Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Moment-Condition Contamination and Pseudo-True Values of Difference-GMM and System-GMM under Persistent AR(1) Measurement Error

sreedevi pn
Pondicherry University
Dr. J. Prabhakara Naik
drnaik.sta@pondiuni.ac.in

Pondicherry University
Asha Ponnada
Pondicherry University

Research Article

Keywords: Dynamic panel data models, Moment-condition contamination, Difference-GMM, System-GMM, Measurement error persistence, Pseudo-true value

Posted Date: November 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11565709/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Moment-Condition Contamination and Pseudo-True Values of Difference-GMM and System-GMM under Persistent AR(1) Measurement Error

Sreedevi P. N.[1], Dr. J. Prabhakara Naik[2]*, and Ponnada Asha[3]

[1,3]Research Scholar [2]Associate Professor
Department of Statistics
Ramanujan School of Mathematical Sciences
Pondicherry University
Puducherry, India

Abstract

This paper studies Difference-GMM and System-GMM estimation of the autoregressive parameter of a dynamic panel when the dependent variable is observed with persistent AR(1) measurement error. Our contribution is a closed-form derivation of how the error contaminates the moment conditions of both estimators. For the lag-s lagged-level instrument of the differenced equation the contamination is $(\phi-1)(\phi-\rho)\phi^{s-2}\sigma_e^2$, and for the lag-$s$ lagged-difference instrument of the level equation it is $(1-\phi)(\phi-\rho)\phi^{s-1}\sigma_e^2$, where ρ is the autoregressive parameter and ϕ and σ_e^2 are the persistence and variance of the error. Both vanish when the error is exactly as persistent as the latent process, because the composite disturbance then collapses to serially uncorrelated noise, and, except at the shallowest lag, when the error is serially independent. We also derive the pseudo-true value of each estimator: for a single instrument it is a weighted average of ρ and ϕ, with the weight on ϕ equal to the noise share of the differenced variance at that instrument's lag. A Monte Carlo study shows that the derived expressions reproduce the simulated biases, that the bias does not vanish as the cross-section or the panel length grows, that the two estimators are biased by almost the same amount under classical error, and that dropping the shallowest instruments removes most of the classical-error bias but not the bias caused by persistent error.

Keywords: Dynamic panel data models, Moment-condition contamination, Difference-GMM, System-GMM, Measurement error persistence, Pseudo-true value

1 Introduction

Dynamic panel data models occupy a central place in applied econometrics precisely because so many economic outcomes are self-referential: today's productivity, output, or inflation rate

*Corresponding author. Email: drnaik.sta@pondiuni.ac.in

is shaped in part by where that same variable stood yesterday. Letting the lagged dependent variable enter as a regressor is a natural way to capture this persistence and adjustment behavior while still allowing for heterogeneity across units that ordinary panel methods cannot absorb. Once a lagged outcome appears on the right-hand side, however, it is automatically correlated with the individual effect, and pooled OLS or fixed-effects estimation becomes biased and inconsistent – a problem that worsens as the number of time periods shrinks. Nickell [1] first characterized this bias formally, and the search for instrumental-variable and GMM-based fixes followed directly from that diagnosis.

The GMM framework of Hansen [2] supplied the tools needed to exploit orthogonality between instruments and the disturbance without requiring a fully specified likelihood. Arellano and Bond [3] turned this into a workable estimator for dynamic panels – Difference-GMM – by first-differencing the model to eliminate the individual effect and then using suitably lagged levels of the dependent variable as internal instruments for the differenced equation. A closely related, near-contemporaneous line of work by Holtz-Eakin et al. [4] developed GMM estimation for panel vector auto-regressions along similar lines. Arellano and Bover [5] and Blundell and Bond [6] later showed that efficiency could be improved substantially, especially when the series is highly persistent, by supplementing the differenced moment conditions with a second set derived from the level equation – the System-GMM approach. Ahn and Schmidt [7] showed that the linear moment conditions exploited by both estimators do not exhaust the information available in the model, additional nonlinear conditions being available in principle. Roodman [8] analyses the problems caused by instrument proliferation in these estimators, including weakened overidentification tests, and proposes collapsing the instrument set; Kiviet et al. [9] revisit the finite-sample accuracy and efficiency of the available GMM techniques under weak identification and instrument proliferation.

All of this machinery, though, is only as good as the orthogonality conditions it leans on. If the dependent variable is measured with error, those conditions can break down, putting both estimation and the diagnostic tests built on top of it at risk. Griliches and Hausman [10] established the classical consequences of errors-in-variables for panel data estimation more broadly, showing that the very transformations used to remove fixed effects can aggravate rather than neutralize measurement error. Biørn [11] traces through how measurement error corrupts the standard GMM setup for dynamic panels and shows that it can undermine the very instrument sets Difference- and System-GMM depend on. Xiao et al. [12] derive instrumental-variable and GMM estimators for panel data with measurement error under error structures more general than the classical case. Lee et al. [13] take up the fixed-effects case directly, showing that conventional estimation stops being tractable once measurement error is present, and propose an alternative route around the problem. Lee et al. [14] extend this line of work with instrumental-variable procedures suited to large-N, large-T panels contaminated by measurement error. Separately, Chowdhury and Russell [15] show that structural instability interacts with dynamic panel GMM's moment conditions in ways that standard specification tests do not reliably detect. Taken together, this body of work makes clear that measurement error is not a peripheral concern for identification, estimation, or inference in dynamic panels.

Closer to the present setting, Biørn [16] reports Monte Carlo evidence on GMM bias and

instrument validity when the variables of a dynamic panel are measured with error, and Biørn [17] sets out the corresponding modeling and GMM estimation framework. Wansbeek [18] and Meijer et al. [19] study GMM and consistent estimation of linear panel data models with measurement error. These contributions are mainly concerned with modeling and estimation under specified error structures; Biørn [27] in particular studies IV and GMM procedures that combine levels and differences when the variables are mismeasured, and the invalidity of the shallowest lagged-level instrument under additive error is familiar from the micro-panel literature [10, 28].

A separate literature studies the finite-sample behavior of dynamic panel estimators. Harris and Mátyás [20] compare a range of IV and GMM estimators in Monte Carlo experiments. Windmeijer [21] derives a finite-sample correction to the estimated variance of two-step GMM estimators; this concerns standard errors rather than the bias of the point estimates, and so does not address the inconsistency studied here. Andrews and Lu [22] propose consistent model and moment selection procedures for GMM with an application to dynamic panel models, which bears on the choice of instrument sets. Hayakawa [23] and Hayakawa and Qi [24] study the asymptotic properties and the weak-instrument problem of the System-GMM estimator, and Harris and Mátyás [25] examine the robustness of dynamic panel estimators to misspecification. The bias documented below differs in kind from these finite-sample problems: it does not vanish as N or T grows (Section 4.3), because it arises from moment conditions that are invalid in the population rather than from weak or numerous instruments.

The case of persistent measurement error, meaning error that is itself serially correlated, has received less attention in the form studied here. Biørn [11, 16, 17] allows the disturbances and the measurement errors of an autoregressive fixed-effects panel equation to have finite memory, and examines how the set of valid instruments and the finite-sample bias of GMM change with that memory. An error that follows an AR(1) process has infinite memory: the contamination of the lag-s moment does not switch off beyond a finite lag but decays geometrically at rate ϕ.

To our knowledge, the closed-form derivation of that contamination is new, and it is the contribution of this paper. We derive, for every instrument lag, the exact value by which AR(1) measurement error violates the moment conditions of Difference-GMM (Proposition 3.1) and of System-GMM (Proposition 3.2). The result is simple: the contamination is proportional to $(\phi - \rho)$, where ρ is the autoregressive parameter and ϕ the persistence of the error, and it decays geometrically with the instrument lag. We then go one step further and derive the pseudo-true value to which each estimator converges (Proposition 3.5), which turns the contamination into a bias formula. A Monte Carlo study compares Difference-GMM and System-GMM under the error and checks the derived expressions against simulated biases.

The paper proceeds as follows. Section 2 sets out the dynamic panel model with AR(1) measurement error and the two estimators. Section 3 derives the contamination of the moment conditions and the pseudo-true values. Section 4 reports the simulation results, and Section 5 concludes.

2 Model and Estimators

2.1 Dynamic Panel with AR(1) Measurement Error

Consider a panel of N units observed over T periods. The latent dynamic panel model is

$$y_{it}^{*} = \rho y_{i,t-1}^{*} + \alpha_i + u_{it}, \qquad i = 1, \ldots, N,\ t = 2, \ldots, T, \tag{1}$$

where y_{it}^{*} is the true, unobserved variable, ρ with $|\rho| < 1$ is the autoregressive parameter, α_i is the individual effect and u_{it} is the idiosyncratic disturbance.

Assumption 2.1 (Structural innovations)**.** The disturbances u_{it} have mean zero and variance σ_u^2, are serially uncorrelated, and are uncorrelated with α_i.

Instead of y_{it}^{*} we observe

$$y_{it} = y_{it}^{*} + e_{it}, \tag{2}$$

where the measurement error follows an AR(1) process,

$$e_{it} = \phi e_{i,t-1} + \eta_{it}, \qquad |\phi| < 1. \tag{3}$$

Assumption 2.2 (Persistent measurement error)**.** The innovation η_{it} is white noise with mean zero and variance σ_η^2. The process $\{e_{it}\}$ is independent of $\{y_{it}^{*}\}$, α_i and $\{u_{it}\}$, including the initial values, and e_{i1} is drawn from the stationary distribution of the AR(1) process, so that

$$\operatorname{Cov}\left(e_{i,t-j}, e_{i,t-k}\right) = \phi^{|j-k|}\sigma_e^2, \qquad \sigma_e^2 = \frac{\sigma_\eta^2}{1-\phi^2}. \tag{4}$$

The case $\phi = 0$ is classical (serially independent) measurement error and $\sigma_e^2 = 0$ is the error-free case. Substituting (2) into (1) gives the observed model

$$y_{it} = \rho y_{i,t-1} + \alpha_i + \varepsilon_{it}, \qquad \varepsilon_{it} = u_{it} + e_{it} - \rho e_{i,t-1}. \tag{5}$$

The composite disturbance ε_{it} is no longer serially uncorrelated: it is an MA(1) process when $\phi = 0$ and an ARMA(1,1) process otherwise. First-differencing removes α_i:

$$\Delta y_{it} = \rho\, \Delta y_{i,t-1} + \Delta\varepsilon_{it}, \qquad \Delta\varepsilon_{it} = \Delta u_{it} + \Delta e_{it} - \rho\, \Delta e_{i,t-1}. \tag{6}$$

2.2 GMM Estimation

Let z_i be a vector of instruments and $\varepsilon_i(\theta)$ the disturbance evaluated at the parameter value θ. GMM rests on the population orthogonality condition

$$\mathbb{E}\left[z_i'\, \varepsilon_i(\theta_0)\right] = 0, \tag{7}$$

and estimates $\theta_0 = \rho$ by minimising a quadratic form in the sample analogue of this moment,

$$\widehat{\theta} = \arg\min_{\theta} \; g_N(\theta)' \, W \, g_N(\theta), \qquad g_N(\theta) = \frac{1}{N} \sum_{i=1}^{N} z_i' \, \varepsilon_i(\theta), \tag{8}$$

for a symmetric positive-definite weighting matrix W. The estimator is consistent only if (7) holds, and this is what measurement error threatens. The Hansen [2] and Sargan [26] overidentification tests and the Arellano–Bond [3] serial-correlation tests are the usual checks on the instruments.

2.3 Difference-GMM

Without measurement error, first-differencing (1) gives

$$\Delta y_{it}^{*} = \rho \, \Delta y_{i,t-1}^{*} + \Delta u_{it}. \tag{9}$$

Because $\Delta y_{i,t-1}^{*}$ and Δu_{it} both contain $u_{i,t-1}$, the differenced regressor is endogenous. Arellano and Bond [3] instrument it with lagged levels, which under Assumption 2.1 are uncorrelated with Δu_{it}:

$$\mathbb{E}\left(y_{i,t-s}^{*} \, \Delta u_{it} \right) = 0, \qquad s \geq 2, \; t = 3, \ldots, T. \tag{10}$$

Difference-GMM is the GMM estimator (8) based on these moments, with the instruments $y_{i1}, \ldots, y_{i,t-2}$ for period t.

2.4 System-GMM

Difference-GMM has weak instruments when the series is highly persistent. Blundell and Bond [6] therefore add the level equation (1), using lagged first differences as instruments. This requires one further assumption.

Assumption 2.3 (Mean stationarity of the initial condition)**.** The first difference of the latent variable is uncorrelated with the individual effect:

$$\mathbb{E}\left(\Delta y_{i2}^{*} \, \alpha_i \right) = 0. \tag{11}$$

Under Assumption 2.3, lagged first differences are valid instruments for the level equation:

$$\mathbb{E}\left[\Delta y_{i,t-s}^{*} \left(\alpha_i + u_{it} \right) \right] = 0, \qquad s \geq 1. \tag{12}$$

System-GMM combines (10) and (12) and is typically more efficient than Difference-GMM when ρ is close to one or T is short. In its standard implementation the level-equation instrument is $\Delta y_{i,t-1}$, that is, $s = 1$.

3 Contamination of the Moment Conditions

Both estimators are consistent only if the orthogonality conditions (10) and (12) hold. Under measurement error the instruments are the observed variables, $y_{i,t-s} = y^*_{i,t-s} + e_{i,t-s}$, and the disturbances are those of the observed model (5)–(6). We therefore evaluate the observed-data counterparts of the two conditions.

3.1 Difference-GMM

For Difference-GMM the condition to check is $\mathbb{E}(y_{i,t-s}\Delta\varepsilon_{it}) = 0$ for $s \geq 2$. Its exact value is given by the following result.

Proposition 3.1 (Contamination of the Difference-GMM moment conditions)**.** *Under Assumptions 2.1 and 2.2, for every $s \geq 2$,*

$$b_s^D \;\equiv\; \mathbb{E}\,(y_{i,t-s}\Delta\varepsilon_{it}) = (\phi-1)(\phi-\rho)\,\phi^{\,s-2}\,\sigma_e^2. \tag{13}$$

Consequently:

(i) for $s \geq 3$ the condition $b_s^D = 0$ holds if and only if $\phi = 0$, $\phi = \rho$ or $\sigma_e^2 = 0$; otherwise it is violated, the contamination decays geometrically in s and its sign reverses as ϕ crosses ρ;

(ii) for $s = 2$ the factor $\phi^{s-2} = 1$ for every ϕ, so $b_2^D = (1-\phi)(\rho-\phi)\sigma_e^2$, which vanishes only at $\phi = \rho$ or $\sigma_e^2 = 0$. In particular, at $\phi = 0$ it equals $\rho\,\sigma_e^2 \neq 0$: the lag-2 instrument is contaminated even by classical error.

Proof. By (6),

$$\mathbb{E}\,(y_{i,t-s}\Delta\varepsilon_{it}) = \mathbb{E}\,(y_{i,t-s}\Delta u_{it}) + \mathbb{E}\,(y_{i,t-s}\Delta e_{it}) - \rho\,\mathbb{E}\,(y_{i,t-s}\Delta e_{i,t-1})\,.$$

The first term is zero: for $s \geq 2$ the instrument depends on u only up to period $t-2$, and the disturbances are serially uncorrelated and unrelated to the error process. By Assumption 2.2, y^* is independent of e, so the other two terms involve only the error. Using $\Delta e_{it} = (\phi-1)e_{i,t-1} + \eta_{it}$ and (4),

$$\mathbb{E}\,(y_{i,t-s}\Delta e_{it}) = \mathbb{E}\,(e_{i,t-s}\Delta e_{it}) = (\phi-1)\phi^{\,s-1}\sigma_e^2, \qquad \mathbb{E}\,(y_{i,t-s}\Delta e_{i,t-1}) = (\phi-1)\phi^{\,s-2}\sigma_e^2,$$

so that $b_s^D = (\phi-1)\phi^{\,s-2}\sigma_e^2(\phi-\rho)$, which is (13). For (i), $s \geq 3$ gives $\phi^{s-2} = 0$ at $\phi = 0$, and $\phi - 1 \neq 0$. For (ii), $\phi^0 = 1$ for every ϕ including $\phi = 0$; directly, at $\phi = 0$ the only surviving term is $-\rho\,\mathbb{E}[e_{i,t-2}(e_{i,t-1} - e_{i,t-2})] = \rho\sigma_e^2$. □

3.2 System-GMM

For System-GMM the level-equation condition is $\mathbb{E}[\Delta y_{i,t-s}(\alpha_i + \varepsilon_{it})] = 0$ for $s \geq 1$. Under Assumptions 2.2 and 2.3, $\Delta y_{i,t-s} = \Delta y^*_{i,t-s} + \Delta e_{i,t-s}$ is uncorrelated with α_i, so α_i drops out and the condition becomes $\mathbb{E}(\Delta y_{i,t-s}\varepsilon_{it}) = 0$.

Proposition 3.2 (Contamination of the System-GMM level moment conditions)**.** *Under Assumptions 2.1–2.3, for every $s \geq 1$,*

$$b_s^L \equiv \mathbb{E}\left(\Delta y_{i,t-s}\varepsilon_{it}\right) = (1-\phi)(\phi-\rho)\,\phi^{\,s-1}\,\sigma_e^2. \tag{14}$$

Consequently:

(i) for $s \geq 2$ the condition $b_s^L = 0$ holds if and only if $\phi = 0$, $\phi = \rho$ or $\sigma_e^2 = 0$, with the same geometric decay and sign reversal as in Proposition 3.1;

(ii) for $s = 1$ the factor $\phi^{s-1} = 1$ for every ϕ, so $b_1^L = (1-\phi)(\phi-\rho)\sigma_e^2$, which vanishes only at $\phi = \rho$ or $\sigma_e^2 = 0$, and at $\phi = 0$ equals $-\rho\,\sigma_e^2 \neq 0$.

Proof. By (5),

$$\mathbb{E}\left(\Delta y_{i,t-s}\varepsilon_{it}\right) = \mathbb{E}\left(\Delta y_{i,t-s}u_{it}\right) + \mathbb{E}\left(\Delta y_{i,t-s}e_{it}\right) - \rho\,\mathbb{E}\left(\Delta y_{i,t-s}e_{i,t-1}\right).$$

The first term is zero because u_{it} is uncorrelated with the past and with the error process. By Assumption 2.2 the remaining terms involve only $\Delta e_{i,t-s}$, and by (4) we have

$$\mathbb{E}\left(\Delta e_{i,t-s}e_{it}\right) = \phi^{\,s}(1-\phi)\sigma_e^2, \qquad \mathbb{E}\left(\Delta e_{i,t-s}e_{i,t-1}\right) = \phi^{\,s-1}(1-\phi)\sigma_e^2.$$

Hence $b_s^L = \phi^{\,s-1}(1-\phi)\sigma_e^2(\phi-\rho)$, which is (14). Claims (i) and (ii) follow as in Proposition 3.1; at $\phi = 0$ and $s = 1$, $\mathbb{E}[(e_{i,t-1} - e_{i,t-2})(e_{it} - \rho e_{i,t-1})] = -\rho\sigma_e^2$. □

3.3 Consequences

Corollary 3.3 (Minimum instrument depth)**.** *Under Assumptions 2.1–2.3 with $\sigma_e^2 > 0$, the shallowest instrument of each set, $y_{i,t-2}$ in the differenced equation and $\Delta y_{i,t-1}$ in the level equation, is contaminated for every $\phi \neq \rho$. Restricting the instruments to $s \geq 3$ in the differenced equation and $s \geq 2$ in the level equation removes the contamination of classical error ($\phi = 0$) entirely, but for $0 < \phi \neq \rho$ the remaining instruments are still contaminated, by a factor ϕ^{s-2} or ϕ^{s-1}.*

Remark 3.4. (a) *Classical error.* For $\phi = 0$ the composite disturbance is MA(1) and its first difference MA(2), so the instrument lagged exactly two periods is never orthogonal to $\Delta\varepsilon_{it}$; Propositions 3.1 and 3.2 show that this is the only failure in the classical case. The conventional Blundell–Bond level instrument $\Delta y_{i,t-1}$ is the shallowest level instrument and is therefore invalid under any AR(1) measurement error with $\phi \neq \rho$ and $\sigma_e^2 > 0$. (b) *The point $\phi = \rho$.* The contamination vanishes at every lag because the cancellation is a common-factor one: at $\phi = \rho$ the composite disturbance reduces to $u_{it} + \eta_{it}$, which is serially uncorrelated, so the observed series is an AR(1) process with a larger innovation variance and the error is observationally equivalent to extra structural noise. It is a knife-edge case that cannot be targeted in practice, because ϕ is unknown. (c) *Non-monotonicity.* Away from $\phi = 0$ and $\phi = \rho$ the contamination grows with σ_e^2 and switches sign as ϕ crosses ρ.

3.4 From Contamination to Bias: Pseudo-true Values

Propositions 3.1 and 3.2 say which moment conditions fail and with what sign. They do not give the bias, because the contamination enters the estimator divided by the relevance of the instrument. We derive both for the collapsed instrument sets used in Section 4. For a set $\mathcal{S}_D$ of difference-equation lags and a set $\mathcal{S}_L$ of level-equation lags, unit i contributes the moments

$$g_{i,s}^{D}(\theta) = \sum_{t} y_{i,t-s}\,(\Delta y_{it} - \theta\,\Delta y_{i,t-1})\,, \quad s \in \mathcal{S}_D;$$
$$g_{i,s}^{L}(\theta) = \sum_{t} \Delta y_{i,t-s}\,(y_{it} - \theta\,y_{i,t-1})\,, \quad s \in \mathcal{S}_L,$$

each summed over the n_s periods for which the instrument is available. Each moment is linear in θ. Under Assumption 3.1 below, let the relevance of the instruments be

$$c_s^D = \mathbb{E}\left[y_{i,t-s}\,\Delta y_{i,t-1}\right] = -\frac{\rho^{s-2}\sigma_u^2}{1+\rho} - (1-\phi)\,\phi^{s-2}\sigma_e^2, \tag{15}$$
$$c_s^L = \mathbb{E}\left[\Delta y_{i,t-s}\,y_{i,t-1}\right] = \frac{\rho^{s-1}\sigma_u^2}{1+\rho} + (1-\phi)\,\phi^{s-1}\sigma_e^2. \tag{16}$$

Assumption 3.1 (Covariance stationarity)**.** The latent process starts from its stationary distribution: $y_{i1}^{*} = \alpha_i/(1-\rho) + v_{i1}$, where v_{i1} has mean zero and variance $\sigma_u^2/(1-\rho^2)$ and is uncorrelated with α_i and with $\{u_{it}\}_{t\geq 2}$. This implies Assumption 2.3.

Proposition 3.5 (Pseudo-true value)**.** *Under Assumptions 2.1, 2.2 and 3.1, let $\tilde{b}$ and $\tilde{c}$ be the vectors with elements $n_s b_s$ and $n_s c_s$ for the moments in the chosen instrument set. Then the population moment vector is $\mathbb{E}[g_i(\theta)] = \tilde{b} - (\theta-\rho)\tilde{c}$, and if the weighting matrix converges to a positive-definite matrix W and $\tilde{c}'W\tilde{c} \neq 0$, the GMM estimator converges in probability to*

$$\theta^{*} = \rho + \frac{\tilde{c}'\,W\,\tilde{b}}{\tilde{c}'\,W\,\tilde{c}}. \tag{17}$$

Proof. Because $\Delta y_{it} = \rho\Delta y_{i,t-1} + \Delta\varepsilon_{it}$, $\mathbb{E}[y_{i,t-s}(\Delta y_{it} - \theta\Delta y_{i,t-1})] = b_s^D - (\theta-\rho)c_s^D$, and likewise for the level block; stationarity makes each expectation the same for every t. For the relevance terms,

$$\Delta y_{i,t-1}^{*} = \rho^{s-2}\Delta y_{i,t-s+1}^{*} + \sum_{j=0}^{s-3}\rho^{j}\Delta u_{i,t-1-j},$$

the innovations are uncorrelated with $y_{i,t-s}^{*}$, and $\mathbb{E}[y_{i,t-s}^{*}\Delta y_{i,t-s+1}^{*}] = (\rho-1)\,\mathrm{Var}(y^{*}) + \mathrm{Cov}(y^{*},\alpha_i) = -\sigma_u^2/(1+\rho)$, the terms in σ_α^2 cancelling; also $\mathbb{E}[e_{i,t-s}\Delta e_{i,t-1}] = \sigma_e^2(\phi^{s-1} - \phi^{s-2})$. The level relevance follows from $\mathbb{E}[\Delta y_{i,t-s}^{*} y_{i,t-1}^{*}] = \rho^{s-1}\sigma_u^2/(1+\rho)$ and $\mathbb{E}[\Delta e_{i,t-s}e_{i,t-1}] = \sigma_e^2(1-\phi)\phi^{s-1}$. Minimising $g'Wg$ over θ gives (17). □

With a single instrument the weighting matrix is irrelevant and the limit takes a simple form. Define the differenced noise-to-signal ratio

$$\kappa = \frac{(1+\rho)(1-\phi)\,\sigma_e^2}{\sigma_u^2}, \tag{18}$$

the ratio of the variance of Δe_{it}, $2(1-\phi)\sigma_e^2$, to that of Δy_{it}^*, $2\sigma_u^2/(1+\rho)$.

Corollary 3.6 (Single-instrument limits)**.** *Under Assumptions 2.1, 2.2 and 3.1, let* $0 < \rho < 1$ *and* $0 \leq \phi < 1$*, with* $0^0 = 1$*. The estimator that uses the single instrument* $y_{i,t-s}$ *(*$s \geq 2$*) converges to*

$$\theta_s^* = \frac{\rho + \phi\,\kappa_s}{1+\kappa_s} = \rho + (\phi-\rho)\frac{\kappa_s}{1+\kappa_s}, \qquad \kappa_s = \kappa\left(\frac{\phi}{\rho}\right)^{s-2}, \tag{19}$$

and the estimator that uses the single instrument $\Delta y_{i,t-s}$ *(*$s \geq 1$*) converges to the same expression with* $\kappa_s = \kappa(\phi/\rho)^{s-1}$*. The two shallowest instruments,* $y_{i,t-2}$ *and* $\Delta y_{i,t-1}$*, have the same limit,* $(\rho+\phi\kappa)/(1+\kappa)$*, and every limit lies between* ρ *and* ϕ*.*

Proof. With one moment, Proposition 3.5 gives $\theta_s^* - \rho = b_s/c_s$. For the differenced equation,

$$\frac{b_s^D}{c_s^D} = \frac{(1-\phi)(\phi-\rho)\phi^{s-2}\sigma_e^2}{\rho^{s-2}\sigma_u^2/(1+\rho) + (1-\phi)\phi^{s-2}\sigma_e^2};$$

dividing numerator and denominator by $\rho^{s-2}\sigma_u^2/(1+\rho)$ gives $(\phi-\rho)\kappa_s/(1+\kappa_s)$. The level equation is identical with $s-1$ in place of $s-2$. □

Remark 3.7. (a) *Classical error.* At $\phi = 0$ the shallowest-instrument limit is $\rho/(1+\kappa)$, the familiar errors-in-variables attenuation, here in the differenced variables: with $\rho = 0.7$, $\sigma_u = 1$ and $\sigma_e = 1$, $\kappa = 1.7$ and the limit is 0.26. Because the two shallowest instruments share the same limit, Difference- and System-GMM are biased by nearly the same amount. (b) *Persistent error.* For $\phi > 0$ the limit is pulled towards ϕ rather than towards zero, which explains the sign reversal at $\phi = \rho$. (c) *Depth.* Since $\kappa_s = \kappa(\phi/\rho)^{s-2}$, deeper instruments carry geometrically less noise relative to signal when $\phi < \rho$ but more when $\phi > \rho$: at $\rho = 0.7$, $\phi = 0.3$ and $\sigma_e = 0.5$ the single-instrument biases for $s = 2, 3, 4$ are -0.092, -0.045 and -0.021, whereas at $\phi = 0.9$ and $\sigma_e = 1$ they are $+0.029$, $+0.036$ and $+0.044$. Dropping the shallowest instruments therefore reduces the pseudo-true bias only when the error is less persistent than the series. (d) *Several moments.* Equation (17) is an affine combination $\theta^* = \sum_j w_j \theta_j^*$ of the single-instrument limits, with weights $w_j = \tilde{c}_j(W\tilde{c})_j/(\tilde{c}'W\tilde{c})$ summing to one. The weights depend on T through n_s, so the limit of the collapsed estimator moves with T although each θ_j^* does not.

4 Simulation Study

4.1 Design

The latent process is generated from (1) with $\rho = 0.70$, $\sigma_u = \sigma_\alpha = 1$ and normal disturbances. The initial value y_{i1}^* is drawn from the stationary distribution, so that Assumptions 2.3 and 3.1 hold, and the error e_{it} from the stationary distribution of (3), with $\sigma_\eta^2 = \sigma_{me}^2(1-\phi^2)$ so that the error variance equals σ_{me}^2 for every ϕ; changes in ϕ therefore isolate persistence from magnitude. The main design has $T = 8$, $N = 500$, $\phi \in \{0, 0.3, 0.6, 0.9\}$ and $\sigma_{me} \in \{0.25, 0.5, 1.0\}$, together with the error-free benchmark, the case $\phi = \rho = 0.7$, and $N \in \{200, 1000\}$ where stated. Each design point uses $R = 1000$ replications.

Both estimators are two-step GMM estimators with the collapsed instrument set of Roodman [8]: Difference-GMM uses the lagged levels $s = 2, \ldots, 5$, and System-GMM adds the level-equation instrument $\Delta y_{i,t-1}$. The first step uses an identity weighting matrix and the second the inverse of the estimated moment covariance matrix. We report the bias, the Monte Carlo standard deviation (MCSD) and the root mean squared error (RMSE) of the estimate of ρ. The simulations were run in R, and the code accompanies the paper; a single random-number seed is set at the start of the program. Design points are independent runs, so a design that appears in more than one table (other than Tables 1 and 4, which share the same runs) is based on different random draws; the standard error of the difference between two such simulated biases is about 0.003, so differences of up to about 0.005 between tables are Monte Carlo error.

4.2 Bias under Persistent Measurement Error

Table 1 reports the main results at $N = 500$. Without measurement error both estimators recover ρ with negligible bias (-0.001 for Difference-GMM and -0.001 for System-GMM), and System-GMM is already markedly more precise (RMSE 0.038 against 0.068). Introducing error changes the picture as Propositions 3.1 and 3.2 predict. The largest biases occur under classical error ($\phi = 0$), where the shallowest instrument is contaminated: at $\sigma_{me} = 0.5$ the bias is -0.206 for Difference-GMM and -0.196 for System-GMM, and at $\sigma_{me} = 1.0$ it is -0.447 and -0.425, so that the average estimate falls to about 0.25–0.28 against a true value of 0.70. The bias shrinks in absolute value as ϕ rises towards ρ (at $\sigma_{me} = 0.5$: -0.094 and -0.089 at $\phi = 0.3$, -0.020 and -0.016 at $\phi = 0.6$), is zero up to Monte Carlo error at $\phi = \rho = 0.7$ (-0.001 and 0.001 at $\sigma_{me} = 1.0$), and turns positive once ϕ exceeds ρ ($+0.027$ and $+0.030$ at $\phi = 0.9$, $\sigma_{me} = 1.0$). The bias is therefore non-monotonic in ϕ and increases with σ_{me} at every ϕ other than ρ.

The two estimators are biased by almost the same amount, because they share the contaminated shallowest instrument (Remark 3.7(a)). System-GMM nevertheless has the lower RMSE in all 16 cells of Table 1, so measurement error erodes but does not remove its efficiency advantage.

4.3 Sample Size and Panel Length

Table 2 shows the bias under classical error ($\sigma_{me} = 0.5$) for $N \in \{200, 500, 1000\}$ and $T \in \{5, 8, 15, 30\}$, beside the pseudo-true value of Proposition 3.5 (Section 4.4). Without measurement error the bias is a small-sample phenomenon: at $N = 1000$ it is at most 0.008 in absolute value for every T. Under classical error it is different. For $T \geq 8$ the bias is essentially unchanged by the sample size once $N \geq 200$ (for Difference-GMM at $T = 8$: -0.211, -0.203 and -0.205 at $N = 200$, 500 and 1000; at $T = 5$ it falls only from -0.245 to -0.215), and it declines only slowly with T, from -0.215 at $T = 5$ to -0.163 at $T = 30$ (Difference-GMM, $N = 1000$), so that almost a quarter of the true value is lost even at $T = 30$. Neither a large cross-section nor a long panel restores consistency while the contaminated shallowest instrument is used. The slow decline in T is not consistency in the making: the pseudo-true value, which does not depend on N, declines in the same way, because T changes only the weights that the collapsed estimator gives to deeper, less contaminated lags (Remark 3.7(d)).

Table 1: Bias and RMSE of Difference-GMM and System-GMM under AR(1) measurement error ($N = 500$, $T = 8$, $\rho = 0.70$, $R = 1000$)

ϕ	σ_{me}	Diff.-GMM bias	Diff.-GMM RMSE	Sys.-GMM bias	Sys.-GMM RMSE
0.0	0.00	−0.001	0.068	−0.001	0.038
0.0	0.25	−0.065	0.096	−0.062	0.074
0.0	0.50	−0.206	0.216	−0.196	0.200
0.0	1.00	−0.447	0.450	−0.425	0.427
0.3	0.25	−0.030	0.073	−0.027	0.048
0.3	0.50	−0.094	0.112	−0.089	0.097
0.3	1.00	−0.218	0.224	−0.212	0.215
0.6	0.25	−0.009	0.069	−0.006	0.040
0.6	0.50	−0.020	0.068	−0.016	0.041
0.6	1.00	−0.043	0.072	−0.042	0.054
0.7	0.25	−0.004	0.067	−0.002	0.038
0.7	0.50	−0.003	0.064	−0.001	0.037
0.7	1.00	−0.001	0.059	0.001	0.036
0.9	0.25	0.000	0.067	0.000	0.037
0.9	0.50	0.005	0.065	0.007	0.039
0.9	1.00	0.027	0.072	0.030	0.048

Note: $\phi = 0$ is classical error and $\phi = 0.7$ equals ρ; the first row ($\sigma_{me} = 0$) is the error-free benchmark.

Table 2: Bias under classical error by sample size N and panel length T ($\phi = 0$, $\sigma_{me} = 0.5$, $\rho = 0.70$, $R = 1000$)

T	Diff. N=200	Diff. N=500	Diff. N=1000	Diff. limit	Sys. N=200	Sys. N=500	Sys. N=1000	Sys. limit
5	−0.245	−0.219	−0.215	−0.213	−0.215	−0.211	−0.207	−0.210
8	−0.211	−0.203	−0.205	−0.204	−0.198	−0.193	−0.195	−0.195
15	−0.184	−0.184	−0.186	−0.187	−0.175	−0.176	−0.176	−0.175
30	−0.163	−0.163	−0.163	−0.164	−0.158	−0.158	−0.158	−0.156

Note: "limit" is the pseudo-true bias $\theta^* - \rho$ of the collapsed two-step estimator (Section 4.4); for $T = 5$ the instrument set contains lags 2–4 only. Without measurement error the bias at $N = 1000$ is −0.008, −0.002, 0.000, 0.000 for Difference-GMM and −0.004, −0.002, 0.000, 0.000 for System-GMM at $T = 5$, 8, 15 and 30.

4.4 The Closed Forms against the Simulations

Table 3 sets the closed forms of Section 3.4 beside the simulated biases at $N = 1000$. The first column is the single-instrument value of Corollary 3.6; the limit columns are the pseudo-true values of Proposition 3.5 for the collapsed two-step estimators, in which the second-moment matrix that defines the two-step weight is evaluated on 10^6 simulated units per design point. The agreement is close: across the 16 cells the largest absolute gap between the limit and the simulated bias is at most 0.005 for Difference-GMM and 0.003 for System-GMM, and the single-instrument formula alone is within 0.005 of the Difference-GMM simulation, so the collapsed estimator puts most of its weight on the shallowest lag. The limits also reproduce the dependence on T in Table 2 (for example -0.213 and -0.164 at $T = 5$ and $T = 30$ for Difference-GMM, against -0.215 and -0.163 simulated at $N = 1000$).

Table 3: Closed-form pseudo-true bias against simulated bias ($N = 1000$, $T = 8$, $\rho = 0.70$, $R = 1000$)

ϕ	σ_{me}	Single-instrument formula	Diff.-GMM limit	Diff.-GMM simulated	Sys.-GMM limit	Sys.-GMM simulated
0.0	0.00	0.000	0.000	−0.003	0.000	−0.003
0.0	0.25	−0.067	−0.061	−0.063	−0.061	−0.061
0.0	0.50	−0.209	−0.204	−0.204	−0.195	−0.195
0.0	1.00	−0.441	−0.450	−0.446	−0.428	−0.426
0.3	0.25	−0.028	−0.026	−0.031	−0.026	−0.027
0.3	0.50	−0.092	−0.090	−0.091	−0.088	−0.087
0.3	1.00	−0.217	−0.216	−0.214	−0.211	−0.209
0.6	0.25	−0.004	−0.004	−0.007	−0.004	−0.005
0.6	0.50	−0.015	−0.015	−0.015	−0.014	−0.015
0.6	1.00	−0.040	−0.040	−0.042	−0.040	−0.041
0.7	0.25	0.000	0.000	−0.003	0.000	−0.002
0.7	0.50	0.000	0.000	−0.001	0.000	−0.002
0.7	1.00	0.000	0.000	−0.002	0.000	0.000
0.9	0.25	0.002	0.002	0.000	0.002	0.001
0.9	0.50	0.008	0.008	0.006	0.008	0.008
0.9	1.00	0.029	0.029	0.026	0.030	0.028

Other values of the latent persistence give the same pattern ($N = 500$, $T = 8$, $\sigma_{me} = 1.0$). Classical error produces Difference-GMM biases of -0.163, -0.290, -0.446 and -0.660 for $\rho = 0.3$, 0.5, 0.7 and 0.9, while at $\phi = \rho$ the bias is -0.001, -0.002, -0.006 and -0.036; for $\rho = 0.9$ Difference-GMM is already biased without any error (-0.047), the familiar weak-instrument problem of near-unit-root series, and System-GMM is accurate at $\phi = \rho$ (-0.013 against -0.014 without error).

4.5 Instrument Depth

Corollary 3.3 suggests removing the shallowest instruments. Table 4 compares the conventional sets with trimmed sets: Difference-GMM with lags 3–5, and System-GMM with difference lags 3–5 and level lags 2–4, so that the lag-2 level and the lag-1 difference are dropped. Under classical error trimming removes almost all of the bias (at $\sigma_{me} = 0.5$, from -0.206 to -0.012 for Difference-GMM and from -0.196 to -0.010 for System-GMM), and because the bias of the conventional estimators dominates, the trimmed estimators also have the smaller RMSE (0.147 and 0.072 against 0.216 and 0.200). For persistent but non-classical error, trimming

removes only part of the bias, because the remaining instruments still carry a factor ϕ in their contamination (at $\phi = 0.3$ the bias falls from -0.089 to -0.043 for System-GMM), and at $\phi = 0.9 > \rho$ the bias is small to begin with and trimming offers no systematic gain, since the pseudo-true value of a deeper instrument is not smaller (Remark 3.7(c)). Trimming also has a price in precision where there is nothing to remove: without error the RMSE rises from 0.068 to 0.119 for Difference-GMM and from 0.038 to 0.058 for System-GMM, and Difference-GMM with lags 3–5 produces explosive estimates ($|\hat{\rho}| \geq 1$) in 1.1% of the replications without error and in up to 6.4% under classical error with $\sigma_{me} = 1.0$. The price is visible in the RMSE as well: for persistent error trimming lowers the RMSE only for System-GMM at $\phi = 0.3$ (0.080 against 0.097), and in the other non-classical rows of Table 4 (Difference-GMM at $\phi = 0.3$, and both estimators at $\phi \geq 0.6$) the conventional estimator has the lower RMSE, so trimming pays only when the classical component of the error is large. Reporting the conventional and the trimmed estimates side by side is therefore a simple diagnostic: a large gap between them signals contaminated shallow instruments.

Table 4: Bias and RMSE of conventional and trimmed estimators ($N = 500$, $T = 8$, $\rho = 0.70$, $R = 1000$)

ϕ	σ_{me}	D 2–5 bias	D 2–5 RMSE	D 3–5 bias	D 3–5 RMSE	S conv. bias	S conv. RMSE	S trim. bias	S trim. RMSE
0.0	0.00	−0.001	0.068	−0.007	0.119	−0.001	0.038	−0.003	0.058
0.0	0.50	−0.206	0.216	−0.012	0.147	−0.196	0.200	−0.010	0.072
0.0	1.00	−0.447	0.450	−0.024	0.211	−0.425	0.427	−0.028	0.109
0.3	0.50	−0.094	0.112	−0.056	0.138	−0.089	0.097	−0.043	0.080
0.6	0.50	−0.020	0.068	−0.023	0.124	−0.016	0.041	−0.017	0.062
0.7	0.50	−0.003	0.064	−0.007	0.116	−0.001	0.037	−0.005	0.056
0.9	0.50	0.005	0.065	0.001	0.115	0.007	0.039	0.004	0.057

Note: D = Difference-GMM; S conv. uses difference lags 2–5 and level lag 1; S trim. uses difference lags 3–5 and level lags 2–4. Share of replications with $|\hat{\rho}| \geq 1$ for Difference-GMM with lags 3–5, in the order of the rows: 1.1%, 2.0%, 6.4%, 0.4%, 0.7%, 0.3%, 0.8%.

5 Conclusion

This paper derives in closed form how persistent AR(1) measurement error in the dependent variable contaminates the moment conditions of Difference-GMM and System-GMM. The contamination of the lag-s instrument is $(\phi - 1)(\phi - \rho)\phi^{s-2}\sigma_e^2$ in the differenced equation and $(1 - \phi)(\phi - \rho)\phi^{s-1}\sigma_e^2$ in the level equation. It is proportional to the difference between the persistence of the error and that of the latent process, vanishes when the two coincide, and, because the exponent degenerates at the shallowest lag, does not vanish at that lag even for classical error. In particular the standard Blundell–Bond level instrument is invalid under any AR(1) measurement error in the dependent variable with $\phi \neq \rho$.

The paper also derives the pseudo-true value of each estimator. For a single instrument it is a weighted average of the true persistence ρ and the error persistence ϕ, the weight on ϕ being the noise share of the differenced variance at the instrument's lag; under classical error it reduces to the familiar attenuation $\rho/(1 + \kappa)$. These limits reproduce the simulated biases, including their dependence on the panel length, and show that deeper instruments reduce the bias only when the error is less persistent than the series.

The simulations compare the two estimators. Under the error both are biased by almost the same amount and the bias does not vanish as the cross-section or the panel length grows, so it is genuine inconsistency and not a small-sample artifact. System-GMM keeps its efficiency advantage, but its extra moment conditions do not protect it from the contamination. Dropping the shallowest instruments removes most of the classical-error bias at a cost in precision, but it does not remove the bias caused by persistent error. We therefore recommend that applied work with variables suspected of measurement error report conventional and trimmed estimates together.

The analysis is confined to the autoregressive panel model without covariates, to measurement error in the dependent variable only and to fixed-T, large-N asymptotics. Extending the derivation to regressors, to errors that are correlated with the latent process, and to formal tests for contaminated instruments are natural next steps.

References

[1] Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica, 49*(6), 1417–1426.

[2] Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica, 50*(4), 1029–1054.

[3] Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies, 58*(2), 277–297.

[4] Holtz-Eakin, D., Newey, W., & Rosen, H. S. (1988). Estimating vector autoregressions with panel data. *Econometrica, 56*(6), 1371–1395.

[5] Arellano, M., & Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics, 68*(1), 29–51.

[6] Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics, 87*(1), 115–143.

[7] Ahn, S. C., & Schmidt, P. (1995). Efficient estimation of models for dynamic panel data. *Journal of Econometrics, 68*(1), 5–27.

[8] Roodman, D. (2009). A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics, 71*(1), 135–158.

[9] Kiviet, J. F., Pleus, M., & Poldermans, R. W. (2017). Accuracy and efficiency of various GMM inference techniques in dynamic micro panel data models. *Econometrics, 5*(1), 14.

[10] Griliches, Z., & Hausman, J. A. (1986). Errors in variables in panel data. *Journal of Econometrics, 31*(1), 93–118.

[11] Biørn, E. (2012a). The measurement error problem in dynamic panel data analysis: Modeling and GMM estimation. Memorandum No. 02/2012, Department of Economics, University of Oslo.

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

[13] Lee, N., Moon, H. R., & Weidner, M. (2012). Analysis of interactive fixed effects dynamic linear panel regression with measurement error. *Economics Letters, 117*(1), 239–242.

[14] Lee, N., Moon, H. R., & Zhou, Q. (2017). Many IVs estimation of dynamic panel regression models with measurement error. *Journal of Econometrics, 200*(2), 251–259.

[15] Chowdhury, R. A., & Russell, B. (2018). The difference, system and 'Double-D' GMM panel estimators in the presence of structural breaks. *Scottish Journal of Political Economy, 65*(3), 271–297.

[16] Biørn, E. (2012b). Panel data dynamics and measurement errors: GMM bias, IV validity and model fit – a Monte Carlo study. Memorandum No. 27/2012, Department of Economics, University of Oslo.

[17] Biørn, E. (2015). Panel data dynamics with mis-measured variables: Modeling and GMM estimation. *Empirical Economics, 48*(2), 517–535.

[18] Wansbeek, T. (2001). GMM estimation in panel data models with measurement error. *Journal of Econometrics, 104*(2), 259–268.

[19] Meijer, E., Spierdijk, L., & Wansbeek, T. (2017). Consistent estimation of linear panel data models with measurement error. *Journal of Econometrics, 200*(2), 169–180.

[20] Harris, M. N., & Mátyás, L. (2004). A comparative analysis of different IV and GMM estimators of dynamic panel data models. *International Statistical Review, 72*(3), 397–408.

[21] Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics, 126*(1), 25–51.

[22] Andrews, D. W. K., & Lu, B. (2001). Consistent model and moment selection procedures for GMM estimation with application to dynamic panel data models. *Journal of Econometrics, 101*(1), 123–164.

[23] Hayakawa, K. (2015). The asymptotic properties of the system GMM estimator in dynamic panel data models when both N and T are large. *Econometric Theory, 31*(3), 647–667.

[24] Hayakawa, K., & Qi, M. (2020). Further results on the weak instruments problem of the system GMM estimator in dynamic panel data models. *Oxford Bulletin of Economics and Statistics, 82*(2), 453–481.

[25] Harris, M. N., & Mátyás, L. (2009). The robustness of estimators for dynamic panel data models to misspecification. *The Singapore Economic Review, 54*(3), 399–426.

[26] Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica, 26*(3), 393–415.

[27] Biørn, E. (2000). Panel data with measurement errors: Instrumental variables and GMM procedures combining levels and differences. *Econometric Reviews, 19*(4), 391–424.

[28] Bond, S. R. (2002). Dynamic panel data models: A guide to micro data methods and practice. *Portuguese Economic Journal, 1*(2), 141–162.