The international platform of Ludwig-Maximilians University's Center for Economic Studies and the Ifo Institute

Working Papers

www.cesifo.org/wp

Consistent Estimation of Linear Panel Data Models with Measurement Error

Erik Meijer
Laura Spierdijk
Tom Wansbeek

CESifo Working Paper No. 5164
Category 12: Empirical and Theoretical Methods
January 2015

ISSN 2364-1428

An electronic version of the paper may be downloaded
- *from the SSRN website:* *www.SSRN.com*
- *from the RePEc website:* *www.RePEc.org*
- *from the CESifo website:* *www.CESifo-group.org/wp*

CESifo
Center for Economic Studies & Ifo Institute

CESifo *Working Paper No. 5164*

Consistent Estimation of Linear Panel Data Models with Measurement Error

Abstract

Measurement error causes a downward bias when estimating a panel data linear regression model. The panel data context offers various opportunities to derive moment conditions that result in consistent GMM estimators. We consider three sources of moment conditions: (i) restrictions on the intertemporal covariance matrix of the errors in the equations, (ii) heteroskedasticity and nonlinearity in the relation between the error-ridden covariate and another, error-free, covariate in the equation, and (iii) nonzero third moments of the covariates. In a simulation study we show that these approaches work well.

JEL-Code: C230, C260.

Keywords: measurement error, panel data, third moments, heteroskedasticity, GMM.

Erik Meijer
University of Southern California
USA
erik.meijer@usc.edu

*Laura Spierdijk**
University of Groningen
The Netherlands
l.spierdijk@rug.nl

Tom Wansbeek
University of Groningen
The Netherlands
t.j.wansbeek@rug.nl

*corresponding author

December 1, 2014

1 Introduction

It has long been recognized that the covariates of interest in a linear regression analysis are often measured with error. A well-known example is the investment model due to Fazzari, Hubbard, and Petersen (1988), where a firm's investments are regressed on a proxy of investment demand such as Tobin's q and cash flows. Measurement error arises because average q, which is observable, is taken as a proxy for marginal q, which is unobservable but desirable from a theoretical point of view; see Almeida, Campello, and Galvao (2010) for an overview of this literature. If not accounted for, measurement errors cause a downward bias in the parameter estimates, known as attenuation bias (e.g., Wansbeek and Meijer, 2000, section 2.3). Here, we reconsider the problem of measurement error in panel data models.

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, generalized by Biørn (2000) to the case where 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) 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.

The literature has focused on moment conditions that exploit structure based on assumptions on 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 other sources. We focus on deriving appropriate moment conditions, for the various cases that we consider. The resulting GMM estimators' consistency and asymptotic normality follow from standard GMM theory. We assess the finite-sample performance of our estimators by means of simulation.

Our derivation of moment conditions starts with the intertemporal covariance matrix of errors in the equations. This matrix is often taken to be highly structured, like in the random effects model. We consider linear restrictions of any form on this matrix and derive all implied moment conditions.

The second set of moment restrictions that we explore consists of restrictions based on an exogenous regressor. Exogenous regressors are not just a complication to be accounted for, but they can also be a source of additional moment conditions that allow for consistent estimation in the presence of measurement error. However, for the method to work well, it is desirable that the relation between the error-ridden regressor and the additional one is heteroskedastic or nonlinear.

The third moments of the error-ridden regressor provide a third source of moment conditions. We thus extend, to the panel data case, the classical literature on exploiting the third moments of the error-ridden regressor. This approach has been developed, for the case of a single cross-section, already by Geary (1942) and has since then been extended by Pal (1980), Dagenais and Dagenais (1997), Lewbel (1996, 1997), and Erickson and Whited (2002). To our best knowledge, application to the panel data case has not yet been addressed in the literature. The only assumption needed is that the third moment of the true value of the regressor for at least one of the time points does not vanish (in addition to independence of measurement errors from the true regressor). Although this assumption cannot be verified as the true value is clouded by measurement error, checking the third moment of the observed value of the regressor should provide enough guidance.

As to our three methods, we propose a variety of instrumental variables (IVs) that can be constructed from the already available data set. Some of these constructs involve the dependent variable (and possibly the error-ridden regressor), and others only the error-ridden regressor. These IVs remain valid when exogenous variables are added to the model. However, when the lagged dependent variable (without measurement error) is added to the model, the IVs involving the dependent variable sometimes need to be adapted. In combination with existing GMM techniques, our moment conditions greatly expand the toolkit of the applied researcher.

The setup of this paper is as follows. The panel data model with measurement error is introduced in section 2. Special attention is given to panel IV estimation and its practical implementation. In section 3, we show how restrictions on the intertemporal covariance matrix of errors in the equations can be used to obtain consistent estimators. Section 4 adapts this to the dynamic panel data model, where such restrictions are commonplace, for example, for deriving the Arellano-Bond estimator. In section 5, we show how the presence of additional exogenous regressors in the model generates additional moment conditions. Section 6 derives the instrumental variables that become available when the third moment of the true regressor is nonzero, and again we use this to obtain consistent estimators. Section 7 discusses the impact of heteroskedasticity on the estimators based on the covariance matrix restrictions and third moments. In Section 8 we adapt the results when fixed individual effects are considered. We present simulation results in Section 9. Section 10 concludes.

2 Basics

We start with the simplest possible panel data regression model and consider extensions in later sections. We derive the basic inconsistency result, consider identification, and discuss panel instrumental variables.

2.1 The model and its implications

Let, for $n = 1, \ldots, N$, the T-vector y_n, with elements $y_{n1}, \ldots, y_{nT}$, depend on the T-vector ξ_n, with analogously denoted elements. Typically, the elements of these vectors are measurements of the same variable at different points in time. The vector ξ_n is unobservable, and instead a proxy x_n is observed:

$$
\begin{aligned}
y_n &= \xi_n \beta + \varepsilon_n \\
x_n &= \xi_n + v_n,
\end{aligned}
$$

with $\xi_n \sim (0, \Sigma_\xi)$, $\varepsilon_n \sim (0, \Sigma_\varepsilon)$, and $v_n \sim (0, \Sigma_v)$, which are mutually independent. Since we do not impose a structure on these matrices as yet, this specification encompasses the random effects model. 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 as we will consider large N, fixed T asymptotics.

Elimination of ξ_n leads to the reduced form

$$
\begin{aligned}
y_n &= x_n \beta + u_n \\
u_n &\equiv \varepsilon_n - v_n \beta.
\end{aligned}
$$

So $\mathbb{E}(x_n' u_n) = -(\operatorname{tr} \Sigma_v)\beta \neq 0$ and hence (pooled) OLS is inconsistent. With $\mathbb{E}(x_n' y_n) = (\operatorname{tr} \Sigma_\xi)\beta$ and $\Sigma_x \equiv \Sigma_\xi + \Sigma_v$, OLS estimation of the reduced form gives

$$
\operatorname*{plim}_{N\to\infty} b_{\text{OLS}} = \operatorname*{plim}_{N\to\infty} \frac{\sum_n x_n' y_n}{\sum_n x_n' x_n} = \frac{\operatorname{tr} \Sigma_\xi}{\operatorname{tr} \Sigma_x} \beta, \tag{1}
$$

which (since $\operatorname{tr} \Sigma_\xi \leq \operatorname{tr} \Sigma_x$) reflects the usual attenuation bias towards zero with measurement error.

2.2 Identification

For the case of a single cross section, that is, the above model with $T = 1$, the model is not identified, and hence does not allow for consistent estimators, if the observations are independent and the regressor is normally distributed, cf. Wansbeek and Meijer (2000, chapter 4). For $T > 1$ the observations are not independent, making an investigation of its identification worthwhile. The complete second-order implications of the model are

$$\mathbb{E}(y_n y_n') = \Sigma_\xi \beta^2 + \Sigma_\varepsilon \tag{2}$$

$$\mathbb{E}(y_n x_n') = \Sigma_\xi \beta \tag{3}$$

$$\mathbb{E}(x_n x_n') = \Sigma_\xi + \Sigma_v. \tag{4}$$

As an informal check on identification, notice that, from (2), we need the information contained in the second moment of y_n to identify Σ_ε since the latter only occurs in (2). Analogously, we need the information contained in the second moment of x_n to identify Σ_v. This leaves us with the covariance of y_n and x_n to identify both β and Σ_ξ, which occur as a product and cannot be disentangled without further information in some form. Hence, dependence of observations in panel data by itself is insufficient for identification.

2.3 Panel instruments

As usual, the way to proceed is through instrumental variables (IVs). Cameron and Trivedi (2005, section 22.2) discuss IV estimation in a panel data context. IVs may be available from outside the model, but under circumstances to be described in the following sections, they are implied by the model itself.

In cross-sections, IV is based on moment conditions of the form $\mathbb{E}(z_n u_n) = 0$, where z_n is a vector of instruments for observation n. This carries over to the panel data context, where the analogous moment condition is $\mathbb{E}(z_{nt} u_{nt}) = 0$. 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 moments like these below. As with the cross-sectional IVs, the panel IVs also need to be correlated with the explanatory variable, that is, $\mathbb{E}(Z_n' x_n) \neq 0$.

With $Z' \equiv (Z_1', \ldots, Z_N')$ and W a weight matrix of order $q \times q$, the basic IV estimator is

$$\hat{\beta}_{\mathrm{IV}} = \frac{\left(\sum_{n=1}^{N} x_n' Z_n\right) W \left(\sum_{n=1}^{N} Z_n' y_n\right)}{\left(\sum_{n=1}^{N} x_n' Z_n\right) W \left(\sum_{n=1}^{N} Z_n' x_n\right)} = \frac{x' Z W Z' y}{x' Z W Z' x}, \tag{5}$$

with the y_n and x_n collected in NT-vectors y and x. The properties of $\hat{\beta}_{\mathrm{IV}}$ and the choice of W follow from basic GMM theory. One choice for W is to take it proportional to $(Z'Z)^{-1}$, leading to 2SLS. Another choice is to take it proportional to a robust estimator of the inverse of the variance of $Z_n' u_n$: first compute the 2SLS estimator $\hat{\beta}_{\mathrm{2SLS}}$ and the residuals $\hat{u}_n = y_n - x_n \hat{\beta}_{\mathrm{2SLS}}$ and then use the inverse of the sample covariance matrix of $Z_n' \hat{u}_n$ as the weight matrix. This gives 3SLS or optimal GMM.

A remarkable practical aspect of panel IV is that it can be estimated in a standard statistical package (like Stata) that offers cross-sectional IV with clustered standard errors. This is mentioned briefly by Cameron and Trivedi (2005, p. 751), but not discussed in detail and therefore perhaps underappreciated. Because this has great practical implications for the implementation of the estimators we propose, we treat this more explicitly here.

Consider the situation where $\mathbb{E}(Z_n' u_n) = 0$ but $\mathbb{E}(z_{nt} u_{nt}) \neq 0$. We think of a panel data set organized in the long form, so each (n, t) combination has one row in the data matrix, and a moment condition of the form $\mathbb{E}(z_{nt} u_{nt}) = 0$ would be considered a single moment condition (of dimension q) when used with an IV command intended for cross-sectional data, not as T moment conditions. The assumption $\mathbb{E}(z_{nt} u_{nt}) \neq 0$ would suggest that standard IV with z_{nt} as instruments for u_{nt} is incorrect. However, the IV estimators are of the same form as (5) and thus are also panel IV estimators and therefore consistent. The reason for this remarkable result is that the IV estimators aggregate the moments $z_{nt}y_{nt}$ and $z_{nt}x_{nt}$ across both n and t, and thus instead of being based on $\mathbb{E}[z_{nt}(y_{nt} - x_{nt}\beta)] = 0$ can be reinterpreted as being based on

$$\mathbb{E}\left[\sum_{t=1}^{T} z_{nt}(y_{nt} - x_{nt}\beta)\right] = 0,$$

or $\mathbb{E}[Z_n'(y_n - x_n\beta)] = 0$ which is the panel IV moment condition. In most cases, one would expect dependence across time within cross-sectional units, and therefore the standard errors need to be clustered, but this is straightforward in software like Stata.

There are, however, two caveats with this implementation. The first is that the procedure breaks down with time-varying sampling weights. For example, suppose that one of the moment conditions is $\mathbb{E}(y_{n2}u_{n1} - y_{n1}u_{n2}) = 0$, so the corresponding column of Z_n has y_{n2} as its first element and $-y_{n1}$ as its second element, and its remaining elements are zero. With time-varying weights, the moment condition that would become implemented is $\mathbb{E}(w_{n1}y_{n2}u_{n1} - w_{n2}y_{n1}u_{n2}) = 0$, which is different, and typically incorrect. Hence, sampling weights that differ across cross-sectional units are fine, but weights that vary within unit cannot be used without adaptation. An example of an adaptation would be to use the average of the sampling weights across time for each individual, because then we obtain $\mathbb{E}[\bar{w}_n(y_{n2}u_{n1} - y_{n1}u_{n2})] = 0$, which is correct if the moment conditions are uncorrelated with the weights. (If this is not the case, one needs to investigate selectivity bias in more detail, but this is not specific to panel IV. Typically, the idea is that the weights remove rather than induce selectivity bias.)

The second, related, caveat is that the method may also break down for unbalanced panel data or missing data. For unbalanced data, moments of the form $\mathbb{E}(y_{n2}u_{n1} - y_{n1}u_{n2}) = 0$ would not pose a problem, because both terms would be missing if, say, the individual was absent in period 2 but not in period 1, and the estimator would aggregate only over time periods that are observed. But, for example, missing data on x_{n2} but not y_{n2} would pose a problem, because the second term would be omitted but the first not if y_{n1} and x_{n1} are both observed. Without such item-missing data, unbalanced panel data still cause problems for moment conditions like $E(y_{n1}u_{n1} - y_{n2}u_{n2}) = 0$. The problems with unbalanced or missing data can be solved by setting the whole affected column of Z_n to zero.

The panel IV theory generalizes immediately to more explanatory variables: if X_n is a $T \times p$ matrix, then the IV estimator becomes

$$\hat{\beta}_{\mathrm{IV}} = \left[\left(\sum_{n=1}^{N} X_n'Z_n\right) W \left(\sum_{n=1}^{N} Z_n'X_n\right)\right]^{-1} \left(\sum_{n=1}^{N} X_n'Z_n\right) W \left(\sum_{n=1}^{N} Z_n'y_n\right) = (X'ZWZ'X)^{-1}X'ZWZ'y,$$

with X the matrix that stacks all the X_n and $W = (Z'Z)^{-1}$ leading to the 2SLS estimator.

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. We start with two motivating examples and then treat the general case.

3.1 Motivating examples

In this section we discuss two examples of models for the errors that lead to linear restrictions on the error covariance matrix, and we show examples of moment conditions that follow from these models. In the examples, we will use that

$$\mathbb{E}[y_{ns}(y_{nt} - x_{nt}\beta)] = \mathbb{E}(\varepsilon_{ns}\varepsilon_{nt}) = (\Sigma_\varepsilon)_{st}. \tag{6}$$

The first example is the random effects model. In this model, it is assumed that the error ε_{it} can be decomposed into a random individual effect α_i and an i.i.d. residual error term w_{it}, that is, $\varepsilon_{it} = \alpha_i + w_{it}$, and $\alpha_i \sim (0, \sigma_\alpha^2)$ and $w_{it} \sim (0, \sigma_w^2)$ are assumed independent of each other and of the other terms in the model. Hence, $(\Sigma_\varepsilon)_{tt} = \sigma_\alpha^2 + \sigma_w^2$ and $(\Sigma_\varepsilon)_{ts} = \sigma_\alpha^2$ for $s \neq t$. From the latter and (6), we obtain $\mathbb{E}[(y_{ns} - y_{nr})(y_{nt} - x_{nt}\beta)] = 0$, with s, r, and t all distinct. Hence, $y_{ns} - y_{nr}$ can be used as an instrument for x_{nt}, provided that $\mathbb{E}[(y_{ns} - y_{nr})x_{nt}] \neq 0$, which is equivalent to $\mathbb{E}[(\xi_{ns} - \xi_{nr})\xi_{nt}] \neq 0$. This is violated if ξ_{nt} itself follows a random effects structure.

In the second example, we assume that $\{\varepsilon_{nt}\}$ is stationary, which implies $(\Sigma_\varepsilon)_{ts} = \pi_{|t-s|}$ for some set of parameters $\pi_0, \ldots, \pi_{T-1}$. Combining this with (6), we obtain $\mathbb{E}[(y_{n,t+k} - y_{n,t-k})(y_{nt} - x_{nt}\beta)] = 0$. Hence, $y_{n,t+k} - y_{n,t-k}$ can be used as an instrument for x_{nt}, again, provided that it is correlated with x_{nt}.

In both cases, panel IV implies moment conditions that do not fit into a standard IV framework, as discussed in section 2.3. For example, in both examples, $\mathbb{E}(\varepsilon_{nt}\varepsilon_{n,t+k}) = \mathbb{E}(\varepsilon_{ns}\varepsilon_{n,s+k})$, which gives rise to moment conditions of the form $\mathbb{E}[y_{n,t+k}(y_{nt} - x_{nt}\beta) - y_{n,s+k}(y_{ns} - x_{ns}\beta)] = 0$.

3.2 Moment conditions

The restrictions we consider are linear and hence can be expressed as $\operatorname{vec}\Sigma_\varepsilon = C_\varepsilon \pi_\varepsilon$, with C_ε known (and of full column rank) and π_ε ($r_\varepsilon \times 1$) unknown. For example, when we have the basic random-effects model, $C_\varepsilon = \left(\iota_{T^2}, \operatorname{vec} I_T\right)$, with in general ι_K a vector of K ones, and $\pi_\varepsilon = (\sigma_\alpha^2, \sigma_w^2)'$.

In stacked form, the moment conditions (2)–(4) with the restrictions inserted can now, with $\sigma_\xi \equiv \operatorname{vec}\Sigma_\xi$ and $\sigma_v \equiv \operatorname{vec}\Sigma_v$, be written as

$$\mathbb{E}\begin{pmatrix} y_n \otimes y_n - \sigma_\xi \beta^2 - C_\varepsilon \pi_\varepsilon \\ x_n \otimes y_n - \sigma_\xi \beta \\ x_n \otimes x_n - \sigma_\xi - \sigma_v \end{pmatrix} = 0. \tag{7}$$

The third element just-identifies σ_v and, conditional on β, the second one just-identifies σ_ξ. Subtracting β times the second element from the first element, the remaining moment condition becomes

$$\mathbb{E}\left[(y_n - x_n\beta) \otimes y_n - C_\varepsilon \pi_\varepsilon\right] = 0. \tag{8}$$

Let the matrix $C_{\varepsilon\perp}$ 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 $C_{\varepsilon\perp}' C_\varepsilon = 0$. We can now isolate the moment conditions for estimating β from (8) as $\mathbb{E}(C_{\varepsilon\perp}'[(y_n - x_n\beta) \otimes y_n]) = 0$, or

$$\mathbb{E}\left[C_{\varepsilon\perp}'(I_T \otimes y_n)(y_n - x_n\beta)\right] = 0. \tag{9}$$

So the (panel) IVs that are implied by the structure on Σ_ε are $Z_n = (I_T \otimes y_n)' C_{\varepsilon\perp}$, with Z_n the IV matrix as described in section 2.3. Finally, conditional on β, π_ε is identified from the complementary transformation of (8):

$$\mathbb{E}\left\{(C_\varepsilon' C_\varepsilon)^{-1} C_\varepsilon' \left[(y_n - x_n\beta) \otimes y_n\right] - \pi_\varepsilon\right\} = 0.$$

3.3 The random effects model reconsidered

By way of illustration, let us reconsider the classical error-components (random effects) model as discussed in section 3.1, with $T = 3$. The error covariance matrix is

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

so $r_\varepsilon = 2$,

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

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

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

As shown in section 3.1, 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. 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 $\mathbb{E}(y_{ns}x_{nt} - y_{nt}x_{ns})\beta = 0$ with $s \neq t$. With $\beta \neq 0$, β drops out of these three moments. Thus, they reflect the assumption of symmetry of Σ_{xy} that is implied by the model.

3.4 Extensions

We have derived these moment conditions in a model with a single regressor, which is too simple for any empirical application. However, our findings remain valid in a quite general setting, with an arbitrary number of regressors, with or without measurement error. To see this, consider the general linear model

$$\begin{aligned} y_n &= \Xi_n\beta + R_n\gamma + \varepsilon_n \\ X_n &= \Xi_n + V_n, \end{aligned}$$

the reduced form of which is

$$\begin{aligned} y_n &= X_n\beta + R_n\gamma + u_n \\ u_n &= \varepsilon_n - V_n\beta, \end{aligned}$$

with X_n a $T \times k$ matrix of observations on k variables with measurement error, true value Ξ_n, measurement error V_n, and R_n a $T \times \ell$ matrix of observations on ℓ correctly measured exogenous variables; β is now a k-vector and γ is an ℓ-vector. Adapting the moment conditions (9), we have

$$\begin{aligned}\mathbb{E}\left[C'_{\varepsilon\perp}(I_T \otimes y_n)(y_n - X_n\beta - R_n\gamma)\right] &= C'_{\varepsilon\perp}\,\mathbb{E}\left[(\varepsilon_n - V_n\beta) \otimes (\Xi_n\beta + R_n\gamma + \varepsilon_n)\right] \\ &= C'_{\varepsilon\perp}\operatorname{vec}\Sigma_\varepsilon \\ &= 0, \end{aligned} \tag{10}$$

due to the exogeneity of R_n. So the moment conditions (9) remain valid after this straightforward adaptation. The panel IVs that are implied by the structure on Σ_ε, $Z_n = (I_T \otimes y_n)'C_{\varepsilon\perp}$, are unaltered.

Another adaptation regards the identification of the regression parameters. The addition of the ℓ parameters in γ can be covered by the ℓ moment conditions

$$\mathbb{E}(R'_n u_n) = 0. \tag{11}$$

As to the identification of β, because of the $T(T-1)/2$ symmetry conditions that do not involve β, the number of informative (on β) moment conditions is $T(T+1)/2 - r_\varepsilon$, so a necessary condition for identification is $T(T+1)/2 - r_\varepsilon \geq k$.

As briefly indicated in section 3.1, when using the structure on Σ_ε, a caveat is in order. To illustrate this, we return to the simplest case of a single regressor and notice that the columns of Z_n are valid IVs only when they are correlated with x_n,

$$\mathbb{E}(Z'_n x_n) = \mathbb{E}[C'_{\varepsilon\perp}(I_T \otimes y_n)x_n] = C'_{\varepsilon\perp}\sigma_\xi\beta \neq 0,$$

so Σ_ξ should not have the same structure as Σ_ε. If the two structures are close, the estimators will have a large variance, possibly too large to be of practical value.

4 The dynamic model

Until now we have considered the static model. The moment conditions (10) do not hold when the lagged dependent variable is included as a regressor; $Z_n = (I_T \otimes y_n)' C_{\varepsilon\perp}$ is no longer a valid set of IV's. To see this, consider the simplest model,

$$y_n = x_n \beta + y_{n,-1} \gamma + u_n, \tag{12}$$

where $y_{n,-1}$ is the vector with t-th element $y_{n,t-1}$. (Note that with this notation, $t = 0$ is the first time point for which y_{nt} is observed.) In the place of (10), we now obtain

$$\begin{aligned}\mathbb{E}\left[C'_{\varepsilon\perp}(I_T \otimes y_n)(y_n - x_n\beta - y_{n,-1}\gamma)\right] &= \mathbb{E}\left\{C'_{\varepsilon\perp}\left[(\varepsilon_n - v_n\beta) \otimes (\xi_n\beta + y_{n,-1}\gamma + \varepsilon_n)\right]\right\} \\ &= C'_{\varepsilon\perp}\,\mathbb{E}(\varepsilon_n \otimes y_{n,-1})\gamma.\end{aligned}$$

Because $y_{n,-1}$ contains elements of ε_n, the last expectation is nonzero. Moreover, this expectation will generally not have the same structure as $\operatorname{vec}\Sigma_\varepsilon$, so that the nonzero elements are not eliminated by premultiplication with $C'_{\varepsilon\perp}$. The extension to more regressors with or without measurement error is straightforward.

Notice that this discussion is about measurement error in x_n, not in y_n. As to the latter, there is a growing literature, see e.g. Meijer, Spierdijk, and Wansbeek (2013), Biørn (2014), Gospodinov, Komunjer, and Ng (2014), and Lee, Moon, and Zhou (2014).

4.1 Example

By way of illustration, consider the random effects example from section 3.3 with the lagged dependent variable added. Still assuming that $\mathbb{E}(\xi_{ns}\varepsilon_{nt}) = 0$ for all s and t, we find that

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

provided that $|\gamma| < 1$. After a bit of algebra, we obtain

$$\mathbb{E}\left[C'_{\varepsilon\perp}(I_T \otimes y_n)(y_n - x_n\beta - y_{n,-1}\gamma)\right] = \gamma\sigma_w^2 \begin{pmatrix} 0 \\ 0 \\ \gamma \\ 0 \\ 1 \\ \gamma \\ 1 \end{pmatrix}. \tag{13}$$

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

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & -1 \end{pmatrix} C'_{\varepsilon\perp}\, \mathbb{E}\left[(I_T \otimes y_n)(y_n - x_n\beta - y_{n,-1}\gamma)\right] = 0,$$

so now

$$Z'_n = \begin{pmatrix} y_{n1} & -y_{n2} & 0 \\ 0 & y_{n2} & -y_{n3} \\ 0 & -y_{n1} & y_{n1} \\ 0 & y_{n1} & -y_{n2} \\ y_{n2} & -y_{n3} & 0 \end{pmatrix}.$$

contains valid instruments. This result is highly case-specific but the point has some generality. Adding the lagged dependent variable to the model invalidates the moment conditions but an adaptation may offer a way out, with fewer moment conditions.

4.2 Arellano-Bond estimation

In the context of dynamic models, the dominant assumption on Σ_ε is that it is a diagonal matrix, possibly with equal elements. This assumption plays a crucial role in the validity of the instruments on which the Arellano-Bond (1991) estimator is based. This estimator takes the model in first differences and uses lags of y as instruments. With measurement error, we get from (12)

$$\Delta y_n = \Delta x_n \beta + \Delta y_{n,-1}\gamma + \Delta u_n,$$

with Δ taking first differences. For example, the first element of Δy_n is $y_{n2} - y_{n1}$. The Arellano-Bond instruments are

$$Z_n' = \begin{pmatrix} y_{n0} & & & & \\ & y_{n0} & & & \\ & y_{n1} & & & \\ & & y_{n0} & & \\ & & y_{n1} & & \\ & & y_{n2} & & \\ & & & \ddots & \\ & & & & y_{n0} \\ & & & & \vdots \\ & & & & y_{n,T-2} \end{pmatrix},$$

of order $T(T-1)/2 \times (T-1)$. Still assuming that the measurement errors in x are independent of the other random terms in the equation. The presence of Δv_n in Δu_n does not affect the validity of the Arellano-Bond instruments. Moreover, these instruments also cover estimation of β; there is no need to search for new instruments when some ($< T(T-1)/2$) regressors are measured with error.

However, this does not necessarily exhaust the set of available instruments. Let y_{n+} be the vector of all observed y_{nt}, so $y_{n+} \equiv (y_{n0}, \ldots, y_{nT})'$ and let Q of order $(T-1) \times (T+1)$

be such that

$$\operatorname{vec} Q = \mathbb{E}\left[(y_{n+} \otimes I_{T-1})\Delta u_n\right].$$

Then

$$Q \equiv \mathbb{E}(\Delta u_n y'_{n+}) = \begin{pmatrix} 0 & \pi_{21} & \pi_{22} & \cdots & \pi_{2,T-1} & \pi_{2T} \\ 0 & 0 & \pi_{32} & \cdots & \pi_{3,T-1} & \pi_{3T} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \pi_{T,T-1} & \pi_{TT} \end{pmatrix},$$

where the πs are functions of γ and the parameters on the diagonal of Σ_ε. In general, there is no linear structure that can be exploited beyond the zeros (which correspond with the Arellano-Bond instruments), unless the elements on the diagonal of Σ_ε are equal. Then the elements on each of the pseudo-diagonals of Q are equal, which can be exploited to generate instruments. As before, these instruments also help to increase the asymptotic efficiency in estimating γ; they are not in any sense specific to the measurement error problem. The latter kind of instruments were first described by Ahn and Schmidt (1995).

Blundell and Bond (1998) spawn yet another set of moment conditions, which usually greatly improve on Arellano-Bond and which can also be used for the estimation of β just like Arellano-Bond. Linear restrictions on Q may carry over to a corresponding matrix for Blundell-Bond.

5 Using an additional exogenous regressor

In the analysis above, the role played by the exogenous variables R_n was restricted. Their only property used, cf. (11), 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 identify and estimate β, even when there are no restrictions on Σ_ε at all. We now turn to this.

When the model contains an exogenous regressor r_n in addition to ξ_n, this can be used to construct additional moment conditions that allow for consistent estimation of the parameters. Intuitively, the reason for this is that the additional regressor also adds exclusion restrictions, because r_{nt} is included in the equation for y_{nt} but not in the equation for y_{ns}, $s \neq t$. Formally, consider the model

$$\begin{aligned} y_n &= \xi_n \beta + r_n \gamma + \varepsilon_n \\ x_n &= \xi_n + v_n, \end{aligned}$$

with r_{nt} also in deviation of the mean at time t. In this context, we define strong exogeneity as

$$\mathbb{E}(r_{ns} u_{nt}) = 0 \qquad \text{for all } s \text{ and } t, \tag{14}$$

where $u_n = y_n - x_n \beta - r_n \gamma$. Weak exogeneity is defined in the same way, except that the orthogonality condition in (14) is only assumed to hold for $s \leq t$. The analysis here will focus on strong exogeneity. The adaptations to weak exogeneity are minor and straightforward. We first consider the advantage of having an exogenous variable available. There appears to be a caveat here, and we can only truly benefit from it when this exogenous variable has a heteroskedastic relation with the variable that was already present in the model.

5.1 Benefit and caveat

In section 3.4, we already showed how γ is identified with the additional moment condition $\mathbb{E}(r_n' u_n) = 0$. However, we can increase the efficiency of the estimators by using all T^2 moment conditions from (14). These conditions include $\mathbb{E}(r_n' u_n) = 0$ as $r_n' u_n = (\operatorname{vec} I_T)'(r_n \otimes u_n)$; the additional moment conditions can be shown to be nonredundant in the sense of Breusch, Qian, Schmidt, and Wyhowski (1999).

But the presence of an additional, exogenous regressor can pay off in a much more general sense. The T^2 moment conditions from (14) by themselves allow for consistent estimation of the model, without recourse to restrictions on the error covariance matrix Σ_ε. This approach is straightforward and seemingly attractive. The Jacobian matrix of the moment conditions is

$$J \equiv \mathbb{E}\left(\frac{\partial(r_n \otimes u_n)}{\partial(\beta, \gamma)}\right) = -(\operatorname{vec} \Sigma_{xr}, \operatorname{vec} \Sigma_r), \tag{15}$$

where $\Sigma_{xr} \equiv \mathbb{E}(x_n r_n')$ and $\Sigma_r \equiv \mathbb{E}(r_n r_n')$. Identification of β and γ from (14) requires J to be of full column rank. This condition will be fulfilled in most cases, but the asymptotic variance of the estimators of β and γ 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 (14). When the model 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 this seemingly attractive approach is not recommended.

5.2 Heteroskedasticity

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

$$\xi_n = Kr_n + \omega_n, \tag{16}$$

where $K \equiv \mathbb{E}(\xi_n r_n')[\mathbb{E}(r_n r_n')]^{-1}$. With $w_n \equiv v_n + \omega_n$ and $\kappa \equiv \operatorname{vec} K$,

$$x_n = Kr_n + w_n = (r_n \otimes I_T)'\kappa + w_n.$$

By the definition of ω_n and the assumption that v_n is independent of r_n, $\mathbb{E}(w_n r_n') = 0$ or, arranged differently,

$$\mathbb{E}(r_n \otimes w_n) = 0. \tag{17}$$

Now consider the situation where the relation (16) 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

$$\mathbb{E}(r_n \otimes \omega_n \otimes \omega_n) \neq 0. \tag{18}$$

Let

$$h_n \equiv \begin{pmatrix} r_n \otimes w_n \\ r_n \otimes u_n \\ r_n \otimes w_n \otimes u_n \end{pmatrix}, \tag{19}$$

where w_n is now shorthand for $x_n - (r_n \otimes I_T)'\kappa$. In view of (14), (17), the various independence assumptions made, and the assumption that r_n is in deviation of its mean, $\mathbb{E}(h_n) = 0$ and hence can be used for GMM estimation of the model parameters $\theta \equiv (\kappa', \beta, \gamma)'$.

There are T^2+2 parameters, and their identification depends on the rank of the matrix $G \equiv \mathbb{E}(\partial h_n/\partial\theta')$. Since

$$\frac{\partial h_n}{\partial \theta'} = -\begin{pmatrix} r_n r_n' \otimes I_T & 0 & 0 \\ 0 & r_n \otimes x_n & r_n \otimes r_n \\ r_n r_n' \otimes I_T \otimes u_n & r_n \otimes w_n \otimes x_n & r_n \otimes w_n \otimes r_n \end{pmatrix},$$

we obtain

$$G = -\begin{pmatrix} \Sigma_r \otimes I_T & 0 & 0 \\ 0 & \operatorname{vec}\Sigma_{xr} & \operatorname{vec}\Sigma_r \\ 0 & q_1 & q_2 \end{pmatrix}, \tag{20}$$

with

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

In view of (18), we conclude that the heteroskedasticity adds T^3 moment conditions. If the projection (16) 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 $q_2 \neq 0$.

Given identification, the parameters can be estimated by (nonlinear) GMM based on h_n. However, the structure of h_n suggests a simpler two-step method: First estimate κ by OLS, which amounts to using only the first moment condition. Second, use the first-step estimate $\hat{\kappa}$ to construct $\hat{w}_n \equiv x_n - (r_n \otimes I_T)'\hat{\kappa}$ and estimate β and γ by panel IV with y_n as dependent variable, x_n and r_n as explanatory variables, and

$$\hat{Z}'_{h,n} \equiv r_n \otimes \begin{pmatrix} 1 \\ \hat{w}_n \end{pmatrix} \otimes I_T \tag{21}$$

as instruments. Although $\hat{w}_n$ is estimated, the IV standard errors are correct. This is due to the fact that the *instruments* are estimated in the first step and not the covariates.

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 $r_{ns}\hat{w}_{nk}$, $s, k = 1, \ldots, T$, as instruments for the t-th time point and zeros for the other time points.

4. With the instruments defined like this, compute a standard IV estimator (2SLS, GMM, LIML, etc.) and use panel-robust standard errors.

5.3 Discussion

It is of interest to compare the Jacobians (15) and (20). The two columns in (15) may be highly collinear, which makes it hard to obtain estimators with decent small-sample properties from an additional regressor. In (20), these two columns are "stretched" with q_1

and q_2, which may decrease collinearity and hence the problem in obtaining satisfactory estimators.

Note that linearity of the regression of x_n on r_n is helpful under heteroskedasticity, because it leads to $q_2 = 0$, whereas 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.

Also for this approach, it is interesting to see how the validity is affected when regressors are added. The IV's 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.

6 Non-zero third moments

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

6.1 Motivating example

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 $r_{nt} = x_{nt} y_{nt}$, that is, it is based on the moment condition

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

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. In a small simulation study, Van Montfort, Mooijaart, and De Leeuw (1987) found that this estimator has reasonably good statistical properties in moderately sized samples.

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

$$\mathbb{E}\left[x_{nr} y_{ns}(y_{nt} - x_{nt}\beta)\right] = \mathbb{E}\left[(\xi_{nr} + v_{nr})(\xi_{ns}\beta + \varepsilon_{ns})(\varepsilon_{nt} - v_{nt}\beta)\right] = 0,$$

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

6.2 Moment conditions

As in the motivating example, we restrict ourselves here to third moments. 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

$$\mathbb{E}(y_n \otimes y_n \otimes y_n) = \lambda_\xi \beta^3 + \lambda_\varepsilon \tag{22a}$$

$$\mathbb{E}(y_n \otimes y_n \otimes x_n) = \lambda_\xi \beta^2 \tag{22b}$$

$$\mathbb{E}(y_n \otimes x_n \otimes x_n) = \lambda_\xi \beta \tag{22c}$$

$$\mathbb{E}(x_n \otimes x_n \otimes x_n) = \lambda_\xi + \lambda_v. \tag{22d}$$

These expressions owe their simplicity to the assumed independence between ξ_n, ε_n, and v_n. As mentioned in the motivating example, 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 (22c) from (22b) gives

$$\mathbb{E}\left[y_n \otimes (y_n - x_n\beta) \otimes x_n\right] = \mathbb{E}\left[(y_n \otimes I_T \otimes x_n)(y_n - x_n\beta)\right] = 0. \tag{23}$$

Hence, $Z_{yx,n} \equiv (y_n \otimes I_T \otimes x_n)'$ is a valid instrument matrix ($\mathbb{E}(Z'_{yx,n}x_n) = \lambda_\xi \beta \neq 0$). In scalar notation, the set of moments (23) is

$$\mathbb{E}\left[y_{ns}x_{nk}(y_{nt} - x_{nt}\beta)\right] = 0, \quad k, s, t = 1, \ldots, T. \tag{24}$$

Thus, (23) amounts to ordinary IV with T^3 instruments, divided in T sets: set t has T^2 instruments $(y_{ns}x_{nk})$, $k, s = 1, \ldots, T$ that multiply with $(y_{nt} - x_{nt}\beta)$ and zeros that multiply with $(y_{n\tau} - x_{n\tau}\beta)$ for $\tau \neq t$.

Although this is the most useful form of the moments for implementation in a standard statistics/econometrics software package, we can linearly transform the moment conditions to highlight that they include a set of symmetry conditions. In particular, let $h_n(k, s, t) \equiv y_{ns} x_{nk}(y_{nt} - x_{nt}\beta)$. Then $h_n(k, s, t) - h_n(t, s, k) = y_{ns}(x_{nk}y_{nt} - x_{nt}y_{nk})$ and $h_n(k, s, t) - h_n(k, t, s) = x_{nk}(x_{nt}y_{ns} - x_{ns}y_{nt})\beta$, which are both symmetry conditions that do not depend on β (assuming $\beta \neq 0$ for the latter). The moments with other permutations of $\{k, s, t\}$ can be similarly transformed, so that we can nonsingularly transform the system of moment conditions (24) to a set of $T(T + 1)(T + 2)/6$ moment conditions

$$\mathbb{E}\left[y_{ns} x_{nk}(y_{nt} - x_{nt}\beta)\right] = 0, \quad k, s, t = 1, \ldots, T; k \leq s \leq t, \tag{25}$$

and a set of $T^3 - T(T + 1)(T + 2)/6$ symmetry conditions that do not involve β.

Conditional on β, (22a) identifies λ_ε, (22d) identifies λ_v, and (22c) identifies λ_ξ. Also, the matrices $Z_{yy,n} \equiv (y_n \otimes y_n \otimes I_T)'$ and $Z_{xx,n} \equiv (I_T \otimes x_n \otimes x_n)'$, of the same order, are valid instrument matrices if the third moments of ε and v, respectively, are assumed to vanish.

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. The method is generally not considered favorably by applied researchers, because the results are often not very robust as the instruments are often weak. In the panel data model context, the situation may be more favorable as 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. We will investigate this issue through simulation in section 9.

When exogenous regressors are added to the model, nothing essential changes and therefore these instruments are still valid. Now let us look at what happens when we add the lagged dependent variable as a regressor. Then

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

This is zero, because we can write $y_{n,-1}$ as an infinite sum of terms of the form $\varepsilon_{n,-j}$ and $\xi_{n,-k}$ 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 $t < 0$) vanish and $Z_{xx,n}$ is still a valid instrument matrix under the assumption that the third moments of v_n vanish.

7 Heteroskedasticity

Apart from the relation between the two regressors in our third approach, we have assumed that relations are homoskedastic. This may be undesirably strong. In our first approach, we considered linear restrictions on the error covariance matrix, that is, $\operatorname{vec}\Sigma_\varepsilon = C_\varepsilon\pi_\varepsilon$, with C_ε a known matrix of constants and π_ε a vector of unknown parameters. With heteroskedasticity, it is most natural to assume that $\operatorname{vec}\Sigma_{\varepsilon n} = C_\varepsilon\pi_{\varepsilon n}$. 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 if the measurement error is heteroskedastic.

The estimators that use third moments do not accommodate heteroskedasticity as easily. The moment condition (22b) is only valid if $\mathbb{E}(\varepsilon_n \otimes \varepsilon_n \otimes \xi_n) = 0$ and $\mathbb{E}(\xi_n \otimes v_n \otimes v_n) = 0$. 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 $\mathbb{E}(\varepsilon_{nt}\varepsilon_{ns}\xi_{nr}) = 0$ and $\mathbb{E}(v_{nt}v_{ns}\xi_{nr}) = 0$ if r, 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 8), even if $\mathbb{E}(\varepsilon_{nt}^2\xi_{nt})$ or $\mathbb{E}(v_{nt}^2\xi_{nt})$ are allowed to be nonzero.

For the estimators that use an additional regressor, the most important assumption is 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. Here too, we will still be able to identify the parameters if the assumption is violated only for a subset of the elements.

If under heteroskedasticity a set of instruments has been selected that identify the parameters, the IV estimators are consistent, but the default IV standard errors are incorrect. However, the default GMM standard errors are robust to heteroskedasticity, so it is straightforward to obtain correct standard errors.

8 Individual effects

We now extend the model to include individual effects,

$$y_n = \xi_n \beta + \iota_T \alpha_n + \varepsilon_n \tag{26}$$

$$x_n = \xi_n + v_n, \tag{27}$$

where α_n is the random, mean-zero individual effect. If it is uncorrelated with ξ_n we have a *random effects* model and it can be subsumed in ε_n. With restrictions on the covariance matrix of the time-varying error term, this potentially leads to restrictions on Σ_ε. This case has been dealt with in section 3.

If α_n is correlated with ξ_n we have a model that is commonly called a *fixed effects* model, which requires a separate treatment. We can for instance employ a matrix B, of order $T \times T$ and rank $T-1$, with property $B'\iota_T = 0$. We put a tilde on a T-vector when it has been pre-multiplied by B'.

We first consider the moment conditions from the restrictions on the covariance matrices. Conceptually the simplest approach is the usual approach with fixed-effects models, which is to transform the model by eliminating the individual effects through some choice of B matrix, and then proceed with the model for transformed variables. In our context, the essential change concerns (8), $\mathbb{E}[(y_n - x_n\beta) \otimes y_n - C_\varepsilon \pi_\varepsilon] = 0$. After transformation, this becomes

$$\mathbb{E}\left[(\tilde{y}_n - \tilde{x}_n\beta) \otimes \tilde{y}_n - (B \otimes B)' C_\varepsilon \pi_\varepsilon\right] = 0. \tag{28}$$

So the IVs are now based on the complement of $(B \otimes B)' C_\varepsilon$ rather than C_ε.

However, there is an avoidable loss of efficiency in this approach. If the individual effects are not correlated with ε_{nt} and v_{nt}, $\mathbb{E}[(\varepsilon_n - v_n\beta) \otimes y_n] = \operatorname{vec}\Sigma_\varepsilon$ still holds. Hence, for estimation we only need to eliminate the α_n from $y_{nt} - x_{nt}\beta$. Consequently,

$$\mathbb{E}\left[(\tilde{y}_n - \tilde{x}_n\beta) \otimes y_n - (B \otimes I_T)' C_\varepsilon \pi_\varepsilon\right] = 0$$

is a larger set of moment conditions, generating more IVs than (28).

The adaptation of the results with an additional exogenous regressor in section 5 to the inclusion of fixed effects is relatively simple. The key element is that in (19), u_n is replaced by $\tilde{u}_n = \tilde{y}_n - \tilde{x}_n\beta$. This eliminates the individual effect, while after this transformation, the analog of (19) still holds. Hence, we then estimate the regressions with the transformed variables, but with the same instruments as in section 5.

We next consider the third moments when there are individual effects. 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 (23)

$$\mathbb{E}\left[\tilde{y}_n \otimes (\tilde{y}_n - \tilde{x}_n\beta) \otimes x_n\right] = \mathbb{E}\left[(\tilde{y}_n \otimes I_{T-1} \otimes x_n)(\tilde{y}_n - \tilde{x}_n\beta)\right] = 0, \tag{29}$$

so that $\tilde{Z}_{yx,n} \equiv (\tilde{y}_n \otimes I_{T-1} \otimes x_n)'$ is a valid instrument matrix. However, if we additionally assume that 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, $\check{Z}_{xx,n} \equiv (I_{T-1} \otimes x_n \otimes x_n)'$ is valid if the third moments of v_n are zero.

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 $\tilde{u}_n$. 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, as discussed in section 4.2.

Not all parameters may be identified in the presence of fixed effects. The main parameter of interest, β, is identified as discussed here, but only transformations of the parameters π_ε in the case of covariance restrictions, or λ_ξ, λ_ε, and λ_v in the case of third order moment estimation may be identified, depending on the restrictions imposed on these parameters by the model. This is similar to the multinomial probit model, where only the covariance matrix of the transformed error terms is identified unless the model sufficiently restricts the covariance matrix of the untransformed errors (e.g., a factor structure).

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

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

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

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

with $\zeta_{nt} \sim \sqrt{\frac{4}{3}}\chi_1^2$ and $\xi_{n0} = \sqrt{\frac{4}{3}}\zeta_{n0}$. 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 1,000, we generate 1,000 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. This setup implies that $\mathbb{V}(\xi_{nt}) = 32/9$ for all t, and thus $\mathbb{C}(y_{nt}, x_{nt}) = 32/9$, $\mathbb{V}(x_{nt}) = 32/9 + 1 = 41/9$, and $\mathbb{V}(y_{nt}) = 0.7 + 32/9 + 2 = 563/90$. Hence, the pooled OLS estimator of β (= 1) converges to $32/41 = 0.78$, with an R^2 of $32^2/(41 \cdot 563/10) = 0.44$. The standard fixed effects ("within") estimator subtracts the within-individual across-time average of y and x before performing OLS. Let $Q_T = I_T - \iota_T\iota_T'/T$ be the associated centering matrix. Then the within estimator converges to $\operatorname{tr}[Q_T\,\mathbb{C}(y_n, x_n')]/\operatorname{tr}[Q_T\,\mathbb{C}(x_n)] = \operatorname{tr}[Q_T\,\mathbb{C}(\xi_n)]/\{\operatorname{tr}[Q_T\,\mathbb{C}(\xi_n)]+(T-1)\sigma_v^2\}$. Using the stationary AR(1) structure of ξ_n, the numerical value of this for our setup is $37/52 = 0.71$, with an R^2 of $\{\operatorname{tr}[Q_T\,\mathbb{C}(\xi_n)]\}^2/(\{\operatorname{tr}[Q_T\,\mathbb{C}(\xi_n)] + (T-1)\sigma_v^2\}\,\{\beta^2 \operatorname{tr}[Q_T\,\mathbb{C}(\xi_n)] + (T-1)\sigma_\varepsilon^2\}) = 0.39$.

Below, we employ estimators based on the moment conditions derived in sections 3–8. Unless stated otherwise, all results are based on the optimally weighted GMM estimator, that is, the estimator based on (5) with $W = (\sum_n \hat{u}_n\hat{u}_n')^{-1}$. Throughout, we take the matrix B' equal to Q_T with the first row left out. We use clustered standard errors that are robust to time-series correlation and heteroskedasticity.

9.1 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 3.3. We leave out the moment conditions that do not involve β, which results in 13 instruments for RE and 9 for FE.[1]

Columns 2–5 of Table 1 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 are weak for $N = 100$, 200 and 500, and strong for $N = 1000$, if we take $F < 10$ as the threshold for weak instruments (e.g., Stock, Wright, and Yogo, 2002).

The simulation results for covariance matrix restrictions are given in Table 2, for random effects (column 2) and fixed effects (column 3). This table reports the average value of the GMM estimator $\hat{\beta}$ over the 1,000 replications ("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

[1]Throughout, the simulation results involving all moment conditions are slightly worse that the ones reported here. They are available upon request.

standard deviation of $\hat{\beta}$ over the replications, but but with sample sizes of 500 and larger, the difference is negligible.

Columns 2–3 of Table 3 report the corresponding rejection rates (in percentages) of the t-test over the 1,000 replications for random and fixed effects, 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 t-test 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 3 shows that the rejection rate of the t-tests is close to 5% when the sample standard deviation is used. For $N \leq 200$, the rejection rates exceed the 5% level when the formula-based standard error is used.

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, 1996). The rejection rates of the t-tests are 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 5, 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 for $N = 100$. In case of FE it even leads to a conservative standard error and a rejection rate slightly below the nominal level.

9.2 Third moments

We now exploit the third moments of the data. We use $T^3 = 125$ instruments for RE and $T(T-1)^2 = 80$ instruments for FE. If we also exploit that the third moments of ε_n and v_n vanish, this adds $2 \times 125 = 250$ instruments for RE, resulting in 375 instruments in total. For FE, the additional number of instruments is $(T-1)^3 + T^2(T-1) = 164$, resulting in 244 instruments in total.

Columns 6–9 of Table 1 give an impression of the strength of these instruments. The R^2s and F statistics are much higher than the ones for the covariance restrictions in columns 2–5.

The simulation results for the third-order restrictions are given in columns 4–9 of Table 2. For the GMM estimator exploiting third moments the weight matrix turns out to be 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).

The GMM estimators based on W_{OPT} and W_{2SLS} show some bias even for $N = 1,000$. Throughout, the FE estimator has a larger bias than the RE estimator. In contrast, for the GMM estimator with $W = I$ there is virtually no bias. As expected, the estimator based on $W = I$ has a larger variance, but the 2SLS estimator is at least as efficient as the asymptotically optimal one. For the 2SLS and identity-weighted estimators, the average formula-based standard error is 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 t-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 3. Even if the formula-based standard errors are replaced by the sample standard deviation of $\hat{\beta}$ across replications, the rejection rates are still too high (particularly for the FE estimator), although they are noticeably better than with the formula-based standard errors. Columns 6–9 show that the test results for the non-optimally weighted GMM estimators are much better, even 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 t-tests, except for the 2SLS estimator. This is illustrated in columns 4–9 of Table 5. Note that, as mentioned in section 9.1, 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 was known.

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).

It appears that 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 no bias, resulting in a considerably smaller (in a relative sense) mean squared error than the covariance restriction estimator.

9.3 Additional regressor

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

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

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

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

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

$$
\begin{aligned}
y_{nt} &= \alpha_n + \xi_{nt}\beta + z_{nt}\gamma + \varepsilon_{nt} \\
x_{nt} &= \xi_{nt} + v_{nt},
\end{aligned}
$$

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 $\gamma = 1$.

From (21) we see that the two-step approach gives us $T(T+1)T = 150$ IVs for RE and $T(T+1)(T-1) = 120$ IVs for FE. Table 4 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 t-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 close to 5% for the larger sample sizes. Again the bootstrap can be used to improve the rejection rates for small values of N, which is illustrated in the third panel of Table 5.

10 Conclusions

We have presented three ways to consistently estimate a linear panel data model with measurement error. We have departed from the existing literature by avoiding the hard-to-justify assumptions on the intertemporal covariance matrix of the measurement errors, and replaced them with various alternative assumptions, which researchers may be more comfortable with. Specifically, we consider restrictions on the intertemporal covariance matrix Σ_ε of the equation errors, exploiting third moments of the data, and using moments that involve an additional regressor to which the error-ridden regressor is heteroskedastically related. For each of these cases, we derive a set of moment conditions that can be used in a GMM procedure to obtain consistent estimators. These are all relatively simple IV estimators, although in the case of an additional regressor, this is a two-step IV estimator. We also adapt these estimators to accommodate fixed effects. The simulation results suggest that our three approaches work well, at least for the particular settings chosen in our simulation study.

When the moment conditions based on at least one of the three approaches have been selected, a standard J-test for over-identifying restrictions can be used to test the validity of the chosen moment conditions. In this way, the moment conditions derived in this paper greatly expand the toolkit of the applied researcher.

There are various directions for further research as to each of our three approaches. As to using restrictions on 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).

As to our second 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).

We have focused on the static panel data model, although we also showed how our results generalize to the dynamic panel data model, provided the measurement error is not in the (lagged) dependent variable. Meijer, Spierdijk, and Wansbeek (2013) explore the effect of measurement error in the dependent variable and propose a consistent estimator, for the simplest version of the model, where the lagged dependent variable is the only regressor and where the structure of the measurement error is kept very simple. One promising area of research is to explore the possibilities offered by the presence of further regressors, which are an underexploited source of instruments in the usual estimation of the dynamic panel data model and which (with caveats, see section 5) can yield instruments to deal with measurement error.

Acknowledgments

The authors are indebted to Ramses Abul Naga, Paul Bekker, participants at the Econometric Society North American Summer Meeting 2013, and two 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).

References

Aasness, J., E. Biørn, and T. Skjerpen (2003), "Distribution of preferences and measurement errors in a disaggregated expenditure system", *Econometrics Journal*, **6**, 374–400.

Ahn, S.C. and P. Schmidt (1995), "Efficient estimation of models for dynamic panel data", *Journal of Econometrics*, **68**, 5–27.

Almeida, H., M. Campello, and A.F. Galvao Jr. (2010), "Measurement errors in investment equations", *Review of Financial Studies*, **23**, 3279–3328.

Altonji, J.G. and L.M. Segal (1996), "Small-sample bias in GMM estimation of covariance structures", *Journal of Business & Economic Statistics*, **14**, 353–366.

Arellano, M. and S. Bond (1991), "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations", *Review of Economic Studies*, **58**, 277–297.

Biørn, E. (2000), "Panel data with measurement errors: instrumental variables and GMM procedures combining levels and differences", *Econometric Reviews*, **19**, 391–424.

Biørn, E. (2003), "Handling the measurement error problem by means of panel data: moment methods applied on firm data", in B. Stigum, editor, *Econometrics and the Philosophy of Economics*, Princeton University Press, Princeton.

Biørn, E. (2014), "Panel data dynamics with mis-measured variables: Modeling and GMM estimation", *Empirical Economics*, forthcoming.

Biørn, E. and J.T. Klette (1998), "Panel data with errors-in-variables: essential and redundant orthogonality conditions in GMM estimation", *Economics Letters*, **59**, 275–282.

Biørn, E. and J.T. Klette (1999), "The labour input response to permanent changes in output: an errors-in-variables analysis based on panel data", *The Journal of Industrial Economics*, **47**, 451–476.

Biørn, E. and J. Krishnakumar (2008), "Measurement errors and simultaneity", in L. Mátyás and P. Sevestre, editors, *The econometrics of panel data*, Springer.

Blundell, R. and S. Bond (1998), "Initial conditions and moment restrictions in dynamic panel data models", *Journal of Econometrics*, **87**, 115–143.

Bramati, M. C. and C. Croux (2007), "Robust estimators for the fixed effects panel data model", *Econometrics Journal*, **10**, 521–540.

Breusch, T., H. Qian, P. Schmidt, and D.J. Wyhowski (1999), "Redundancy of moment conditions", *Journal of Econometrics*, **91**, 89–111.

Cameron, C. and P.K. Trivedi (2005), *Microeconometrics*, Cambrigde University Press, Cambridge, UK.

Dagenais, M.G. and D.L. Dagenais (1997), "Higher moment estimators for linear regression models with errors in the variables", *Journal of Econometrics*, **76**, 193–221.

Erickson, T. and T.M. Whited (2002), "Two-step GMM estimation of the errors-in-variables model using high-order moments", *Econometric Theory*, **18**, 776–799.

Fazzari, S.M., R.G. Hubbard and B.C. Petersen (1988), "Financing constraints and corporate investment", *Brookings Papers on Economic Activity 1*, 141–195.

Geary, R.C. (1942), "Inherent relations between random variables", *Proceedings of the*

Royal Irish Academy A, **47**, 63–67.

Gospodinov, N., I. Komunjer, and S. Ng (2014), *Minimum Distance Estimation of Dynamic Models with Errors-In-Variables*, Working paper.

Griliches, Z. and J.A. Hausman (1986), "Errors in variables in panel data", *Journal of Econometrics*, **31**, 93–118.

Hall, P. and J.L. Horowitz (2012), "Bootstrap critical values for tests based on generalized-method-of-moments estimators", *Econometrica*, **64**, 891–916.

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

Lewbel, A. (1996), "Demand estimation with expenditure measurement errors on the left and right hand side", *The Review of Economics and Statistics*, **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", *Journal of Business & Economic Statistics*, **30**, 67–80.

Meijer, E., L. Spierdijk, and T.J. Wansbeek (2013), "Measurement error in the linear dynamic panel data model", in B.C. Sutradhar, editor, *Longitudinal data analysis subject to measurement error, missing values and/or outliers*, Springer, New York.

Mooijaart, A. and A. Satorra (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.

Pal, M. (1980), “Consistent moment estimators of regression coefficients in the presence of errors in variables”, *Journal of Econometrics*, **14**, 349–364.

Shao, J., Z. Xiao, and R. Xu (2011), “Estimation with unbalanced panel data having covariate measurement error”, *Journal of Statistical Planning and Inference*, **141**, 800–808.

Stock, J. H., J. H. Wright, and M. Yogo (2002), “A survey of weak instruments and weak identification in generalized method of moments”, *Journal of Business and Economic Statistics*, **20**, 518–529.

Wagenvoort, R. and R. Waldmann (2002), “On B-robust instrumental variable estimation of the linear model with panel data”, *Journal of Econometrics*, **106**, 297–324.

Wansbeek, T.J. (2001), “GMM estimation in panel data models with measurement error”, *Journal of Econometrics*, **104**, 259–268.

Wansbeek, T.J. and R.H. Koning (1991), “Measurement error and panel data”, *Statistica Neerlandica*, **45**, 85–92.

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

Xiao, Z., J. Shao and M. Palta (2010a), “Instrumental variable and GMM estimation for panel data with measurement error”, *Statistica Sinica*, **20**, 1725–1747.

Xiao, Z., J. Shao and M. Palta (2010b), “GMM in linear regression for longitudinal data with multiple covariates measured with error", *Journal of Applied Statistics*, **37**, 791–805.

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

Table 1: **Strength of instruments**

	cov. restr.				*3rd order mom.*			
	RE		FE		RE		FE	
N	$\bar{R}^2$	F	$\bar{R}^2$	F	$\bar{R}^2$	F	$\bar{R}^2$	F
100	0.05	3.14	0.04	3.09	0.43	12.06	0.37	9.03
200	0.04	4.41	0.03	4.71	0.43	22.77	0.36	16.44
500	0.04	8.22	0.03	9.78	0.42	53.76	0.36	37.92
1000	0.03	14.67	0.03	18.13	0.42	103.93	0.35	72.97

Table 2: **Covariance restrictions and 3rd order moments: bias and variance**

	cov. restr.		*3rd order mom.*					
			W_{OPT}		W_{2SLS}		$W = I$	
	RE	FE	RE	FE	RE	FE	RE	FE
***N* = 100**								
avg. $\hat{\beta}$	97	96	96	91	98	93	100	99
sample $\sigma(\hat{\beta})$	187	208	75	94	72	88	96	109
avg. $\hat{\sigma}(\hat{\beta})$	135	160	27	33	71	84	84	89
***N* = 200**								
avg. $\hat{\beta}$	98	97	97	94	99	96	100	100
sample $\sigma(\hat{\beta})$	133	150	53	63	49	59	66	72
avg. $\hat{\sigma}(\hat{\beta})$	116	131	27	33	52	62	64	67
***N* = 500**								
avg. $\hat{\beta}$	99	99	98	96	100	98	100	100
sample $\sigma(\hat{\beta})$	90	94	35	42	33	40	43	46
avg. $\hat{\sigma}(\hat{\beta})$	85	92	23	28	34	41	43	44
***N* = 1000**								
avg. $\hat{\beta}$	100	99	99	98	100	99	100	100
sample $\sigma(\hat{\beta})$	66	68	24	28	23	28	28	31
avg. $\hat{\sigma}(\hat{\beta})$	63	67	19	22	25	29	31	32

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 1,000.

Table 3: **Covariance restrictions and 3rd order moments: rejection rates (in %)**

	cov. restr.		*3rd order mom.*					
			W_{OPT}		W_{2SLS}		$W = I$	
	RE	FE	RE	FE	RE	FE	RE	FE
***N* = 100**								
% rejections (sample $\sigma(\hat{\beta})$)	5	6	7	16	6	12	5	5
% rejections ($\hat{\sigma}(\hat{\beta})$)	16	13	55	68	7	15	12	15
***N* = 200**								
% rejections (sample $\sigma(\hat{\beta})$)	6	5	9	17	6	10	5	5
% rejections ($\hat{\sigma}(\hat{\beta})$)	10	9	42	51	5	9	8	9
***N* = 500**								
% rejections (sample $\sigma(\hat{\beta})$)	5	5	9	14	5	7	5	5
% rejections ($\hat{\sigma}(\hat{\beta})$)	6	5	28	35	5	8	6	8
***N* = 1000**								
% rejections (sample $\sigma(\hat{\beta})$)	5	4	7	12	6	6	5	5
% rejections ($\hat{\sigma}(\hat{\beta})$)	6	5	18	24	5	6	3	5

Table 4: **Exogenous regressors: bias, variance, and rejection rates**

	RE		FE		RE		FE	
	β	γ	β	γ	β	γ	β	γ
		N = 100				**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
		N = 200				**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 1,000. The rejection rates are in %.

Table 5: **Rejection rates (in %) for $N = 100$**

	cov. restr.		*3rd order 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	7	16	6	12	5	5	13	13	6	6
% rejections ($\hat{\sigma}(\hat{\beta})$)	16	13	55	68	7	15	12	15	18	18	10	7
% rejections (bootstrap $\hat{\sigma}(\hat{\beta})$)	5	3	9	23	7	18	7	9	15	15	8	5