Measurement Error in the Linear Dynamic Panel Data Model

Erik Meijer, Laura Spierdijk, and Tom Wansbeek

Abstract We study measurement error in the simplest dynamic panel data model without covariates. We start by investigating the first-order effects, on the most commonly used estimator, of the presence of measurement error. As was to be expected, measurement error renders this estimator inconsistent. However, with a slight adaptation, the estimator can be made consistent. This approach to consistent estimation is ad hoc and we next develop a systematic approach to consistent estimation. We show how to obtain the most efficient estimator from this class of consistent estimators. We illustrate our findings through an empirical example.

1 Introduction

In econometrics, the analysis of panel data is a rapidly expanding research area. Frequently, the models formulated are dynamic models in the sense that the lagged dependent variables is among the regressors. Especially the linear dynamic panel data model (LDPDM) is hugely popular. This model typically has both the lagged dependent variable and an individual effect on the right-hand side. Its estimation is not entirely straightforward since the least-squares estimator is inconsistent.

E. Meijer
RAND Corporation, Santa Monica, CA 90401, USA
e-mail: meijer@rand.org

L. Spierdijk • T. Wansbeek (✉)
University of Groningen, 9700 AV Groningen, The Netherlands
e-mail: l.spierdijk@rug.nl; t.j.wansbeek@rug.nl

Consistent estimation of the LDPDM has inspired many researchers, and the number of publications on the topic is still growing. The leading idea is to transform the model equation into first differences over time, and next use the twice lagged-dependent variable as an instrumental variable (IV). Under the crucial assumption that the error term is not correlated over time, it is easy to see that this approach gives a consistent estimator of the regression coefficient. The idea is due to Anderson and Hsiao (1981, 1982). Arellano and Bond (1991) pointed out that all preceding values of the dependent variable, not just the directly preceding one, can be used as IVs, leading to more IVs and smaller asymptotic variance, if there are more than three periods. The estimator due to Arellano and Bond (1991) has found application on a very large scale.

Little attention has been paid to issues around measurement error in the LDPDM. This comes not entirely as a surprise, as the situation for the much simpler static model is not vastly more favorable. But at least, there is a line of literature for the static model. The pioneering contribution there is Griliches and Hausman (1986). Much of the literature for the static model is reviewed by Meijer et al. (2012), where also a number of new ways are described for consistent estimation.

As to the literature on the LDPDM with 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 et al. (2008), in their study of the dynamics of students' test scores, construct consistent estimators through IVs 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 IVs based on the absence of correlation between regressors and disturbances for some combinations of time indices.

In this paper we contribute to the literature on the LDPDM with measurement error in two ways. In the first place we derive, in Sect. 2, the effect of measurement error on the Arellano–Bond (AB) estimator. For the simple case of a panel with three waves, we investigate the inconsistency of this estimator in the presence of measurement error. We provide some interpretation and further elaboration in Sect. 3. Next, we move over to consistent estimation. In Sect. 4 we consider a wide class of estimators that are consistent, and in Sect. 5, we study efficiency within this class. An illustrative example is given in Sect. 6. In Sect. 7 we make some concluding remarks.

2 The Effect of Measurement Error

In this section we consider the simplest possible LDPDM and investigate the effect of measurement error when it is estimated in the usual way. The model is represented by the following two equations,

$$\eta_{nt} = \gamma \eta_{n,t-1} + \alpha_n + \varepsilon_{nt} \tag{1}$$

$$y_{nt} = \eta_{nt} + v_{nt}, \tag{2}$$

for $n = 1, \ldots, N$ and $t = 1, \ldots, T$. In this model, η_{nt} is an unobserved variable, according to (1) subject to an autoregressive process of order one. The error term in (1) consists of two components, a time-constant one, α_n, and a time-varying one, ε_{nt}. The link between the unobserved variable η_{nt} and the observed variable y_{nt} is given by the measurement equation (2), where v_{nt} represents the measurement error. All variables have mean zero over n, possibly after demeaning per time period thus accounting for fixed time effects. The parameter of interest in the autocorrelation parameter γ. We restrict ourselves to the case where $-1 < \gamma < 1$.

It is assumed that α_n, ε_{nt}, and v_{nt} are uncorrelated over n. Moreover, ε_{nt} and v_{nt} are assumed uncorrelated over t. This is quite a simplification but, somewhat surprisingly, these assumptions are commonly made. The various error terms are taken homoskedastic, with means zero and variances σ_α^2, σ_ε^2, and σ_v^2, respectively. As is usual in econometrics, these parameters, in particular the absolute or relative measurement error variance, are taken to be unknown. Finally, it is assumed that the process has been going on since minus infinity, and $-1 < \gamma < 1$, so that the distributions of all variables are stationary.

We take N to be large relative to T and hence, in our asymptotic results, keep T fixed and let N go to infinity. So our perspective is cross-sectional. For a time-series perspective on measurement error, see Aigner et al. (1984, Sect. 6) and Buonaccorsi (2010, Chap. 12).

Even if η_{nt} would be observed, estimation of (1) is not straightforward as (1) implies that α_n is correlated with $\eta_{n\tau}$ for all τ, including the case $\tau = t - 1$. So, since $\eta_{n,t-1}$ is the regressor, the regression model (1) has an error term that is correlated with the regressor. Hence least squares gives an inconsistent result for the parameter of interest, γ.

The Anderson–Hsiao (AH) estimator (Anderson and Hsiao 1981, 1982) first transforms (1) into first differences over time, and next uses $\eta_{n,t-2}$ as an IV. Under the crucial assumption that ε_{nt} is not correlated over time, it is easy to see that this approach gives a consistent estimator of γ. The Arellano–Bond (AB) estimator (Arellano and Bond 1991) is a generalization that uses all preceding values of the dependent variable, not just the directly preceding one, as IVs, leading to more instruments and smaller asymptotic variance, if the number of observed periods is larger than three. In the derivations in this section we restrict ourselves, for reasons of tractability and emphasis on the essentials, to three periods, so our estimator is the AH estimator.

This is all about the case where η_{nt} is observed and not clouded by measurement error. If there is measurement error, so if (2) enters the stage, the consistency of the AH estimator is at peril. We now have measurement error twice in the model. It evidently enters both the dependent variable and the regressor. As is well known from the measurement error literature, measurement error in the dependent variable does not affect consistency, but measurement error in a regressor does, in most cases in the form of a bias towards zero of the estimator. Here the measurement error in both variables comes from the same source, that is, from (2), and the effect of measurement error is not straightforward.

In order to gain insight into the effect of measurement error on the AH estimator, we transform the model into first differences over time and substitute out the unobserved variable η from the model. This gives us

$$y_{nt} - y_{n,t-1} = \gamma(y_{n,t-1} - y_{n,t-2}) + u_{nt}, \tag{3}$$

where the error term u_{nt} is defined as

$$u_{nt} \equiv (\varepsilon_{nt} - \varepsilon_{n,t-1}) + (v_{nt} - v_{n,t-1}) - \gamma(v_{n,t-1} - v_{n,t-2}).$$

The AH estimator of γ is obtained by estimating (3) with $y_{n,t-2}$ as the IV. In the presence of measurement error this estimator is not consistent, because the IV is not valid as it is not orthogonal to the error term in (3):

$$\begin{aligned}
\mathrm{E}(y_{n,t-2}u_{nt}) &= \mathrm{E}\left\{(\eta_{n,t-2} + v_{n,t-2})[(\varepsilon_{nt} - \varepsilon_{n,t-1}) \right.\\
&\quad \left. + (v_{nt} - v_{n,t-1}) - \gamma(v_{n,t-1} - v_{n,t-2})]\right\}\\
&= \gamma\sigma_v^2.
\end{aligned}$$

In order to derive the probability limit of the AH estimator, we first notice that, through repeated substitution, the unobserved variable can be expressed as

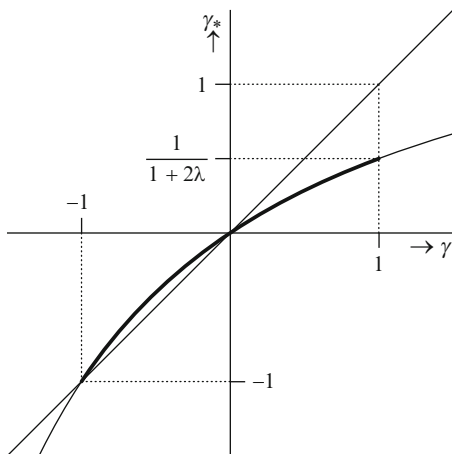$$\eta_{nt} = \frac{\alpha_n}{1 - \gamma} + \sum_{s=0}^{\infty} \gamma^s \varepsilon_{n,t-s}. \tag{4}$$

Hence

$$\mathrm{E}(\eta_{nt}\eta_{n,t-\tau}) = \frac{\sigma_\alpha^2}{(1 - \gamma)^2} + \gamma^\tau \frac{\sigma_\varepsilon^2}{1 - \gamma^2}. \tag{5}$$

Consequently,

$$\begin{aligned}
\omega_\tau &\equiv \mathrm{E}(y_{nt}y_{n,t-\tau})\\
&= \frac{\sigma_\alpha^2}{(1 - \gamma)^2} + \gamma^\tau \frac{\sigma_\varepsilon^2}{1 - \gamma^2} + I(\tau = 0)\sigma_v^2, \tag{6}
\end{aligned}$$

Fig. 1 Probability limit of
the Anderson–Hsiao
estimator with $\lambda = \frac{1}{2}$

where $I(\cdot)$ is the indicator function, which is 1 if its argument is true and 0 otherwise.
The AH estimator is given by

$$\hat{\gamma} = \frac{\frac{1}{N}\sum_n y_{n,t-2}(y_{nt} - y_{n,t-1})}{\frac{1}{N}\sum_n y_{n,t-2}(y_{n,t-1} - y_{n,t-2})}.$$

Let $\lambda \equiv \sigma_v^2/\sigma_\varepsilon^2$ be the ratio of the measurement error variance to the equation error
variance. Under weak assumptions, the probability limit of the AH estimator is

$$\gamma_* \equiv \mathrm{plim}_{N\to\infty}\hat{\gamma}$$
$$= \frac{\omega_2 - \omega_1}{\omega_1 - \omega_0}$$
$$= \frac{(\gamma^2 - \gamma)\frac{1}{1-\gamma^2}}{(\gamma-1)\frac{1}{1-\gamma^2} - \lambda}$$
$$= \frac{\gamma}{1 + (1+\gamma)\lambda}. \tag{7}$$

This result is depicted in Fig. 1. Clearly, measurement error causes the estimator to
be biased towards zero. The bias towards zero is a well-known phenomenon from
the literature on measurement error in a single cross-section.

The figure is made for the case of $\lambda = \frac{1}{2}$, so $\sigma_v^2 = \frac{1}{2}\sigma_\varepsilon^2$. With decreasing
measurement error, so with decreasing λ, the hyperbola will become closer to the
45° line. It is also striking how asymmetric the biasing effect is: the bias is much
larger for positive values of γ (which are arguably more likely in most applications)
than for negative values, both absolutely and in a relative sense.

3 Interpretation and Elaboration

Another view of the result (7) can be obtained as follows. Due to the stationarity and the absence of serial correlation in the measurement errors, and with Δ denoting the first-difference operator, we have

$$
\begin{aligned}
E\sum_n y_{n,t-2}(y_{nt}-y_{n,t-1}) &= E\sum_n \eta_{n,t-2}(\eta_{nt}-\eta_{n,t-1}) \\
&= \gamma E\sum_n \eta_{n,t-2}(\eta_{n,t-1}-\eta_{n,t-2}) \\
&= -\tfrac{1}{2}\gamma\sum_n (\eta_{n,t-1}-\eta_{n,t-2})^2 \\
&= -\tfrac{1}{2}\gamma\sigma^2_{\Delta\eta}
\end{aligned}
$$

and

$$
\begin{aligned}
E\sum_n y_{n,t-2}(y_{n,t-1}-y_{n,t-2}) &= -\frac{1}{2}E(y_{n,t-1}-y_{n,t-2})^2 \\
&= -\tfrac{1}{2}\sigma^2_{\Delta y}.
\end{aligned}
$$

We thus obtain

$$
\gamma_* = \frac{\sigma^2_{\Delta\eta}}{\sigma^2_{\Delta y}}\gamma.
$$

The bias factor is the "reliability" of Δy as a proxy for $\Delta\eta$. The situation closely resembles the situation in the classical measurement error model for a single cross-section, where the same result holds but then in levels, not differences. Also, in that case the reliability does not mathematically depend on γ, because it is the reliability of the exogenous variable. In the LDPDM, it depends on γ, causing the curvature depicted in Fig. 1.

As mentioned above, the AH estimator is a special instance of the AB estimator. Nowadays, researchers often use the "systems" generalized method-of-moments (GMM) estimator (Arellano and Bover 1995; Blundell and Bond 1998), which combines the building blocks of the AB estimator with those that can be derived when the correlation between y_{nt} and the individual effect α_n does not depend on t. Usually, the systems GMM estimator greatly outperforms the AB estimator. In the present setup we have from (4) that

$$
E(y_{nt}\alpha_n) = \frac{\sigma^2_\alpha}{1-\gamma}, \tag{8}
$$

establishing this equicorrelation here. It can be used to estimate γ in a way that in a sense is the mirror image of AB. There, previous values of y are used as IV for a

model in first differences. Here, we keep the model in levels but use previous values of y in first-difference form as IV. Then

$$\tilde{\gamma} = \frac{\frac{1}{N}\sum_n (y_{n,t-1} - y_{n,t-2})y_{nt}}{\frac{1}{N}\sum_n (y_{n,t-1} - y_{n,t-2})y_{n,t-1}}$$

$$\xrightarrow{p} \frac{\omega_1 - \omega_2}{\omega_0 - \omega_1}$$

$$= \gamma_*.$$

Thus, the inconsistency is the same as with the AH estimator.

So, with measurement error, we encounter the inconsistency issue well known from the cross-sectional case. A major difference, though, is that the latter case, in its simplest form of linearity, normality, and independence of observations (LIN), results in an identification problem that precludes the existence of a consistent estimator (e.g., Wansbeek and Meijer 2000, p. 79). In the LDPDM, LIN does not apply and the situation is more favorable. In fact, a consistent estimator is easily found; instead of using $y_{n,t-2}$ as an IV, we can use $y_{n,t-3}$ (assuming $T > 3$). (Bond et al. 2001, make the same observation.) We call this estimator the Anderson–Hsiao lagged (AHL) estimator. Its probability limit is

$$\text{plim}_{N\to\infty}\hat{\gamma}_{\text{AHL}} = \frac{\text{plim}_{N\to\infty}\frac{1}{N}\sum_n y_{n,t-3}(y_{nt} - y_{n,t-1})}{\text{plim}_{N\to\infty}\frac{1}{N}\sum_n y_{n,t-3}(y_{n,t-1} - y_{n,t-2})}$$

$$= \frac{\omega_3 - \omega_2}{\omega_2 - \omega_1}$$

$$= \frac{(\gamma^3 - \gamma^2)\frac{1}{1-\gamma^2}}{(\gamma^2 - \gamma)\frac{1}{1-\gamma^2}}$$

$$= \gamma.$$

So AHL is a consistent estimator, due to the assumed lack of correlation over time of the measurement error. Analogously, the Arellano–Bond lagged (ABL) estimator is obtained by removing $y_{n,t-2}$ from the list of IVs of the Arellano–Bond estimator, which is also a consistent estimator in our setup. Arellano and Bond (1991) mention this estimator in the context of autocorrelation resulting from a moving average process in the errors, so it serves a dual purpose.

This approach to consistent estimation is of course somewhat ad hoc. Moreover, it breaks down when the measurement errors are autocorrelated. To gain some insight here, we assume that the measurement errors are subject to an autoregressive process of order one, AR(1), so

$$v_{nt} = \rho v_{n,t-1} + w_{nt},$$

with w_{nt} white noise with variance σ_w^2. Instead of (6), we now have

$$\omega_\tau = \frac{\sigma_\alpha^2}{(1-\gamma)^2} + \gamma^\tau \frac{\sigma_\varepsilon^2}{1-\gamma^2} + \rho^\tau \frac{\sigma_w^2}{1-\rho^2}.$$

With λ redefined as $\lambda \equiv \sigma_w^2/\sigma_\varepsilon^2$, the probability limit of the AH estimator now becomes

$$\gamma_* = \frac{(\gamma^2-\gamma)\frac{\sigma_\varepsilon^2}{1-\gamma^2} + (\rho^2-\rho)\frac{\sigma_w^2}{1-\rho^2}}{(\gamma-1)\frac{\sigma_\varepsilon^2}{1-\gamma^2} + (\rho-1)\frac{\sigma_w^2}{1-\rho^2}}$$

$$= \frac{\gamma(1+\rho) + \rho(1+\gamma)\lambda}{(1+\rho) + (1+\gamma)\lambda}.$$

So the effect of measurement error is now more complicated. The measurement error in the dependent variable, which in the classical case has no effect on the estimator of the regression coefficient, now plays a role due to its correlation with the measurement error in the regressor. The estimator is consistent ($\gamma_* = \gamma$) if $\rho = \gamma$ or $\lambda = 0$. The estimator has a positive bias if $\rho > \gamma$ and a negative bias if $\rho < \gamma$. In the most likely case that $0 < \rho < \gamma$, we see the usual attenuation bias towards zero. Note that we can write $\gamma_* = \phi\gamma + (1-\phi)\rho$, with

$$\phi = \frac{\frac{\sigma_\varepsilon^2}{1+\gamma}}{\frac{\sigma_\varepsilon^2}{1+\gamma} + \frac{\sigma_w^2}{1+\rho}},$$

so γ_* is a weighted average of γ and ρ. Although the weights themselves depend on γ and ρ, they are always between 0 and 1.

Using a similar derivation, it follows that the AHL estimator, or in general using previous values of y as an IV, does not yield a consistent estimator anymore. We now turn to a more systematic approach to consistent estimation, for general values of T. We first investigate what consistency implies and derive a class of consistent estimators. We next consider issues of optimality.

4 Consistent Estimation

Our approach to consistent estimation extends Wansbeek and Bekker (1996) by taking measurement error into account. They derive an instrumental variable that is linear in the values of the dependent variable across time and that results in an IV estimator 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) and find that Wansbeek and Bekker's estimator "generally outperformed all other estimators when T was moderate *in all of the situations that an applied researcher might encounter*" [italics in original]. We adapt the Wansbeek

and Bekker (1996) approach also in another way, in that we assume stationarity throughout.

We now turn to the model and derive our estimator. For general T, it is convenient to move over to matrix notation. With observations y_{nt}, $n = 1, \ldots, N$, $t = 0, \ldots T$, we define

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

For η, v, and ε, we use analogous notation. Note that the number of observed periods is $T + 1$ now, as opposed to the T used before. The model can now be written as

$$
\eta_n = \gamma \eta_{n,-1} + \alpha_n \iota_T + \varepsilon_n, \tag{9}
$$

where ι_T is a T-vector of ones, and $\varepsilon_n \sim (0, \sigma_\varepsilon^2 I_T)$. The measurement equation is

$$
y_n = \eta_n + v_n, \tag{10}
$$

with

$$
v_n = \rho v_{n,-1} + w_n,
$$

where $w_n \sim (0, \sigma_w^2 I_T)$, thus allowing for measurement errors correlated over time according to an AR(1) process.

One way to estimate the model parameters consistently is through GMM. From (5) we obtain

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

where V_γ is the AR(1) correlation matrix of order $(T + 1) \times (T + 1)$, that is, the matrix whose (t,s)th element is $\gamma^{|t-s|}$. So the second-order implication of the model for the observations, taking the measurement error into account, is

$$
\begin{aligned}
\Sigma_y &\equiv \mathrm{E}(y_{n,+} y_{n,+}') \\
&= \mathrm{E}\left((\eta_{n,+} + v_{n,+})(\eta_{n,+} + v_{n,+})'\right) \tag{11} \\
&= \Sigma_\eta + \frac{\sigma_w^2}{1-\rho^2} V_\rho \\
&= \frac{\sigma_\alpha^2}{(1-\gamma)^2} \iota_{T+1} \iota_{T+1}' + \frac{\sigma_\varepsilon^2}{1-\gamma^2} V_\gamma + \frac{\sigma_w^2}{1-\rho^2} V_\rho. \tag{12}
\end{aligned}
$$

From this, an essential identification problem with the model is immediately clear. The model is locally identified but not globally. The parameter set $(\sigma_\varepsilon^2, \gamma)$ can be interchanged with the parameter set (σ_w^2, ρ) as they play a symmetric role in Σ_y, and the data do not provide sufficient information to tell which is which. Hence we restrict ourselves to the case where the measurement error has no autocorrelation and have $\rho = 0$ from now on.

The GMM estimator of the parameters is obtained by minimizing the distance between

$$\sigma_y \equiv \text{vech } \Sigma_y,$$

the vector containing the non-redundant elements of Σ_y, and its sample counterpart

$$s_y \equiv \text{vech } S_y,$$

where $S_y \equiv \sum y_{n,+} y'_{n,+} / N$. For an appropriate choice of the 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 in this case is that it will be cumbersome to compute as Σ_y depends on the parameter of interest, γ, in a highly nonlinear way. Hence we consider a simpler way to obtain a consistent estimator, focusing on γ. The price for this simplicity is that this estimator does not exploit all the structure imposed by the model on Σ_y and hence will be asymptotically inefficient.

As a start, we eliminate η_n from the model by substitution from (10) into (9) to obtain

$$y_n = \gamma y_{n,-1} + \upsilon_n \tag{13}$$

$$\upsilon_n \equiv \alpha_n \iota_T + \varepsilon_n + v_n - \gamma v_{n,-1}. \tag{14}$$

We consider IV estimation of γ. As an IV, we consider a general linear function of $y_{n,+}$ of the form $A' y_{n,+}$ for some $(T+1) \times T$-matrix A. Below we will also use the form

$$a \equiv \text{vec } A.$$

Given A, our IV estimator of γ is

$$\hat{\gamma} = \frac{\sum_n y'_{n,+} A y_n}{\sum_n y'_{n,+} A y_{n,-1}}$$

$$= \gamma + \frac{\sum_n y'_{n,+} A \upsilon_n}{\sum_n y'_{n,+} A y_{n,-1}}. \tag{15}$$

We now investigate the conditions under which this estimator exists and, if so, if it is consistent. In order to do so, we need the following notation. Let $C_0' \equiv (I_T, 0_T)$

and let $C'_1, \ldots, C'_T$ 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. For example, for $T = 3$, we have

$$C'_0 = \begin{pmatrix} 1\,0\,0\,0 \\ 0\,1\,0\,0 \\ 0\,0\,1\,0 \end{pmatrix}; \ C'_1 = \begin{pmatrix} 0\,1\,0\,0 \\ 0\,0\,1\,0 \\ 0\,0\,0\,1 \end{pmatrix}; \ C'_2 = \begin{pmatrix} 0\,0\,1\,0 \\ 0\,0\,0\,1 \\ 0\,0\,0\,0 \end{pmatrix}; \ C'_3 = \begin{pmatrix} 0\,0\,0\,1 \\ 0\,0\,0\,0 \\ 0\,0\,0\,0 \end{pmatrix}.$$

Next, let

$$C \equiv (\operatorname{vec} C_0, \ldots, \operatorname{vec} C_T).$$

We now consider the requirements that A has to satisfy.

In the first place, A should be such that $\hat{\gamma}$ exists. More precisely, the expression for $\hat{\gamma}$ should be meaningful in the sense that neither numerator nor denominator is identically equal to zero. For example, if $T = 2$ and

$$A_1 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \\ -1 & 0 \end{pmatrix} \text{ and } A_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \\ 0 & 0 \end{pmatrix}$$

we have $y'_{n,+} A_1 y_n = 0$ and $y'_{n,+} A_2 y_{n,-1} = 0$. To exclude such cases, consider

$$y_{n,+} \otimes y_{n,+} = D_{T+1}(y_{n,+} \bar{\otimes} y_{n,+}),$$

where the bar over the Kronecker product indicates the omission of duplicate elements. The duplication matrix D_{T+1}, of order $(T+1)^2 \times (T+1)(T+2)/2$ restores them (see, e.g., Magnus and Neudecker 1986). Next, let

$$F_\tau \equiv (C'_\tau \otimes I_{T+1}) D_{T+1}, \quad \tau = 0, 1,$$

and note that

$$s_y = \frac{1}{N} \sum_n (y_{n,+} \bar{\otimes} y_{n,+}).$$

Using $y_{n,-1} = C'_0 y_{n,+}$ and $y_n = C'_1 y_{n,+}$, we can now write the estimator of γ as

$$\begin{aligned} \hat{\gamma} &= \frac{a' \sum_n (y_n \otimes y_{n,+})}{a' \sum_n (y_{n,-1} \otimes y_{n,+})} \\ &= \frac{a' (C'_1 \otimes I_{T+1}) D_{T+1} \sum_n (y_{n,+} \bar{\otimes} y_{n,+})}{a' (C'_0 \otimes I_{T+1}) D_{T+1} \sum_n (y_{n,+} \bar{\otimes} y_{n,+})} \\ &= \frac{a' F_1 s_y}{a' F_0 s_y}. \end{aligned} \tag{16}$$

So for a meaningful estimator, we should have both $F_0'a \neq 0$ and $F_1'a \neq 0$.

We next turn to consistency. From (14) and (15) we see that it requires

$$
\begin{aligned}
0 &= \mathrm{E}\left(y_{n,+}' A \upsilon_n\right) \\
&= \mathrm{E}\left(\alpha_n y_{n,+}' A \iota_T\right) + \mathrm{tr}\left(\mathrm{E}\left[(\varepsilon_n + v_n - \gamma v_{n,-1}) y_{n,+}'\right] A\right),
\end{aligned}
\tag{17}
$$

where tr indicates the trace. First, we have the equicorrelation property from (8),

$$
\mathrm{E}(\alpha_n y_{n,+}) = c \cdot \iota_{T+1}
$$

with $c = \sigma_\alpha^2/(1-\gamma)$. So one requirement for consistency is $\iota_{T+1}' A \iota_T = 0$ or

$$
\iota_{T(T+1)}' a = 0.
\tag{18}
$$

Since

$$
\mathrm{E}(\varepsilon_n y_{n,+}') = \sigma_\varepsilon^2 \sum_{\tau=1}^{T} \gamma^{\tau-1} C_\tau', \ \mathrm{E}(v_n y_{n,+}') = \sigma_v^2 C_1', \ \mathrm{E}(v_{n,-1} y_{n,+}') = \sigma_v^2 C_0',
\tag{19}
$$

we conclude from (17) and (19) that consistency is obtained when we let A be such that $\mathrm{tr}(C_t'A) = 0$ or $(\mathrm{vec}\, C_t)'a = 0$ for $t = 0, \ldots, T$. This means that a should satisfy

$$
C'a = 0_{T+1}.
\tag{20}
$$

Any estimator of the form (16) that satisfies (18), (20), and the existence conditions is consistent.

5 Efficient Estimation

To find an estimator that is not only consistent but also asymptotically efficient we have to distinguish between two kinds of efficiency, which we may label as local and global. We call $\hat{\gamma}$ locally efficient if it is in the class of estimators defined by $\hat{\gamma}$, with a properly restricted. A globally efficient estimator is as efficient as the GMM estimator discussed at the beginning of the previous section. There we saw that GMM on the covariance matrix was a daunting task. This task is greatly simplified when we already have estimators that are consistent without any further optimality qualities. We can then adapt these estimators such that we get asymptotically efficient estimators in just a single step. This approach is called *linearized GMM* (see, e.g., Wansbeek and Meijer 2000, Sect. 9.3). It requires initial consistent estimators not just of γ but also of the other model parameters, σ_α^2, σ_ε^2, and σ_v^2. Given a consistent estimator of γ, such estimators can be easily

constructed by employing some appropriately chosen moment conditions implied by the structure of Σ_y.

We now turn to local efficiency. To that end we write the restrictions on a, which are all linear, in the condensed form $a = Qb$, where Q is full column rank, and b can be chosen freely, subject to the numerator and denominator of $\hat{\gamma}$ not becoming identically zero. So we now have

$$\hat{\gamma} = \frac{a' \sum_n (y_n \otimes y_{n,+})}{a' \sum_n (y_{n,-1} \otimes y_{n,+})}$$

$$= \frac{b' Q' F_1 s_y}{b' Q' F_0 s_y}.$$

With

$$\Psi_y \equiv \mathrm{plim}_{N \to \infty} \frac{1}{N} \sum_n \left[(y_{n,+} \bar{\otimes} y_{n,+} - s_y)(y_{n,+} \bar{\otimes} y_{n,+} - s_y)' \right],$$

we obtain

$$\mathrm{AVar}(\hat{\gamma}) = \frac{b' \Phi_y b}{(b' Q' F_0 \sigma_y)^2}, \tag{21}$$

where $\Phi_y \equiv Q' F_1 \Psi_y F_1' Q$. If Φ_y were nonsingular, we could use the Cauchy–Schwarz inequality to derive the lower bound $(\sigma_y' F_0' Q \Phi_y^{-1} Q' F_0 \sigma_y)^{-1}$ of the asymptotic variance, with equality for $b = \Phi_y^{-1} Q' F_0 \sigma_y$. Hence, (16) would become

$$\hat{\gamma} = \frac{\hat{b}' Q' F_1 s_y}{\hat{b}' Q' F_0 s_y}$$

$$= \frac{s_y' F_0' Q \hat{\Phi}_y^{-1} Q' F_1 s_y}{s_y' F_0' Q \hat{\Phi}_y^{-1} Q' F_0 s_y},$$

where $\hat{b}$ denotes b with sample counterparts for Ψ_y and σ_y substituted. Unfortunately, however, it turns out that Φ_y is singular, and it is not immediately clear whether there exists a feasible optimal estimator, and if so, what this estimator would be. We leave this problem for future research. However, from the analysis here, an appealing consistent estimator is obtained by replacing the regular inverse with the Moore–Penrose generalized inverse. Thus, we propose the estimator

$$\hat{\gamma}_{\mathrm{MP}} \equiv \frac{s_y' F_0' Q \hat{\Phi}_y^{+} Q' F_1 s_y}{s_y' F_0' Q \hat{\Phi}_y^{+} Q' F_0 s_y}$$

$$= \frac{s_y' F_0' Q (Q' F_1 \hat{\Psi}_y F_1' Q)^{+} Q' F_1 s_y}{s_y' F_0' Q (Q' F_1 \hat{\Psi}_y F_1' Q)^{+} Q' F_0 s_y}, \tag{22}$$

and we call this the Moore–Penrose (MP) estimator. Its asymptotic variance can be estimated by the sample counterpart of (21), that is

$$\widehat{\text{AVar}}(\hat{\gamma}_{\text{MP}}) = \frac{1}{s_y' F_0' Q (Q' F_1 \hat{\Psi}_y F_1' Q)^+ Q' F_0 s_y}.$$

6 Illustrative Example

To illustrate the application of these estimators, we study the persistence in household wealth in the Health and Retirement Study (HRS; Juster and Suzman 1995). The HRS started in 1992 with a sample of individuals born in 1931–1941 and their spouses and interviewed them biennially afterward. Over time, additional cohorts have been added. We select all individuals who participated in all ten waves from 1992 to 2010 and who were either single across all waves or married to the same spouse across all waves. (We treat cohabitation the same as marriage, as is common in HRS analyses.) Because wealth is reported at the household level, we select only one respondent per household. This leaves us with a sample of 2,668 households.

We use the RAND version of the HRS, version L (St. Clair et al. 2011), including the imputations, and study total household wealth excluding the second home (HwATOTA), because information about the second home is not available in all waves. We compute the inverse hyperbolic sine transform of this variable and then subtract the wave-specific average, which captures macro effects and age effects. We then estimate the simple LDPDM for this transformed variable. We computed the standard Anderson–Hsiao estimator (by 2SLS), the Arellano-Bond estimator (using two-step GMM), the consistent "lagged" versions of these introduced earlier (AHL and ABL), and the MP estimator.

Table 1 shows the results. We clearly see the attenuation in the AH and AB estimators. Unfortunately, the standard errors increase substantially for the AHL and ABL estimators, compared to the AH and AB estimators. Meijer and Wansbeek (2000) showed this phenomenon for a cross-sectional regression model, but here it is even more dramatic. The MP estimate is close to the AH estimate, but its standard error is much smaller, though still almost four times as large as the standard errors of the (inconsistent) AH and AB estimators. Nevertheless, the MP estimate is highly

Table 1 Estimates of γ for transformed household wealth in the HRS

	AH	AB	AHL	ABL	MP
$\hat{\gamma}$	0.107***	0.119***	0.424	0.264*	0.417***
(s.e.)	(0.022)	(0.020)	(0.224)	(0.112)	(0.078)

AH Anderson–Hsiao, *AB* Arellano–Bond, *AHL* Anderson–Hsiao lagged, *ABL* Arellano–Bond lagged, *MP* Moore–Penrose estimator (22)
$^* p < 0.05,^{**} p < 0.01,^{***} p < 0.001$

significant and indicates a much stronger persistence in household wealth than we would conclude from the standard AH and AB estimators.

7   Discussion

Measurement error is a common problem in economic data, and this may have especially grave consequences in dynamic models. We study this and show the inconsistency of standard estimators for dynamic panel data models. We then develop a characterization of a class of consistent estimators and study efficiency within this class. Based on efficiency considerations, we propose an estimator, the Moore–Penrose (MP) estimator that has attractive statistical properties, although we have not been able to conclude whether it is the most efficient estimator in its class.

We apply the theory to the study of persistence of household wealth. We show that the attenuation bias of estimators that do not take measurement error into account can be quite large, and that our proposed estimator is much more efficient than two consistent estimators that are ad-hoc adaptations of the Anderson–Hsiao and Arellano–Bond estimators.

The results here are still quite limited. The set of model specifications needs to be expanded. Adding exogenous covariates is relatively straightforward, and weakly exogenous covariates can also be accommodated without much trouble. Our derivations thus far assume homoskedasticity, which is too strong in many economic applications. Relaxing this assumption adds restrictions that the estimator must satisfy, but does not conceptually change much. As indicated by Arellano and Bond (1991), a moving average process of the errors can be accommodated by dropping the first few lags of the dependent variable. Within our framework, this translates into additional linear restrictions. Although in the example, our estimator appears to work well, further efficiency gains may be obtained by GMM estimation based on (12). Fan et al. (2012) pursue such an approach for the static panel data model with measurement error and obtain even better results with a generalized quasi-likelihood-based estimator. We leave the development of the specifics to further research.

References

Ahn, S.C., Schmidt, P.: Efficient estimation of models for dynamic panel data. J. Econom. **68**, 5–27 (1995)

Aigner, D.J., Hsiao, C., Kapteyn, A., Wansbeek, T.: Latent variable models in econometrics. In: Griliches, Z., Intriligator, M.D. (eds.) Handbook of Econometrics, vol. 2, pp. 1321–1393. North-Holland, Amsterdam (1984)

Altonji, J.G., Siow, A.: Testing the response of consumption to income changes with (noisy) panel data. Q. J. Econ. **101**, 293–328 (1987)

Anderson, T.W., Hsiao, C.: Estimation of dynamic models with error components. J. Am. Stat. Assoc. **77**, 598–606 (1981)

Anderson, T.W., Hsiao, C.: Formulation and estimation of dynamic models using panel data. J. Econom. **18**, 47–82 (1982)

Antman, F., McKenzie, D.: Poverty traps and nonlinear income dynamics with measurement error and individual heterogeneity. J. Dev. Stud. **43**, 1057–1083 (2007)

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

Arellano, M., Bover, O.: Another look at the instrumental variable estimation of error-components models. J. Econom. **68**, 29–51 (1995)

Biørn, E.: The measurement error problem in dynamic panel data analysis: modeling and GMM estimation, Working paper, University of Oslo, 2012

Blundell, R., Bond, S.: Initial conditions and moment restrictions in dynamic panel data models. J. Econom. **87**, 115–143 (1998)

Bond, S., Hoeffler, A., Temple, J.: GMM estimation of empirical growth models, CEPR Working Paper 3048, 2001

Buonaccorsi, J.: Measurement Error, Models, Methods, and Applications. Chapman & Hall, Boca Raton (2010)

Chen, J., Ni, S., Podgursky, M.: Estimating dynamic panel data models with measurement errors with an application to school evaluation based on student test scores. In: Proceedings of the 2008 Meeting of the American Statistical Association, pp. 951–957 (2008)

Fan, Z., Sutradhar, B.C., Rao, R.P.: Bias corrected generalized method of moments and generalized quasi-likelihood inferences in linear models for panel data with measurement error. Sankhyā B. **74**, 126–148 (2012)

Griliches, Z., Hausman, J.A.: Errors in variables in panel data. J. Econom. **31**, 93–118 (1986)

Harris, M.N., Mátyás, L.: Performance of the operational Wansbeek-Bekker estimator for dynamic panel data models. Appl. Econ. Lett. **7**, 149–153 (2000)

Juster, F.T., Suzman, R.: An overview of the Health and Retirement Study. J. Hum. Resour. **30**, S7–S56 (1995)

Komunjer, I., Ng, S.: Measurement errors in dynamic models, Working paper, 2011

Magnus, J.R., Neudecker, H.: Symmetry, 0–1 matrices and Jacobians: A review. Econom. Theory **2**, 157–190 (1986)

Meijer, E., Wansbeek, T.J.: Measurement error in a single regressor. Econ. Lett. **69**, 277–284 (2000)

Meijer, E., Spierdijk, L., Wansbeek, T.J.: Point and set identification in linear panel data models with measurement error. Working Paper WR-941, RAND Corporation, Santa Monica, 2012

St. Clair, P., et al.: RAND HRS Data Documentation, Version L. RAND Corporation, Santa Monica (2011)

Wansbeek, T.J., Bekker, P.A.: On IV, GMM and ML in a dynamic panel data model. Econ. Lett. **51**, 145–152 (1996)

Wansbeek, T.J., Kapteyn, A.: Simple estimators for dynamic panel data models with errors in the variables. In: Bewley, R., Van Hoa, T. (eds.) Contributions to Consumer Demand and Econometrics. McMillan, London (1992)

Wansbeek, T.J., Meijer, E.: Measurement Error and Latent Variables in Econometrics. North-Holland, Amsterdam (2000)