysis is based on a Compustat firm-level panel data set for the 1992–1995 period, with  $n = 737$  and  $T = 4$ .

We consider the linear panel regression model given by

$$y_{it} = \alpha_i + \delta_t + \beta_q q_{it} + \beta_{c_{it}} + \beta_{f_{it}} f_i + \varepsilon_{it}, \quad (23)$$

where  $y_{it}$  denotes the ratio of investments to the replacement value of the capital stock for firm  $i$  in year  $t$ ,  $\alpha_i$  a firm fixed effect,  $\delta_t$  a year fixed effect,  $q_{it}$  the proxy of marginal Tobin's  $q$ ,  $c_{it}$  cash flow divided by the replacement value of the capital stock, and  $f_i$  a 0–1 variable indicating whether a firm is financially constrained or not. The indicator

<sup>9</sup> The difference curve for the earnings variable is highly non-monotonic. Unreported examples in the setting of Table 1 confirm that non-monotonic patterns are indeed possible for certain model configurations, such as non-classical ME model with a nonzero correlation between the measurement and regression errors.

variable  $f_i$  is constructed on the basis of a firm's lack of bond rating and does not vary over time; its own marginal effect is therefore contained in the fixed effect  $\alpha_i$ .

In the presence of measurement error in  $q$ , the FE estimator of (23) will typically be inconsistent due to a lack of strict exogeneity of the proxy of marginal  $q$ . We estimate the linear panel regression model in (23) after differencing over a time span of  $j = 1, 2, 3$  years.<sup>10</sup> Detailed estimation results are given in the lower panel of Table 5. This table also reports the estimation results based on the FE estimator. At the 5% significance level, our Wald test fails to reject the null hypothesis  $H_0 : \beta_{q,j} = \beta_{q,j+1}, \beta_{c,j} = \beta_{c,j+1}; \beta_{fc,j} = \beta_{fc,j+1}$  for  $j = 1, 2$ ; see again Table 6.

We note that the sup-Wald test rejects  $\tilde{H}_0$  at the 5% level, as shown in Table 6. As noted in Sect. 3.4.4, it remains unclear what the rejection of  $\tilde{H}_0$  means for the inconsistency of the FE estimator.

We conclude that our Wald test finds no evidence against the consistency of the FE estimator. As mentioned in Sect. 3.3, we should remain aware of the possibility that the test may have low power in certain cases. Low power could also arise from limited data variability due to taking differences, yielding coefficient estimates with relatively large standard errors. In such a scenario, our test could fail to reject in the presence of misspecification. This explanation does not seem very likely in the present case, though. The strong significance of the estimated coefficients in the lower panel of Table 5 suggests that the time-differenced data still contain a sufficient amount of variation. Another possibility is that the inconsistencies in the differences estimators do not depend on  $j$ .

Given these considerations and the rejection by the sup-Wald test, it remains important to look for other evidence against the FE estimator's consistency, such as coefficient signs that are unlikely from an economic perspective. Here, we find the coefficient signs that we would expect on the basis of economic theory: Tobin's  $q$  and the cash flow variable both have a positive effect on the expected investment rate, which is smaller if firms are financially constraint. In sum, also these additional investigations do not find evidence against the consistency of the FE estimator.

## 6 Conclusion

The FE estimator is widely used to estimate the linear panel regression model. Under large- $n$  and fixed- $T$  panel data asymptotics, we have developed a test to validate a sufficient condition for the FE estimator's consistency using a stacked regression framework. Our test takes the familiar form of a panel-robust Wald test. Because our test is asymptotically equivalent to a specific GMM test for overidentifying restrictions, our approach also fits in the familiar setting of GMM estimation and specification testing.

We have shown that our Wald test will generally test a different null hypothesis than the strict exogeneity test of Wooldridge (2010) and the extension proposed by Su et al. (2016). Our Wald test is specifically tailored for testing a sufficient condition for the FE estimator's consistency. The other two tests, by contrast, consider more

<sup>10</sup> This data set of Erickson and Whited (2000) is available at <http://toni.marginalq.com/publications.html>.

generic strict exogeneity conditions that are neither sufficient nor necessary for the FE estimator's consistency.

The Wald test's finite-sample properties have been investigated in a simulation study, where we continued the comparison with the strict exogeneity test of Su et al. (2016). Our Wald test has been shown to possess good finite-sample properties, especially if the estimator of the covariance matrix is based on a panel bootstrap. We have also illustrated the test in two applications to existing studies from the literature.

If our tests reject, it is still possible that the FE estimator is consistent. Further testing, as discussed below, would be needed to exclude this possibility. If our test does not reject, there is no evidence against the FE estimator's consistency. Although this is the most favorable outcome, researchers should still be aware of the possibility that the test may have low power in certain cases. It therefore remains important to look for other evidence against the FE estimator, such as coefficient signs and magnitudes that are unlikely from an economic perspective. Researchers should also recognize the possibility that low power could arise from limited data variability due to taking differences, yielding coefficient estimates with relatively large standard errors. Hence, although our test does not require IVs, it should be used in combination with additional analysis.

As usual, finding a well-specified model remains to a large extent a case-by-case puzzle without guaranteed success, depending on, e.g., prior information and the availability of valid and strong instruments. As a general guideline, it is nevertheless useful to consider the existing literature on the selection of moment conditions. Here a distinction is made between (i) separating valid moment conditions from invalid ones and (ii) the elimination of redundant conditions; i.e., conditions that do not contribute to a reduction in the GMM estimator's variance (Okui 2009). Various consistent selection procedures have been proposed, including methods that add a penalty term to the usual  $J$ -statistic for overidentification (Andrews 1999). Ideally, a fully integrated selection procedure for moment conditions should include our Wald test's moment conditions, as well as the more generic exogeneity conditions underlying the sup-Wald test of Su et al. (2016). We leave the development of such an integrated approach as a topic for future research.

A final topic for further research relates to the panel dimensions. The asymptotic distribution of our test statistic has been derived under large- $n$  and fixed- $T$  asymptotics, making the test suitable when  $n \gg T$ . This was the format in the classical panel data literature, but there has been increasing attention to panel data where  $n$  and  $T$  are of a different relative size, requiring different asymptotics. A first step would be to investigate the asymptotic behavior of our Wald test statistic for  $n$  fixed and  $T \rightarrow \infty$ , or for  $n \rightarrow \infty$  and  $T \rightarrow \infty$  jointly in some way.

**Acknowledgements** Laura Spierdijk would like to thank Terra McKinnish for sharing her data. She is also grateful to Tom Wansbeek and Erik Meijer for providing useful comments and suggestions.

**Funding** Laura Spierdijk gratefully acknowledges financial support by a Vidi grant (452.11.007) in the 'Vernieuwingsimpuls' program of the Netherlands Organization for Scientific Research (NWO). Her work was also supported by the Netherlands Institute for Advanced Study in the Humanities and Social Sciences (NIAS-KNAW). The usual disclaimer applies.

**Data availability** Available from the author upon request

## Declarations

**Conflicts of interest** The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Appendix

### A. FE estimator is weighted average of difference estimators

The full derivation of (3) writes as

$$\begin{aligned}\hat{\beta}_{FE} &= \left( \sum_{i=1}^n \mathbf{X}'_i \Delta_T \mathbf{X}_i \right)^{-1} \sum_{i=1}^n \mathbf{X}'_i \Delta_T \mathbf{y}_i = \left( \sum_{i=1}^n \sum_{t=1}^{T-1} \mathbf{X}'_i \Delta_t \mathbf{X}_i \right)^{-1} \sum_{i=1}^n \sum_{j=1}^{T-1} \mathbf{X}'_i \Delta_j \mathbf{y}_i \\ &= \sum_{j=1}^{T-1} \left( \sum_{i=1}^n \sum_{t=1}^{T-1} \mathbf{X}'_i \Delta_t \mathbf{X}_i \right)^{-1} \sum_{i=1}^n \mathbf{X}'_i \Delta_j \mathbf{y}_i = \sum_{j=1}^{T-1} \mathbf{W}_j \left( \sum_{i=1}^n \mathbf{X}'_i \Delta_j \mathbf{X}_i \right)^{-1} \sum_{i=1}^n \mathbf{X}'_i \Delta_j \mathbf{y}_i \\ &= \sum_{j=1}^{T-1} \mathbf{W}_j \hat{\beta}_j,\end{aligned}\quad (\text{A.1})$$

where

$$\mathbf{W}_j = \left( \sum_{i=1}^n \sum_{t=1}^{T-1} \mathbf{X}'_i \Delta_t \mathbf{X}_i \right)^{-1} \sum_{i=1}^n \mathbf{X}'_i \Delta_j \mathbf{X}_i \quad (\text{A.2})$$

$$\sum_{j=1}^{T-1} \mathbf{W}_j = \mathbf{I}_k, \quad \mathbf{W}_j \text{ symmetric and positive definite.} \quad (\text{A.3})$$

### B. Motivating examples: calculations

This appendix makes use of a few elementary properties of stationary AR(1) processes, which we summarize here for completeness. Assume that  $x_{it}$  and  $z_{it}$  are generated by

stationary AR(1) processes, such that

$$x_{it} = \rho x_{i,t-1} + \theta_{it} \quad [0 < \rho < 1] \quad (\text{B.1})$$

$$z_{it} = \delta z_{i,t-1} + \eta_{it} \quad [0 < \delta < 1]. \quad (\text{B.2})$$

We assume that  $\mathbb{E}(\theta_{it}) = \mathbb{E}(\eta_{it}) = 0$ ,  $\mathbb{E}(\theta_{it}^2) = \sigma_\theta^2$  and  $\mathbb{E}(\eta_{it}^2) = \sigma_\eta^2$  for all  $i$  and  $t$ . We also assume that  $\mathbb{Cov}(\theta_{mt}, \eta_{is}) = 0$  for  $m \neq n$ ,  $\mathbb{Cov}(\theta_{is}, \eta_{it}) = 0$  for  $s \neq t$ , and  $\mathbb{Cov}(\theta_{it}, \eta_{it}) = \sigma_{\theta\eta}$ . Lastly, we assume that  $\mathbb{Cov}(\theta_{mt}, \varepsilon_{is}) = \mathbb{Cov}(\eta_{mt}, \varepsilon_{is}) = 0$  for all  $m, i, s, t$ .

For  $k \geq 1$ , we can write

$$x_{it} = \rho^k x_{i,t-k} + \sum_{\ell=0}^{k-1} \rho^\ell \theta_{i,t-\ell}, \quad z_{it} = \delta^k z_{i,t-k} + \sum_{\ell=0}^{k-1} \delta^\ell \eta_{i,t-\ell}. \quad (\text{B.3})$$

By letting  $k \rightarrow \infty$ , we find

$$x_{it} = \sum_{\ell=0}^{\infty} \rho^\ell \theta_{i,t-\ell}, \quad z_{it} = \sum_{\ell=0}^{\infty} \delta^\ell \eta_{i,t-\ell}. \quad (\text{B.4})$$

Using these alternative formulations for  $x_{it}$  and  $z_{it}$ , we find for  $j \geq 0$ ,

$$\text{Var}(x_{it}) = \sigma_\theta^2 / (1 - \rho^2) \equiv \sigma_x^2, \quad \text{Var}(z_{it}) = \sigma_\eta^2 / (1 - \delta^2) \equiv \sigma_z^2, \quad (\text{B.5})$$

$$\mathbb{Cov}(x_{it}, x_{i,t-j}) = \rho^j \sigma_x^2, \quad \mathbb{Cov}(z_{it}, z_{i,t-j}) = \delta^j \sigma_z^2. \quad (\text{B.6})$$

We also have

$$\begin{aligned} \mathbb{Cov}(x_{it}, z_{i,t-j}) &= \delta^{-j} \sum_{\ell=j}^{\infty} (\delta\rho)^\ell \mathbb{Cov}(\theta_{i,t-\ell}, \eta_{i,t-\ell}) = \delta^{-j} \sigma_{\theta\eta} \\ &\times \sum_{\ell=j}^{\infty} (\delta\rho)^\ell = \rho^j \sigma_{\theta\eta} / (1 - \delta\rho). \end{aligned} \quad (\text{B.7})$$

Similarly, we find

$$\mathbb{Cov}(x_{i,t-j}, z_{it}) = \delta^j \sigma_{\theta\eta} / (1 - \delta\rho). \quad (\text{B.8})$$

### B.1. (Non-)Classical measurement error

We start with the errors-in-variables model and allow for non-classical measurement error, with classical measurement as a special case. We will derive the inconsistency in both cases.

**Model** Consider the linear panel regression model with measurement error, given by

$$y_{it} = \alpha_i + \beta \xi_{it} + \varepsilon_{it} \quad (\text{B.9})$$

$$x_{it} = \xi_{it} + v_{it}, \quad (\text{B.10})$$

where  $n = 1, \dots, n$  and  $t = 1, \dots, T$ . We assume that  $(\varepsilon_{it})$  is i.i.d. with  $\mathbb{E}(\varepsilon_{it}) = 0$  and  $\mathbb{E}(\varepsilon_{it}^2) = \sigma_\varepsilon^2$  for all  $i$  and  $t$ . Regarding  $(\xi_{it})$  and  $(v_{it})$ , we assume that they are generated by stationary AR(1) processes, such that

$$\xi_{it} = \rho \xi_{i,t-1} + \theta_{it} \quad [0 < \rho < 1] \quad (\text{B.11})$$

$$v_{it} = \delta v_{i,t-1} + \eta_{it} \quad [0 < \delta < 1]. \quad (\text{B.12})$$

We assume that  $\mathbb{E}(\theta_{it}) = \mathbb{E}(\eta_{it}) = 0$ ,  $\mathbb{E}(\theta_{it}^2) = \sigma_\theta^2$  and  $\mathbb{E}(\eta_{it}^2) = \sigma_\eta^2$  for all  $i$  and  $t$ . Furthermore, we assume that  $\text{Cov}(\theta_{mt}, \eta_{is}) = 0$  for  $m \neq n$ ,  $\text{Cov}(\theta_{is}, \eta_{it}) = 0$  for  $s \neq t$ ,  $\text{Cov}(\theta_{it}, \eta_{it}) = \sigma_{\theta\eta}$  and  $\text{Cov}(\theta_{mt}, \varepsilon_{is}) = 0$  for all  $m, i, s, t$ . Lastly, we assume that  $\text{Cov}(\varepsilon_{mt}, \eta_{is}) = 0$  for  $m \neq i$ ,  $\text{Cov}(\varepsilon_{is}, \eta_{it}) = 0$  for all  $s, t$ . If  $\sigma_{\theta\eta} \neq 0$ , we have a form of non-classical measurement error.

**Inconsistency** We first show that the FE estimator will usually be inconsistent. Let  $\mathbf{X}_i = (x_{i1}, \dots, x_{iT})'$  and  $\mathbf{y}_i = (y_{i1}, \dots, y_{iT})'$ . With  $\mathbf{A}$  the  $T \times T$  centering matrix we obtain

$$\text{plim}_{n \rightarrow \infty} \hat{\beta}_{\text{FE}} = \text{plim}_{n \rightarrow \infty} \frac{\sum_{i=1}^n \mathbf{x}_i' \mathbf{A} \mathbf{y}_i}{\sum_{i=1}^n \mathbf{x}_i' \mathbf{A} \mathbf{X}_i} = \frac{\text{tr}[\mathbf{A}(\boldsymbol{\Sigma}_\xi + \boldsymbol{\Sigma}_{\xi v})]}{\text{tr}[\mathbf{A}(\boldsymbol{\Sigma}_\xi + \boldsymbol{\Sigma}_v)]} \beta, \quad (\text{B.13})$$

where  $\boldsymbol{\Sigma}_v$  contains the covariances  $\text{Cov}(v_{ns}, v_{nt})$  and  $\boldsymbol{\Sigma}_{\xi v}$  the covariances  $\text{Cov}(\xi_{ns}, v_{nt})$ . This yields the inconsistency

$$\text{plim}_{n \rightarrow \infty} \hat{\beta}_{\text{FE}} - \beta = \frac{\text{tr}[\mathbf{A}(\boldsymbol{\Sigma}_{\xi v} - \boldsymbol{\Sigma}_v)]}{\text{tr}[\mathbf{A}(\boldsymbol{\Sigma}_\xi + \boldsymbol{\Sigma}_v)]} \beta, \quad (\text{B.14})$$

The inconsistency will typically be nonzero if at least  $\boldsymbol{\Sigma}_v \neq \mathbf{0}$ .

We now turn to the estimators  $\hat{\beta}_j$  that are obtained after taking differences over time span  $j$ . It holds that

$$\begin{aligned} \text{plim}_{n \rightarrow \infty} \hat{\beta}_j &= \frac{\text{Cov}(y_{it} - y_{i,t-j}, x_{it} - x_{i,t-j})}{\text{Var}(x_{it} - x_{i,t-j})} \\ &= \frac{\text{Cov}(\beta(x_{it} - x_{i,t-j}) - \beta(v_{it} - v_{i,t-j}) + \varepsilon_{it} - \varepsilon_{i,t-j}, x_{it} - x_{i,t-j})}{\text{Var}(x_{it} - x_{i,t-j})} \\ &= \beta + \frac{\text{Cov}(\varepsilon_{it} - \varepsilon_{i,t-j}, x_{it} - x_{i,t-j}) - \beta \text{Cov}(v_{it} - v_{i,t-j}, x_{it} - x_{i,t-j})}{\text{Var}(x_{it} - x_{i,t-j})}. \end{aligned} \quad (\text{B.15})$$

Under the given assumptions, the numerator in (B.15) reduces to

$$\begin{aligned} &\text{Cov}(\varepsilon_{it} - \varepsilon_{i,t-j}, v_{it} - v_{i,t-j}) - \beta [\text{Var}(v_{it} - v_{i,t-j}) \\ &+ \text{Cov}(v_{it} - v_{i,t-j}, \xi_{it} - \xi_{i,t-j})] = \\ &-2\beta[\sigma_\eta^2(1 - \delta^j)/(1 - \delta^2) + \sigma_{\theta\eta}(1 - (\delta^j + \rho^j)/2)/(1 - \delta\rho)]. \end{aligned} \quad (\text{B.16})$$

Furthermore, the denominator can be written as

$$\begin{aligned} & \text{Var}(\xi_{it} - \xi_{i,t-j} + v_{it} - v_{i,t-j}) = \\ & \text{Var}(\xi_{it} - \xi_{i,t-j}) + \text{Var}(v_{it} - v_{i,t-j}) + 2\text{Cov}(\xi_{it} - \xi_{i,t-j}, v_{it} - v_{i,t-j}) = \\ & 2[\sigma_\theta^2(1 - \rho^j)/(1 - \rho^2) + \sigma_\eta^2(1 - \delta^j)/(1 - \delta^2) \\ & + 2\sigma_{\theta\eta}(1 - (\delta^j + \rho^j)/2)/(1 - \delta\rho)]. \end{aligned} \quad (\text{B.17})$$

The inconsistency thus boils down to

$$\begin{aligned} \text{plim}_{n \rightarrow \infty} \hat{\beta}_j - \beta &= \frac{-\beta[\sigma_\eta^2(1 - \delta^j)/(1 - \delta^2) + \sigma_{\theta\eta}(1 - (\delta^j + \rho^j)/2)/(1 - \delta\rho)]}{\sigma_\theta^2(1 - \rho^j)/(1 - \rho^2) + \sigma_\eta^2(1 - \delta^j)/(1 - \delta^2) + 2\sigma_{\theta\eta}(1 - (\delta^j + \rho^j)/2)/(1 - \delta\rho)} \\ &= \frac{-\beta[\sigma_v^2(1 - \delta^j) + \sigma_{\xi v}(1 - (\delta^j + \rho^j)/2)]}{\sigma_\xi^2(1 - \rho^j) + \sigma_v^2(1 - \delta^j) + 2\sigma_{\xi v}(1 - (\delta^j + \rho^j)/2)}. \end{aligned} \quad (\text{B.18})$$

Because

$$(1 - \delta^j)(1 - \rho^{j+1}) > (1 - \delta^{j+1})(1 - \rho^j) \quad (\text{B.19})$$

if and only if  $\delta < \rho$ , it is readily seen that the inconsistency's magnitude decreases with  $j$  if and only if  $\delta < \rho$ . For  $\delta > \rho$ , the magnitude of the inconsistency is increasing and for  $\delta = \rho$  the inconsistency does not depend on  $j$ . For both classical and non-classical measurement error, the inconsistency does not vanish for larger values of  $j$ .

## B.2. Omitted variables

The second source of endogeneity that we consider is an omitted variable.

*Model* Consider the linear panel regression model with two regressors, given by

$$y_{it} = \alpha_i + \beta x_{it} + \gamma z_{it} + \varepsilon_{it}, \quad (\text{B.20})$$

where  $i = 1, \dots, n$  and  $t = 1, \dots, T$ . We assume that  $(\varepsilon_{it})$  is i.i.d. with  $\mathbb{E}(\varepsilon_{it}) = 0$  and  $\mathbb{E}(\varepsilon_{it}^2) = \sigma_\varepsilon^2$  for all  $i$  and  $t$ . Regarding the explanatory variables, we assume that  $x_{it}$  and  $z_{it}$  are generated by stationary AR(1) processes, such that

$$x_{it} = \rho x_{i,t-1} + \theta_{it} \quad [0 < \rho < 1] \quad (\text{B.21})$$

$$z_{it} = \delta z_{i,t-1} + \eta_{it} \quad [0 < \delta < 1]. \quad (\text{B.22})$$

We assume that  $\mathbb{E}(\theta_{it}) = \mathbb{E}(\eta_{it}) = 0$ ,  $\mathbb{E}(\theta_{it}^2) = \sigma_\theta^2$  and  $\mathbb{E}(\eta_{it}^2) = \sigma_\eta^2$  for all  $i$  and  $t$ . Furthermore, we assume that  $\text{Cov}(\theta_{mt}, \eta_{is}) = 0$  for  $m \neq n$ ,  $\text{Cov}(\theta_{is}, \eta_{it}) = 0$  for  $s \neq t$ , and  $\text{Cov}(\theta_{it}, \eta_{it}) = \sigma_{\theta\eta}$ . Lastly, we assume that  $\text{Cov}(\theta_{mt}, \varepsilon_{is}) = \text{Cov}(\eta_{mt}, \varepsilon_{is}) = 0$  for all  $m, n, s, t$ .

We estimate the omitted variable regression

$$y_{it} = \alpha_i + \beta x_{it} + \varepsilon_{it}, \quad (\text{B.23})$$

and are interested in the probability limit of  $\widehat{\beta}_j$ , the estimator of  $\beta$  based on the model after taking differences over time span  $j$ .

*Inconsistency* We first show that the FE estimator for  $\beta$  will usually be inconsistent. Using similar matrix notation as before, we obtain

$$\text{plim}_{n \rightarrow \infty} \widehat{\beta}_{\text{FE}} = \text{plim}_{n \rightarrow \infty} \frac{\sum_{i=1}^n \mathbf{x}'_i \mathbf{A} \mathbf{y}_i}{\sum_{i=1}^n \mathbf{x}'_i \mathbf{A} \mathbf{X}_i} = \beta + \frac{\text{tr}(\mathbf{A} \boldsymbol{\Sigma}_{zx})}{\text{tr}(\mathbf{A} \boldsymbol{\Sigma}_x)} \gamma, \quad (\text{B.24})$$

yielding the inconsistency

$$\text{plim}_{n \rightarrow \infty} \widehat{\beta}_{\text{FE}} - \beta = \frac{\text{tr}(\mathbf{A} \boldsymbol{\Sigma}_{zx})}{\text{tr}(\mathbf{A} \boldsymbol{\Sigma}_x)} \gamma. \quad (\text{B.25})$$

The inconsistency will be nonzero for  $\gamma \neq 0$  and  $\boldsymbol{\Sigma}_{zx} \neq \mathbf{0}$ .

We now turn to the estimators  $\widehat{\beta}_j$  that are obtained after taking differences over time span  $j$ . It holds that

$$\text{plim}_{n \rightarrow \infty} \widehat{\beta}_j = \frac{\text{Cov}(y_{it} - y_{i,t-j}, x_{it} - x_{i,t-j})}{\text{Var}(x_{it} - x_{i,t-j})}. \quad (\text{B.26})$$

Under the given assumptions, the numerator reduces to

$$\begin{aligned} \text{Cov}(y_{it} - y_{i,t-j}, x_{it} - x_{i,t-j}) &= \beta \text{Var}(x_{it} - x_{i,t-j}) \\ &+ \gamma \text{Cov}(x_{it} - x_{i,t-j}, z_{it} - z_{i,t-j}) = \\ &\beta \text{Var}(x_{it} - x_{i,t-j}) + \gamma [2 \text{Cov}(x_{it}, z_{it}) - \text{Cov}(x_{i,t-j}, z_{it}) - \text{Cov}(x_{it}, z_{i,t-j})] = \\ &2[\beta \sigma_\theta^2 (1 - \rho^j) / (1 - \rho^2) + \gamma \sigma_{\theta\eta} (1 - (\delta^j + \rho^j) / 2) / (1 - \delta\rho)]. \end{aligned} \quad (\text{B.27})$$

For the denominator, we find

$$\begin{aligned} \text{Var}(x_{it} - x_{i,t-j}) &= 2[\text{Var}(x_{it}) - \text{Cov}(x_{it}, x_{i,t-j})] \\ &= 2\sigma_x^2 (1 - \rho^j) = 2\sigma_\theta^2 (1 - \rho^j) / (1 - \rho^2). \end{aligned} \quad (\text{B.28})$$

The probability limit then becomes

$$\begin{aligned} \text{plim}_{n \rightarrow \infty} \widehat{\beta}_j &= \frac{\beta \sigma_\theta^2 (1 - \rho^j) / (1 - \rho^2) + \gamma \sigma_{\theta\eta} (1 - (\delta^j + \rho^j) / 2) / (1 - \delta\rho)}{\sigma_\theta^2 (1 - \rho^j) / (1 - \rho^2)} \\ &= \beta + \frac{\gamma \sigma_{\theta\eta} [1 - (\delta^j + \rho^j) / 2] / (1 - \delta\rho)}{\sigma_\theta^2 (1 - \rho^j) / (1 - \rho^2)}. \end{aligned} \quad (\text{B.29})$$

The inconsistency thus boils down to

$$\text{plim}_{n \rightarrow \infty} \hat{\beta}_j - \beta = \frac{\gamma \sigma_{\theta\eta} [1 - (\delta^j + \rho^j)/2] / (1 - \delta\rho)}{\sigma_\theta^2 (1 - \rho^j) / (1 - \rho^2)} = \frac{\gamma \sigma_{xz} [1 - (\delta^j + \rho^j)/2]}{\sigma_x^2 (1 - \rho^j)}. \quad (\text{B.30})$$

As a sanity check on the above expression, we notice that the inconsistency is zero for  $\sigma_{\theta\eta} = 0$ . The inconsistency should be zero in this particular case, because  $\sigma_{\theta\eta} = 0$  implies that  $x_{it}$  and  $z_{it}$  are uncorrelated.

Because

$$(1 - \rho^{j+1})(1 - (\delta^j + \rho^j)/2) > (1 - \rho^j)(1 - (\delta^{j+1} + \rho^{j+1})/2) \quad (\text{B.31})$$

if and only if  $\delta < \rho$ , is readily seen that  $\text{plim}_{n \rightarrow \infty} |\hat{\beta}_j - \beta| > \text{plim}_{n \rightarrow \infty} |\hat{\beta}_{j+1} - \beta|$  if and only if  $\delta < \rho$ . The inconsistency's magnitude is increasing for  $\delta > \rho$  and for  $\delta = \rho$  the inconsistency does not depend on  $j$ . We note that the inconsistency does not vanish for larger values of  $j$ .

### B.3. Simultaneity

The third source of endogeneity that we consider is simultaneity.

*Model* We consider the simultaneous equations model given by the structural equations

$$y_{it} = \beta_i + \beta x_{it} + \varepsilon_{it} \quad (\text{B.32})$$

$$x_{it} = \alpha_i + \alpha y_{it} + u_{it}. \quad (\text{B.33})$$

We assume that  $(\varepsilon_{it})$  is i.i.d. with  $\mathbb{E}(\varepsilon_{it}) = 0$  and  $\text{Var}(\varepsilon_{it}) = \sigma_\varepsilon^2$ , independent of  $(u_{it})$ . Here  $(u_{it})$  is a stationary AR(1) process defined by

$$u_{it} = \rho u_{i,t-1} + \theta_{it} \quad [0 < \rho < 1], \quad (\text{B.34})$$

with  $\mathbb{E}(\theta_{it}) = 0$ ,  $\mathbb{E}(\theta_{it}^2) = \sigma_\theta^2$  and  $\text{Cov}(\theta_{mt}, \varepsilon_{is}) = \mathbf{0}$  for all  $m, i, t, s$ .

Solving the two equations yields the reduced forms

$$y_{it} = \frac{\beta_i + \beta \alpha_i}{1 - \alpha\beta} + \frac{\beta u_{it} + \varepsilon_{it}}{1 - \alpha\beta} \quad (\text{B.35})$$

$$x_{it} = \frac{\alpha_i + \alpha \beta_i}{1 - \alpha\beta} + \frac{u_{it} + \alpha \varepsilon_{it}}{1 - \alpha\beta}. \quad (\text{B.36})$$

We estimate (B.32) in  $j$ th differences, thereby ignoring (B.33). We are interested in the probability limit of  $\hat{\beta}_j$ , the estimator of  $\beta$  based on the model in  $j$ th differences. We want to know how the inconsistency depends on  $j$ .

**Inconsistency** We first show that the FE estimator for  $\beta$  will usually be inconsistent. Using similar matrix notation as before, we obtain

$$\text{plim}_{n \rightarrow \infty} \hat{\beta}_{\text{FE}} = \text{plim}_{n \rightarrow \infty} \frac{\sum_{i=1}^n \mathbf{x}_i' \mathbf{A} \mathbf{y}_i}{\sum_{i=1}^n \mathbf{x}_i' \mathbf{A} \mathbf{X}_i} = \frac{\text{tr}(\mathbf{A} \boldsymbol{\Sigma}_u) \beta + \text{tr}(\mathbf{A} \boldsymbol{\Sigma}_\varepsilon) \alpha}{\text{tr}(\mathbf{A} \boldsymbol{\Sigma}_u) + \text{tr}(\mathbf{A} \boldsymbol{\Sigma}_\varepsilon) \alpha^2}, \quad (\text{B.37})$$

yielding the inconsistency

$$\text{plim}_{n \rightarrow \infty} \hat{\beta}_{\text{FE}} - \beta = \frac{\text{tr}(\mathbf{A} \boldsymbol{\Sigma}_\varepsilon) \alpha (1 - \alpha \beta)}{\text{tr}(\mathbf{A} \boldsymbol{\Sigma}_u) + \text{tr}(\mathbf{A} \boldsymbol{\Sigma}_\varepsilon) \alpha^2}. \quad (\text{B.38})$$

The inconsistency will be nonzero if  $\alpha \neq 0$  and  $\alpha \neq 1/\beta$ .

We now turn to the estimators  $\hat{\beta}_j$  that are obtained after taking differences over time span  $j$ . The probability limit of the resulting estimator for  $\beta$  equals

$$\begin{aligned} \text{plim}_{n \rightarrow \infty} \hat{\beta}_j &= \frac{\text{Cov}(y_{it} - y_{i,t-j}, x_{it} - x_{i,t-j})}{\text{Var}(x_{it} - x_{i,t-j})} \\ &= \frac{[\alpha \text{Var}(\varepsilon_{it} - \varepsilon_{i,t-j}) + \beta \text{Var}(u_{it} - u_{i,t-j})] / (1 - \alpha \beta)^2}{\text{Var}(x_{it} - x_{i,t-j})} \\ &= \frac{\alpha \text{Var}(\varepsilon_{it} - \varepsilon_{i,t-j}) + \beta \text{Var}(u_{it} - u_{i,t-j})}{\alpha^2 \text{Var}(\varepsilon_{it} - \varepsilon_{i,t-j}) + \text{Var}(u_{it} - u_{i,t-j})}. \end{aligned} \quad (\text{B.39})$$

Under the given assumption, this reduces to

$$\begin{aligned} \text{plim}_{n \rightarrow \infty} \hat{\beta}_j &= \frac{2[\alpha \sigma_\varepsilon^2 + \beta \sigma_\theta^2 (1 - \rho^j) / (1 - \rho^2)]}{2[\alpha^2 \sigma_\varepsilon^2 + \sigma_\theta^2 (1 - \rho^j) / (1 - \rho^2)]} = \frac{\alpha \sigma_\varepsilon^2 + \beta \sigma_\theta^2 (1 - \rho^j) / (1 - \rho^2)}{\alpha^2 \sigma_\varepsilon^2 + \sigma_\theta^2 (1 - \rho^j) / (1 - \rho^2)} \\ &= \frac{\alpha \sigma_\varepsilon^2 + \beta \sigma_u^2 (1 - \rho^j)}{\alpha^2 \sigma_\varepsilon^2 + \sigma_u^2 (1 - \rho^j)}. \end{aligned}$$

This gives the inconsistency

$$\text{plim}_{n \rightarrow \infty} \hat{\beta}_j - \beta = \frac{\alpha \sigma_\varepsilon^2 (1 - \alpha \beta)}{\alpha^2 \sigma_\varepsilon^2 + \sigma_u^2 (1 - \rho^j)}. \quad (\text{B.40})$$

The inconsistency is positive for  $\alpha(1 - \alpha \beta) > 0$  and negative for  $\alpha(1 - \alpha \beta) < 0$ . Its magnitude decreases with  $j$  for  $0 < \rho < 1$ . We note that the inconsistency does not vanish for larger values of  $j$ .

## References

- Amini S, Delgado M, Henderson D, Parmeter C (2012) Fixed vs random: the Hausman test four decades later. In: Baltagi B, Carter Hill R, Newey W, White H (eds) *Essays in Honor of Jerry Hausman* (Advances in Econometrics), vol 29. Emerald Group Publishing Limited, Bingley, pp 479–513
- Andrews D (1999) Consistent moment selection procedures for generalized method of moments estimation. *Econometrica* 67:543–563

- Arellano M, Bond S (1991) Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Rev Econ Stud* 58:277–297
- Baltagi B, Bresson G, Pirotte A (2003) Fixed effects, random effects or Hausman-Taylor? A pretest estimator. *Econ Lett* 79:361–369
- Blundell R, Bond S, Devereux M, Schiantarelli F (1992) Investment and Tobin's  $Q$ : evidence from company panel data. *J Econom* 51:233–257
- Breusch T, Godfrey L (1986) Data transformation tests. *Econ J* 96:47–58
- Cameron A, Trivedi P (2005) *Microeconometrics methods and applications*. Cambridge University Press, New York
- Davidson R, Godfrey L, MacKinnon J (1985) A simplified version of the differencing test. *Int Econ Rev* 26:639–647
- Erickson T, Whited T (2000) Measurement error and the relationship between investment and  $q$ . *J Polit Econ* 108:1027–1057
- Goolsbee A (2000) The importance of measurement error in the cost of capital. *Natl Tax J* 53:215–228
- Griliches Z, Hausman J (1986) Errors in variables in panel data. *J Econom* 31:93–118
- Hall A (2005) *Generalized method of moments*. Oxford University Press, Oxford
- Harris R, Tzavalis E (1999) Inference for unit roots in dynamic panels where the time dimension is fixed. *J Econom* 91:201–226
- Hausman J (1978) Specification tests in econometrics. *Econometrica* 46:1251–1271
- Hayakawa K (2019) Alternative over-identifying restriction test in the GMM estimation of panel data models. *Econom Stat* 10:71–95
- Joshi R, Wooldridge J (2019) Correlated random effects models with endogenous explanatory variables and unbalanced panels. *Ann Econ Stat* 134:243–268
- Lloyd C (2005) Estimating test power adjusted for size. *J Stat Comput Simul* 75(11):921–933
- McKinnish T (2008) Panel data models and transitory fluctuations in the explanatory variable. In: Fomby T, Hill R, Millimet D, Smith J, Vytalil E (eds) *Advances in Econometrics*. Vol. 21: Modelling and Evaluating Treatment Effects in Econometrics. pp 335–358
- Mundlak Y (1961) Empirical production function free of management bias. *J Farm Econ* 43:44–56
- Mundlak Y, Hoch I (1965) Consequences of alternative specifications in estimation of Cobb-Douglas production functions. *Econometrica* 33:814–828
- Newey W (1985) Generalized method of moments specification testing. *J Econom* 29:229–256
- Newey W, McFadden D (1994) Large sample estimation and hypothesis testing. In: Griliches Z, Intriligator M (eds) *Handbook of Econometrics*, vol 4. North Holland
- Newey W, West K (1987) Hypothesis testing with efficient method of moments estimation. *Int Econ Rev* 28:777–787
- Okui R (2009) The optimal choice of moments in dynamic panel data models. *J Econom* 151:1–16
- Parente P, Santos Silva J (2012) A cautionary note on tests of overidentifying restrictions. *Econ Lett* 115:314–317
- Plosser C, Schwert G, White H (1982) Differencing as a test of specification. *Int Econ Rev* 23:535–552
- Sarafidis V, Wansbeek T (2012) Cross-sectional dependence in panel data analysis. *Economet Rev* 31:483–531
- Su L, Zhang Y, Wei J (2016) A practical test for strict exogeneity in linear panel data models with fixed effects. *Econ Lett* 147:27–31
- Thursby J (1989) A comparison of several specification error tests for a general alternative. *Int Econ Rev* 30:217–230
- Verbeek M, Nijman T (1992) Testing for selectivity bias in panel data models. *Int Econ Rev* 33:681–703
- Wooldridge J (2010) *Econometric analysis of cross section and panel data*, 2nd edn. MIT Press, Cambridge
- Wooldridge JM (1995) Selection corrections for panel data models under conditional mean independence assumptions. *J Econom* 68:115–132