WORKING P A P E R

Point and Set Identification in Linear Panel Data Models with Measurement Error

ERIK MEIJER, LAURA SPIERDIJK AND
TOM WANSBEEK

WR-941

March 2012

This paper series made possible by the NIA funded RAND Center for the Study of Aging (P30AG012815) and the NICHD funded RAND Population Research Center (R24HD050906).

This product is part of the RAND Labor and Population working paper series. RAND working papers are intended to share researchers' latest findings and to solicit informal peer review. They have been approved for circulation by RAND Labor and Population but have not been formally edited or peer reviewed. Unless otherwise indicated, working papers can be quoted and cited without permission of the author, provided the source is clearly referred to as a working paper. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors. **RAND®** is a registered trademark.

LABOR AND POPULATION

Point and set identification in linear panel data models with measurement error

Erik Meijer

Laura Spierdijk

Tom Wansbeek*

March 9, 2012

Abstract

The rich dependency structure of panel data can be exploited to generate moment conditions that can be used to identify linear regression models in the presence of measurement error. We add to a small body of literature on this topic by showing how heteroskedasticity and nonlinear relationships between the error-ridden regressors and error-free regressors lead to identifying moment conditions in a static panel data setting, how suitably chosen linear combinations of lagged and lead values of the dependent variable can be used as instrumental variables in a dynamic panel data with measurement errors, and how reverse regression can be generalized to the panel data setting, thereby giving bounds on regression coefficients in the absence of point identification.

JEL: C23, C26

1 Introduction

Measurement error in panel data is an issue that has been the subject of a number of papers since the pioneering contribution of Griliches and Hausman (1986). Spierdijk and Wansbeek (2012), referred as SW further on, review much of this literature. In general, the richness of panel data can offer a way out of the identification problem that hampers consistent estimation of a (linear) regression model based on a single cross section.

Up till now, the literature has focused on exploiting restrictions on the $T \times T$ covariance matrix of the measurement errors over the T time periods covered by the panel data set. These restrictions provide the identifying information needed. However, as SW argue, these restrictions are often poorly justified and they suggest to consider other elements of the model beyond the

*Meijer: RAND Corporation, meijer@rand.org, Spierdijk and Wansbeek: Faculty of Economics and Business, University of Groningen, l.spierdijk@rug.nl, t.j.wansbeek@rug.nl.

measurement error covariance matrix. Elsewhere, sources of identification may be found that have better justification. Hence SW considers consistent estimation based on moment conditions following from restrictions on the $T \times T$ covariance matrix of the errors in the equations. This matrix is generally taken to be highly restricted, spawning many instruments, as SW shows.

Yet these restrictions, although more or less generically imposed, are also often poorly justified, leading to a search of instruments not based on restrictions on the parameter space. Such instruments can be derived from the third moments of the incorrectly measured regressor. If these third moments are not zero, valid instruments are easily constructed, as shown by SW. The “credibility” of these instruments is much higher since, somewhat informally speaking, the third moments are nonzero almost everywhere in the set of all distribution functions, whereas the restrictions on the covariance matrices are a set of measure zero in the space of the parameters of the covariance matrices. In principle, this may give rise to a many weak instruments problem, but the simulations in SW show promise in using a limited set of principal components of the many instruments.

A third source of instruments can come from an exogenous variable if present in the model. SW shows how the time dimension in the data can be exploited to derive instruments.

In this paper we cover three topics supplementing the analysis by SW. First, in section 2, we show how heteroskedasticity in a model that includes an exogenous variable can be exploited for consistent estimation. We extend, to the panel data model case, results due to Lewbel (2011).

Next, we consider consistent estimation in a dynamic panel data model with the lagged dependent variable as its only regressor. The only variable in this model is measured with error. For the case without measurement error, Wansbeek and Bekker (1996) derive an estimator that is asymptotically efficient in the class of all consistent estimators that are linear functions of the variable. In section 3 we consider the case where measurement error is present.

Third, we return to the static case and address the issue of what to do when none of the assumptions leading to moment conditions and consistent estimation are deemed justified, and the model is not identified. However, when a parameter is not “point identified”, it may still be “set identified”, that is, there may an interval whose bounds are consistently estimable and that limits the true value of the parameters. In section 4 we show how such an interval can be constructed by extending the notion of reverse regression to the present case.

2 Exploiting heteroskedasticity

SW describe how the third moments of the observables, if non-zero, can be exploited to derive consistent estimators in a panel data model with measurement error. Their approach extends, to the panel data case, ideas that have been known and elaborated for the case of a single cross section, cf. Wansbeek and Meijer (2000), ch.6. Third moments also play a role in an approach

recently put forward by Lewbel (2011), for a single cross section. We extend his approach to the panel data case.

The model is a linear model with two regressors, one of which being subject to measurement error:

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

with y_n, x_n, z_n and ξ_n having mean zero, and ε_n and v_n random with mean zero, independent of each other and of z_n and ξ_n . The reduced form obtained by substituting out ξ_n is

$$y_n = x_n \beta + z_n \gamma + u_n,$$

with $u_n = \varepsilon_n - v_n \beta$. We consider the projection of ξ_n on z_n ,

$$\xi_n = \Lambda z_n + \omega_n, \quad (1)$$

so, with $w_n \equiv v_n + \omega_n$ and $\lambda \equiv \text{vec} \Lambda$,

$$\begin{aligned} x_n &= \Lambda z_n + w_n \\ &= (z_n \otimes I_T)' \lambda + w_n. \end{aligned}$$

By construction and the independence of z_n and v_n , there holds $E(z_n \omega_n') = E(z_n w_n') = 0$ or, arranged differently,

$$E(z_n \otimes \omega_n) = 0.$$

Now consider the situation where the relation (1) between ξ_n and z_n is heteroskedastic and $E(\omega_n \omega_n' | z_n)$ is a function of z_n . Thus, in general,

$$E(z_n \otimes \omega_n \otimes \omega_n) \neq 0. \quad (2)$$

Now let

$$h_n \equiv \begin{pmatrix} z_n \otimes u_n \\ z_n \otimes w_n \\ z_n \otimes u_n \otimes w_n \end{pmatrix},$$

where $u_n = y_n - x_n \beta - z_n \gamma$ and $w_n = x_n - (z_n \otimes I_T)' \lambda$. In view of the assumptions made, $E(h_n) = 0$ and hence can be used for generalized method of moments (GMM) estimation of the model parameters $\theta \equiv (\beta, \gamma, \lambda)'$. There are $T^2 + 2$ parameters, and their identification depends on the rank of the matrix $G \equiv E(\partial h_n / \partial \theta')$. Since

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

we obtain

$$G = - \begin{pmatrix} \text{vec}\Sigma_{xz} & \text{vec}\Sigma_z & 0 \\ 0 & 0 & \Sigma_z \otimes I_T \\ q_1 & q_2 & 0 \end{pmatrix},$$

with

$$\begin{aligned} \Sigma_{xz} &\equiv E(x_n z'_n) \\ &= \Lambda \Sigma_z \\ \Sigma_z &\equiv E(z_n z'_n) \\ q_1 &\equiv E(z_n \otimes x_n \otimes w_n) \\ &= E(z_n \otimes (\Lambda z_n + v_n + \omega_n) \otimes (v_n + \omega_n)) \\ &= (I_T \otimes \Lambda \otimes I_T) E(z_n \otimes z_n \otimes \omega_n) + E(z_n \otimes \omega_n \otimes \omega_n) \\ q_2 &\equiv E(z_n \otimes z_n \otimes w_n) \\ &= E(z_n \otimes z_n \otimes (v_n + \omega_n)) \\ &= E(z_n \otimes z_n \otimes \omega_n). \end{aligned}$$

In view of (2), we conclude that the heteroskedasticity adds T^3 moment conditions. If the projection (1) can be strengthened to $E(\xi_n | z_n) = \Lambda z_n$, then $E(z_n \otimes z_n \otimes \omega_n) = 0$, q_1 simplifies to $E(z_n \otimes \omega_n \otimes \omega_n)$, and $q_2 = 0$. Conversely, if $E(z_n \otimes z_n \otimes \omega_n) \neq 0$, the regression of ξ_n on z_n must be nonlinear and $q_2 \neq 0$.

It is of interest to compare the panel data case with $T \geq 2$ with the case of a single cross section, $T = 1$. In the latter case and without heteroskedasticity, G is a 2×3 matrix, obviously of rank 2, and the model is not identified. With heteroskedasticity, one valid moment condition is added to the model, rendering it just-identified. In the panel data case, without heteroskedasticity, G has $2T^2$ rows and $T^2 + 2$ columns and is of full column rank. The orthogonality condition $E(z_n \otimes u_n) = 0$ suffices for consistent estimation, as was noted and elaborated by SW. So in that sense the presence of heteroskedasticity is not needed.

Still, heteroskedasticity, if present, can be helpful in obtaining estimators with better small-sample properties. As is apparent from the two elements in the first row of G , the estimators of β and γ will have poor small-sample properties when x_n and z_n have a similar correlation structure over time ($\Lambda \approx c \cdot I_T$), a case that is not unlikely to happen in practice. Additional moment conditions then can be helpful, like those spawned by heteroskedasticity as in (2). Note that linearity of the regression of x_n on z_n is helpful here, 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 $E(z_n \otimes z_n \otimes \omega_n) \neq 0$ and $E(z_n \otimes \omega_n \otimes \omega_n) = 0$ is unlikely to be approximately met, so the additional moment conditions still add value.

The conditions $E(z_n \otimes u_n \otimes w_n) = 0$ involve third moments. Each of the additional moment conditions may contribute little, but there are many of them even for low T . For a related case with many moments based on third-order considerations in a panel data setting, SW find, in a simulation study, excellent results after condensing the multitude of instruments by using only the first few principal components of them.

3 Measurement error in a dynamic model

The linear dynamic panel data model is a hugely popular model, and in particular the estimator due to Arellano and Bond (1991) has found widespread application in economics and beyond. This estimator is an instrumental-variables (IV) estimator, the consistency of which is based on the assumption that the error terms in the equation, apart from the individual effect, are uncorrelated over time. Hence lagged values of the dependent variable can serve as instruments for the model in first-difference form. Ahn and Schmidt (1995) show how this assumed lack of correlation can be exploited to derive more orthogonality conditions than the ones underlying the Arellano-Bond estimator. Also, see Blundell and Bond (1998).

Wansbeek and Bekker (1996) presented a generalized approach to consistent estimation in the simple linear dynamic panel data model without exogenous regressors, by considering the covariance matrix over time of the only variable that is present in the model. This covariance matrix is structured through the model. The authors derive an IV estimator that is linear in the variable and that has minimal asymptotic variance. Harris and Mátyás (2000) extend this approach to include exogenous regressors. They compare the estimator thus defined with the Arellano-Bond estimator and some estimators based on Ahn and Schmidt (1995), to find that this estimator “generally outperformed all other estimators when T was moderate *in all of the situations that an applied researcher might encounter*” [italics in original].

In this section we extend the Wansbeek-Bekker estimator to the case where the variable of the model contains measurement error. We thus add to the literature on the dynamic panel data model, which is remarkably scarce given the popularity of the model without measurement error. An early contribution is Wansbeek and Kapteyn (1992). For the model without exogenous regressors, they derive the probability limit of the within estimator and the OLS estimator after first-differencing the model, and suggest to use the result to construct a consistent estimator of the autoregressive parameter. In an empirical study on income dynamics, Antman and McKenzie (2007) consider a dynamic panel data model where the current value depends on a cubic function of the lagged value. Their estimator is based on outside information on the reliability of the income variable, that is, on the ratio of the true variance and the observed variance. Chen, Ni and Podgursky (2008), in their study of the dynamics of students’ test scores, construct consistent estimators through instruments derived from within the model, adapting an approach due to Altonji and

Siow (1987) to the dynamic case. Komunjer and Ng (2011) consider a VARX model with all variables contaminated by measurement error and exploit the dynamics of the model for consistent estimation. Biørn (2012) presents a thorough treatment of the topic, with instruments based on the absence of correlation between regressors and disturbances for some combinations of time indices.

We now turn to the model and derive our estimator. With observations y_{nt} , $n = 1, \dots, N$, $t = 1, \dots, T$, we define

$$y_n \equiv \begin{pmatrix} y_{n1} \\ \vdots \\ y_{nT} \end{pmatrix} \quad y_{n,-1} \equiv \begin{pmatrix} y_{n0} \\ \vdots \\ y_{n,T-1} \end{pmatrix} \quad y_{n,+} \equiv \begin{pmatrix} y_{n0} \\ \vdots \\ y_{nT} \end{pmatrix}.$$

We assume that y_n satisfies

$$y_n = \eta_n + v_n, \quad (3)$$

where η_n is the unobserved true value and v_n is i.i.d. measurement error, $v_n \sim (0, \sigma_v^2 I_T)$. For η_n and v_n we use notation analogous to the one for y_n . The model is

$$\eta_n = \gamma \eta_{n,-1} + \alpha_n \iota_T + \varepsilon_n, \quad (4)$$

where $\alpha_n \sim (0, \sigma_\alpha^2)$, ι_T is a T -vector of ones, and $\varepsilon_n \sim (0, \sigma_\varepsilon^2 I_T)$; ε_n and α_n are independent and $|\gamma| < 1$. The assumptions that v_n and ε_n have the same variance over t and are uncorrelated over t is strong. However, leaving their correlation structure free would result in a highly underidentified model.

One way to estimate the model is through GMM. If we assume that the process has an infinite past, it follows that

$$\eta_{nt} = \frac{1}{1 - \gamma} \alpha_n + \sum_{\tau=0}^{\infty} \gamma^\tau \varepsilon_{n,t-\tau}.$$

So

$$\begin{aligned} \Sigma_\eta &\equiv E(\eta_{n,+} \eta'_{n,+}) \\ &= \frac{\sigma_\alpha^2}{(1 - \gamma)^2} \iota_{T+1} \iota'_{T+1} + \frac{\sigma_\varepsilon^2}{1 - \gamma^2} V_\gamma, \end{aligned}$$

where V_γ is the AR(1) matrix of order $(T + 1) \times (T + 1)$, i.e. the matrix whose (t, τ) th element is $\gamma^{|t-\tau|}$. So the second-order implication of the model is

$$\begin{aligned} \Sigma_y &\equiv E(y_{n,+} y'_{n,+}) \\ &= E((\eta_{n,+} + v_{n,+})(\eta_{n,+} + v_{n,+})') \\ &= \Sigma_\eta + \sigma_v^2 I_{T+1}. \end{aligned}$$

The GMM estimator of the parameters is obtained by minimizing the distance between the vector containing the non-redundant elements of Σ_y and its sample counterpart $S_y \equiv \sum y_{n,+} y'_{n,+} / N$. By an appropriate choice of weight matrix in the distance function, the GMM estimator is asymptotically efficient among all estimators based on S_y .

A drawback of the GMM estimator is that it may be cumbersome to compute. We now consider a simpler alternative, where we focus on estimating the parameter of most interest, γ . Moreover, the properties of this estimator do not hinge upon the assumption of an infinite past. We first discuss a simple way to derive a consistent estimator of γ , and then consider issues of optimality.

As a start, we eliminate η_n from the model by substitution from (3) into (4) to obtain

$$y_n = \gamma y_{n,-1} + \alpha_n \iota_T + (\varepsilon_n + v_n - \gamma v_{n,-1}). \quad (5)$$

We consider estimation of γ by IV. As an instrument, we consider a general linear function of $y_{n,+}$ of the form $A' y_{n,+}$ for some non-random $(T+1) \times T$ -matrix A . Below we will also use the form $a \equiv \text{vec} A$. In order to eliminate the individual effect, we let A be such that $A \iota_T = 0_{T+1}$. Given A , the IV estimator of γ is:

$$\hat{\gamma} \equiv \frac{\sum_n y'_{n,+} A y_n}{\sum_n y'_{n,+} A y_{n,-1}}. \quad (6)$$

We want to impose such structure on A that $\hat{\gamma}$ is consistent. Premultiplying (5) by $y'_{n,+} A$ gives

$$y'_{n,+} A y_n = \gamma y'_{n,+} A y_{n,-1} + y'_{n,+} A (\varepsilon_n + v_n - \gamma v_{n,-1}).$$

So consistency requires

$$\begin{aligned} 0 &= E(y'_{n,+} A (\varepsilon_n + v_n - \gamma v_{n,-1})) \\ &= \text{tr}(E[(\varepsilon_n + v_n - \gamma v_{n,-1}) y'_{n,+}] A). \end{aligned} \quad (7)$$

Let $C'_0 \equiv (I_T, 0_T)$ and let $C'_1, \dots, C'_T$ be a series of matrices of order $T \times (T+1)$, where $C'_1 \equiv (0_T, I_T)$, in C'_2 the ones are moved one position to the right, and so on, ending with C'_T , which is zero, except for its $(1, T+1)$ element. Since

$$\begin{aligned} E(\varepsilon_n y'_{n,+}) &= \sigma_\varepsilon^2 \sum_{t=1}^T \gamma^{t-1} C'_t \\ E(v_n y'_{n,+}) &= \sigma_v^2 \sum_{t=1}^T \gamma^{t-1} C'_t \\ E(v_{n,-1} y'_{n,+}) &= \sigma_v^2 \sum_{t=0}^T \gamma^t C'_t, \end{aligned}$$

we conclude from (7) that consistency is obtained when we let A be such that $\text{tr } C'_t A = 0$ or $(\text{vec } C_t)' a = 0$ for $t = 0, \dots, T$. This means, for

$$C \equiv (\text{vec } C_0, \dots, \text{vec } C_T),$$

of order $T(T+1) \times (T+1)$, that a should satisfy $C'a = 0_{T+1}$. Any A satisfying this and $A\iota_T = 0_{T+1}$ leads to a consistent estimator of γ .

To find an estimator that is not just consistent but also asymptotically efficient we proceed as follows. When A is such that $\hat{\gamma}$ is consistent, there holds

$$\text{avar}(\hat{\gamma}) = \frac{a'Va}{a'qa}, \quad (8)$$

where, with $u_n \equiv \varepsilon_n + v_n - \gamma v_{n-1}$,

$$\begin{aligned} V &\equiv \text{var}(u_n \otimes y_{n,+}) \\ q &\equiv E(y_{n,-1} \otimes y_{n,+}). \end{aligned}$$

We are free to normalize A (or a). In view of (8), an obvious normalization is $q'a = 1$, leaving us with the task of minimizing $a'Va$ over a subject to the normalization $q'a = 1$, the $T+1$ restrictions $(\iota'_T \otimes I_{T+1})a = 0_{T+1}$, i.e. the vectorized form of $A\iota_T = 0_{T+1}$, and the $T+1$ restrictions $C'a = 0$ leading to consistency. With

$$R \equiv (q, \iota'_T \otimes I_{T+1}, C),$$

of order $T(T+1) \times (2T+3)$, we minimize $a'Va$ subject to $R'a = e_1$. The Lagrangian is $a'Va - 2\mu'(R'a - e_1)$, so the first-order condition is $Va - R\mu = 0$. A solution is

$$\begin{aligned} a &= (I_{T+1} - M(MVM)^+MV)R(R'R)^{-1}e_1 \\ \mu &= (R'R)^{-1}R'QR(R'R)^{-1}e_1, \end{aligned}$$

where

$$\begin{aligned} M &\equiv I_{T+1} - R(R'R)^{-1}R' \\ Q &\equiv V - VM(MVM)^+MV, \end{aligned}$$

so $QM = MQ = 0$. The solution is easy to verify since

$$\begin{aligned} Va &= QR(R'R)^{-1}e_1 \\ R\mu &= (I - M)QR(R'R)^{-1}e_1 = Va. \end{aligned}$$

So the asymptotic variance of $\hat{\gamma}$ is $a'Va = e'_1(R'R)^{-1}R'QR(R'R)^{-1}e_1 = e'_1\mu$.

In order to make the procedure operational we need consistent estimates of V and q , the latter so since R depends on q . As to V , we can estimate it by the sample variance of $\hat{u}_n \otimes y_{n,+}$. So we proceed in two steps. In the first step we estimate the model by some choice of A such that $\hat{\gamma}$ is consistent, and derive consistent estimates for the other parameters from $\hat{\gamma}$. We plug these values into V and R , leading to $\hat{V}$ and $\hat{R}$. In the second step, we compute a using $\hat{V}$ for V and $\hat{R}$ for R .

4 Set identification

We return to the static case and address the issue of what to do when none of the assumptions leading to moment conditions and consistent estimation are deemed justified, and the model is not identified. However, when a parameter is not “point identified”, it may still be “set identified”, that is, there may be an interval whose bounds are consistently estimable and that the true value of the parameters is in the limit. Such set identification is well-known in the simple linear model with measurement error, where direct and reverse regression provide such an interval. Wansbeek and Meijer (2000), ch.3, provides an elaborate treatment of this topic that originated in the 1930s. We now extend this to the panel data case.

Consider a simple panel data model with a single regressor, ξ_n , of order $T \times 1$, which is unobservable. 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 $v_n \sim (0, \Sigma_v)$, independent. We do not impose any restrictions on Σ_ε or Σ_v . As a result, we may ignore the presence of individual effects. If these are taken random, the implied GLS structure is captured by the general structure of Σ_ε . If they are taken fixed, we assume that they have been eliminated by, for example, the within transformation, and the variables above are transformed variables.

From the model above, we eliminate ξ_n to obtain

$$y_n = \beta x_n + u_n,$$

with $u_n \equiv \varepsilon_n - v_n \beta$. Since $E(x_n u_n') = -\beta \Sigma_v \neq 0$, OLS gives an inconsistent result:

$$\text{plim}_{N \rightarrow \infty} b_{\text{OLS}} = \text{plim}_{N \rightarrow \infty} \frac{\sum_n x_n' y_n}{\sum_n x_n' x_n} = \frac{\text{tr } \Sigma_\xi}{\text{tr } \Sigma_\xi + \text{tr } \Sigma_v} \beta. \quad (9)$$

So there is a bias towards zero. For the reverse regression for this case we have

$$\text{plim}_{N \rightarrow \infty} b_{\text{REV}} = \text{plim}_{N \rightarrow \infty} \frac{\sum_n y_n' y_n}{\sum_n x_n' y_n} = \beta + \frac{\text{tr } \Sigma_\varepsilon}{\beta \text{tr } \Sigma_\xi}, \quad (10)$$

which is away from zero. So also in the panel data case, in the limit, β is between the results of the direct and reverse regressions. We now extend this for additional regressors and individual effects.

We start by considering the bounds on β from OLS and reverse regression, and first assess the effect of the additional regressors Z_n only. We assume Z_n to be uncorrelated with the error in the equation and the measurement error. Collect the y_n , x_n and Z_n in y , x , and Z , of order $NT \times 1$,

$NT \times 1$, and $NT \times k$, respectively. Project Z out by $M_Z \equiv I_{NT} - Z(Z'Z)^{-1}Z'$, and do OLS and reverse regression on the transformed variables $M_Z y$ and $M_Z x$. Let $\sigma_{Z\xi} \equiv E(Z'_n \xi_n)$ and $\Sigma_Z \equiv E(Z'_n Z_n)$. The presence of additional regressors leads to an adaptation of (9) and (10) where $\sigma'_{Z\xi} \Sigma_Z \sigma_{Z\xi}$ has to be subtracted from Σ_ξ while Σ_ε and Σ_v remain unaffected. As a result, the bound widens.

We now assess the effect of individual effects only. We eliminate them by premultiplying the model by B' , where B is any matrix such that $B' \iota_T = 0$. Hence, in (9) and (10), we have to replace $\Sigma_\xi, \Sigma_\varepsilon$ and Σ_v by $\Sigma_{\tilde{\xi}} \equiv B' \Sigma_\xi B$, $\Sigma_{\tilde{\varepsilon}} \equiv B' \Sigma_\varepsilon B$ and $\Sigma_{\tilde{v}} \equiv B' \Sigma_v B$, respectively. In the empirically likely case that the regressor is more correlated over time than the error in the equation and the measurement error, the variance of ξ is reduced much more than the variance of ε and v and hence the bound widens. This occurs a fortiori in the full model with both additional regressors and individual effects. Yet, the bound might be tight enough to be useful.

It is of some interest to notice that the bounds are parameters that, by definition, can be consistently estimated. (For readability, avoiding lots of tildes, we look at the case without individual effects.) Let us call these parameters β_{OLS} and β_{REV} . They are estimated from the systems

$$\begin{aligned} E(x'_n y_n - x'_n x_n \beta_{OLS} - x'_n Z_n \pi_*) &= 0 \\ E(Z'_n y_n - Z'_n x_n \beta_{OLS} - Z'_n Z_n \pi_*) &= 0 \end{aligned}$$

and

$$\begin{aligned} E(y'_n x_n - y'_n y_n / \beta_{REV} - y'_n Z_n \pi^*) &= 0 \\ E(Z'_n x_n - Z'_n y_n / \beta_{REV} - Z'_n Z_n \pi^*) &= 0, \end{aligned}$$

with π_* and π^* irrelevant parameters. By GMM we can find the asymptotic variance of the upper and lower bound and their difference in a straightforward way.

Up till now we have used the panel in its entirety and have derived a single lower and upper bound. However, we can exploit the richness of panel data and derive a variety of lower and upper bounds. To focus on essentials, we return to the simple case we started with and abstract from individual effects and additional regressors. We notice that we can compute a lower and an upper bound per wave. This provides us with T lower bounds and T upper bounds. We can then select the highest lower bound and the lowest upper bound, and have an interval that consistently bounds β .

To take this one step further, let W be a positive semidefinite $T \times T$ weight matrix, and consider estimators

$$\text{plim}_{N \rightarrow \infty} b_{OLS, W} = \text{plim}_{N \rightarrow \infty} \frac{\sum_n x'_n W y_n}{\sum_n x'_n W x_n} = \frac{\text{tr } W \Sigma_\xi}{\text{tr } W \Sigma_\xi + \text{tr } W \Sigma_v} \beta \quad (11)$$

$$\text{plim}_{N \rightarrow \infty} b_{REV, W} = \text{plim}_{N \rightarrow \infty} \frac{\sum_n y'_n W y_n}{\sum_n x'_n W y_n} = \beta + \frac{\text{tr } W \Sigma_\varepsilon}{\beta \text{tr } W \Sigma_\xi}. \quad (12)$$

Clearly, for any p.s.d W these estimators bound β . The question naturally arises as to the optimal choice of W .

To tackle the question, write $W = AA'$, with A to be determined. Let $a = \text{vec}A$. Then

$$\begin{aligned} b_{\text{OLS}, W} &= \frac{\sum_n \text{tr}(AA' y_n x_n')}{\sum_n \text{tr}(AA' x_n x_n')} \\ &= \frac{\text{tr}(AA' S_{xy})}{\text{tr}(AA' S_x)} \\ &= \frac{a'(I_T \otimes S_{xy})a}{a'(I_T \otimes S_x)a}. \end{aligned}$$

Now, in order to obtain the highest lower bound from (11), we have to maximize $a'(I_T \otimes S_{xy})a/a'(I_T \otimes S_x)a$ over a . By normalizing a such that $a'(I_T \otimes S_x)a = 1$, the solution follows from the eigenequation

$$[I_T \otimes (S_{xy} - \lambda S_x)]a = 0. \quad (13)$$

To find a maximum, a should be taken to be the eigenvector corresponding to the largest eigenvalue. Let b be a vector satisfying $(S_{xy} - \lambda S_x)b = 0$. Then (13) is satisfied for $a = e \otimes b$, where e is an arbitrary vector. Conversely, any solution of (13) must be of the form $a = e \otimes b$. Thus, the solution for A is found by first computing the eigenvector b of the system $(S_{xy} - \lambda S_x)b = 0$ corresponding to the largest eigenvalue, subject to $b'S_x b = 1$. Then choose an arbitrary e subject to $e'e = 1$, and choose $a = e \otimes b$, which results in $A = be'$ and $W = bb'$. Then, the estimator with the highest lower bound for β is

$$\begin{aligned} b_{\text{OLS}, W}^{\max} &= \frac{\sum_n x_n' b b' y_n}{\sum_n x_n' b b' x_n} \\ &= \frac{\sum_n (b' x_n)(b' y_n)}{\sum_n (b' x_n)^2} \\ &= \frac{\sum_n \tilde{x}_n \tilde{y}_n}{\sum_n \tilde{x}_n^2}, \end{aligned}$$

where $\tilde{x}_n = b' x_n$ is the optimally weighted combination of the elements of x_n and $\tilde{y}_n = b' y_n$ is the corresponding combination of the elements of y_n .

Correspondingly, in order to find the lowest upper bound, we see from (12) that we have to minimize $a'(I_T \otimes S_y)a/a'(I_T \otimes S_{xy})a$ or, equivalently, to maximize $a'(I_T \otimes S_{xy})a/a'(I_T \otimes S_y)a$. The eigenequation then is

$$[I_T \otimes (S_{xy} - \lambda S_y)]a = 0,$$

with again a corresponding with the largest eigenvalue, and the solution is obtained analogous to the solution for the highest lower bound. Notice that S_{xy} is asymmetric, although its expectation is symmetric. To avoid computational problems, we may replace S_{xy} by $(S_{xy} + S_{yx})/2$.

References

- Ahn, S.C. and P. Schmidt (1995), “Efficient estimation of models for dynamic panel data”, *Journal of Econometrics*, **68**, 5–27.
- Altonji, J.G. and A. Siow (1987), “Testing the response of consumption to income changes with (noisy) panel data”, *The Quarterly Journal of Economics*, **101**, 293–328.
- Antman, F. and D. McKenzie (2007), “Poverty traps and nonlinear income dynamics with measurement error and individual heterogeneity”, *Journal of Development Studies*, **43**, 1057–1083.
- 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. (2012), *The measurement error problem in dynamic panel data analysis: modeling and GMM estimation*, Working paper, University of Oslo.
- Blundell, R. and S. Bond (1998), “Initial conditions and moment restrictions in dynamic panel data models”, *Journal of Econometrics*, **87**, 115–143.
- Chen, J., S. Ni, and M. Podgursky (2008), “Estimating dynamic panel data models with measurement errors with an application to school evaluation based on student test scores”, *Proceedings of the 2008 meeting of the American Statistical Association*, 951–957.
- Griliches, Z. and J.A. Hausman (1986), “Errors in variables in panel data”, *Journal of Econometrics*, **31**, 93–118.
- Harris, M.N. and L. Mátyás (2000), “Performance of the operational Wansbeek-Bekker estimator for dynamic panel data models”, *Applied Economics Letters*, **7**, 149–153.
- Komunjer, I. and S. Ng (2011), *Measurement errors in dynamic models*, Working paper.
- Lewbel, A. (2011), “Using heteroskedasticity to identify and estimate mismeasured and endogenous regressor models”, *Journal of Business & Economic Statistics*, forthcoming.
- Spierdijk, L. and T.J. Wansbeek (2012), *A comprehensive approach to measurement error in panel data*, Working paper, University of Groningen.
- Wansbeek, T.J. and P.A. Bekker (1996), “On IV, GMM and ML in a dynamic panel data model”, *Economics Letters*, **51**, 145–152.
- Wansbeek, T.J. and A. Kapteyn (1992), “Simple estimators for dynamic panel data models with errors in the variables”, in R. Bewley and Tran Van Hoa, editors, *Contributions to consumer demand and econometrics*, McMillan, London.
- Wansbeek, T.J. and E. Meijer (2000), *Measurement error and latent variables in econometrics*, North-Holland, Amsterdam.