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

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2017

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Consistent estimation of linear panel data models with measurement error

Erik Meijer, Laura Spierdijk, Tom Wansbeek

University of Southern California, United States
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: (i) restrictions on the intertemporal covariance matrix of errors in the equations, (ii) third moments of regressors with error, and (iii) 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.

http://dx.doi.org/10.1016/j.jeconom.2017.06.003
0304-4076/© 2017 Elsevier B.V. All rights reserved.
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

\[ y_{nt} = \alpha_n + \gamma y_{n,t-1} + \xi_{nt} \beta + \epsilon_{nt}, \]

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

\[ x_{nt} = \xi_{nt} + v_{nt}. \]

The reduced form obtained by the elimination of \( \xi_{nt} \) is

\[ y_{nt} = \alpha_n + \gamma y_{n,t-1} + x_{nt} \beta + \epsilon_{nt}, \]

where

\[ \epsilon_{nt} = \epsilon_{nt} - v_{nt} \beta. \]

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 \( y_{nt} \) also observed. We assume \( N > T \) and thus we will use large \( N \), fixed \( T \) asymptotics.

It is notionally convenient to define stacked versions for the \( nt \)th individual: \( y_n = (y_{n1}, \ldots, y_{nT})' \), \( x_n = (x_{n1}, \ldots, x_{nT})' \), and so forth. We denote the intertemporal covariance matrix of the error terms by \( \Sigma_\epsilon \equiv \text{Var}(\epsilon_{nt}) \).

Initially, we assume that \( \epsilon_{nt} \) and the measurement errors \( v_{nt} \) are independent of the regressors \( (\Sigma_n, \Sigma_v) \) 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 \( \alpha_n \) from the equation and assuming a specific structure for \( \Sigma_\epsilon \), or allowing \( \Sigma_\epsilon \) 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 \( (\sigma_\epsilon)^2 \), \( \sigma_v^2 \) and \( \sigma_\epsilon v \) covariances and third moments of the random variables in the model \((\epsilon_{nt}, v_{nt}, \tilde{\epsilon}_{nt}, \tilde{v}_{nt})\). 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 \( E(z_{nt}u_n) = 0 \), where \( z_{nt} \) is a \( q \times 1 \) vector of instruments for observation \( n \). This carries over to the panel data context, where the analogous moment conditions are \( E(z_{nt}u_{nt}) = 0 \). However, in panel data contexts, we can expand this to moment conditions of the form \( E(Z'_{nt}u_{nt}) = 0 \), 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 \( E(z_{nt}u_{nt} - z_{nt}u_{nt}) = 0 \) (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, generally denoted by \( X_n \); specifically, \( E(Z'_{nt}X_n) \) must have full column rank. With the \( y_{nt}, X_n, Z_n \) stacked in \( y, X, Z \), respectively, so \( XZ = \sum_n X_nZ_n \), and \( W \) a weight matrix of order \( q \times q \), the IV estimator is

\[ \hat{\beta}_W = (X'ZWZ'X)^{-1}X'ZWZ'y. \]

The consistency and asymptotic normality of \( \hat{\beta}_W \) 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 \( \Sigma_\epsilon \)

We consider linear restrictions that we may be willing to impose on \( \Sigma_\epsilon \), the covariance matrix of the errors in the model equations. The restrictions we consider are linear and hence can be expressed as vec \( \Sigma_\epsilon = C \pi_\epsilon \), with \( C \) known (and of full column rank) and \( \pi_\epsilon (r \times 1) \) 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, \( y_n = \xi_n \beta + \epsilon_n \), \( x_n = \xi_n + v_n \), and consider the random-effects model, where \( \epsilon_{nt} = \alpha_n + w_{nt} \), with \( \alpha_n \sim (0, \sigma_\alpha^2) \)
now a random effect and $v_{int} \sim (0, \sigma^2_w)$ i.i.d. Then

$$
\Sigma_e = 
\begin{pmatrix}
\sigma^2_u + \sigma^2_w & \sigma^2_u + \sigma^2_w & \sigma^2_u + \sigma^2_w \\
\sigma^2_u + \sigma^2_w & \sigma^2_u + \sigma^2_w & \sigma^2_u + \sigma^2_w \\
\sigma^2_u + \sigma^2_w & \sigma^2_u + \sigma^2_w & \sigma^2_u + \sigma^2_w
\end{pmatrix},
$$

$$
C_e = 
\begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1
\end{pmatrix},
\quad \pi_e = (\sigma^2_u, \sigma^2_w),
$$

and $r_e = 2$. Hence, with $u_e = y_e - x_a \beta$,

$$
\mathbb{E}(u_e \otimes y_e - C_e \pi_e) = \mathbb{E}[(\epsilon_e - v_a \beta) \otimes (\xi_e + \epsilon_e) - C_e \pi_e] = 0. \tag{2}
$$

Let $C_{r e}$ be a complement of $C_e$, that is, a matrix of order $T^2 \times (T^2 - r_e)$ and rank $T^2 - r_e$ such that $C_{r e} C_e = 0$. Premultiplication of (2) by $C_{r e}$ gives

$$
C_{r e} \mathbb{E}(u_e \otimes y_e) = C_{r e} \mathbb{E}[(I_r \otimes y_e)(y_e - x_a \beta)] = 0. \tag{3}
$$

So the (panel) IVs that are implied by the structure on $\Sigma_e$ are $Z_e = (I_r \otimes y_e)' C_{r e}$. The requirement that $\mathbb{E}(Z_e' X_e)$ have full column rank translates into $C_{r e} \mathbb{E}(\xi_e \otimes \xi_e) \neq 0$, so $\xi_{int}$ should not itself follow a random effects structure.

For one valid but otherwise arbitrary choice of $C_{r e}$, we obtain

$$
Z'_e = C_{r e}' (I_r \otimes y_e) =
\begin{pmatrix}
1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
\times
\begin{pmatrix}
y_{n1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
y_{n2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
y_{n3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
y_{n4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
y_{n5} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
y_{n6} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
y_{n7} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
y_{n8} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
y_{n9} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
y_{n10} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix},
$$

with $T = 3$, we have $r_e = 3$,

$$
C'_e =
\begin{pmatrix}
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0
\end{pmatrix},
\quad \pi_e = (\pi_0, \pi_1),
$$

and $C_{r e}$ and $Z_e$ 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 $y_e = \mathcal{E}_n \beta + \mathcal{R}_e \delta + \epsilon_e$ and $X_e = \mathcal{E} + V_e$, adapting (2) is straightforward:

$$
C_{r e}' \mathbb{E}(u_e \otimes y_e - C_{r e}) = C_{r e}' \mathbb{E}[(\epsilon_e - v_a \beta) \otimes (\mathcal{E}_n \beta + \mathcal{R}_e \delta + \epsilon_e)] = 0 \tag{5}
$$

due to the exogeneity of $\mathcal{R}_n$. So the moment conditions (3) remain valid; the panel IVs that are implied by the structure on $\Sigma_e$, $Z_e = (I_r \otimes y_e)' C_{r e}$, are unaltered. The addition of the $\ell$ parameters in $\delta$ can be covered by the $\ell$ moment conditions

$$
\mathbb{E}(R'_e u_e) = 0, \tag{6}
$$

so the entire instrument set available for estimating the general static model is $(\mathcal{E}_n, \mathcal{R}_n)$. The number of moment conditions from the structure on $\Sigma_e$ is $T^2 - r_e$.

**Heteroskedasticity.** In the static model the restrictions $\text{vec} \, \Sigma_e = C_{r e} \pi_e$ with $\pi_e$ constant but unknown parameters, imply homoskedasticity. This may be undesirably strong. However, this can be easily relaxed by allowing $\pi_e$ to be different for different individuals, and thus writing it as $\pi_e$ in which may depend in an arbitrary way on $\xi_e$ or even $v_e$. The instruments that we use are still valid under this relaxation, because they operate in the space orthogonal to $\xi_e$, in which $\pi_e$ 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 $t$ a $T$-vector of ones, the static model becomes

$$
y_e = t y_a + \mathcal{E}_n \beta + \mathcal{R}_e \delta + \epsilon_e.
$$

The fixed effect $a_e$ can be eliminated from the model by any matrix $B$ with property $B' t = 0$; typical choices for $B$ are the centering matrix $B_t = I_T - t t' / T$ (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

$$
\mathbb{E}[t_u \otimes y_e - (B \otimes B)' C_{r e} \pi_e] = 0, \tag{7}
$$

with $t_u \equiv y_e - \tilde{\mathcal{E}}_e \beta - \tilde{\mathcal{R}}_e \delta$. So the IVs are now based on the complement of $(B \otimes B)' C_{r e}$ rather than $C_{r e}$.\footnote{If we assume that $\mathbb{E}(a_e V_e) = 0$ and $\mathbb{E}(a_e v_e) = 0$, there is a slight loss of efficiency in this approach since $\mathbb{E}((\epsilon_e - v_e \beta) \otimes y_e) = \text{vec} \, \Sigma_e$ still holds. Hence, for estimation we only need to eliminate the $a_e$ from $y_e - \mathcal{E}_n \beta - \mathcal{R}_e \delta$. Consequently, $\mathbb{E}[(\tilde{y}_e - \tilde{\mathcal{E}}_e \beta - \tilde{\mathcal{R}}_e \delta) \otimes y_e - (B \otimes t') C_{r e} \pi_e] = 0$ 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
\[
y_{nt} = \alpha_n + \sum_{j=0}^{T-1} \gamma_j \xi_{n,t-j},
\]
where \(\alpha_n\) is an infinite sum containing contemporaneous and lagged versions of \(\xi_n\) and \(r_{nt}\), as well as \(\alpha_n\) (in the fixed effects model). Hence,
\[
d \equiv C_{n\ell}' E(u_{nt} \otimes y_{nt}) = C_{n\ell}' \sum_{j=1}^{\infty} E(\varepsilon_{n,t} \otimes \varepsilon_{n,-j}) \neq 0,
\]
where the notation \(\varepsilon_{n,-j} = (\varepsilon_{n,-1} t \ldots, \varepsilon_{n,-T})'\) denotes the \(j\)-periods lagged version of the vector \(\varepsilon_n\). So \((I_T \otimes \varepsilon_n)C_{n\ell}'\) 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
\[
E(u_{nt} y_{nt}) = E(\varepsilon_{nt} y_{nt}) = \sum_{j=0}^{\infty} E(\varepsilon_{nt} \varepsilon_{n,t-j}) y_j = \sigma_\varepsilon^2 \gamma_{s-t} + I(t \leq s) \sigma_\varepsilon^2 \gamma_{s-t},
\]
provided that \(|\gamma| < 1\). Therefore,
\[
d = E(Z_{nt} u_{nt}) = E(y_{n1} y_{n2} y_{n3}) = \sigma_\varepsilon^2 \begin{pmatrix} \frac{\gamma_1}{2} & \frac{\gamma_2}{2} & \frac{\gamma_3}{2} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ \frac{\gamma_1}{2} & \frac{\gamma_2}{2} & \frac{\gamma_3}{2} \end{pmatrix} \neq 0.
\]
So the moment conditions do not apply anymore. However, they still can be exploited to some extent since
\[
\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \ 0 & 1 & 0 & 0 & 0 & 0 & 0 \ 0 & 0 & 1 & 0 & 0 & 0 & 0 \ 0 & 0 & 0 & 1 & 0 & 0 & 0 \ 0 & 0 & 0 & 0 & 1 & 0 & 0 \ 0 & 0 & 0 & 0 & 0 & 1 & 0 \ 0 & 0 & 0 & 0 & 0 & 0 & 1 \ \end{pmatrix}
\]
\[
\begin{pmatrix} y_{n1} & -y_{n2} & 0 \ -y_{n2} & y_{n3} & 0 \ 0 & -y_{n1} & y_{n1} \ y_{n2} & y_{n3} & 0 \ \end{pmatrix}
\]
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 \(d = C_{n\ell} \alpha_n\), where the elements of \(\alpha_n\) are functions of the model parameters, and \(C_n\) does not depend on parameters. If \(C_n\) has at least \(k\) more rows than columns, this may identify the coefficients of interest.

We simply replace \(C_{n\ell}\) by \(C_n \alpha_n\) 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^{FE} \equiv ((B \otimes B)' C_{\ell}\) \(E(u_{nt} \otimes y_{nt}) = C_{\ell}' \pi^{FE} \alpha_n^{FE},
\]
for some known constant matrix \(C_{\ell}\) and some vector \(\pi^{FE}\) that may depend on parameters. The resulting IVs are then \((I_{T-1} \otimes \mathbf{y}_{n})' (B \otimes B)' C_{\ell}\) \(C_{\ell}\).

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 \(\xi_{nt}\) has only one element, suppose that \(E(\varepsilon_{nt}) = \lambda \neq 0\) and that \(\xi_{nt}, \varepsilon_{nt}\), and \(r_{nt}\) 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
\[
\hat{\beta} = \frac{\sum_{n=1}^{N} \sum_{t=1}^{T} y_{nt} y_{nt}^{2}}{\sum_{n=1}^{N} \sum_{t=1}^{T} y_{nt}^{2}}
\]
This is an IV estimator with instrument \(z_{nt} = x_{nt} y_{nt}\), that is, it is based on the moment condition
\[
E[\{x_{nt} y_{nt} (y_{nt} - x_{nt} \beta)\} = E[\{x_{nt} + v_{nt} \xi_{nt} \beta + \varepsilon_{nt} (\varepsilon_{nt} - v_{nt} \beta)\} = 0,
\]
whereas \(E[\{x_{nt} y_{nt} \varepsilon_{nt}\}] = \beta \lambda \neq 0\), provided that \(\beta\) is nonzero. Note that the independence assumption ensures that expressions like \(E(\xi_{nt} E_{nt} \varepsilon_{nt})\) 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 \(E(\xi_{np} E_{nt} \varepsilon_{nt}) = \lambda_{pnt} \neq 0\) and that \(\xi_{nt}, \varepsilon_{nt}\), and \(\varepsilon_{nt}\) are stochastically independent of each other. Then
\[
E[\{x_{np} y_{nt} (y_{nt} - x_{nt} \beta)\} = E[\{x_{np} + v_{ns} \xi_{nt} \beta + \varepsilon_{nt} (\varepsilon_{nt} - v_{nt} \beta)\} = 0,
\]
while \(E[\{x_{np} y_{nt} \varepsilon_{nt}\}] = \beta \lambda_{pnt} \neq 0\). Hence, \((x_{np} y_{nt})\) 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 \(\lambda_{s} \equiv E(\xi_{nt} \otimes \xi_{n} \otimes \xi_{n}) \neq 0\), the latter meaning that at least one element of \(\lambda_{s}\) is nonzero. Let \(\lambda_{c} \equiv E(\xi_{nt} \otimes \varepsilon_{nt} \otimes \varepsilon_{nt})\) and \(\lambda_{v} \equiv E(\varepsilon_{nt} \otimes \varepsilon_{nt} \otimes \varepsilon_{nt})\) be defined accordingly.

The third moments can now be written as
\[
E(y_{nt} \otimes y_{nt} \otimes y_{nt}) = \lambda_{c} \beta^3 + \lambda_{v}
\]
\[
E(y_{nt} \otimes y_{nt} \otimes x_{nt}) = \lambda_{c} \beta^2
\]
\[
E(y_{nt} \otimes x_{nt} \otimes x_{nt}) = \lambda_{c} \beta
\]
\[
E(x_{nt} \otimes x_{nt} \otimes x_{nt}) = \lambda_{c} + \lambda_{v}.
\]
These expressions owe their simplicity to the assumed independence between \(\xi_{nt}, \varepsilon_{nt}\), and \(v_{nt}\). The independence assumption implies homoskedasticity:
\[
E(\varepsilon_{nt} \otimes \varepsilon_{nt} \otimes \varepsilon_{nt}) = E(v_{nt} \otimes v_{nt} \otimes v_{nt}) = 0.
\]
Subtracting \(\beta\) times (9c) from (9b) gives
\[
E[\{y_{nt} \otimes (y_{nt} - x_{nt} \beta) \otimes x_{nt}\} = E[\{y_{nt} \otimes I_{T} \otimes x_{nt}(y_{nt} - x_{nt} \beta)\} = 0.\]
Thus, (10) amounts to ordinary IV with alizesto of identifications added more recently, all pertaining to the case of a single vec transform the set of momentsto obtain a set of so the balance is not clear a priori. As an example, we can linearly Of course, more instruments mean more signal but also more noise, E moment conditions that depend on \( \beta \) and a set of \( T - T(T + 1)(T + 2)/6 \) symmetry conditions that do not involve \( \beta \), for example which may not improve finite-sample statistical properties. We will investigate the many instruments issue through simulation in Section 6.

Under the assumption that \( \epsilon_n \) and \( v_n \) are independent of each other and of \( \xi_n \), 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.

\textbf{Heteroskedasticity.} In the static model the estimators that use third moments do not accommodate heteroskedasticity easily. The moment condition (9b) is only valid if \( \mathbb{E}(\epsilon_n \otimes \epsilon_n \otimes \xi_n) = 0 \) 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 \( \mathbb{E}(\epsilon_n \otimes \epsilon_n \otimes \epsilon_n) = 0 \) and \( \mathbb{E}(v_n \otimes v_n \otimes v_n) = 0 \) if \( p, s, r \) and \( t \) are all distinct. This would identify \( \beta \) 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 \( \mathbb{E}(\epsilon_n \otimes \xi_n) \) and \( \mathbb{E}(v_n \otimes \epsilon_n) \) are allowed to be nonzero. The moment conditions (11) or (12) then apply only to \( (p, s, t) \) with \( p, s, r, t \) all distinct.

\textbf{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):

\[
\mathbb{E}\left[ y_n \otimes (y_n - \bar{y}_n \beta) \otimes \xi_n \right] = \mathbb{E}\left[ (y_n \otimes I_{T-1} \otimes \xi_n) (y_n - \bar{y}_n \beta) \right] = 0,
\]

so that \( \bar{Z}_{\text{ne},n} \equiv (y_n \otimes I_{T-1} \otimes \xi_n) \) is a valid instrument matrix. However, if the individual effect \( \alpha_n \) is independent of \( \epsilon_n \) and \( v_n \), then the instrument matrix \( \bar{Z}_{\alpha,n} \equiv (y_n \otimes I_{T-1} \otimes \xi_n) \) is also valid and gives us more instruments. An analogous analysis shows that, under the assumption that the third moments of \( \epsilon_n \) vanish, \( \bar{Z}_{\text{pe},n} \equiv (y_n \otimes I_{T-1} \otimes \xi_n) \) is a valid instrument matrix, and \( \bar{Z}_{\text{ve},n} \equiv (y_n \otimes I_{T-1} \otimes \xi_n) \) is valid if \( v_n \) is independent of \( \xi_n \) and \( \alpha_n \). Finally, \( \bar{Z}_{\alpha,n} \equiv (I_{T-1} \otimes \xi_n \otimes \xi_n) \) is valid if the third moments of \( \alpha_n \) are
Zero. Again, with multiple regressors, we can replace \( x_n \) in the instruments by \( \text{vec} X_n \), 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 \( (y_n - x_n \beta) \). 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,

\[
E(Z'_{xy}u_n) = E(y_n \otimes u_n \otimes x_n) = E \left[ \left( y_{n-1}' \gamma + \xi_n \beta + \epsilon_n \right) \otimes \left( \xi_n + v_n \right) \right] = E \left[ y_{n-1} \otimes \left( \epsilon_n - v_n \beta \right) \otimes \left( \xi_n + v_n \right) \right] \gamma. \]

This is zero, because we can write \( y_{n-1}' \) as an infinite sum of terms of the form \( \epsilon_{n-1} \) and \( \xi_{n-1} \) and in the resulting triple products there is always at least one mean-zero factor that is independent of the others. Thus, \( Z_{xy}' \) is still a valid instrument matrix. Analogously, \( Z_{xy,n} \) is still a valid instrument matrix under the assumption that the third moments of \( \epsilon_n \) (for all triples of time points, including \( t \leq 0 \)) vanish and \( Z_{x,n} \) is still a valid instrument matrix under the assumption that the third moments of \( v_n \) 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., \( \ell \)) 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 \( \beta \). 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:

\[
y_n = \xi_n \beta + \gamma_n \delta + \epsilon_n. \tag{14}\]

The assumption that \( \gamma_n, \epsilon_n, v_n \) are mutually independent implies that

\[
E[(r_{n,t-1}, r_{nt})(y_n - x_n \beta - \gamma_n \delta)] = E[(r_{n,t-1}, r_{nt})(\xi_n - v_n \beta)] = 0. \]

Thus, the main criterion for the validity of \((r_{n,t-1}, r_{nt})\)' as instruments is satisfied. Identification of \( \beta \) and \( \delta \) from this moment condition requires that the coefficient of \( r_{n,t-1} \) in the linear projection of \( x_n \beta \) (or, equivalently given our assumptions, \( \xi_n \beta \)) on \( r_{n,t-1} \) and \( r_{nt} \) is nonzero. Thus, \( r_{n,t-1} \) should be excluded from the equation for \( y_n \) but should not be excluded from the equation for \( x_n \beta \) 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 \( \xi_n = r_n \kappa + \omega_n \), with \( E(\omega_n | r_n) = 0 \) but \( E(\omega_n^2 | r_n) \neq 0 \). Because we take all variables in deviation of their time-mean, this implies that the relation between \( \xi_n \) and \( r_n \) is heteroskedastic. Then, if \( v_n \) and \( \epsilon_n \) are independent of \( \xi_n \) and \( r_n \),

\[
E[r_n(x_n - r_n \kappa)(y_n - x_n \beta - \gamma_n \delta)] = E[r_n(y_n + \omega_n)(\xi_n - v_n \beta)] = 0 \quad \text{and} \quad E[r_n(x_n - r_n \kappa)(\xi_n, r_n)] = E[r_n(y_n + \omega_n)(\kappa, 1)] + E[r_n(y_n + \omega_n)^2(1, 0)]= E(\omega_n^2 | r_n)(1, 0). \]

This shows that, if \( \kappa \) were known, \( r_n(x_n - r_n \kappa) \) would be a valid instrument for \( x_n \), and with \( r_n \) as an instrument for itself, would jointly identify \( \beta \) and \( \gamma \), where in this case the heteroskedasticity of the relation between \( \xi_n \) and \( r_n \) 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 \( \kappa \), but can estimate it consistently by OLS of \( x_n \) on \( r_n \), and replacing \( \kappa \) 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

\[
E(r_{nt} u_n) = 0 \quad \text{for all } n \text{ and } t, \tag{15}\]

where \( u_n \equiv y_n - x_n \beta - v_n \delta \equiv \epsilon_n - v_n \beta \). In matrix notation, we can write (15) as \( E(r_n \otimes u_n) = 0 \), or \( E[(r_n \otimes I_r)u_n] = 0 \). Hence, \( Z_n = (r_n \otimes I_r)^\top \) is a valid set of instruments. Identification of \( \beta \) and \( \delta \) then requires that

\[
J = E[Z_n(x_n, r_n)] = (\text{vec } X_n, \text{vec } R_n) \tag{16}\]

has full column rank, where \( \text{vec } X_n \equiv E(x_n r_n) \) and \( \text{vec } R_n \equiv E(r_n r_n) \). This condition will be fulfilled in most cases, but the asymptotic variance of the estimators of \( \beta \) and \( \delta \) 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 \( x_n = c + r_n + u_n \) with \( E(r_n u_n') = 0 \), \( \Sigma_{X_n} = \Sigma_{R_n} \) 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 \( \xi_n \) on \( r_n \),

\[
\xi_n = K r_n + \omega_n, \tag{17}\]

where \( K \equiv E(\xi_n | r_n) \) is the inverse of \( E(r_n r_n') \). With \( w_n \equiv v_n + \omega_n, x_n = K r_n + w_n \). Now consider the situation where the relation (17) between \( \xi_n \) and \( r_n \) is heteroskedastic, so that \( E(\omega_n w_n' | r_n) \) is a function of \( r_n \). We make the slightly stronger assumption that \( E(r_n \otimes u_n \otimes \omega_n) \neq 0 \). Let

\[
h_n \equiv \left( r_n \otimes u_n \otimes w_n \right)' \equiv Z_n u_n, \tag{18}\]

where \( Z_n \equiv r_n' \otimes (1, w_n' \otimes I_r) \). In view of (15) and the various independence assumptions made (esp. that \( v_n \) is independent of \( r_n \)) and the assumption that \( r_n \) is in deviation of its mean, \( E(h_n) = 0 \). Hence,
if $w_n$ were observed, $Z_n$ would be a valid set of instruments that could be used for consistent estimation of the coefficients $\beta$ and $\delta$. Because $w_n$ is not observed, this cannot be used directly. However, we can replace it by $\tilde{w}_n$, 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 $\tilde{w}_n$ 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 $\tilde{Z}_n$.

Identification depends on the rank of the matrix

$$G = \mathbb{E}[Z'_n(x_n, r_n)] = \begin{pmatrix} \text{vec } \Sigma_n & \text{vec } \Sigma_r \\ q_1 & q_2 \end{pmatrix},$$

(19)

where

$$\Sigma_{nt} = K \Sigma,$$

$$q_1 = \mathbb{E}(r_n \otimes w_n \otimes x_n)$$

$$= \mathbb{E}(r_n \otimes (v_n + \omega_n) \otimes (Kr_n + v_n + \omega_n))$$

$$= (I \otimes I \otimes K)\mathbb{E}(r_n \otimes \omega_n \otimes r_n) + \mathbb{E}(r_n \otimes \omega_n \otimes \omega_n)$$

$$q_2 = \mathbb{E}(r_n \otimes w_n \otimes r_n)$$

$$= \mathbb{E}(r_n \otimes \omega_n \otimes r_n).$$

If the projection (17) can be strengthened to $\mathbb{E}(\xi_n | r_n) = Kr_n$, then $\mathbb{E}(r_n \otimes \omega_n \otimes r_n) = 0$, and $q_1$ 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 $\xi_n$ 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 $r_{nt}, s = 1, \ldots, T$. 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 $\tilde{u}_{nt}$.
3. Create $T$ sets of instruments. Instrument set $t$ consists of $r_{nt}, s = 1, \ldots, T$ and all products $r_{nt}\tilde{w}_{nt}, p = 1, \ldots, T$, as instruments for the $t$th 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 $r_{nt}$ now a vector, and $u_{nt} = y_{nt} - x_{nt}\beta - r_{nt}\delta = \epsilon_n - \nu_n\beta$, with $x_{nt}, \nu_n, \beta$ and $\delta$ now also vectors. In (18), $r_{nt}$ and $w_{nt}$ 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 $x_{jt}$ 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 $T \cdot T$ regressors each. Step 2 gives $T \cdot k$ residuals $\tilde{w}_{nt,j}$ for each individual. In step 3, instrument set $t$ consists of $r_{nt,i}$ and $r_{nt,i}\tilde{w}_{nt,j}, p = 1, \ldots, T, i = 1, \ldots, T, j = 1, \ldots, k$.

**Heteroskedasticity.** Because $y_{nt}$ is not included in the instruments derived in this section, heteroskedasticity of $\epsilon_n$ 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 $\mathbb{E}(r_n \otimes \nu_n \otimes \nu_n) = 0$. Thus, the regression of $\epsilon_n$ 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 $\mathbb{E}(r_n \otimes \nu_{nt} \otimes \nu_{nt}) = 0$ 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 $x_{nt}$. 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 $u_n = y_n - y_{t-1} + x_{nt}\beta - r_{nt}\delta = \epsilon_n - \nu_n\beta$ and (18) still holds. The rank condition for identification now applies to the matrix

$$G_D = \mathbb{E}[Z'_n(y_{n-1}, x_n, r_n)] = \begin{pmatrix} \text{vec } \Sigma_{n} & \text{vec } \Sigma_{n} & \text{vec } \Sigma_{n} \\ q_0 & q_1 & q_2 \end{pmatrix},$$

(20)

where $\Sigma_{n} = \mathbb{E}(y_{n-1}r'_{nt})$ and $q_0 = \mathbb{E}(r_n \otimes \nu_n \otimes \nu_{n-1})$. 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

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

with $\xi_{nt} \sim N(0, \sigma^2_\xi)$ 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 $\sigma^2_v = 0.7, \beta = 1, \sigma^2_\beta = 2$, and $\sigma^2_\delta = 1$. The $\xi_{nt}$ are generated according to

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

with $\xi_{nt} \sim \sqrt{\frac{4}{7}} \chi^2_1$ and $\epsilon_{n0} = \sqrt{\frac{4}{7}} \xi_{n0}$. This choice of $\xi_{nt}$ implies that the third moment of $\xi_{nt}$ 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 $\xi_{nt}$ 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
We leave out the moment conditions that do not involve matrix \(N\) or fixed effects (FE) estimation. With RE, the individual effect \(z\) therejectionrates of the errors, as opposed to the unbiased estimators. Table 4 shows that theseinstruments. The adjusted \(\bar{\beta}\) is, of course, not a test that can be used in practice, but it is ative 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 ("\(\bar{\sigma}(\hat{\beta})\)"), and the bootstrap (bootstrap \(\bar{\sigma}(\hat{\beta})\)). The bootstrap results in a considerable improvement of the rejection rates relative to the formula-based standard error.

A panel bootstrap based on the centered moment conditions can be used to estimate standard deviations that represent a second-order improvement relative to the formula-based asymptotic standard errors. 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 ("\(\bar{\sigma}(\hat{\beta})\)"), and the bootstrap (bootstrap \(\bar{\sigma}(\hat{\beta})\)). 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 relative 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

\(^2\) The R code is available upon request.

\(^3\) In preliminary simulations, we found that including these moment conditions leads to worse finite-sample properties.
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 \( \beta \), 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 \( R^2 \)'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 \( W_{\text{opt}} \) (columns 4–5), the 2SLS weight matrix \( W_{\text{2SLS}} \) (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 \( W_{\text{opt}} \), 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 \( \beta \). 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 \( \beta \) 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 \( x_{ni} \) (levels) instead of \( x_{nt} \) (within-differences), our simulations are based on \( x_{nt} \); see also the appendix with supplementary material. Using \( x_{nt} \) 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 \( W = W_{\text{2SLS}} \). 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 \( \xi_{nt} \) in the previous simulations: \( e_{nt} \sim \sqrt{\frac{1}{n}} \chi^2_{1} \) for \( t = 0, \ldots, T \) (\( T = 5 \)), \( r_{n0} = \sqrt{\frac{r}{2}} e_{n0} \), and \( r_{nt} = 0.5 r_{nt-1} + e_{nt} \).

We then compute \( \omega_{nt} = r_{nt} \xi_{nt} \), where \( \xi_{nt} \sim N(0, \sigma^2_{\xi}) \) (i.i.d.), so that \( \mathbb{E}(r_{nt} \xi_{nt}) = 0 \) for all \( t \) and \( s \), but \( \mathbb{E}(r_{nt} \xi_{nt}^2) = \mathbb{E}(r_{nt}^2) \mathbb{E}(\xi_{nt}^2) \neq 0 \), so that \( \mathbb{E}(r_{nt} \xi_{nt} \xi_{nt} \omega_{nt}) \neq 0 \). The regressor \( \xi_{nt} \) is then generated according to

\[
\xi_{nt} = \kappa_{1} r_{nt} + \kappa_{2} r_{nt-1} + \omega_{nt}.
\]

This satisfies the setup in Section 5 with \( K \neq c I \), but \( q_2 = 0 \) and \( q_1 = \mathbb{E}(r_{nt} \xi_{nt} \omega_{nt}) \neq 0 \). The model is completed by the system

\[
y_{nt} = \alpha_{nt} \xi_{nt} + \beta + r_{nt} y_{nt} + \epsilon_{nt}
\]

\[
x_{nt} = \xi_{nt} + v_{nt}.
\]

\(^{4}\) These results are available upon request.

**Table 4**

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

<table>
<thead>
<tr>
<th>cov. restr.</th>
<th>( W_{\text{opt}} )</th>
<th>( W_{\text{2SLS}} )</th>
<th>( W = I )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N = 100 )</td>
<td>( % ) rejection rate (sample ( \sigma(\hat{\beta}) ) )</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>( N = 200 )</td>
<td>( % ) rejection rate (sample ( \sigma(\hat{\beta}) ) )</td>
<td>15</td>
<td>11</td>
</tr>
<tr>
<td>( N = 500 )</td>
<td>( % ) rejection rate (sample ( \sigma(\hat{\beta}) ) )</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>( N = 1000 )</td>
<td>( % ) rejection rate (sample ( \sigma(\hat{\beta}) ) )</td>
<td>6</td>
<td>5</td>
</tr>
</tbody>
</table>

1. The model is completed by the system

\[
y_{nt} = \alpha_{nt} \xi_{nt} + \beta + r_{nt} y_{nt} + \epsilon_{nt}
\]

\[
x_{nt} = \xi_{nt} + v_{nt}.
\]

\(^{4}\) These results are available upon request.
with again $v_{nt} \sim N(0, \sigma^2_v)$, $\omega_n \sim N(0, \sigma^2_\epsilon)$, and $u_{nt} \sim N(0, \sigma^2_\eta)$. As in the previous simulation, we choose $\sigma^2_\epsilon = 0.7, \beta = 1, \sigma^2_v = 2$, and $\sigma^2_\eta = 1$. The additional parameters are $\sigma^2_\gamma = 1, \kappa_1 = \kappa_2 = 1/\sqrt{3} = 0.577$, 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 $\beta$ 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 ($\Sigma_\epsilon$), 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.

### Table 5

<table>
<thead>
<tr>
<th></th>
<th>RE</th>
<th>FE</th>
<th>RE</th>
<th>FE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N = 100$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N = 500$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N = 1000$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg. $\hat{\beta}$</td>
<td>97</td>
<td>103</td>
<td>103</td>
<td>102</td>
</tr>
<tr>
<td>Sample $\sigma(\hat{\beta})$</td>
<td>31</td>
<td>56</td>
<td>32</td>
<td>61</td>
</tr>
<tr>
<td>% Rejections</td>
<td>13</td>
<td>6</td>
<td>13</td>
<td>6</td>
</tr>
<tr>
<td>Avg. $\hat{\gamma}$</td>
<td>34</td>
<td>59</td>
<td>36</td>
<td>64</td>
</tr>
<tr>
<td>% Rejections</td>
<td>18</td>
<td>10</td>
<td>18</td>
<td>7</td>
</tr>
</tbody>
</table>

### Table 6

<table>
<thead>
<tr>
<th></th>
<th>RE</th>
<th>FE</th>
<th>RE</th>
<th>FE</th>
<th>RE</th>
<th>FE</th>
<th>RE</th>
<th>FE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cov. restr.</td>
<td>5</td>
<td>6</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>3rd mom.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RE</td>
<td>FE</td>
<td>RE</td>
<td>FE</td>
</tr>
<tr>
<td>$W_{opt}$</td>
<td>15</td>
<td>11</td>
<td>55</td>
<td>45</td>
<td>8</td>
<td>9</td>
<td>13</td>
<td>17</td>
</tr>
<tr>
<td>$W_{SLS}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RE</td>
<td>FE</td>
<td>RE</td>
<td>FE</td>
</tr>
<tr>
<td>$W = I$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RE</td>
<td>FE</td>
<td>RE</td>
<td>FE</td>
</tr>
<tr>
<td>Ex. restr.</td>
<td>5</td>
<td>4</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>6</td>
</tr>
</tbody>
</table>

### Nonlinear moment conditions

As to $\Sigma_\eta$, we considered linear restrictions only. A researcher may be willing to restrict $\Sigma_\eta$ 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, $\eta$, say, in a nonlinear way. One may proceed by using a consistent but inefficient estimate of $\eta$, 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. Wagenaar 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 $\Sigma_\eta$, offers a useful framework. When the nonclassical measurement error is not restricted and $\epsilon_{nt}$ and $v_{nt}$ 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 $\Sigma_\eta$. 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 $\xi_{nt} = x_{nt} + \nu_{nt}$, with $E(\nu_{nt} \mid x_{nt}) = 0$. Thus, $\xi$ and $x$ switch roles. Hyslop and Imbens (2001) argued that in many cases, this model can be more appropriate than the classical measurement error model. In this model, $E(y_{nt} \mid x_{nt}) = x_{nt} \beta$, 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...
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,

An issue related to moment selection is the impact of selecting
moment conditions on post-selection inference such as hypothe-
sis 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 distribu-
tion 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 instru-
ments. 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 Crudd (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 \( N \rightarrow \infty \)
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 \( T \rightarrow \infty \) or \( N \rightarrow \infty \)
and \( T \rightarrow \infty \) 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., Bjørn, E., Skjerpøy, T., 2003. Distribution of preferences and measure-
ment errors in a disaggregated expenditure system. Econom. J. 6, 374–400.
J. Econometrics 68, 5–27.
alternative assumptions and simplified estimation. J. Econometrics 76, 309–
321.
Altrov, J.C., Segal, L.M., 1996. Small-sample bias in GMM estimation of covariance
Angrist, J.D., Krueger, A.B., 1991. Does compulsory school attendance affect school-
ing and earnings? Q. J. Econ. 106, 979–1014.
Bekker, P.A., Crudd, F., 2015. Jackknife instrumental variable estimation with het-
eroskedasticity. 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.
Bjørn, E., 2000. Panel data with measurement errors: instrumental variables and
GMM procedures combining levels and differences. Econometric Rev. 19, 391–
424.
Bjørn, E., 2003. Handling the measurement error problem by means of panel data:
method moments applied on firm data. In: Stigum, B. (Ed.), Econometrics and
Bjørn, E., 2015. Panel data dynamics with mis-measured variables: Modeling and
GMM estimation. Empir. Econ. 48, 517–534.
Bjørn, E., Klette, J.T., 1998. Panel data with errors-in-variables: essential and re-
Bjørn, 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.
In: Heckman, J., Leamer, E.E. (Eds.), Handbook of Econometrics 5. North-
Holland, Amsterdam, pp. 3705–3843.
earnings data: do two wrongs make a right? J. Labor Econ. 9, 1–24.
Brattani, M.C., Croux, C., 2007. Robust estimators for the fixed effects panel data
Cameron, A.C., Trivedi, P.K., 2005. Microeconometrics. Cambridge University Press,
Cambridge, UK.
sion models with errors in the variables. J. Econometrics 76, 193–221.
69, 1161–1191.
Erickson, T., White, T.M., 2002. Two-step GMM estimation of the errors-in-
Geary, R.C., 1942. Inherent relations between random variables. Proc. R. Ir. Acad. A
Gospodinov, N., Komunjer, I., Ng, S., 2014. Minimum Distance Estimation of Dy-
31, 93–118.
Hall, P., Horowitz, J.L., 2012. Bootstrap critical values for tests based on generalized-
of moments estimation and entropy-based moment selection. J. Econometrics
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) 138, 488–512.
Hyslop, D.R., Imbens, G.W., 2001. Bias from classical and other forms of measure-
Kim, B., Solon, G., 2005. Implications of mean-reverting measurement error for
Kollár, I., 1967. On characterizing the gamma and the normal distribution. Pacific
J. Math. 20, 69–76.