Consistent estimation of linear panel data models with measurement error

Meijer, Erik; Spierdijk, Laura; Wansbeek, Thomas

Published in:
Journal of Econometrics

DOI:
10.1016/j.jeconom.2017.06.003

Journal of Econometrics 200 (2017) 169–180

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Consistent estimation of linear panel data models with measurement error

Erik Meijer [a], Laura Spierdijk [b,*], Tom Wansbeek [b]

[a] University of Southern California, United States
[b] University of Groningen, Netherlands

ARTICLE INFO

Article history:
Available online 1 July 2017

JEL classification:
C23, C26

Keywords:
Measurement error
Panel data
Third moments
Heteroskedasticity
GMM

ABSTRACT

Measurement error causes a bias towards zero when estimating a panel data linear regression model. The panel data context offers various opportunities to derive instrumental variables allowing for consistent estimation. We consider three sources of moment conditions: (i) restrictions on the covariance matrix of the errors in the equations, (ii) nonzero third moments of the regressors, and (iii) heteroskedasticity and nonlinearity in the relation between the error-ridden regressor and another, error-free, regressor. In simulations, these approaches appear to work well.

1. Introduction

Covariates of interest in a linear regression analysis are often measured with error. If not accounted for, measurement error causes a bias towards zero in the parameter estimates; see, for example, Wansbeek and Meijer (2000) for a comprehensive treatment. In this paper we consider measurement error in panel data models and group a number of results that are based on instrumental variables.

Since the seminal article by Griliches and Hausman (1986), several papers have discussed the topic of measurement error in panel data models. Wansbeek and Koning (1991) present a simple approach for the case where the intertemporal covariance matrix of the measurement errors is scalar (i.e., proportional to the identity matrix). For the more general case where this matrix is diagonal, Biørn and Klette (1998) present a generalized method of moments (GMM) approach. This is further generalized by Biørn (2000) to the case where only some off-diagonal elements of the intertemporal covariance matrix of the measurement errors are zero. Wansbeek (2001) presents a general GMM approach based on linear restrictions of any form on this matrix, which is extended by Shao et al. (2011) to the case of unbalanced panel data. Xiao et al. (2007) correct an error in Wansbeek (2001) and identify cases in which a single-step approach in GMM is already optimal. Xiao et al. (2010a, 2010b) and provide several extensions, including the presence of multiple covariates measured with error. Biørn and Klette (1999), Aasness et al. (2003), Biørn (2003), and Biørn and Krishnakumar (2008) provide further applications and context. Meijer et al. (2014) provide a recent overview.

The literature has focused on moment conditions that exploit assumptions about the intertemporal covariance matrix of the measurement errors. Because the measurement errors are not observed, these assumptions may be hard to justify. We therefore consider GMM estimation based on moment conditions from three other sources: (1) restrictions on the intertemporal covariance matrix of errors in the equations, (2) third moments of regressors with error, and (3) exogenous regressors.

Various of the above-mentioned papers have exploited zero restrictions on the error covariance matrix. We extend this to general linear structures. The use of the third moment goes back to Geary (1942). It has been extended by Pal (1980), Dagenais and Dagenais (1997), Lewbel (1996, 1997), and Erickson and Whited (2002), and we further extend this to exploit the additional information available in panel data. As to exogenous regressors, we extend and elaborate the recent approach due to Lewbel (2012) exploiting heteroskedasticity or nonlinearity in the relation between an error-ridden regressor and a correctly measured, exogenous one. Our main contribution is to collect these approaches, extend or adapt them to the panel data context, and present them in a unified way. This leads to estimators that are easy to use by the applied researcher. They may also prove helpful to theoretical researchers who want to build on this and who are provided with a template for further extensions.

* Corresponding author.
E-mail addresses: erik.meijer@usc.edu (E. Meijer), l.spierdijk@rug.nl (L. Spierdijk), t.j.wansbeek@rug.nl (T. Wansbeek).

http://dx.doi.org/10.1016/j.jeconom.2017.06.003

The setup of this paper is as follows. The panel data model with measurement error is introduced in Section 2. Section 3 explains how restrictions on the intertemporal covariance matrix of errors in the equations can be used to obtain consistent estimators. Section 4 shows how to generate instruments when the third moment of the regressor is nonzero, and Section 5 demonstrates how correctly measured exogenous variables can be used to generate instruments. We present simulation results in Section 6, with special attention to panel IV estimation and its practical implementation. Finally, Section 7 discusses remaining issues and further challenges. An online appendix with supplementary material is available.

2. Model and estimation basics

In this section, we introduce our general model setup and assumptions and then discuss instrumental variables estimation for panel data in some generality, which has not received much attention in the form we discuss it, and thus is also useful more generally.

Model and assumptions. We consider the linear regression models for panel data model. The most general form we consider is

$$y_{nt} = \alpha_n + \gamma y_{n,t-1} + \xi_{nt}' \beta + r_{nt}'\delta + \varepsilon_{nt},$$

where y_{nt} is the dependent variable for cross-sectional unit n at time t, α_n is a fixed effect, ξ_{nt} and r_{nt} are vectors of (k and ℓ, respectively) regressors, and ε_{nt} is the error term, which has mean zero. The regressors ξ_{nt} may be measured with error, so instead we observe x_{nt}, which is related to ξ_{nt} by

$$x_{nt} = \xi_{nt} + v_{nt}.$$

The reduced form obtained by the elimination of ξ_{nt} is

$$y_{nt} = \alpha_n + \gamma y_{n,t-1} + x_{nt}' \beta + r_{nt}'\delta + u_{nt},$$

where

$$u_{nt} = \varepsilon_{nt} - v_{nt}' \beta.$$

The data set consists of N cross-sectional units, which we will generally call *individuals*. Unless otherwise specified, we assume we have a balanced panel with each individual observed at the same T time points, with y_{n0} also observed. We assume $N \gg T$ and thus we will use large N, fixed T asymptotics.

It is notationally convenient to define stacked versions for the nth individual: $y_n \equiv (y_{n1}, \ldots, y_{nT})'$, $X_n \equiv (x_{n1}, \ldots, x_{nT})'$, and so forth. We denote the intertemporal covariance matrix of the error terms by $\Sigma_\varepsilon \equiv \mathbb{E}(\varepsilon_n\varepsilon_n')$.

Initially, we assume that ε_n and the measurement errors V_n are independent of the regressors (Ξ_n, R_n) and of each other, which implies strict exogeneity of the regressors, classical measurement error, and homoskedasticity. However, we will discuss to what extent these assumptions can be relaxed. We will also discuss the random effects model, which we do by dropping α_n from the equation and assuming a specific structure for Σ_ε, or allowing Σ_ε to be arbitrary.

Throughout, we take all variables in deviations from their means per time period, thus implicitly handling fixed time effects. In order not to burden the notation unduly, this is left implicit in the following. This is without loss of generality because of the large N, fixed T asymptotics.

In general, the dynamic panel data model is more complicated to deal with than the static model, sometimes much more so, spawning a huge literature. Our analysis is also complicated by the presence of the lagged dependent variable. Hence, we will begin each of our cases by considering the static version ($\gamma = 0$) first and then indicate what adaptations are required when the lagged dependent variable enters the model. When doing so, we will assume that y_{nt} and hence $y_{n,t-1}$ do not contain measurement error. The case where the dependent variable is measured with error is the topic of a separate, growing literature; see Meijer et al. (2013), Biørn (2015), Gospodinov et al. (2014), and Lee et al. (2014).

As with the cross-sectional measurement error model, our model is not identified without additional information or further assumptions. The assumptions we consider are restrictions on the (co)variances and third moments of the random variables in the model (ε_{nt}, v_{nt}, ξ_{nt}, and r_{nt}). These restrictions are used to derive moment conditions, which define panel instrumental variables. Thus, we will obtain consistent estimators of the coefficients of this model and various special cases of it by instrumental variables techniques for panel data. Below, we discuss the technique generally, which turns out to be a generalization of instrumental variables estimation for cross sections.

Panel IV estimation. The idea behind instrumental variables (IVs) in a panel data setting is the same as in the usual cross-sectional case but there are a few things specific to panel data that we would like to point out, expanding the discussion of Cameron and Trivedi (2005, Section 22.2).

In cross-sections, IV estimation is based on moment conditions of the form $\mathbb{E}(z_n u_n) = 0$, where z_n is a q-vector of instruments for observation n. This carries over to the panel data context, where the analogous moment conditions are $\mathbb{E}(z_{nt}u_{nt}) = 0$. However, in panel data contexts, we can expand this to moment conditions of the form $\mathbb{E}(Z_n'u_n) = 0$, with Z_n now a matrix of order $T \times q$ and u_n now a T-vector. For example, this allows moment conditions of the form $\mathbb{E}(z_{ns}u_{nt} - z_{nt}u_{ns}) = 0$ (for some $s \neq t$), which do not fit in the standard (cross-sectional) IV structure. We will encounter moment conditions like these below. As with the cross-sectional IVs, the panel IVs also need to be correlated with the regressors, generically denoted by X_n; specifically, $\mathbb{E}(Z_n'X_n)$ must have full column rank. With the y_n, X_n, and Z_n stacked in y, X, and Z, respectively, so $X'Z = \sum_n X_n'Z_n$, and W a weight matrix of order $q \times q$, the IV estimator is

$$\hat{\beta}_{\text{IV}} = (X'ZWZ'X)^{-1}X'ZWZ'y. \tag{1}$$

The consistency and asymptotic normality of $\hat{\beta}_{\text{IV}}$ follows from standard GMM theory (Hansen, 1982). GMM theory additionally provides an asymptotically optimal choice of W and specification tests when there are more instruments than regressors. The literature also offers heteroskedasticity–robust and cluster–robust standard errors, and ways to handle unbalanced panel data and sampling weights correctly (see, e.g., Cameron and Trivedi, 2005, Chapter 22).

3. Restrictions on Σ_ε

We consider linear restrictions that we may be willing to impose on Σ_ε, the covariance matrix of the errors in the model equations. The restrictions we consider are linear and hence can be expressed as vec $\Sigma_\varepsilon = C_\varepsilon \pi_\varepsilon$, with C_ε known (and of full column rank) and π_ε ($r_\varepsilon \times 1$) unknown. We start with two motivating examples and then treat the model in general, first for the static case and then adapt it for the case that the model is dynamic. Our approach is inspired by Ahn and Schmidt (1995, 1997), but reframes and generalizes their results to arbitrary linear restrictions.

Motivating examples. In the first example, we take $T = 3$, and have only a single regressor, $y_n = \xi_n \beta + \varepsilon_n$, $x_n = \xi_n + v_n$, and consider the random-effects model, where $\varepsilon_{nt} = \alpha_n + w_{nt}$, with $\alpha_n \sim (0, \sigma_\alpha^2)$

now a random effect and $w_{nt} \sim (0, \sigma_w^2)$ i.i.d. Then

$$\Sigma_\varepsilon = \begin{pmatrix} \sigma_\alpha^2+\sigma_w^2 & \sigma_\alpha^2 & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2+\sigma_w^2 & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \sigma_\alpha^2+\sigma_w^2 \end{pmatrix},$$

$$C'_\varepsilon = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}, \quad \pi_\varepsilon = \begin{pmatrix} \sigma_w^2 \\ \sigma_\alpha^2 \end{pmatrix},$$

and $r_\varepsilon = 2$. Hence, with $u_n = y_n - x_n\,\beta$,

$$\mathbb{E}(u_n \otimes y_n - C_\varepsilon \pi_\varepsilon) = \mathbb{E}\,[(\varepsilon_n - \upsilon_n\,\beta) \otimes (\xi_n\,\beta + \varepsilon_n) - C_\varepsilon\pi_\varepsilon] = 0. \quad (2)$$

Let $C_{\varepsilon\perp}$ be a complement of C_ε, that is, a matrix of order $T^2\times(T^2-r_\varepsilon)$ and rank $T^2 - r_\varepsilon$ such that $C'_{\varepsilon\perp}C_\varepsilon = 0$. Premultiplication of (2) by $C'_{\varepsilon\perp}$ gives

$$C'_{\varepsilon\perp}\mathbb{E}(u_n \otimes y_n) = C'_{\varepsilon\perp}\mathbb{E}\,[(I_T \otimes y_n)(y_n - x_n\,\beta)] = 0. \quad (3)$$

So the (panel) IVs that are implied by the structure on Σ_ε are $Z_n = (I_T \otimes y_n)'C_{\varepsilon\perp}$. The requirement that $\mathbb{E}(Z'_nX_n)$ have full column rank translates into $C'_{\varepsilon\perp}\mathbb{E}(\xi_n \otimes \xi_n) \neq 0$, so ξ_{nt} should not itself follow a random effects structure.

For one valid but otherwise arbitrary choice of $C_{\varepsilon\perp}$, we obtain

$$Z'_n = C'_{\varepsilon\perp}(I_T \otimes y_n)$$
$$= \begin{pmatrix} 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \end{pmatrix}$$
$$\times \begin{pmatrix} y_{n1} & 0 & 0 \\ y_{n2} & 0 & 0 \\ y_{n3} & 0 & 0 \\ 0 & y_{n1} & 0 \\ 0 & y_{n2} & 0 \\ 0 & y_{n3} & 0 \\ 0 & 0 & y_{n1} \\ 0 & 0 & y_{n2} \\ 0 & 0 & y_{n3} \end{pmatrix}$$
$$= \begin{pmatrix} y_{n1} & -y_{n2} & 0 \\ 0 & y_{n2} & -y_{n3} \\ y_{n3} & -y_{n1} & 0 \\ 0 & y_{n1} & -y_{n2} \\ y_{n2} & -y_{n1} & 0 \\ y_{n3} & 0 & -y_{n1} \\ 0 & y_{n3} & -y_{n2} \end{pmatrix}.$$

Instrument validity can be checked directly. The expectation of each of the rows of the latter matrix multiplied by $(y_{n1}-x_{n1}\,\beta, y_{n2}-x_{n2}\,\beta, y_{n3}-x_{n3}\,\beta)'$ is zero, yielding seven moment conditions, cf. (3). The first two exploit the homoskedasticity, that is, the equality of the diagonal elements of Σ_ε. The third and the fourth exploit the equality of the off-diagonal elements of Σ_ε. The final three are of the form

$$\mathbb{E}(y_{ns}x_{nt} - y_{nt}x_{ns})\beta = 0, \quad s \neq t, \quad (4)$$

and reflect the symmetry of $\Sigma_{xy} \equiv \mathbb{E}(x_n\,y'_n)$ implied by the model. With $\beta \neq 0$, β drops out of these three moments, and we are left with moment conditions that do not depend on parameters. Including such moment conditions increases the asymptotic efficiency of the IV procedure (Qian and Schmidt, 1999), but the contribution of the moment conditions spawned by symmetry is probably very modest.

In the second example, we assume that $\{\varepsilon_{nt}\}$ is stationary, which implies $(\Sigma_\varepsilon)_{ts} = \pi_{|t-s|}$ for some set of parameters $\pi_0, \ldots, \pi_{T-1}$. With $T = 3$, we have $r_\varepsilon = 3$,

$$C'_\varepsilon = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}, \quad \pi_\varepsilon = \begin{pmatrix} \pi_0 \\ \pi_1 \\ \pi_2 \end{pmatrix},$$

and $C_{\varepsilon\perp}$ and Z_n for this case follow readily. There are six moment conditions, two exploiting equality along the main diagonal and one along the subdiagonal; the contribution of symmetry through three moment conditions that do not depend on parameters is the same as in the first example. This example also covers the case of $T = 4$ with fixed effects eliminated by taking first differences.

Static model. For the general model but still without the lagged dependent variable as a regressor and without fixed effects, so $y_n = \Xi_n\,\beta + R_n\delta + \varepsilon_n$ and $X_n = \Xi_n + V_n$, adapting (2) is straightforward:

$$C'_{\varepsilon\perp}\mathbb{E}(u_n \otimes y_n - C_{\varepsilon\perp})$$
$$= C'_{\varepsilon\perp}\mathbb{E}\,[(\varepsilon_n - V_n\,\beta) \otimes (\Xi_n\,\beta + R_n\delta + \varepsilon_n)] = 0 \quad (5)$$

due to the exogeneity of R_n. So the moment conditions (3) remain valid; the panel IVs that are implied by the structure on Σ_ε, $Z_n = (I_T \otimes y_n)'C_{\varepsilon\perp}$, are unaltered. The addition of the ℓ parameters in δ can be covered by the ℓ moment conditions

$$\mathbb{E}(R'_n u_n) = 0, \quad (6)$$

so the entire instrument set available for estimating the general static model is (Z_n, R_n). The number of moment conditions from the structure on Σ_ε is $T^2 - r_\varepsilon$.

Heteroskedasticity. In the static model the restrictions vec $\Sigma_\varepsilon = C_\varepsilon\pi_\varepsilon$, with π_ε constant but unknown parameters, imply homoskedasticity. This may be undesirably strong. However, this can be easily relaxed by allowing π_ε to be different for different individuals, and thus writing it as $\pi_{\varepsilon n}$, which may depend in an arbitrary way on ξ_n or even υ_n. The instruments that we use are still valid under this relaxation, because they operate in the space orthogonal to C_ε, in which π_ε is eliminated, and this carries over to the heteroskedastic case. The estimators are also still consistent (but inefficient) if the measurement error is heteroskedastic, because they do not use the homoskedasticity of the measurement errors in any way.

Fixed effects. We now consider fixed effects. With ι_T a T-vector of ones, the static model becomes

$$y_n = \iota_T\alpha_n + \Xi_n\,\beta + R_n\,\delta + \varepsilon_n.$$

The fixed effect α_n can be eliminated from the model by any matrix B with property $B'\iota_T = 0$; typical choices for B are the centering matrix $Q_T = I_T - \iota_T\iota'_T/T$ (or a subset of $T-1$ rows of it) or the $T \times (T-1)$ matrix that transforms a T vector in first differences. We put a tilde on a vector or matrix when it has been premultiplied by B'. After transformation by B we can proceed as before. That is, we now start from

$$\mathbb{E}\left[\tilde{u}_n \otimes \tilde{y}_n - (B \otimes B)'C_\varepsilon\pi_\varepsilon\right] = 0, \quad (7)$$

with $\tilde{u}_n \equiv \tilde{y}_n - \tilde{X}_n\,\beta - \tilde{R}_n\,\delta$. So the IVs are now based on the complement of $(B \otimes B)'C_\varepsilon$ rather than C_ε.[1]

[1] If we assume that $\mathbb{E}(\alpha_nV_n) = 0$ and $\mathbb{E}(\alpha_n\varepsilon_n) = 0$, there is a slight loss of efficiency in this approach since $\mathbb{E}[(\varepsilon_n - V_n\,\beta) \otimes y_n] = \text{vec}\,\Sigma_\varepsilon$ still holds. Hence, for estimation we only need to eliminate the α_n from $y_n - \Xi_n\,\beta - R_n\delta$. Consequently, $\mathbb{E}\left[(\tilde{y}_n - \tilde{X}_n\,\beta - \tilde{R}_n\,\delta) \otimes y_n - (B \otimes I_T)'C_\varepsilon\pi_\varepsilon\right] = 0$ is a larger set of moment conditions, generating more instruments than (7).

Dynamic model. When the lagged dependent variable is included as a regressor ($\gamma \neq 0$), the moment conditions (5) do not hold anymore. To see this, write the reduced form of the dynamic model as $y_{nt} = a_{nt} + \sum_{j=0}^{\infty}\gamma^j \varepsilon_{n,t-j}$, where a_{nt} is an infinite sum containing contemporaneous and lagged versions of ξ_{nt} and r_{nt}, as well as α_n (in the fixed effects model). Hence,

$$d \equiv C'_{\varepsilon\perp}\mathbb{E}(u_n \otimes y_n) = C'_{\varepsilon\perp}\sum_{j=1}^{\infty}\mathbb{E}(\varepsilon_n \otimes \varepsilon_{n,-j}) \neq 0,$$

where the notation $\varepsilon_{n,-j} = (\varepsilon_{n,1-j}, \ldots, \varepsilon_{n,T-j})'$ denotes the j-periods lagged version of the vector ε_n. So $(I_T \otimes y_n)'C_{\varepsilon\perp}$ is no longer a set of valid instruments.

By way of illustration, let us consider again the random-effects model with $T = 3$ from Section 3, with the lagged dependent variable added on the right hand side. Then

$$\begin{aligned}\mathbb{E}(u_{nt}\, y_{ns}) &= \mathbb{E}(\varepsilon_{nt}\, y_{ns}) \\ &= \sum_{j=0}^{\infty}\mathbb{E}(\varepsilon_{nt}\varepsilon_{n,s-j})\gamma^j = \frac{\sigma_\alpha^2}{1-\gamma} + I(t \leq s)\sigma_w^2\gamma^{s-t},\end{aligned}$$

provided that $|\gamma| < 1$. Therefore,

$$d = \mathbb{E}(Z_n'u_n) = \mathbb{E}\begin{pmatrix} y_{n1}u_{n1} - y_{n2}u_{n2} \\ y_{n2}u_{n2} - y_{n3}u_{n3} \\ y_{n3}u_{n1} - y_{n1}u_{n2} \\ y_{n1}u_{n2} - y_{n2}u_{n3} \\ y_{n2}u_{n1} - y_{n1}u_{n2} \\ y_{n3}u_{n1} - y_{n1}u_{n3} \\ y_{n3}u_{n2} - y_{n2}u_{n3} \end{pmatrix} = \sigma_w^2\begin{pmatrix} 0 \\ 0 \\ \gamma^2 \\ 0 \\ \gamma \\ \gamma^2 \\ \gamma \end{pmatrix}$$

$$= \gamma\sigma_w^2\begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ \gamma \end{pmatrix} \neq 0.$$

So the moment conditions do not apply anymore. However, they still can be exploited to some extent since

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & -1 \end{pmatrix} d = 0,$$

so now

$$Z_n' = \begin{pmatrix} y_{n1} & -y_{n2} & 0 \\ 0 & y_{n2} & -y_{n3} \\ 0 & -y_{n1} & y_{n1} \\ 0 & y_{n1} & -y_{n2} \\ y_{n2} & -y_{n3} & 0 \end{pmatrix}. \tag{8}$$

contains valid instruments.

This result is specific for the random effects model, but the approach is general. We can always write d in the form $d = C_\theta\pi_\theta$, where the elements of π_θ are functions of the model parameters and C_θ does not depend on parameters. If C_θ has at least k more rows than columns, this may identify the coefficients of interest. We simply replace $C_{\varepsilon\perp}$ by $C_{\varepsilon\perp}C_{\theta\perp}$ and then proceed as in the static case. Thus, adding the lagged dependent variable to the model invalidates (some of) the moment conditions, but an adaptation may offer a way out, much of the same form as before, but with fewer moment conditions.

For the fixed effects model, the same approach can be followed as in Section 3, which is to eliminate the individual effects and proceed with the transformed data. This gives

$$d^{\mathrm{FE}} \equiv \left((B \otimes B)'C_\varepsilon\right)'_\perp \mathbb{E}(\tilde{u}_n \otimes \tilde{y}_n) = C_\theta^{\mathrm{FE}}\pi_\theta^{\mathrm{FE}},$$

for some known constant matrix C_θ^{FE} and some vector π_θ^{FE} that may depend on parameters. The resulting IVs are then $(I_{T-1} \otimes \tilde{y}_n)'\left((B \otimes B)'C_\varepsilon\right)_\perp C_{\theta\perp}^{\mathrm{FE}}$.

The first panel of Table 1 lists the IVs and the underlying assumptions for the various cases.

4. Nonzero third moments

We can obtain instruments from within the model when the mismeasured variables are not normally distributed. Then, higher moments contain additional information that can help with identification and estimation. This has received considerable attention in the literature on cross-sectional models, as cited in the introduction. We start with a motivating example and then treat the general case.

Motivating examples. In the static model $y_{nt} = \xi_{nt}\,\beta + \varepsilon_{nt}$ where ξ_{nt} has only one element, suppose that $\mathbb{E}(\xi_{nt}^3) = \lambda \neq 0$ and that ξ_{nt}, v_{nt}, and ε_{nt} are stochastically independent of each other. Under these assumptions, Pal (1980), in a cross-sectional context, discusses a consistent estimator that in the panel context translates into

$$\hat{\beta} = \frac{\sum_{n=1}^{N}\sum_{t=1}^{T} x_{nt}\, y_{nt}^2}{\sum_{n=1}^{N}\sum_{t=1}^{T} x_{nt}^2\, y_{nt}}.$$

This is an IV estimator with instrument $z_{nt} = x_{nt}\, y_{nt}$, that is, it is based on the moment condition

$$\begin{aligned}&\mathbb{E}\left[(x_{nt}\, y_{nt})(y_{nt} - x_{nt}\,\beta)\right] \\ &\quad = \mathbb{E}\left[(\xi_{nt} + v_{nt})(\xi_{nt}\,\beta + \varepsilon_{nt})(\varepsilon_{nt} - v_{nt}\,\beta)\right] = 0,\end{aligned}$$

whereas $\mathbb{E}[(x_{nt}\, y_{nt})x_{nt}] = \beta\lambda \neq 0$, provided that β is nonzero. Note that the independence assumption ensures that expressions like $\mathbb{E}(\xi_{nt}\varepsilon_{nt}^2)$ are zero. Lack of independence, in this case heteroskedasticity, would invalidate this moment condition, see Section 4. In a small simulation study, Van Montfort et al. (1987) found that this estimator has reasonably good statistical properties in moderately sized cross-sectional samples.

The panel data context implies additional moment conditions. Suppose that $\mathbb{E}(\xi_{np}\xi_{ns}\xi_{nt}) = \lambda_{pst} \neq 0$ and that ξ_n, v_n, and ε_n are stochastically independent of each other. Then

$$\begin{aligned}&\mathbb{E}\left[x_{np}\, y_{ns}(y_{nt} - x_{nt}\,\beta)\right] \\ &\quad = \mathbb{E}\left[(\xi_{np} + v_{np})(\xi_{ns}\,\beta + \varepsilon_{ns})(\varepsilon_{nt} - v_{nt}\,\beta)\right] = 0,\end{aligned}$$

while $\mathbb{E}[(x_{np}\, y_{ns})x_{nt}] = \beta\lambda_{pst} \neq 0$. Hence, $(x_{np}\, y_{ns})$ is also a valid instrument for x_{nt}.

Static model. As in the motivating example, we continue with the static model $y_{nt} = \xi_{nt}\,\beta + \varepsilon_{nt}$ with only one regressor, but we now study third moments more generally. Let $\lambda_\xi \equiv \mathbb{E}(\xi_n \otimes \xi_n \otimes \xi_n) \neq 0$, the latter meaning that at least one element of λ_ξ is nonzero. Let $\lambda_\varepsilon \equiv \mathbb{E}(\varepsilon_n \otimes \varepsilon_n \otimes \varepsilon_n)$ and $\lambda_v \equiv \mathbb{E}(v_n \otimes v_n \otimes v_n)$ be defined accordingly.

The third moments can now be written as

$$\mathbb{E}(y_n \otimes y_n \otimes y_n) = \lambda_\xi\,\beta^3 + \lambda_\varepsilon \tag{9a}$$

$$\mathbb{E}(y_n \otimes y_n \otimes x_n) = \lambda_\xi\,\beta^2 \tag{9b}$$

$$\mathbb{E}(y_n \otimes x_n \otimes x_n) = \lambda_\xi\,\beta \tag{9c}$$

$$\mathbb{E}(x_n \otimes x_n \otimes x_n) = \lambda_\xi + \lambda_v. \tag{9d}$$

These expressions owe their simplicity to the assumed independence between ξ_n, ε_n, and v_n. The independence assumption implies homoskedasticity:

$$\mathbb{E}(\varepsilon_n \otimes \varepsilon_n \otimes \xi_n) = \mathbb{E}(v_n \otimes v_n \otimes \xi_n) = 0.$$

Subtracting β times (9c) from (9b) gives

$$\mathbb{E}\left[y_n \otimes (y_n - x_n\,\beta) \otimes x_n\right] = \mathbb{E}\left[(y_n \otimes I_T \otimes x_n)(y_n - x_n\,\beta)\right] = 0. \tag{10}$$

Table 1
Summary of IVs and underlying assumptions.

Single regressor			
Static model	$y_n = \xi_n \beta + r_n\delta + \epsilon_n$		
Dynamic model	$y_n = \gamma y_{n,-1} + \xi_n \beta + r_n\delta + \epsilon_n$		
Measurement error model	$x_n = \xi_n + v_n$		
Basic assumptions	ϵ_n, v_n, and (ξ_n, r_n) mutually independent		
	cov.restr.	*3rd mom.*	*exog. regr.*
Specific assumptions	$\mathbb{E}(\epsilon_n \otimes \epsilon_n) = C_\epsilon \pi_\epsilon$	$\mathbb{E}(\xi_n \otimes \xi_n \otimes \xi_n) \neq 0$	$\mathbb{E}(r_n \otimes \omega_n \otimes \omega_n) \neq 0$
Instrumental variables ($= Z_n$)			
Static model (+ RE)	$(I_T \otimes y_n)' C_{\epsilon\perp}$	$(y_n \otimes I_T \otimes x_n)'$	$r_n' \otimes (1, \hat{w}_n') \otimes I_T$
Static model + heteroskedastic ϵ_n	$(I_T \otimes y_n)' C_{\epsilon\perp}$	case specific	$r_n' \otimes (1, \hat{w}_n') \otimes I_T$
Static model + heteroskedastic v_n	$(I_T \otimes y_n)' C_{\epsilon\perp}$	case specific	case specific
Static model + FE	$(I_{T-1} \otimes \tilde{y}_n)' \Big((B \otimes B)' C_\epsilon\Big)_\perp$	$(\tilde{y}_n \otimes I_{T-1} \otimes x_n)'$	$r_n' \otimes (1, \hat{w}_n') \otimes I_{T-1}$
Dynamic model (+ RE)	$(I_T \otimes y_n)' C_{\epsilon\perp} C_{\theta\perp}$	$(y_n \otimes I_T \otimes x_n)'$	$r_n' \otimes (1, \hat{w}_n') \otimes I_T$
Dynamic model + FE	$(I_{T-1} \otimes \tilde{y}_n)' \left((B \otimes B)' C_\epsilon\right)_\perp C_{\theta\perp}^{\mathrm{FE}}$	$(\tilde{y}_n \otimes I_{T-1} \otimes x_n)'$	$r_n' \otimes (1, \hat{w}_n') \otimes I_{T-1}$
Multiple regressors			
Static model	$y_n = \Xi_n \beta + R_n\delta + \epsilon_n$		
Dynamic model	$y_n = \gamma y_{n,-1} + \Xi_n \beta + R_n\delta + \epsilon_n$		
Measurement error model	$X_n = \Xi_n + V_n$		
Basic assumptions	ϵ_n, V_n, and (Ξ_n, R_n) mutually independent		
Instrumental variables	The corresponding IVs are obtained by replacing x_n, r_n, and $\hat{w}_n$ in the single-regressor IVs by $\operatorname{vec} X_n$, $\operatorname{vec} R_n$, and $\operatorname{vec} \hat{W}_n$, respectively.		

Notes: B is a transformation matrix with property $B'\iota_T = 0$; $\tilde{y}_n = B'y_n$; ω_n is the error term in the linear projection of ξ_n on r_n; $\hat{w}_n$ is the vector of OLS residuals from the regression of x_n on r_n. For covariance restrictions and 3rd moments, r_n (R_n if multiple regressors) is an additional IV. This is already part of the listed IVs for exogenous regressors. See Sections 3–5 for detailed definitions and derivations.

Hence, $Z_{yx,n} \equiv (y_n \otimes I_T \otimes x_n)'$ is a valid instrument matrix, because $\mathbb{E}(Z_{yx,n}' x_n) = \lambda_\xi \beta \neq 0$. In scalar notation, the set of moments (10) is

$$\mathbb{E}\left[x_{np}\, y_{ns}(y_{nt} - x_{nt}\,\beta)\right] = 0, \quad p, s, t = 1, \ldots, T. \tag{11}$$

Thus, (10) amounts to ordinary IV with T^3 instruments, divided in T sets: set t has T^2 instruments $(x_{np}\, y_{ns})$, $p, s = 1, \ldots, T$ that multiply with $(y_{nt} - x_{nt}\,\beta)$ and zeros that multiply with $(y_{n\tau} - x_{n\tau}\,\beta)$ for $\tau \neq t$.

Conditional on β, (9c) identifies λ_ξ, (9a) identifies λ_ε, and (9d) identifies λ_v. Also, the matrices $Z_{yy,n} \equiv (y_n \otimes y_n \otimes I_T)'$ and $Z_{xx,n} \equiv (I_T \otimes x_n \otimes x_n)'$ are valid instrument matrices if the third moments of ε and v, respectively, are assumed to vanish.

In the more general model with multiple explanatory variables with or without measurement error, (11) straightforwardly generalizes to

$$\begin{aligned}&\mathbb{E}\left[x_{npj}\, y_{ns}(y_{nt} - x_{nt}'\,\beta - r_{nt}'\delta)\right] = 0,\\ &p, s, t = 1, \ldots, T;\, j = 1, \ldots, k.\end{aligned} \tag{12}$$

This corresponds with the instrument matrix $Z_{yx,n} = (y_n \otimes I_T \otimes \operatorname{vec} X_n)'$. Furthermore, as in Section 3, R_n is also a valid set of instruments.

As stated before, using third moments to estimate measurement error models has a long history, with a number of publications added more recently, all pertaining to the case of a single cross-section. In the panel data model context, the number of instruments generated is $O(T^3)$. So there is no dearth of instruments. Of course, more instruments mean more signal but also more noise, so the balance is not clear a priori. As an example, we can linearly transform the set of moments to obtain a set of $T(T+1)(T+2)/6$ moment conditions

$$\mathbb{E}\left[x_{np}\, y_{ns}(y_{nt} - x_{nt}\,\beta)\right] = 0, \quad 1 \leq p \leq s \leq t \leq T,$$

that depend on β and a set of $T^3 - T(T+1)(T+2)/6$ symmetry conditions that do not involve β, for example

$$\mathbb{E}(x_{np}\, y_{ns} y_{nt} - x_{nt}\, y_{ns} y_{np}) = 0, \quad 1 \leq p < t \leq T, \quad 1 \leq s \leq T,$$

which may not improve finite-sample statistical properties. We will investigate the many instruments issue through simulation in Section 6.

Under the assumption that ε_n and v_n are independent of each other and of ξ_n, identification of the linear measurement error model under nonnormality is related to Kotlarski (1967) theorem. See Hansen et al. (2004), who discuss the details. Kotlarski's proof uses characteristic functions to show the identification of distributions. Van Montfort et al. (1989) used these relations between characteristic functions to estimate the cross-sectional measurement error model. This would be an alternative to using higher order moments in our panel data case as well.

Heteroskedasticity. In the static model the estimators that use third moments do not accommodate heteroskedasticity easily. The moment condition (9b) is only valid if $\mathbb{E}(\varepsilon_n \otimes \varepsilon_n \otimes \xi_n) = 0$ and (9c) is only valid if $\mathbb{E}(\xi_n \otimes v_n \otimes v_n) = 0$. Under arbitrary heteroskedasticity, the third moments do not identify the regression coefficient anymore. We can, however, allow some form of heteroskedasticity, as long as enough elements of these third moment vectors are zero. For example, we may be willing to assume that $\mathbb{E}(\varepsilon_{nt}\varepsilon_{ns}\xi_{np}) = 0$ and $\mathbb{E}(v_{nt} v_{ns}\xi_{np}) = 0$ if p, s, and t are all distinct. This would identify β if $T \geq 3$ in the random effects situation and $T \geq 4$ in the fixed effects situation (to be discussed in Section 4), even if $\mathbb{E}(\varepsilon_{nt}^2 \xi_{nt})$ and $\mathbb{E}(v_{nt}^2 \xi_{nt})$ are allowed to be nonzero. The moment conditions (11) or (12) then apply only to (p, s, t) with p, s, and t all distinct.

Fixed effects. We next consider the third moments with fixed effects in the static model. Again, we need to transform the regression equation to eliminate the individual effect. We can eliminate the individual effect also from y_n in the instrument matrix to obtain the analog of (10):

$$\begin{aligned}&\mathbb{E}\left[\tilde{y}_n \otimes (\tilde{y}_n - \tilde{x}_n\,\beta) \otimes x_n\right]\\ &\quad = \mathbb{E}\left[(\tilde{y}_n \otimes I_{T-1} \otimes x_n)(\tilde{y}_n - \tilde{x}_n\,\beta)\right] = 0,\end{aligned} \tag{13}$$

so that $\tilde{Z}_{yx,n} \equiv (\tilde{y}_n \otimes I_{T-1} \otimes x_n)'$ is a valid instrument matrix. However, if the individual effect α_n is independent of ε_n and v_n, then the instrument matrix $\check{Z}_{yx,n} \equiv (y_n \otimes I_{T-1} \otimes x_n)'$ is also valid and gives us more instruments. An analogous analysis shows that, under the assumption that the third moments of ε_n vanish, $\tilde{Z}_{yy,n} \equiv (\tilde{y}_n \otimes \tilde{y}_n \otimes I_{T-1})'$ is a valid instrument matrix, and $\check{Z}_{yy,n} \equiv (y_n \otimes y_n \otimes I_{T-1})'$ is valid if α_n is independent of ε_n and v_n. Finally, $\tilde{Z}_{xx,n} \equiv (I_{T-1} \otimes x_n \otimes x_n)'$ is valid if the third moments of v_n are zero. Again, with multiple regressors, we can replace x_n in the instruments by vec X_n, and we can add R_n to the set of instruments.

In most linear panel data models with fixed effects, the analysis starts by transforming the model by some choice of B to eliminate the individual effects. Nearly always, this is done by taking first differences or by using the "within" transformation. Since panel data often evolve only slowly over time, this step takes out quite a bit of the variation in the data, to the detriment of the precision of the estimates. The striking feature of the analysis here is the presence, in the final result, of the untransformed variables in the instruments, though not in $(\tilde{y}_n - \tilde{x}_n \beta)$. This is analogous to the Arellano–Bond estimator for the dynamic panel data model, where a model in first differences is estimated by IVs in levels.

Dynamic model. When we add the lagged dependent variable as a regressor,

$$\begin{aligned}\mathbb{E}(Z'_{yx,n}u_n) &= \mathbb{E}(y_n \otimes u_n \otimes x_n)\\ &= \mathbb{E}\left[(y_{n,-1}\gamma + \xi_n \beta + \varepsilon_n) \otimes (\varepsilon_n - \upsilon_n \beta) \otimes (\xi_n + \upsilon_n)\right]\\ &= \mathbb{E}\left[y_{n,-1} \otimes (\varepsilon_n - \upsilon_n \beta) \otimes (\xi_n + \upsilon_n)\right]\gamma.\end{aligned}$$

This is zero, because we can write $y_{n,-1}$ as an infinite sum of terms of the form $\varepsilon_{n,-j}$ and $\xi_{n,-i}$ and in the resulting triple products there is always at least one mean-zero factor that is independent of the others. Thus, $Z_{yx,n}$ is still a valid instrument matrix. Analogously, $Z_{yy,n}$ is still a valid instrument matrix under the assumption that the third moments of ε_n (for all triples of time points, including $t \le 0$) vanish and $Z_{xx,n}$ is still a valid instrument matrix under the assumption that the third moments of υ_n vanish. With fixed effects, the instruments derived for the static model are still valid in the dynamic model for the same reason.

The second panel of Table 1 summarizes the IVs and the underlying assumptions for the various cases.

5. Exogenous regressors

In Sections 3 and 4, the role played by the exogenous variables R_n was limited. Their only property used, cf. (6), was contemporaneous lack of correlation with u_n. This suffices to obtain the number of additional moment conditions (i.e., ℓ) equal to the number of additional regression coefficients. However, we can exploit the exogeneity of R_n to obtain more moment conditions, which can be used to help identify and estimate β. We now turn to this.

Motivating examples. Consider the simplest static model with one regressor subject to measurement error and one additional, correctly measured, exogenous regressor:

$$y_{nt} = \xi_{nt}\beta + r_{nt}\delta + \varepsilon_{nt}. \quad (14)$$

The assumption that r_n, ε_n, and υ_n are mutually independent implies that

$$\begin{aligned}&\mathbb{E}[(r_{n,t-1}, r_{nt})'(y_{nt} - x_{nt}\beta - r_{nt}\delta)]\\ &\quad = \mathbb{E}[(r_{n,t-1}, r_{nt})'(\varepsilon_{nt} - \upsilon_{nt}\beta)] = 0.\end{aligned}$$

Thus, the main criterion for the validity of $(r_{n,t-1}, r_{nt})'$ as instruments is satisfied. Identification of β and γ from this moment condition requires that the coefficient of $r_{n,t-1}$ in the linear projection of x_{nt} (or, equivalently given our assumptions, ξ_{nt}) on $r_{n,t-1}$ and r_{nt} is nonzero. Thus, $r_{n,t-1}$ should be excluded from the equation for y_{nt} but should not be excluded from the equation for x_{nt} and in fact contribute significantly in that equation. It is conceivable that this can be derived from economic theory in some cases, but in many cases, theory will not give such strong contrasting predictions. Even if this assumption is technically correct, it will often be the case that $r_{n,t-1}$ will be a weak instrument (after controlling for r_{nt}).

Now suppose that $\xi_{nt} = r_{nt}\kappa + \omega_{nt}$, with $\mathbb{E}(\omega_{nt} \mid r_{nt}) = 0$ but $\mathbb{E}(\omega_{nt}^2 r_{nt}) \neq 0$. Because we take all variables in deviation of their time-mean, this implies that the relation between ξ_{nt} and r_{nt} is heteroskedastic. Then, if υ_n and ε_n are independent of ξ_n and r_n,

$$\begin{aligned}&\mathbb{E}[r_{nt}(x_{nt} - r_{nt}\kappa)(y_{nt} - x_{nt}\beta - r_{nt}\delta)]\\ &\quad = \mathbb{E}[r_{nt}(\upsilon_{nt} + \omega_{nt})(\varepsilon_{nt} - \upsilon_{nt}\beta)] = 0\end{aligned}$$

and

$$\begin{aligned}&\mathbb{E}[r_{nt}(x_{nt} - r_{nt}\kappa)(x_{nt}, r_{nt})] = \mathbb{E}[r_{nt}(\upsilon_{nt} + \omega_{nt})r_{nt}(\kappa, 1)]\\ &\quad + \mathbb{E}[r_{nt}(\upsilon_{nt} + \omega_{nt})^2(1, 0)]\\ &\quad = \mathbb{E}(\omega_{nt}^2 r_{nt})(1, 0).\end{aligned}$$

This shows that, if κ were known, $r_{nt}(x_{nt} - r_{nt}\kappa)$ would be a valid instrument for x_{nt}, and with r_{nt} as an instrument for itself, would jointly identify β and γ, where in this case the heteroskedasticity of the relation between ξ_{nt} and r_{nt} ensures that the rank condition on the instruments is satisfied. This is a special case of a result due to Lewbel (2012). We generally do not know κ, but can estimate it consistently by OLS of x_{nt} on r_{nt}, and replacing κ by a consistent estimator in the instrument does not affect consistency. We now generalize Lewbel's result to the panel data setting.

Static model. The motivating examples showed that the exogenous variable can be used by itself as an additional instrument at different time points, and also as an instrument in combination with the residual of the regression of x on r. Before allowing more regressors, we study this in full generality still for the same model (14). The exogeneity of r_n implies the moment conditions

$$\mathbb{E}(r_{ns}u_{nt}) = 0 \quad \text{for all } s \text{ and } t, \quad (15)$$

where $u_{nt} = y_{nt} - x_{nt}\beta - r_{nt}\delta = \varepsilon_{nt} - \upsilon_{nt}\beta$. In matrix notation, we can write (15) as $\mathbb{E}(r_n \otimes u_n) = 0$, or $\mathbb{E}[(r_n \otimes I_T)u_n] = 0$. Hence, $Z_n = (r_n \otimes I_T)'$ is a valid set of instruments. Identification of β and δ then requires that

$$J \equiv \mathbb{E}[Z'_n(x_n, r_n)] = (\text{vec}\, \Sigma_{xr}, \text{vec}\, \Sigma_r) \quad (16)$$

has full column rank, where $\Sigma_{xr} \equiv \mathbb{E}(x_n r'_n)$ and $\Sigma_r \equiv \mathbb{E}(r_n r'_n)$. This condition will be fulfilled in most cases, but the asymptotic variance of the estimators of β and δ depends on the degree of collinearity of the two columns of J, and in many cases of empirical relevance this degree will be high, leading to imprecise and unreliable results. In particular, when $x_n = c\, r_n + w_n$ with $\mathbb{E}(r_n w'_n) = 0$, $\Sigma_{xr} = c\,\Sigma_r$ and consequently the rank of J is 1 and the model is not identified from (15). When the relation deviates somewhat from this, the model is identified but the estimators have a large variance in a wide set of reasonable parameter values. This showed up clearly in various simulation exercises that we performed. Hence we do not recommend this seemingly attractive approach.

The presence of an additional regressor can be helpful as soon as the relation between x_n and r_n is more complex, in particular when it is heteroskedastic. We will now elaborate this point, generalizing results from Lewbel (2012) to our setting. Consider the linear projection of ξ_n on r_n,

$$\xi_n = K r_n + \omega_n, \quad (17)$$

where $K \equiv \mathbb{E}(\xi_n r'_n)\left[\mathbb{E}(r_n r'_n)\right]^{-1}$. With $w_n \equiv \upsilon_n + \omega_n$, $x_n = K r_n + w_n$. Now consider the situation where the relation (17) between ξ_n and r_n is heteroskedastic, so that $\mathbb{E}(\omega_n \omega'_n \mid r_n)$ is a function of r_n. We make the slightly stronger assumption that $\mathbb{E}(r_n \otimes \omega_n \otimes \omega_n) \neq 0$. Let

$$h_n \equiv \begin{pmatrix} r_n \otimes u_n \\ r_n \otimes w_n \otimes u_n \end{pmatrix} = Z'_{hn}u_n, \quad (18)$$

where $Z_{hn} = r'_n \otimes (1, w'_n) \otimes I_T$. In view of (15) and the various independence assumptions made (esp. that υ_n is independent of r_n), and the assumption that r_n is in deviation of its mean, $\mathbb{E}(h_n) = 0$. Hence, if w_n were observed, Z_{hn} would be a valid set of instruments that could be used for consistent estimation of the coefficients β and δ. Because w_n is not observed, this cannot be used directly. However, we can replace it by $\hat{w}_n$, which is the vector of OLS residuals from the regression of x_n on r_n. This is a case of "generated instruments", and thus proceeding with $\hat{w}_n$ instead of w_n gives consistent estimators, and the IV standard errors are correct (Wooldridge, 2010 Section 6.1.2), provided that the coefficients would be identified with the hypothetical instruments Z_{hn}.

Identification depends on the rank of the matrix

$$G \equiv \mathbb{E}[Z'_{hn}(x_n, r_n)] = \begin{pmatrix} \text{vec}\, \Sigma_{xr} & \text{vec}\, \Sigma_r \\ q_1 & q_2 \end{pmatrix}, \quad (19)$$

where

$$\begin{aligned} \Sigma_{xr} &= K\Sigma_r \\ q_1 &\equiv \mathbb{E}(r_n \otimes w_n \otimes x_n) \\ &= \mathbb{E}\,(r_n \otimes (v_n + \omega_n) \otimes (Kr_n + v_n + \omega_n)) \\ &= (I_T \otimes I_T \otimes K)\mathbb{E}(r_n \otimes \omega_n \otimes r_n) + \mathbb{E}(r_n \otimes \omega_n \otimes \omega_n) \\ q_2 &\equiv \mathbb{E}(r_n \otimes w_n \otimes r_n) \\ &= \mathbb{E}(r_n \otimes \omega_n \otimes r_n). \end{aligned}$$

If the projection (17) can be strengthened to $\mathbb{E}(\xi_n \mid r_n) = Kr_n$, then $\mathbb{E}(r_n \otimes \omega_n \otimes r_n) = 0$, q_1 simplifies to $\mathbb{E}(r_n \otimes \omega_n \otimes \omega_n)$, and $q_2 = 0$. If $\mathbb{E}(r_n \otimes \omega_n \otimes r_n) \neq 0$, the regression of ξ_n on r_n must be nonlinear and $q_2 \neq 0$.

It is of interest to compare (16) and (19). The two columns in (16) may be highly collinear, which makes it hard to obtain estimators with decent small-sample properties from an additional regressor. In (19), these two columns are supplemented by q_1 and q_2, respectively, which may decrease collinearity and hence the problem in obtaining satisfactory estimators. Linearity of the regression of x_n on r_n is helpful under heteroskedasticity, because it leads to $q_2 = 0$. If nonlinearity is more important than heteroskedasticity, the additional moment conditions are less helpful, although a situation in which $\mathbb{E}(r_n \otimes r_n \otimes \omega_n) \neq 0$ and $\mathbb{E}(r_n \otimes \omega_n \otimes \omega_n) = 0$ is unlikely to be approximately met, so the additional moment conditions still add value.

Summarizing, this procedure consists of the following steps:

1. For $t = 1, \ldots, T$, regress x_{nt} on r_{ns}, $s = 1, \ldots, T$. So these are T separate regressions with in each regression the values of the additional regressor at *all* T time points as regressors (T regressions with T regressors each).
2. For each of these T regressions, compute the residual $\hat{w}_{nt}$.
3. Create T sets of instruments. Instrument set t consists of r_{ns}, $s = 1, \ldots, T$ and all products $r_{ns}\hat{w}_{np}$, $s, p = 1, \ldots, T$, as instruments for the tth time point and zeros for the other time points.
4. With the instruments defined like this, compute the panel IV estimator and use appropriate GMM standard errors.

When there are multiple regressors with or without measurement error, the moments (15) are still valid, with r_{ns} now a vector, and $u_{nt} = y_{nt} - x'_{nt}\beta - r'_{nt}\delta = \varepsilon_{nt} - v'_{nt}\beta$, with x_{nt}, v_{nt}, β, and δ now also vectors. In (18), r_n and w_n are then replaced by vec R_n and vec W_n, respectively, where W_n is the matrix in which the (t, j) element is $x_{nt,j}$ minus its projection on vec R_n. Step 1 of the procedure above then consists of $T \cdot k$ regressions of $x_{nt,j}$ on vec R_n, that is, with $T \cdot \ell$ regressors each. Step 2 gives $T \cdot k$ residuals $\hat{w}_{nt,j}$ for each individual. In step 3, instrument set t consists of $r_{ns,i}$, and $r_{ns,i}\hat{w}_{np,j}$, $s, p = 1, \ldots, T$, $i = 1, \ldots, \ell$, $j = 1, \ldots, k$.

Heteroskedasticity. Because y_n is not included in the instruments derived in this section, heteroskedasticity of ε_n has no effect on the validity of these instruments. However, the same does not hold for the measurement errors. The validity of the instruments depends on the assumption that $\mathbb{E}(r_n \otimes v_n \otimes v_n) = 0$. Thus, the regression of ξ_n on r_n is required to be heteroskedastic, but the measurement error must be homoskedastic with respect to r_n. If this assumption is violated, this approach is not valid without adaptation. If the assumption is violated for only a subset of the elements, we can still use a subset of the instruments for identification. For example, we may be willing to assume that $\mathbb{E}(r_n \otimes v_{nt} \otimes v_{ns}) = 0$ for $s \neq t$ but not for $s = t$. The offending elements should then be removed from (18). The procedure then is still straightforward and very similar to the one described above, but the notation becomes a bit cumbersome.

Fixed effects. For including fixed effects, the key element is that in (18), u_n is replaced by $\tilde{u}_n = \tilde{y}_n - \tilde{x}_n\beta - \tilde{r}_n\delta$. This eliminates the individual effect, while after this transformation, the analog of (18) still holds. Hence, we then estimate the regressions with the transformed variables, but with the same instruments. This extends immediately to the generalization with more regressors.

Dynamic model. The IVs are based on x_n and r_n only and do not involve y_n. Hence, their validity is not affected when regressors are added to the model, even if they include the lagged dependent variable. Specifically, u_n is now redefined as $u_n = y_n - y_{n,-1}\gamma - x_n\beta - r_n\delta = \varepsilon_n - v_n\beta$ and (18) still holds. The rank condition for identification now applies to the matrix

$$G_D \equiv \mathbb{E}[Z'_{hn}(y_{n,-1}, x_n, r_n)] = \begin{pmatrix} \text{vec}\, \Sigma_{yr} & \text{vec}\, \Sigma_{xr} & \text{vec}\, \Sigma_r \\ q_0 & q_1 & q_2 \end{pmatrix}, \quad (20)$$

where $\Sigma_{yr} \equiv \mathbb{E}(y_{n,-1}r'_n)$ and $q_0 \equiv \mathbb{E}(r_n \otimes w_n \otimes y_{n,-1})$. In general, G_D will have full column rank, ensuring identification of the regression coefficients in the dynamic model. Analogously, the instruments in the dynamic model with fixed effects are the same as in the static model with fixed effects.

The third panel of Table 1 lists the IVs and the underlying assumptions for the various cases.

6. Simulations

To get an impression of the performance of the various estimators proposed in the previous sections, we conducted some simulations. We generated data largely following a well-known setup originally due to Nerlove (1971) and subsequently used by various other researchers. This setup has

$$y_{nt} = \alpha_n + \xi_{nt}\beta + \varepsilon_{nt},$$

with $\varepsilon_{nt} \sim N(0, \sigma^2_\varepsilon)$ and $\alpha_n \sim N(0, \sigma^2_\alpha)$. We introduce measurement errors by

$$x_{nt} = \xi_{nt} + v_{nt},$$

with $v_{nt} \sim N(0, \sigma^2_v)$. We let $\sigma^2_\alpha = 0.7$, $\beta = 1$, $\sigma^2_\varepsilon = 2$, and $\sigma^2_v = 1$. The ξ_{nt} are generated according to

$$\xi_{nt} = 0.5\xi_{n,t-1} + \zeta_{nt},$$

with $\zeta_{nt} \sim \sqrt{\tfrac{4}{3}}\,\chi^2_1$ and $\xi_{n0} = \sqrt{\tfrac{4}{3}}\,\zeta_{n0}$. This choice of ζ_{nt} implies that the third moment of ξ_{nt} is nonzero, which is exploited in the estimators based on third moments. For $N = 100$, 200, 500, and 1000, we generated 1000 data sets, all with $T = 5$. In each sample, y_{nt} and x_{nt} are centered by subtracting their sample averages across n before further estimation.

Below, we employ estimators based on the moment conditions derived in Sections 3–5. Unless stated otherwise, all results are based on the optimally weighted GMM estimator, that is, the

Table 2
Strength of instruments.

	cov. restr.				*3rd mom.*			
	RE		FE		RE		FE	
N	$\bar{R}^2$	*F*	$\bar{R}^2$	*F*	$\bar{R}^2$	*F*	$\bar{R}^2$	F
100	0.06	3.6	0.07	4.71	0.43	12.06	0.25	7.87
200	0.05	5.03	0.06	7.51	0.43	22.77	0.24	13.92
500	0.04	9.21	0.06	15.64	0.43	53.76	0.23	31.45
1000	0.04	16.01	0.05	29.36	0.42	103.93	0.23	60.28

estimator based on (1) with $W = (\sum_n \hat{u}_n \hat{u}_n')^{-1}$, with $\hat{u}_n$ based on the initial GMM estimator with $W = I$. Throughout, we take the matrix B' equal to the centering matrix Q_T with the first row left out. We use clustered standard errors that are robust to time-series correlation and heteroskedasticity. All simulations have been done in R version 3.2.0.[2]

Covariance restrictions. The number of instruments from the structure on Σ_ε depends on whether we consider random effects (RE) or fixed effects (FE) estimation. With RE, the individual effect α_n is subsumed in the error term ε_{nt}. These terms have covariance matrix Σ_ε. Its structure has already been illustrated in Section 3. We leave out the moment conditions that do not involve β,[3] and thus use 13 instruments for RE and 8 for FE. We refer to the online appendix with supplementary material for the exact set of moment conditions used in the simulations.

Columns 2–5 of Table 2 give an impression of the strength of these instruments. The adjusted R^2 of the regression of x on the instruments, denoted by $\bar{R}^2$, is typically low, which is not unusual for panel data. The F statistics indicate that the instruments tend to be weak for $N = 100$ and 200, and strong for $N = 500$ and 1000, if we take $F < 10$ as the threshold for weak instruments (e.g., Stock et al., 2002).

The simulation results for covariance matrix restrictions are given in Table 3, for RE (column 2) and FE (column 3). This table reports the average value of the GMM estimator $\hat{\beta}$ over the 1000 simulation runs ("avg. $\hat{\beta}$"), the sample standard deviation of $\hat{\beta}$ over the replications ("sample $\sigma(\hat{\beta})$"), and the average (formula-based) standard error of $\hat{\beta}$ according to the usual asymptotic theory ("avg. $\hat{\sigma}(\hat{\beta})$"). For sample sizes larger than $N = 200$, the bias of $\hat{\beta}$ is negligible for both RE and FE estimators. A comparison of the average formula-based standard error and the sample standard deviation computed over the replications sheds light on the accuracy of the asymptotic standard error as an approximation of the finite-sample standard deviation. The average formula-based standard error of both the RE and FE estimators is smaller than the sample standard deviation of $\hat{\beta}$ over the replications, but for sample sizes of 500 and larger, the difference is small.

Columns 2–3 of Table 4 report the corresponding rejection rates (in percentages) of the z-tests over the 1000 replications for RE and FE, where the null hypothesis $H_0 : \beta = 1$ is rejected if and only if $|\hat{\beta} - 1|/\hat{\sigma}(\hat{\beta}) > 1.96$. Thus, the rejection rates should be close to 5%. The standard error used in the z-tests is either the formula-based standard error ("$\hat{\sigma}(\hat{\beta})$"), or the sample standard deviation as measured over the replications ("sample $\sigma(\hat{\beta})$"). The latter is, of course, not a test that can be used in practice, but it serves here to illustrate the impact of the less accurate standard errors, as opposed to the unbiased estimators. Table 4 shows that the rejection rates of the z-tests are close to 5% when the sample standard deviation is used. Especially for $N \leq 200$, the rejection rates exceed the nominal level when the formula-based standard error is used.

2 The R code is available upon request.

3 In preliminary simulations, we found that including these moment conditions leads to worse finite-sample properties.

Table 3
Covariance restrictions and 3rd moments: bias and variance.

	cov. restr.		*3rd mom.*					
			W_{OPT}		W_{2SLS}		$W = I$	
	RE	FE	RE	FE	RE	FE	RE	FE
***N* = 100**								
avg. $\hat{\beta}$	96	97	96	97	98	97	100	99
sample $\sigma(\hat{\beta})$	174	203	75	106	72	100	96	116
avg. $\hat{\sigma}(\beta)$	132	164	27	47	71	92	84	92
***N* = 200**								
avg. $\hat{\beta}$	98	99	97	98	99	99	100	100
sample $\sigma(\hat{\beta})$	130	142	53	74	49	70	66	82
avg. $\hat{\sigma}(\beta)$	112	131	27	43	52	68	64	71
***N* = 500**								
avg. $\hat{\beta}$	99	100	98	99	100	100	100	100
sample $\sigma(\hat{\beta})$	87	91	35	46	33	47	43	51
avg. $\hat{\sigma}(\beta)$	81	90	23	33	34	45	43	47
***N* = 1000**								
avg. $\hat{\beta}$	100	100	99	99	100	100	100	100
sample $\sigma(\hat{\beta})$	62	65	24	31	23	32	28	35
avg. $\hat{\sigma}(\beta)$	60	65	19	26	25	33	31	34

Notes: To facilitate reading, avg. $\hat{\beta}$ has been multiplied by 100, whereas sample $\sigma(\hat{\beta})$ and avg. $\hat{\sigma}(\hat{\beta})$ have been multiplied by 1000.

A panel bootstrap based on the recentered moment conditions can be used to estimate standard deviations that represent a second-order improvement relative to the formula-based asymptotic standard errors (Hall and Horowitz, 2012). The rejection rates of the z-tests are substantially improved when the bootstrap standard errors are used instead of the formula-based standard errors. The use of bootstrap standard errors is particularly useful for smaller values of N, when the downward bias in the formula-based standard errors is relatively large. The improvement in rejection rates thanks to the bootstrap is illustrated in columns 2–3 of Table 6 for $N = 100$, where we display the rejection rates based on the sample standard deviation as measured over the replications ("sample $\sigma(\hat{\beta})$"), the formula-based standard error ("$\hat{\sigma}(\hat{\beta})$"), and the bootstrap (bootstrap $\hat{\sigma}(\hat{\beta})$). The bootstrap results in a considerable improvement of the rejection rates relative to the formula-based standard error.

In the cases where we have weak instruments according to the F statistic ($N = 100$ or 200), our simulation results do not improve when we resort to continuous updating estimation (CUE) as proposed by Hansen et al. (1996) or regularized CUE (RCUE) developed by Hausman et al. (2011). CUE is known as the GMM-equivalent of LIML (Hausman et al., 2011). Unlike LIML, CUE is consistent in the presence of heteroskedasticity and autocorrelation if based on a heteroskedasticity and autocorrelation robust weighting matrix. However, it exhibits relatively large dispersion and suffers from the no-moment problem. We have therefore also implemented RCUE. Like CUE, RCUE has a reduced bias relatively to GMM, especially in the presence of weak instruments. Furthermore, it features less dispersion than CUE and does not suffer from the no-moment problem (Hausman et al., 2011). Our simulations confirm that CUE results in less bias than GMM estimation but increased dispersion, while RCUE shows more bias but less dispersion. Neither of the two

Table 4
Covariance restrictions and 3rd moments: rejection rates (in %).

	cov. restr.		*3rd mom.*					
			W_{OPT}		W_{2SLS}		$W = I$	
	RE	FE	RE	FE	RE	FE	RE	FE
***N* = 100**								
% rejection rate (sample $\sigma(\hat{\beta})$)	5	6	8	6	6	7	5	5
% rejection rate ($\hat{\sigma}(\beta)$)	15	11	55	45	8	9	13	17
***N* = 200**								
% rejection rate (sample $\sigma(\hat{\beta})$)	5	6	9	7	6	5	5	5
% rejection rate ($\hat{\sigma}(\beta)$)	9	7	42	29	6	7	9	11
***N* = 500**								
% rejection rate (sample $\sigma(\hat{\beta})$)	5	5	9	5	5	5	5	5
% rejection rate ($\hat{\sigma}(\beta)$)	6	5	28	18	5	7	6	8
***N* = 1000**								
% rejection rate (sample $\sigma(\hat{\beta})$)	6	6	8	6	6	6	5	6
% rejection rate ($\hat{\sigma}(\beta)$)	6	5	18	11	5	5	4	6

alternative estimators uniformly outperforms the standard GMM estimators. More detailed results are available in the appendix with supplementary material.

Nonzero third moments. We now exploit the third moments of the data. Again we leave out the moment conditions that do not involve β, so we use 35 IVs for RE and 20 for FE. The online appendix with supplementary material describes the exact set of moment conditions used in the simulations.

Columns 6–9 of Table 2 give an impression of the strength of these instruments. The $\bar{R}^2$s and F statistics are much higher than the ones for the covariance restrictions in columns 2–5. Yet we might still have weak instruments for $N = 100$ according to the F statistic.

The simulation results for the third-moment restrictions are given in columns 4–9 of Table 3. For the GMM estimator that exploits third moments the weight matrix turns out crucial. We consider three different weight matrices: the asymptotically optimal weight matrix W_{OPT} (columns 4–5), the 2SLS weight matrix W_{2SLS} (columns 6–7), and the identity matrix I (columns 8–9).

As expected, the estimator based on $W = I$ has a larger variance than the other two. For the 2SLS and identity-weighted estimators, the average formula-based standard error tends to be relatively close to the sample standard deviation measured over the replications. For the optimally-weighted GMM estimator the difference between the two is much larger.

The combination of some bias in the estimator and a large downward bias in the standard error results in rejection rates of the two-sided z-test for the null hypothesis $H_0 : \beta = 1$ that are far too high for the GMM estimator based on W_{OPT}, as shown in columns 4 and 5 of Table 4. If the formula-based standard errors are replaced by the sample standard deviation of $\hat{\beta}$ across replications, the rejection rates improve substantially. Columns 6–9 show that the test results for the non-optimally weighted GMM estimators tend to be better than those based on the optimally-weighted GMM estimators, especially when the formula-based standard errors are used.

Again we can use the bootstrap to obtain more accurate estimates of the standard deviation of $\hat{\beta}$. The resulting bootstrap standard errors improve the rejection rates of the z-tests based on the formula-based standard errors. This is illustrated in columns 4–9 of Table 6. Note that, as mentioned in Section 6, the test that uses the sample standard deviation of β across replications is not available in practice. It should be viewed as a hypothetical test that would be obtained if the correct standard error were known.

For $N = 100$ we have weak instruments according to the F statistic. Again our simulation results do not improve when we resort to (R)CUE. More detailed results are available in the appendix with supplementary material.

The standard deviations of the third-moment estimator are much smaller than for the estimators based on covariance restrictions. The third-moment GMM estimator with $W = I$ has relatively little bias, resulting in a considerably smaller (in a relative sense) mean squared error than the covariance restriction estimator.

The non-optimally weighted GMM estimators perform better in terms of bias and rejection rates than the GMM estimator based on the asymptotically optimal weight matrix. If the asymptotic distribution is a reasonable approximation of the exact distribution, optimal weighting is to be preferred. However, there is ample evidence that, especially when the sample used is not too large, the approximation can be poor. GMM estimators may then be severely biased and inference based on them can be highly unreliable. One possible cause is the imprecision of the weight matrix based on higher-order moments (Mooijaart and Satorra, 2012). Another cause is due to the fact that the data are used twice, to construct both the instruments and the weight matrix, inducing a correlation between the two. This correlation leads to a negative bias in the case of covariance structures, as shown by Altonji and Segal (1996). See also the discussion in Wansbeek and Meijer (2000, p. 274).

Although the optimal moment conditions involve x_n (levels) instead of $\tilde{x}_n$ (within-differences), our simulations are based on $\tilde{x}_n$; see also the appendix with supplementary material. Using x_n results in additional moment conditions, which only improve the rejection rates of the GMM estimator for $W = I$, while they worsen the rejection rates for $W = W_{\text{OPT}}$ and $W = W_{\text{2SLS}}$.[4] Again the aforementioned imprecision of the weight matrix based on higher-order moments and the correlation between the instruments and the weight matrix are likely to account for these results.

Exogenous regressor. To study the estimator that exploits the presence of an additional regressor, we start by simulating the regressor r_{nt} analogous to ξ_{nt} in the previous simulations: $e_{nt} \sim \sqrt{\frac{4}{3}}\,\chi_1^2$ for $t = 0, \ldots, T$ $(T = 5)$, $r_{n0} = \sqrt{\frac{4}{3}}\, e_{n0}$, and

$$r_{nt} = 0.5 r_{n,t-1} + e_{nt}.$$

We then compute $\omega_{nt} = r_{nt}\zeta_{nt}$, where $\zeta_{nt} \sim N(0, \sigma_\zeta^2)$ (i.i.d.), so that $\mathbb{E}(r_{nt}\omega_{ns}) = 0$ for all t and s, but $\mathbb{E}(r_{nt}\omega_{nt}^2) = \mathbb{E}(r_{nt}^3)\mathbb{E}(\zeta_{nt}^2) \neq 0$, so that $\mathbb{E}(r_n \otimes \omega_n \otimes \omega_n) \neq 0$. The regressor ξ_{nt} is then generated according to

$$\xi_{nt} = \kappa_1 r_{nt} + \kappa_2 r_{n,t-1} + \omega_{nt}.$$

This satisfies the setup in Section 5 with $K \neq c\, I_T$, but $q_2 = 0$ and $q_1 = \mathbb{E}(r_n \otimes \omega_n \otimes \omega_n) \neq 0$. The model is completed by the system

$$y_{nt} = \alpha_n + \xi_{nt}\,\beta + r_{nt}\gamma + \varepsilon_{nt}$$
$$x_{nt} = \xi_{nt} + v_{nt},$$

[4] These results are available upon request.

Table 5
Exogenous regressors: bias, variance, and rejection rates.

	RE		FE		RE		FE	
	β	γ	β	γ	β	γ	β	γ
	$\boldsymbol{N = 100}$				$\boldsymbol{N = 500}$			
avg. $\hat{\beta}$	97	103	97	102	100	100	100	100
sample $\sigma(\hat{\beta})$	31	56	32	61	16	26	17	28
% rejections	13	6	13	6	6	4	6	4
avg. $\hat{\sigma}(\hat{\beta})$	34	59	36	64	16	26	17	29
% rejections	18	10	18	7	7	5	8	6
	$\boldsymbol{N = 200}$				$\boldsymbol{N = 1000}$			
avg. $\hat{\beta}$	98	101	98	101	100	100	100	100
sample $\sigma(\hat{\beta})$	24	41	25	44	12	18	12	20
% rejections	10	6	9	6	5	6	6	5
avg. $\hat{\sigma}(\hat{\beta})$	24	42	25	45	12	19	13	21
% rejections	11	8	10	6	6	8	8	7

Notes: To facilitate reading, avg. $\hat{\beta}$ has been multiplied by 100, whereas sample $\sigma(\hat{\beta})$ and avg. $\hat{\sigma}(\hat{\beta})$ have been multiplied by 1000. The rejection rates are in %.

with again $\varepsilon_{nt} \sim N(0, \sigma_\varepsilon^2)$, $\alpha_n \sim N(0, \sigma_\alpha^2)$, and $v_{nt} \sim N(0, \sigma_v^2)$. As in the previous simulation, we choose $\sigma_\alpha^2 = 0.7$, $\beta = 1$, $\sigma_\varepsilon^2 = 2$, and $\sigma_v^2 = 1$. The additional parameters are $\sigma_\zeta^2 = 1$, $\kappa_1 = \kappa_2 = 1/\sqrt{3} = 0.577$, and $\gamma = 1$.

We have 150 moment conditions for RE and 120 for FE. Table 5 shows the simulation results for the optimally-weighted GMM estimator based on the restrictions that follow from the presence of an exogenous regressor. This table reports the average values of $\hat{\beta}$ and $\hat{\gamma}$ over the replications, the sample standard deviations over the replications, the average formula-based standard errors, and the rejection rates corresponding to the two z-tests $H_0 : \beta = 1$ and $H_0 : \gamma = 1$ (for both types of standard errors). The bias in the estimators is small, particularly for $N = 500$ and $N = 1,000$. Throughout, the average formula-based standard error is close to the sample standard deviation over the replications. The rejection rates are approximately nominal for the larger sample sizes. Again the bootstrap can be used to improve the rejection rates of the z-tests based on the formula-based standard errors for small values of N, which is illustrated in the third panel of Table 6.

7. Discussion

We have presented three approaches to find instruments in a linear panel data model with measurement error. We avoided the hard-to-justify assumptions on the measurement error structure and replaced them with assumptions that researchers may be more comfortable with. Specifically, we consider restrictions on the covariance matrix of the equation errors (Σ_ε), exploit third moments, and use exogenous regressors. For each of these cases, we derive simple IV estimators. The simulation results suggest that our three approaches work well, at least for the particular settings chosen in our simulation study, and subject to some qualifications regarding the implementation. We thus have expanded the toolkit of the applied researcher. Yet there are many openings for further research. We mention a number of them.

Nonlinear moment conditions. As to Σ_ε, we considered linear restrictions only. A researcher may be willing to restrict Σ_ε in a nonlinear way, for example, by imposing some ARMA structure. Then, the elements of the error covariance matrix are functionally dependent on a few underlying parameters, η, say, in a nonlinear way. One may proceed by using a consistent but inefficient estimate of η, which can often easily be constructed, and improve on it by linearized GMM, cf. Wansbeek and Meijer (2000, Section 9.3).

Robust estimation. As to the approach using third moments, the presence of outliers will often have a negative impact on estimator quality. This also defines a direction for further research, by bringing the literature on robust estimation in panel data models to bear on measurement error, cf. Wagenvoort and Waldmann (2002) and Bramati and Croux (2007).

Nonclassical measurement error. Our results were based on the assumed independence of regressor and measurement error, "classical" measurement error, contrasting with "non-classical measurement error", which is the case where the two are not independent. The importance of this was argued by Bound and Krueger (1991), who compared self-reported wage data with (supposedly true) administrative data and found the measurement errors in self-reported data to be negatively correlated with the true data. Bound et al. (2001) provide an overview of this phenomenon of "mean reversion" in general, and Kim and Solon (2005) discuss it in the panel data setting.

Recently, Meijer et al. (2014) discussed consistent estimation with nonclassical measurement error. The first approach in the current paper, exploiting restrictions on Σ_ε, offers a useful framework. When the nonclassical measurement error is not restricted and ξ_{nt} and v_{ns} can be correlated freely for all s and t, there is an even larger identification problem than in the classical case. However, linear restrictions on these correlations, like exclusion restrictions, can be exploited in much the same way as restrictions on Σ_ε. Elaboration of this approach and discussion of nonclassical measurement error for the other two approaches is a topic for further research.

Nonclassical measurement error is not a problem when the Berkson model applies. Adapted to the panel data context, the key equation of this model is

$$\xi_{nt} = x_{nt} + v_{nt},$$

with $\mathbb{E}(v_{nt} \mid x_{nt}) = 0$. Thus, ξ and x switch roles. Hyslop and Imbens (2001) argued that in many cases, this model may be more appropriate than the classical measurement error model. In this model, $\mathbb{E}(y_{nt} \mid x_{nt}) = x'_{nt}\beta$, and thus OLS is consistent.

Moment selection. Another issue is the selection of moments. Our three approaches for finding instruments are not mutually exclusive and can, under the appropriate assumptions, be combined. For each of the sources, especially third moments, the number of IVs can be large, which may negatively affect the finite-sample performance of the estimators. Furthermore, some of the moment conditions might not be valid. Methods for choosing among a set of IVs then become relevant. We can distinguish two different selection methods. The first method has the goal to separate valid moment conditions from invalid ones, while the second approach

Table 6
Rejection rates (in %) for $\boldsymbol{N = 100}$.

	cov. restr.		*3rd mom.*						*ex. regr.*			
			W_{opt}		W_{2SLS}		$W = I$		β		γ	
	RE	FE	RE	FE	RE	FE	RE	FE	RE	FE	RE	FE
% rejections (sample $\sigma(\hat{\beta})$)	5	6	8	6	6	7	5	5	13	13	6	6
% rejections ($\hat{\sigma}(\hat{\beta})$)	15	11	55	45	8	9	13	17	18	18	10	7
% rejections (bootstrap $\hat{\sigma}(\hat{\beta})$)	5	4	9	8	7	8	7	7	15	16	8	6

aims to eliminate redundant conditions, that is, conditions that do not contribute to a reduction in the variance of the GMM estimator (Okui, 2009). Various consistent selection procedures have been proposed in the literature, including methods that add a penalty term to the usual J-statistic for overidentification (Andrews, 1999). A worthwhile topic for further research is to extend, to a panel data environment, the literature on instrument selection like the lasso, cf. Belloni et al. (2012).

An issue related to moment selection is the impact of selecting moment conditions on post-selection inference such as hypothesis testing. Okui (2009) considers the linear dynamic panel data model. He proves that, under certain conditions, a specific form of moment selection does not affect the asymptotic distribution of the GMM-based persistence parameter if the moment choice set is finite and independent of the sample size. In several other cases, where such theoretical results are not available, Monte Carlo simulations have provided favorable results for the conventional asymptotic confidence interval that simply ignores that any moment selection has taken place; see, for example, Donald and Newey (2001), Hall and Peixe (2003), Hall et al. (2007) and Okui (2009). In the more general context of post-selection estimators, however, Leeb and Pötscher (2005) warn that even a consistent model selection procedure does not guarantee the correctness of post-selection inference that ignores the pre-testing phase by assuming that the true model is known in advance. This means that the occasional evidence based on Monte Carlo simulations should be interpreted with some caution and that further research into the asymptotic distribution of post-selection estimators is required.

Weak instruments. As an alternative, one might bypass the choice among non-redundant moment conditions and use all valid instruments. One might then end up with many weak instruments, a topic whose relevance was forcefully put on the agenda by Angrist and Krueger (1991). Bekker (1994) revitalized limited-information maximum likelihood (LIML) in this context and extended LIML to ensure many-instruments consistency. This has been generalized to the heteroskedastic case; see, e.g., Hansen et al. (1996) and Hausman et al. (2011) for the (R)CUE estimators that were also used in Section 6, or Bekker and Crudu (2015) for a jackknife estimator.

Large-T asymptotics. A final topic for further research concerns the dimensions of the panel. Our findings are based on $N \gg T$ justifying large-N asymptotics. This was the format in the classical panel data literature, but there is increasing attention to panel data where N and T are of a different relative size, asking for different asymptotics. This may lead to different estimators for the measurement error case, but a first step would be to investigate the asymptotic behavior, for N fixed and $T \to \infty$ or $N \to \infty$ and $T \to \infty$ jointly in some way, of the estimators derived above.

Acknowledgments

We are indebted to Ramses Abul Naga, Paul Bekker, participants at the Econometric Society North American Summer Meeting 2013, and four anonymous referees for their very helpful comments. 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).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.jeconom.2017.06.003.

References

Aasness, J., Biørn, E., Skjerpen, T., 2003. Distribution of preferences and measurement errors in a disaggregated expenditure system. Econom. J. 6, 374–400.
Ahn, S.C., Schmidt, P., 1995. Efficient estimation of models for dynamic panel data. J. Econometrics 68, 5–27.
Ahn, S.C., Schmidt, P., 1997. Efficient estimation of dynamic panel data models: alternative assumptions and simplified estimation. J. Econometrics 76, 309–321.
Altonji, J.G., Segal, L.M., 1996. Small-sample bias in GMM estimation of covariance structures. J. Bus. Econom. Statist. 14, 353–366.
Andrews, D.W.K., 1999. Consistent moment selection procedures for generalized method of moments estimation. Econometrica 67, 543–563.
Angrist, J.D., Krueger, A.B., 1991. Does compulsory school attendance affect schooling and earnings? Q. J. Econ. 106, 979–1014.
Bekker, P.A., 1994. Alternative approximations to the distributions of instrumental variable estimators. Econometrica 62, 657–681.
Bekker, P.A., Crudu, F., 2015. Jackknife instrumental variable estimation with heteroskedasticity. J. Econometrics 185, 332–342.
Belloni, A., Chen, D., Chernozhukov, V., Hansen, C., 2012. Sparse models and methods for optimal instruments with an application to eminent domain. Econometrica 80, 2369–2429.
Biørn, E., 2000. Panel data with measurement errors: instrumental variables and GMM procedures combining levels and differences. Econometric Rev. 19, 391–424.
Biørn, E., 2003. Handling the measurement error problem by means of panel data: moment methods applied on firm data. In: Stigum, B. (Ed.), Econometrics and the Philosophy of Economics. Princeton University Press, Princeton.
Biørn, E., 2015. Panel data dynamics with mis-measured variables: Modeling and GMM estimation. Empir. Econ. 48, 517–535.
Biørn, E., Klette, J.T., 1998. Panel data with errors-in-variables: essential and redundant orthogonality conditions in GMM estimation. Econom. Lett. 59, 275–282.
Biørn, E., Klette, J.T., 1999. The labour input response to permanent changes in output: an errors-in-variables analysis based on panel data. J. Ind. Econ. 47, 451–476.
Biørn, E., Krishnakumar, J., 2008. Measurement errors and simultaneity. In: Mátyás, L., Sevestre, P. (Eds.), The Econometrics of Panel Data. Springer, New York.
Bound, J., Brown, C., Mathiowetz, N., 2001. Measurement error in survey data. In: Heckman, J.J., Leamer, E.E. (Eds.), Handbook of Econometrics 5. North-Holland, Amsterdam, pp. 3705–3843.
Bound, J., Krueger, A.B., 1991. The extent of measurement error in longitudinal earnings data: do two wrongs make a right? J. Labor Econ. 9, 1–24.
Bramati, M.C., Croux, C., 2007. Robust estimators for the fixed effects panel data model. Econom. J. 10, 521–540.
Cameron, A.C., Trivedi, P.K., 2005. Microeconometrics. Cambridge University Press, Cambridge, UK.
Dagenais, M.G., Dagenais, D.L., 1997. Higher moment estimators for linear regression models with errors in the variables. J. Econometrics 76, 193–221.
Donald, S.G., Newey, W.K., 2001. Choosing the number of instruments. Econometrica 69, 1161–1191.
Erickson, T., Whited, T.M., 2002. Two-step GMM estimation of the errors-in-variables model using high-order moments. Econometric Theory 18, 776–799.
Geary, R.C., 1942. Inherent relations between random variables. Proc. R. Ir. Acad. A 47, 63–67.
Gospodinov, N., Komunjer, I., Ng, S., 2014. Minimum Distance Estimation of Dynamic Models with Errors-In-Variables. Working Paper.
Griliches, Z., Hausman, J.A., 1986. Errors in variables in panel data. J. Econometrics 31, 93–118.
Hall, P., Horowitz, J.L., 2012. Bootstrap critical values for tests based on generalized-method-of-moments estimators. Econometrica 64, 891–916.
Hall, A.R, Inoue, A., Jana, K., Shin, C., 2007. Information in generalized method of moments estimation and entropy-based moment selection. J. Econometrics 138, 488–512.
Hall, A.R, Peixe, F.P.M., 2003. A consistent method for the selection of relevant instruments. Econometric Rev. 22, 269–287.
Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.
Hansen, L.P., Heaton, J., Yaron, A., 1996. Finite-Sample properties of some alternative GMM estimators. J. Bus. Econ. Stat. 14, 262–280.
Hansen, K.T., Heckman, J.J., Mullen, K.J., 2004. The effect of schooling and ability on achievement test scores. J. Econometrics 121, 39–98.
Hausman, J., Lewis, R., Menzel, K., Newey, W., 2011. Properties of the CUE estimator and a modification with moments. J. Econometrics (ISSN: 0304-4076) 165, 45–57.
Hyslop, D.R., Imbens, G.W., 2001. Bias from classical and other forms of measurement error. J. Bus. Econ. Stat. 19, 475–481.
Kim, B., Solon, G., 2005. Implications of mean-reverting measurement error for longitudinal studies of wages and employment. Rev. Econ. Stat. 87, 193–196.
Kotlarski, I., 1967. On characterizing the gamma and the normal distribution. Pacific J. Math. 20, 69–76.

Lee, N., Moon, H.R., Zhou, Q., 2014. Many IVs Estimation of Dynamic Panel Regression Models with Measurement Error. Working paper.

Leeb, H., Pötscher, B.M., 2005. Model selection and inference: Facts and fiction. Econometric Theory 21, 21–59.

Lewbel, A., 1996. Demand estimation with expenditure measurement errors on the left and right hand side. Rev. Econ. Stat. 78, 718–725.

Lewbel, A., 1997. Constructing instruments for regressions with measurement error when no additional data are available, with an application to patents and R&D. Econometrica 65, 1201–1213.

Lewbel, A., 2012. Using heteroskedasticity to identify and estimate mismeasured and endogenous regressor models. J. Bus. Econ. Stat. 30, 67–80.

Meijer, E., Spierdijk, L., Wansbeek, T.J., 2013. Measurement error in the linear dynamic panel data model. In: Sutradhar, B.C. (Ed.), Longitudinal Data Analysis Subject To Measurement Error, Missing Values and/Or Outliers. Springer, New York, pp. 77–92.

Meijer, E., Spierdijk, L., Wansbeek, T.J., 2014. Measurement error in panel data. In: Baltagi, B.H. (Ed.), The Oxford Handbook of Panel Data Econometrics. Oxford University Press, Oxford, UK, pp. 325–362.

Mooijaart, A., Satorra, A., 2012. Moment testing for interaction terms in structural equation modeling. Psychometrika 77, 65–84.

Nerlove, M., 1971. Further evidence on the estimation of dynamic economic relations from a time series of cross sections. Econometrica 39, 359–382.

Okui, R., 2009. The optimal choice of moments in dynamic panel data models. J. Econometrics 151, 1–16.

Pal, M., 1980. Consistent moment estimators of regression coefficients in the presence of errors in variables. J. Econometrics 14, 349–364.

Qian, H., Schmidt, P., 1999. Improved instrumental variables and generalized method of moments estimators. J. Econometrics 91, 145–169.

Shao, J., Xiao, Z., Xu, R., 2011. Estimation with unbalanced panel data having covariate measurement error. J. Statist. Plann. Inference 141, 800–808.

Stock, J.H., Wright, J.H., Yogo, M., 2002. A survey of weak instruments and weak identification in generalized method of moments. J. Bus. Econ. Stat. 20, 518–529.

Van Montfort, K., Mooijaart, A., De Leeuw, J., 1987. Regression with error in variables: Estimators based on third order moments. Stat. Neerl. 41, 223–239.

Van Montfort, K., Mooijaart, A., De Leeuw, J., 1989. Estimation of regression coefficients with the help of characteristic functions. J. Econometrics 41, 267–278.

Wagenvoort, R., Waldmann, R., 2002. On B-robust instrumental variable estimation of the linear model with panel data. J. Econometrics 106, 297–324.

Wansbeek, T.J., 2001. GMM estimation in panel data models with measurement error. J. Econometrics 104, 259–268.

Wansbeek, T.J., Koning, R.H., 1991. Measurement error and panel data. Stat. Neerl. 45, 85–92.

Wansbeek, T.J., Meijer, E., 2000. Measurement error and latent variables in econometrics. North-Holland, Amsterdam.

Wooldridge, J.M., 2010. Econometric analysis of cross section and panel data, second ed.. MIT Press, Cambridge, MA.

Xiao, Z., Shao, J., Palta, M., 2010a. Instrumental variable and GMM estimation for panel data with measurement error. Statist. Sinica 20, 1725–1747.

Xiao, Z., Shao, J., Palta, M., 2010b. GMM in linear regression for longitudinal data with multiple covariates measured with error. J. Appl. Stat. 37, 791–805.

Xiao, Z., Shao, J., Xu, R., Palta, M., 2007. Efficiency of GMM estimation in panel data models with measurement error. Sankhyā 69, 101–111.