

Minimum Distance Estimation of Dynamic Models with Errors-In-Variables

Nikolay Gospodinov*

Ivana Komunjer[†]

Serena Ng[‡]

July 31, 2014

Abstract

Empirical analysis often involves using inexact measures of desired predictors. The bias created by the correlation between the problematic regressors and the error term motivates the need for instrumental variables estimation. This paper considers a class of estimators that can be used when external instruments may not be available or are weak. The idea is to exploit the relation between the parameters of the model and the least squares biases. In cases when this mapping is not analytically tractable, a special algorithm is designed to simulate the latent predictors without completely specifying the processes that induce the biases. The estimators perform well in simulations of the autoregressive distributed lag model and the dynamic panel model. The methodology is used to re-examine the Phillips curve in which the real activity gap is latent.

JEL Classification: C1, C3

Keywords: Measurement Error, Minimum Distance, Simulation Estimation, Dynamic Panel.

*Federal Reserve Bank of Atlanta. Email: Nikolay.Gospodinov@atl.frb.org

[†]University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093. Email: komunjer@ucsd.edu

[‡](Corresponding author) Columbia University, 420 W. 118 St. MC 3308, New York, NY 10027. Email: serena.ng@columbia.edu

The authors acknowledge financial support from the National Science Foundation SES-0962473 and SES-0962431.

1 Introduction

Empirical analysis often involves using incorrectly measured data that complicates identification of the behavioral parameters and testing of economic hypothesis.¹ The problem is acute in cross-section and survey data when errors in data collection and reporting are inevitable and new estimation methods continue to be developed. The literature on measurement error in time series data is smaller but the problem is no less important. The real time estimates which underlie economic decisions can differ from the revised estimates that researchers used for analysis. We do not observe variables such as the state of economy, potential output, or natural rate of unemployment, and filtered series are often used as proxies. Except by coincidence, the latent processes will not be the same as the constructed ones with differences that can be correlated over time. Orphanides and van Norden (2002) and Orphanides and Williams (2002) find that misperceptions or measurement errors can be quite persistent. Ermini (1993) shows that allowing for serially uncorrelated measurement errors changes the measure of persistence in consumption growth. Falk and Lee (1990) suggest that measurement errors can explain rejections of the permanent income hypothesis. Nalewalk (2010) shows that the income (GDI) and product (GDP) side of output growth exhibit rather different fluctuations over the past 25 years and that the GDI shows a steeper downturn in 2007-2009 than GDP. Aruoba, Diebold, Nalewaik, Schorfheide, and Song (2013) find that the series filtered from GDP and GDI are less volatile but more persistent than the two contaminated measures. Sargent (1989) allows the data collected to have serially correlated errors and shows that identification of the deep parameters of an accelerator model is affected by how the data are reported.

This paper is concerned with dynamic models that can generically be represented by

$$y_t = W_t' \gamma + V_t, \tag{1}$$

where y_t is a scalar dependent variable, W_t is a vector of observed regressors that may include lags of the dependent variable, and V_t is a latent disturbance term. The crucial feature of the model is that some components of W_t are mismeasured which causes W_t and V_t to become correlated. In consequence, the model (1) contains the regression parameters of interest γ as well as the nuisance parameters δ . The correlation between V_t and W_t can sometimes be mitigated by resorting to instrumental variables (IVs), such as in purely static models in which measurement errors are known to be uncorrelated over time. More often than not, however, lags of W_t cannot be used as instruments. This is the case for distributed lag models with autocorrelated measurement errors, and also true of dynamic panel models in which fixed effects are eliminated by demeaning the observables. Not only is the least squares estimator (OLS) biased, but so is the instrumental variable estimator (IV) because the lagged variables are no longer valid instruments.

¹Wilcox (1992) discusses the issues in consumption measurements especially at the monthly level.

The goal of this paper is to show that identification of γ is still possible when external instruments are unavailable or that the validity of lagged variables as instruments is in doubt. The possibility of using internal information to turn an unidentified problem into an identifiable one was noted in Goldberger (1972) and Hsiao (1977). Our main insight is that the least squares (OLS) bias depends on the nuisance parameters δ that characterize W_t . While OLS does not consistently estimate γ , there is in general enough information in the serially correlated regressors and the least squares residuals about the measurement error process to permit identification of γ . Persistence in the regressors is important here because the least squares residuals will be serially correlated if the regressors are serially correlated. In a way, our approach is to combine information in several sample estimates whose bias is magnified by persistence of the regressors.

Specifically, we propose to identify $\theta = (\gamma', \delta')'$ from a vector of *auxiliary statistics* $\hat{\psi}(\theta)$ that includes the OLS estimates and moments formed from the least squares residuals. The role of the auxiliary statistics $\hat{\psi}(\theta)$ is to provide a mapping from the parameter space of θ to the parameter space of an auxiliary model. We study the probability limit of $\hat{\psi}(\theta)$ as the sample size grows to infinity. Following Gourieroux, Monfort, and Renault (1993), the mapping $\psi : \theta \rightarrow \psi(\theta)$ will be referred to as the *binding function*. Local identifiability of θ requires that the binding function is locally one-to-one. A sufficient (though not necessary) condition for the latter is that the matrix of partial derivatives $\partial\psi/\partial\theta$ is of full column rank. The parameter θ is globally identifiable over the entire parameter space Θ whenever ψ is invertible on $\psi(\Theta)$. In this case, a natural estimator for θ is obtained as $\hat{\theta} = \psi^{-1}(\hat{\psi})$.

In simple models where the mapping between θ and the auxiliary statistics can be derived analytically, we have a classical minimum distance (CMD) estimator that is \sqrt{T} consistent and asymptotically normal, where T is the sample size. Our CMD estimator is similar in the spirit to the ones proposed in Lewbel (2012) and Erickson (2001) who considered identification of parameters in a linear regression model without additional instruments. Lewbel (2012) uses the fact that under heteroskedasticity of the errors, the product of the regression and measurement error are uncorrelated with an exogenous variable. Erickson (2001) considered identification using higher order moments. Schennach and Hu (2013) also considered identification without side information, but their focus is non- and semi-parametric models. Our emphasis is on combining individually biased estimators without making assumptions about normality or homoskedasticity in a linear regression setting.

In more complex models such as when lags of the dependent variable and the regressors are involved, the binding function will not be tractable. We use Monte-Carlo methods to approximate this mapping. The resulting simulated minimum distance (SMD) estimator is consistent to the extent that the simulator can reconstruct those variations in the contaminated regressors that are

orthogonal to the measurement errors. Our simulation estimator differs from the ones considered in Smith (1993), Gouriéroux, Monfort, and Renault (1993), and Gallant and Tauchen (1996). These estimators treat the predictors as exogenous which can be held fixed in the simulations. The exogeneity assumption is not appropriate in measurement error models because the parameters in the marginal distribution of the covariates and those of the conditional model are not variation free in the sense of Engle, Hendry, and Richard (1983). Thus, even though the correctly measured predictors could have been held fixed, the mismeasured ones cannot. Our SMD estimator is designed to use limited information from the joint distribution of the data to simulate the model.

The paper proceeds as follows. After stating the assumptions for analysis, Section 2 uses the simple regression model to motivate the choice of auxiliary statistics. Section 3 turns to autoregressive distributed lag models with possibly serially correlated measurement errors. Section 4 extends the analysis to dynamic panel models. Classical minimum distance estimation is not feasible in models with more general dynamics as the model-implied moments do not have closed form expressions. This motivates the need for simulation estimation, which is presented in Section 5. The same section presents Monte Carlo simulation evidence and a Phillips-curve application. Last section concludes. Technical proofs are relegated to an Appendix.

As a matter of notation, we use $\Gamma_x(j) \equiv E[x_t x_{t-j}]$ to generically denote the autocovariance of a covariance stationary mean-zero time series x_t . We use $\Gamma_{xy}(j, k) \equiv E[x_{t-j} y_{t-k}]$ to denote the covariance between x_t and another mean-zero covariance stationary process y_t . If $E(x_t) = 0$, $E(x_t^2) = \sigma_x^2 \equiv \Gamma_x(0)$, and $\Gamma_x(j) = 0$ for $j \geq 1$, then x_t is a white noise. In this case, we write $x_t \sim WN(0, \sigma_x^2)$.

2 The Econometric Framework

Consider the autoregressive distributed lag ADL(p,q) model with scalar predictor x_t :

$$\alpha(L)y_t = \beta(L)x_t + u_t, \quad (2)$$

where $\alpha(L) = 1 - \sum_{i=1}^p \alpha_i L^i$, $\beta(L) = \sum_{i=0}^q \beta_i L^i$, and L is the lag operator. Instead of x_t , we only observe a contaminated variable X_t :

$$X_t = x_t + \epsilon_t. \quad (3)$$

Now let $\alpha = (\alpha_1, \dots, \alpha_p)'$, $\beta = (\beta_0, \beta_1, \dots, \beta_q)'$ and $\gamma = (\alpha', \beta')'$. In the context of the general dynamic model in (1), $W_t = (y_{t-1}, \dots, y_{t-p}, X_t, X_{t-1}, \dots, X_{t-q})'$, and the error term becomes

$$V_t = u_t - \beta(L)\epsilon_t.$$

As is well known, identification in these models is impossible without further assumptions when the predictors are serially uncorrelated or normally distributed.² But as Goldberger (1972, p.996) pointed out, measurement errors in the exogenous regressors need not destroy identification provided that the model is overidentified. One approach is to use additional information from N variables. Bai and Ng (2010) assume that there are mismeasured indicators $(X_{1t}, \dots, X_{Nt})'$ of the latent predictor x_t , each characterized by $X_{it} = \pi_i x_t + \epsilon_{it}$ with $\text{cov}(\epsilon_{it}, x_t) = 0$, and N is large. The IV approach makes use of $N = 2$ variables: one X_{1t} to replace x_t and a second one X_{2t} that is correlated with X_{1t} but uncorrelated with ϵ_{1t} . A second approach is to drop the normality assumption. Pal (1980), Dagenais and Dagenais (1997), Lewbel (1997) and Spierdijk and Wansbeek (2011) exploit heteroskedasticity, excess skewness and kurtosis for identification without relying on instruments. Our approach falls in the third category of exploiting the bias in the sample estimates. Assumptions on the latent variables of the model, u_t , ϵ_t , and x_t , are as follows.

Assumption A .

- (a) $\{(x_t, \epsilon_t)'\}$ is covariance stationary with $E(x_t) = 0$, $E(\epsilon_t) = 0$, $E(x_t \epsilon_\tau) = 0$ for every (t, τ) .
- (b) $u_t \sim WN(0, \sigma_u^2)$. For every (t, τ) , $E(u_t x_\tau) = 0$ and $E(u_t \epsilon_\tau) = 0$.
- (c) The covariance matrix of $(y_{t-1}, \dots, y_{t-p}, x_t, x_{t-1}, \dots, x_{t-q})$ is nonsingular.
- (d) Measurement error autocovariances $\Gamma_\epsilon(1), \dots, \Gamma_\epsilon(q^*)$ are well approximated by $\sigma_\epsilon^2 f(\phi)$ where $f(\cdot)$ is a function parameterized by $\phi = (\phi_1, \dots, \phi_m)'$.

We assume in (a) that the model is dynamically correctly specified and that u_t is serially uncorrelated. The latent regressor x_t and measurement error ϵ_t are assumed to have zero mean. The intercept is therefore suppressed in (2). The measurement error ϵ_t is assumed to be classical, i.e. orthogonal to the latent regressor x_t at all leads and lags. The true regressor x_t is observed with error whenever $\sigma_\epsilon^2 \neq 0$. The measurement error ϵ_t is allowed to be serially correlated. The model disturbance u_t in (b) is assumed to be white noise and orthogonal to both x_t and ϵ_t . Serial correlation in u_t and its correlation with x_t can be seen as due to the omission of relevant lags of y_t and x_t in (2). Thus (b) subsumes that all the relevant regressors have been included in the model. Assumption (c) is standard for least squares analysis. Note however that this condition involves latent variables x_t, \dots, x_{t-q} . As we shall demonstrate next, the dynamics of V_t depends on the lag order q which affects the identifiability of α and β . For $q = 0$, identification can be established by leaving the dynamic structure of ϵ_t unspecified. For values $q \geq 1$, however, this approach leads to a

²See Reiersøl (1950) and Aigner, Hsiao, Kapteyn, and Wansbeek (1984, p.1324) for an overview of identification conditions in measurement error models. Gillard (2010) presents an overview of approaches to handle the errors-in-variables (EIV) problem from different fields.

proliferation of nuisance parameters in the model. In those cases, we reparameterize the model as in (d). While this assumption does not parametrically specify the dynamic structure of ϵ_t explicitly, we require that its first q^* autocorrelations can be parsimoniously approximated by m parameters, where both m and q^* depend on the choice of auxiliary statistics and will be further discussed.

An implication of the presence of measurement error is that $E(V_t X_{t-j}) \neq 0$ for $0 \leq j \leq q$. Moreover, \hat{V}_t is generally serially correlated even if ϵ_t is white noise. Let δ denote the nuisance parameters characterizing V_t . In the case of white noise measurement errors, $\delta = (\sigma_u^2, \sigma_\epsilon^2)'$. When ϵ_t is serially correlated, $\delta = (\sigma_u^2, \sigma_\epsilon^2, \Gamma_\epsilon(1), \dots, \Gamma_\epsilon(q^*))'$ where q^* depends on auxiliary statistics used. The parameters of the model are $\theta = (\alpha', \beta', \delta')'$. The objective of the exercise is to identify and consistently estimate θ from suitable choice of auxiliary statistics. While α and β are the parameters of interest, joint identification of δ makes it possible to assess the severity of measurement errors. IV estimators work around the attenuation bias and can only be silent about the properties of ϵ_t .

The idea behind our proposed estimator is to exploit the bias in the OLS estimator for identification of θ . Specifically, if $\text{plim}_{T \rightarrow \infty} \hat{\gamma} - \gamma = \text{bias}(\theta)$ is non-degenerate and

$$\hat{V}_t = y_t - W_t' \hat{\gamma} = V_t - W_t' \text{bias}(\theta),$$

then moments formed from \hat{V}_t must also be a function of the bias and hence depend on θ . We will be specifically interested in $\hat{\Gamma}_{\hat{V}}(j) = \frac{1}{T} \sum_{t=j+1}^T \hat{V}_t \hat{V}_{t-j}$ and $\hat{\Gamma}_{\hat{V}X}(j, 0) = \frac{1}{T} \sum_{t=j+1}^T \hat{V}_{t-j} X_t$. Their probability limits as $T \rightarrow \infty$ are, respectively,

$$\Gamma_{\hat{V}}(j) = \Gamma_V(j) - \left(\Gamma_{VW}(j, 0) + \Gamma_{VW}(0, j) \right)' \text{bias}(\theta) + \text{bias}(\theta)' \Gamma_W(j) \text{bias}(\theta) \quad (4)$$

$$\Gamma_{\hat{V},X}(j, 0) = \Gamma_{VX}(j, 0) - \Gamma_{WX}(j, 0)' \text{bias}(\theta). \quad (5)$$

Observe that these moments use \hat{V}_t instead of V_t and $\text{plim } \hat{\Gamma}_{\hat{V}}(0) \neq \Gamma_V(0)$, $\text{plim } \hat{\Gamma}_{\hat{V}}(1) \neq \Gamma_V(1)$, and $\text{plim } \hat{\Gamma}_{\hat{V}X}(1, 0) \neq \Gamma_{VX}(1, 0)$. The biased sample estimates are functions of least squares bias and thus useful for identification of θ . To layout the framework for analysis, we use the simple regression model with no lagged dependent variables to motivate the choice of auxiliary statistics. It will be shown that θ is identified whether or not measurement errors are serially correlated. We then extend the analysis to allow for lagged dependent variables.

2.1 The Simple Regression Model

The scalar regression model

$$y_t = x_t \beta + u_t$$

with $X_t = x_t + \epsilon_t$ and $V_t = u_t - \beta \epsilon_t$ is well studied under the assumption that x_t and ϵ_t are *iid*. We consider this model in a time series context when x_t is serially correlated, and under

different assumptions about ϵ_t . In the case when ϵ_t is white noise, it holds, as in the *iid* case, that $\Gamma_X(0) = \Gamma_x(0) + \sigma_\epsilon^2$. The parameters of this model are

$$\theta_A \equiv (\beta, \sigma_u^2, \sigma_\epsilon^2)'. \quad (6)$$

The least squares estimator has attenuation bias given by

$$\text{plim}_{T \rightarrow \infty} (\hat{\beta} - \beta) = -\frac{\beta \sigma_\epsilon^2}{\Gamma_X(0)} \equiv \text{bias}_A(\theta).$$

Our estimator exploits two features of time series data. The first is the relation between the autocovariance of X_t (observed) and x_t (latent). Even though X_t and x_t have different autocorrelations, they have the same autocovariance at lag $j \geq 1$. That is, if ϵ_t is white noise,

$$\Gamma_X(j) = \Gamma_x(j) \quad \text{for all } j \geq 1.$$

The second is the properties of the least squares residuals $\hat{V}_t = y_t - X_t \hat{\beta} = V_t - X(\hat{\beta} - \beta)$. By (4), the autocovariances for this model are given by

$$\text{plim}_{T \rightarrow \infty} \hat{\Gamma}_{\hat{V}}(j) = \Gamma_V(j) - \text{bias}_A \left(\Gamma_{VX}(0, j) + \Gamma_{VX}(j, 0) \right) + \text{bias}_A(\theta)^2 \Gamma_X(j).$$

With white noise measurement error, $\Gamma_V(j)$, $\Gamma_{VX}(0, j)$ and $\Gamma_{VX}(j, 0)$ are zero for $j = 1$. However, $\text{bias}_A(\theta) \neq 0$ and $\Gamma_{VX}(0, 0) = -\beta \sigma_\epsilon^2$. Thus, even though V_t is serially uncorrelated when ϵ_t is white noise, \hat{V}_t has non-zero autocovariances whenever X_t is serially correlated. A researcher who is not aware of the presence of measurement error might estimate a model with additional lags of X_t . The ability to identify σ_ϵ^2 can lead to the choice of a more appropriate dynamic model.

Consider identifying θ_A from a vector of three auxiliary statistics:³

$$\hat{\psi}_A \equiv \begin{pmatrix} \hat{\beta} \\ \hat{\Gamma}_{\hat{V}}(0) \\ \hat{\Gamma}_{\hat{V}}(1) \end{pmatrix} \xrightarrow{p} \begin{pmatrix} \beta + \text{bias}_A(\theta) \\ \sigma_u^2 + \beta^2 \sigma_\epsilon^2 + \text{bias}_A(\theta) \beta \sigma_\epsilon^2 \\ \text{bias}_A(\theta)^2 \Gamma_X(1) \end{pmatrix} \equiv \psi_A(\theta). \quad (7a)$$

As pointed out in Spierdijk and Wansbeek (2011), $\Gamma_X(0)$, $\Gamma_Y(0)$ and $\Gamma_{XY}(0, 0)$ are needed to identify θ_A . The proposed $\hat{\psi}_A$ is implicitly a function of these moments. Its choice is guided by the fact that $\hat{\beta}$ is a sufficient statistic for β in the absence of measurement error, and the properties of \hat{V}_t are completely characterized by its second moments. A useful way to think about $\hat{\psi}_A$ is that its components $\hat{\beta}$, $\hat{\Gamma}_{\hat{V}}(0)$ and $\hat{\Gamma}_{\hat{V}}(1)$ are biased for β , $\Gamma_V(0)$ and $\Gamma_V(1)$, and we exploit the biases to identify θ . This contrasts with the IV estimator which works around the correlation between X_t and y_t and cannot identify σ_ϵ^2 .

³Note that $\text{bias}_A \theta \Gamma_X(0) = -\beta \sigma_\epsilon^2$. Thus, $\hat{\Gamma}_{\hat{V}}(0) \xrightarrow{p} \sigma_u^2 + \beta^2 \sigma_\epsilon^2 + 2\text{bias}_A(\theta) \beta \sigma_\epsilon^2 + \text{bias}_A(\theta)^2 \Gamma_X(0) = \sigma_u^2 + \beta^2 \sigma_\epsilon^2 + \text{bias}_A(\theta) \beta \sigma_\epsilon^2$.

Substituting in $\text{bias}_A(\theta) = -\beta \frac{\sigma_\epsilon^2}{\Gamma_X(0)}$, the binding function simplifies to

$$\psi_A(\theta) = \begin{pmatrix} \psi_{A1} \\ \psi_{A2} \\ \psi_{A3} \end{pmatrix} = \begin{pmatrix} \beta - \beta \frac{\sigma_\epsilon^2}{\Gamma_X(0)} \\ \sigma_u^2 + \beta^2 \sigma_\epsilon^2 - \frac{(\beta \sigma_\epsilon^2)^2}{\Gamma_X(0)} \\ \left(\frac{\beta \sigma_\epsilon^2}{\Gamma_X(0)} \right)^2 \Gamma_X(1) \end{pmatrix}. \quad (7b)$$

Given that (7b) is a system of three equations and three unknowns, the parameters are locally identified if the matrix of partial derivatives $\partial \psi_A / \partial \theta_A$ has a non-zero determinant. To establish global identifiability of θ_A , we need to show that the mapping $\theta_A \mapsto \psi_A(\theta_A)$ is invertible.

Lemma 1 *Suppose $y_t = x_t \beta + u_t$, $X_t = x_t + \epsilon_t$, Assumptions A(a)-(c) hold, and $\epsilon_t \sim WN(0, \sigma_\epsilon^2)$. Then:*

- (a) $(\beta = 0, \sigma_u^2)$ is globally identified from ψ_A .
- (b) $(\beta \neq 0, \sigma_u^2)$ is globally identified from ψ_A if and only if $\Gamma_X(1) \neq 0$.
- (c) $\theta_A = (\beta, \sigma_\epsilon^2, \sigma_u^2)$ is globally identified from ψ_A if and only if (i) $\beta \neq 0$ and (ii) $\Gamma_X(1) \neq 0$.

Lemma 1, proved in the Appendix, gives the necessary and sufficient conditions for global identification. If β alone is of interest, global identification only requires X_t to be serially correlated.⁴ But $\beta \neq 0$ is necessary to also identify σ_ϵ^2 because the regression residuals would have no information about β if X_t has no role in the regression model. Identification strength is determined by $\Gamma_X(1)$, which measures the persistence of the mismeasured regressor X_t .

The question then arises as to whether θ_A can be identified at $\beta = 0$ when the nuisance parameter $\delta_A = \sigma_\epsilon^2$ is not identifiable. This is possible with additional restrictions on the dynamic structure of the latent process x_t . For example, suppose that

$$\Gamma_x(j) = \phi^j \Gamma_x(0)$$

for two consecutive values of $j \geq 1$, a condition that holds if x_t has an autoregressive structure. Since ϵ_t is white noise, it also holds that $\Gamma_X(j) = \phi^j \Gamma_x(0)$ for $j \geq 1$. From $\phi = \frac{\Gamma_X(2)}{\Gamma_X(1)}$ when $j = 2$ and $\Gamma_X(0) = \Gamma_x(0) + \sigma_\epsilon^2$, we have

$$\sigma_\epsilon^2 = \Gamma_X(0) - \frac{\Gamma_X(1)^2}{\Gamma_X(2)}.$$

We can use σ_ϵ^2 to assess the severity of measurement error prior to any regression analysis. The result is specific to the fact that ϵ_t is a white noise process.

⁴This contrasts with Reiersøl (1950), Pal (1980), Erickson, Jiang, and Whited (2014) in which the identification results exclude the important special case of $\beta = 0$. The reason is that they consider identification of the entire parameter vector $\theta_A = (\beta, \sigma_u^2, \sigma_\epsilon^2)'$, while parts (a) and (b) of our result apply to (β, σ_u^2) alone.

2.2 Correlated Measurement Errors

The above analysis for the simple regression model with $\epsilon_t \sim WN(0, \sigma_\epsilon^2)$ highlights the possibility of combining several biased estimators for identification. Nonetheless, X_{t-k} could have been used as instrument under the assumptions of the analysis for $k > q$. The practical interest of our approach is in situations when lags of X_t may not be valid instruments. In such cases we will need to augment θ_A and $\hat{\psi}_A$ with additional parameters and statistics. We now show how this can be achieved when x_t in the simple regression model above is contaminated by a serially correlated measurement error.

Serial correlation in the measurement error has the important implication that X_{t-1} is no longer a valid instrument. Though longer lags could be valid, they will likely have weak correlation with X_t . To begin, we define

$$\text{plim } T \rightarrow \infty \hat{\beta} - \beta = -\beta \frac{\sigma_\epsilon^2}{\Gamma_X(0)} \equiv \text{bias}_B(\theta_B)$$

to distinguish it from $\text{bias}_A(\theta_A)$ even though least squares bias is numerically identical whether or not ϵ_t is serially correlated. This distinction is important because θ_B will be of higher dimension when ϵ_t is correlated as we shall now show.

When ϵ_t is serially correlated, the autocovariance of the least squares residuals $\hat{V}_t = y_t - X_t \hat{\beta}$ is affected by the dynamics of ϵ_t in two ways. First, $V_t = u_t - \beta \epsilon_t$ is serially correlated because ϵ_t is serially correlated. As a consequence, $\Gamma_{VX}(0, 1) = E(u_t - \beta \epsilon_t)(x_{t-1} + \epsilon_{t-1}) = -\beta \Gamma_\epsilon(1) \neq 0$. Second, with ϵ_t serially correlated we now have

$$\Gamma_X(j) = \Gamma_x(j) + \Gamma_\epsilon(j), \quad j \geq 1$$

instead of $\Gamma_X(j) = \Gamma_x(j)$. It follows from (4) that

$$\text{plim } T \rightarrow \infty \hat{\Gamma}_{\hat{V}}(1) = \beta^2 \Gamma_\epsilon(1) + 2\beta \Gamma_\epsilon(1) \text{bias}_B(\theta_B) + \text{bias}_B(\theta_B)^2 \Gamma_X(1).$$

Serial correlation in ϵ_t thus introduces the additional nuisance parameter $\Gamma_\epsilon(1)$, leading to

$$\theta_B \equiv (\beta, \sigma_u^2, \sigma_\epsilon^2, \Gamma_\epsilon(1))'. \quad (8a)$$

Since there is an additional parameter compared to θ_A , an additional auxiliary statistic is required. We consider

$$\text{plim } T \rightarrow \infty \hat{\Gamma}_{\hat{V}X}(1, 0) = \frac{1}{T} \sum_{t=1}^T (V_{t-1} - (\hat{\beta} - \beta)X_{t-1})X_t \xrightarrow{p} -\beta \Gamma_\epsilon(1) - \text{bias}_B(\theta) \Gamma_X(1),$$

because its probability limit depends on the same parameters as the limit of $\Gamma_{\hat{V}}(0)$ and $\hat{\Gamma}_V(1)$.⁵ The auxiliary statistic

$$\hat{\psi}_B \equiv (\hat{\beta}, \hat{\Gamma}_{\hat{V}}(0), \hat{\Gamma}_{\hat{V}}(1), \hat{\Gamma}_{VX}(1, 0))'$$

⁵An alternative is $\hat{\Gamma}_{\hat{V}X}(0, 1)$ which in this case has the same probability limit as $\hat{\Gamma}_{\hat{V}X}(1, 0)$.

has binding function given by

$$\psi_B(\theta_B) = \begin{pmatrix} \beta + \text{bias}_B(\theta) \\ \sigma_u^2 + \beta^2 \sigma_\epsilon^2 + \text{bias}_B(\theta) \beta \sigma_\epsilon^2 \\ \beta^2 \Gamma_\epsilon(1) + 2\beta \Gamma_\epsilon(1) \text{bias}_B(\theta) + \text{bias}_B(\theta)^2 \Gamma_X(1) \\ -\beta \Gamma_\epsilon(1) - \text{bias}_B(\theta) \Gamma_X(1) \end{pmatrix} \quad (8b)$$

Lemma 2 Consider the model $y_t = x_t \beta + u_t$, $X_t = x_t + \epsilon_t$. Under Assumptions A(a)-(c):

- (a) $(\beta = 0, \sigma_u^2)$ is globally identified from ψ_B ;
- (b) $(\beta \neq 0, \sigma_u^2)$ is globally identified from ψ_B if $\Gamma_x(1) \neq 0$;
- (c) $\theta_B = (\beta, \sigma_u^2, \Gamma_\epsilon(0), \Gamma_\epsilon(1))$ is globally identified from ψ_B if (i) $\Gamma_x(1) \neq 0$ and (ii) $\beta \neq 0$.

The proof, given in the Appendix, involves inverting the binding function and showing that a unique solution exists. Serial correlation in the latent regressor x_t is needed. Note that when ϵ_t is white noise, as was the case in the previous section, then $\Gamma_x(1) = \Gamma_X(1)$ and the condition reduces to that in Lemma 1. It is worth pointing out that $\Gamma_x(1) \neq 0$ is sufficient to globally identify β and σ_u^2 , irrespective of whether or not $\beta = 0$. However, $\Gamma_\epsilon(0)$ and $\Gamma_\epsilon(1)$ can only be identified if $\beta \neq 0$. As in the case when ϵ_t is white noise, the only way we can learn about the measurement error is by looking at the regressor X_t .

The assumption that $\Gamma_x(1) \neq 0$ is not directly testable. However, when $\beta = 0$, we can use the fact that $\Gamma_X(1) = \Gamma_x(1) + \Gamma_\epsilon(1)$. An IV estimator with X_{t-1} as instrument has

$$\text{plim } \hat{\beta}_{IV} = \beta \left(1 - \frac{\Gamma_\epsilon(1)}{\Gamma_X(1)} \right) = \beta \frac{\Gamma_x(1)}{\Gamma_X(1)}$$

which is zero if and only if $\Gamma_x(1) = 0$. Indirect evidence of whether the latent regressor is correlated can be gleaned from the IV estimate, even though it is biased for β .

3 Autoregressive Distributed Lag Models

The idea of using the OLS biases to achieve identification extends to more general models. The ADL(p,q) model expressed in terms of the observables is

$$(1 - \alpha_1 L - \dots - \alpha_p L^p) y_t = (\beta_0 + \beta_1 L + \dots + \beta_q L^q) X_t + V_t.$$

The error term $V_t = u_t - \beta_0 \epsilon_t - \dots - \beta_q \epsilon_{t-q}$ is generally serially correlated. For instance, even if ϵ_t is white noise, V_t is an MA(q) process. This serial correlation can confound the determination of the lag orders p and q .⁶ Furthermore, any X_{t-j} for $j \leq q$ is an invalid instrument because it

⁶We assume that y_t is observed without error. ARMA models when y_t is observed with error is studied in Komunjer and Ng (2014).

is correlated with ϵ_{t-j} . Serial correlation in ϵ_t makes it even more difficult to find instruments that are both strong and relevant. While X_{t-q+1} and further lags are valid instruments, they may only be weakly correlated with the regressors if they are stationary and ergodic as assumed. We consider estimation that does not directly involve internal or external instruments, beginning with the ADL(1,0) model.

3.1 ADL(1,0) model

The ADL(1,0) model is represented by $y_t = \alpha y_{t-1} + \beta x_t + u_t$, and we observe $X_t = x_t + \epsilon_t$. Here, $\gamma = (\alpha, \beta)$, $W_t = (y_{t-1}, X_t)'$. Assuming that $\Gamma_W(0)$ is nonsingular, the least squares bias is given by

$$\text{plim}_{T \rightarrow \infty} \hat{\gamma} - \gamma = \begin{pmatrix} \beta \frac{\sigma_\epsilon^2 \Gamma_{yX}(1,0)}{\Gamma_y(0) \Gamma_X(0) - \Gamma_{yX}(1,0)^2} \\ -\beta \frac{\sigma_\epsilon^2 \Gamma_y(0)}{\Gamma_y(0) \Gamma_X(0) - \Gamma_{yX}(1,0)^2} \end{pmatrix} = \text{bias}_C(\theta).$$

The parameters of the model are $\theta_C = (\alpha, \beta, \sigma_u^2, \sigma_\epsilon^2(0), \Gamma_\epsilon(1))'$, and the auxiliary statistics are $\hat{\psi}_C = (\hat{\alpha}, \hat{\beta}, \hat{\Gamma}_{\hat{V}}(0), \hat{\Gamma}_{\hat{V}}(1), \hat{\Gamma}_{\hat{V}X}(1,0))'$. Identification requires invertibility of the probability limit $\psi_C(\theta_C)$ of $\hat{\psi}_C(\theta_C)$. The analysis can be simplified by noting that σ_u^2 will be identified from $\Gamma_{\hat{V}}(0)$. Thus, we only need to consider identification of

$$\theta_C^- = (\gamma, \Gamma_\epsilon(0), \Gamma_\epsilon(1))$$

from

$$\hat{\psi}_C^- = (\hat{\alpha}, \hat{\beta}, \hat{\Gamma}_{\hat{V}}(1), \hat{\Gamma}_{\hat{V}X}(1,0))'.$$

As before, θ_C^- is globally identified from ψ_C^- if the binding function $\psi_C^-(\theta_C^-)$ is invertible. Consider then the system of equations $\psi_C^-(\theta_C^-) = \psi_C^-$ to solve: by plugging the first two equations into the last two, and pre-multiplying the first two equations by the nonsingular matrix $\Gamma_W(0)$, this system is equivalent to:

$$\begin{aligned} \Gamma_{Wy}(0,0) &= \Gamma_W(0) \begin{pmatrix} \psi_{C1}^- \\ \psi_{C2}^- \end{pmatrix} \\ \Gamma_{Wy}(0,0) \Gamma_W(0)^{-1} \Gamma_{Wy}(1,0) &= (\psi_{C1}^- \quad \psi_{C2}^-) \Gamma_W(1) \begin{pmatrix} \psi_{C1}^- \\ \psi_{C2}^- \end{pmatrix} - \psi_{C3}^- \\ \Gamma_{yX}(1,0) &= \psi_{C4}^- + (\psi_{C1}^- \quad \psi_{C2}^-) \Gamma_{WX}(1,0) \end{aligned} \tag{9}$$

The system of 4 equations in 4 unknowns in (9) has the important feature that only the left-hand side of (9) depends on θ . The right hand side consists either of $(\psi_{C1}^-, \dots, \psi_{C4}^-)$ or the elements in $\Gamma_W(0)$ and $\Gamma_W(1)$ for which sample estimates are available. Global identifiability of θ_C^- from ψ_C^- holds if it can be established that the system (9) has a unique solution in θ_C^- .

Lemma 3 *Consider the model $y_t = \alpha y_{t-1} + \beta x_t + u_t$, $X_t = x_t + \epsilon_t$. Under Assumptions A(a)-(c):*

- (a) $(\alpha, \beta = 0)$ is globally identified from ψ_C^- ;
- (b) $(\alpha, \beta \neq 0)$ is globally identified from ψ_C^- if (i) $\Gamma_x(1) \neq 0$ and (ii) $\psi_{C3}^- + \psi_{C2}^- \psi_{C4}^- \neq 0$;
- (c) $\theta_C^- = (\alpha, \beta, \Gamma_\epsilon(0), \Gamma_\epsilon(1))$ is globally identified if from ψ_C^- if (i) $\Gamma_x(1) \neq 0$, (ii) $\psi_{C3}^- + \psi_{C2}^- \psi_{C4}^- \neq 0$, and (iii) $\beta \neq 0$.

Compared to the assumptions in Lemma 2, the additional restriction that $\psi_{C3}^- + \psi_{C2}^- \psi_{C4}^- \neq 0$ is needed. To understand this restriction, note that

$$\psi_{C3}^- + \psi_{C2}^- \psi_{C4}^- = \Gamma_{\hat{V}}(1) + (\text{plim } T \rightarrow \infty \hat{\beta}) \Gamma_{\hat{V}X}(1, 0).$$

The restriction $\psi_{C3}^- + \psi_{C2}^- \psi_{C4}^- \neq 0$ ensures that there is independent information in $\Gamma_{\hat{V}}(1)$ and $\Gamma_{\hat{V}X}(1, 0)$.

3.2 An Order Condition for the ADL(p,q) Model

For the ADL(p,q) model, $\gamma = (\alpha_1, \dots, \alpha_p, \beta_0, \dots, \beta_q)$ and $W_t = (y_{t-1}, \dots, y_{t-p}, X_t, \dots, X_{t-q})$. The parameters of the model are

$$\theta_C = (\gamma, \sigma_u^2, \Gamma_\epsilon(0), \dots, \Gamma_\epsilon(q+k))',$$

The OLS estimator has asymptotic bias

$$\text{plim } T \rightarrow \infty \hat{\gamma} - \gamma = \Gamma_W(0)^{-1} \Gamma_{VW}(0, 0) \equiv \text{bias}_C(\theta_C)$$

and the auxiliary statistics are $\hat{\psi}_C(\theta_C) = (\hat{\gamma}, \hat{\Gamma}_{\hat{V}}(0), \dots, \hat{\Gamma}_{\hat{V}}(k), \hat{\Gamma}_{\hat{V}X}(1, 0), \dots, \hat{\Gamma}_{\hat{V}X}(k, 0))'$.

Exact global identification requires inverting the binding function $\psi_C(\theta) = \text{plim } T \rightarrow \infty \hat{\psi}_C(\theta_C)$ and establishing that it has a unique solution. This is already a nontrivial problem for the ADL(1,0) model. For the ADL(1,1), there are six nonlinear equations in six unknowns to be solved. However, we can still provide a useful order condition.

The parameters entering $\text{bias}_C(\theta_C)$ are those appearing in the covariances $\Gamma_{\hat{V}X}(0, i)$ for $0 \leq i \leq q$. Since $X_t = x_t + \epsilon_t$ and $V_t = u_t - \beta_0 \epsilon_t - \dots - \beta_q \epsilon_{t-q}$, the parameters $\Gamma_\epsilon(i)$ with $0 \leq i \leq q$ will appear in the probability limit of the moments. This implies that in addition to the $p + q + 2$ parameters of (γ, σ_u^2) in the ADL(p,q) model, there are now $q + 1$ nuisance parameters. At least $(p + q + 2) + (q + 1)$ auxiliary statistics are needed. The OLS estimator provides $p + q + 1$ statistics; the variance of the OLS residuals $\hat{\Gamma}_{\hat{V}}(0)$ provides another. But we still need $q + 1$ auxiliary statistics. By construction of the least squares estimator, we have $\hat{\Gamma}_{\hat{V}X}(0, i) = 0$, $0 \leq i \leq q$. So the moments left to consider are $\hat{\Gamma}_{\hat{V}X}(k, 0)$, $\hat{\Gamma}_{\hat{V}}(k)$, and $\hat{\Gamma}_{\hat{V}X}(0, q + k)$ for $k \geq 1$. The last one is less desirable unless X_t is strongly persistent. Thus we focus on $\hat{\Gamma}_{\hat{V}X}(k, 0)$ and $\hat{\Gamma}_{\hat{V}}(k)$.

To determine k , observe that in the ADL(p,q) model, $\hat{\Gamma}_{\hat{V}_X}(1,0)$ and $\hat{\Gamma}_{\hat{V}}(1)$ depend on parameters of ADL(p,q) ie. (γ, σ_u^2) , nuisance parameters already appearing in the OLS bias ie. $(\Gamma_\epsilon(0), \dots, \Gamma_\epsilon(q))$, and a new nuisance parameter, $\Gamma_\epsilon(q+1)$. Thus, when $k=1$, the inclusion of two auxiliary statistics $\hat{\Gamma}_{\hat{V}_X}(1,0)$ and $\hat{\Gamma}_{\hat{V}}(1)$ increases the nuisance parameters by one. This suggests the following simple rule: compare the total number of auxiliary statistics ($2k$) $\hat{\Gamma}_{\hat{V}_X}(1,0), \dots, \hat{\Gamma}_{\hat{V}_X}(k,0)$ and $\hat{\Gamma}_{\hat{V}}(1), \dots, \hat{\Gamma}_{\hat{V}}(k)$, with the total number of nuisance parameters $\Gamma_\epsilon(0), \dots, \Gamma_\epsilon(q)$ and $\Gamma_\epsilon(q+1), \dots, \Gamma_\epsilon(q+k)$ ($q+1+k$). The first has to be larger than the second, i.e. $2k \geq q+1+k$. Setting

$$k = q + 1$$

satisfies the rule. The condition works for the simple ADL(0,0) model with correlated errors. Indeed, we have defined $\theta_B = (\beta, \sigma_u^2, \Gamma_\epsilon(0), \Gamma_\epsilon(1))$ and $\hat{\psi}_B = (\hat{\beta}, \hat{\Gamma}_{\hat{V}}(0), \hat{\Gamma}_{\hat{V}}(1), \hat{\Gamma}_{\hat{V}_X}(1,0))$ which corresponds to $k=1$ for that model. The same condition holds for the ADL(1,0). We have used $\hat{\psi}_C \equiv (\hat{\alpha}, \hat{\beta}, \hat{\Gamma}_{\hat{V}}(0), \hat{\Gamma}_{\hat{V}}(1), \hat{\Gamma}_{\hat{V}_X}(1,0))$ with $\theta_C = (\alpha, \beta, \sigma_u^2, \Gamma_\epsilon(0), \Gamma_\epsilon(1))$.

For higher order ADL(p,q), letting $k = q+1$ rapidly leads to a large number of new nuisance parameters to include. For example, even in the ADL(1,1) model we would need to identify $\theta = (\alpha, \beta_0, \beta_1, \sigma_u^2, \Gamma_\epsilon(0), \Gamma_\epsilon(1), \Gamma_\epsilon(2), \Gamma_\epsilon(3))$ from $\hat{\psi} = (\hat{\alpha}, \hat{\beta}_0, \hat{\beta}_1, \hat{\Gamma}_{\hat{V}}(0), \hat{\Gamma}_{\hat{V}}(1), \hat{\Gamma}_{\hat{V}}(2), \hat{\Gamma}_{\hat{V}_X}(1,0), \hat{\Gamma}_{\hat{V}_X}(2,0))$. But a large number of nuisance parameters would make estimation difficult. To avoid proliferation of nuisance parameters, we use Assumption A(d) which requires that $\Gamma_\epsilon(1), \dots, \Gamma_\epsilon(q^*)$ are well approximated by ϕ_1, \dots, ϕ_m . Thus consider the ADL(p,q) model with $p+q+1$ parameters γ and nuisance parameters $\delta = (\sigma_u^2 \ \sigma_\epsilon^2 \ \phi)'$, where $\phi = (\phi_1, \dots, \phi_m)'$ characterizes $\Gamma_\epsilon(1), \dots, \Gamma_\epsilon(q^*)$. If Assumption A(d) holds, a necessary condition for

$$\theta_D = (\gamma', \delta')'$$

to be identified from

$$\hat{\psi}_D = (\hat{\gamma}, \hat{\Gamma}_{\hat{V}}(0), \dots, \hat{\Gamma}_{\hat{V}}(k), \hat{\Gamma}_{\hat{V}_X}(1,0), \dots, \hat{\Gamma}_{\hat{V}_X}(k,0))' \quad (10)$$

is

$$k \geq (m+1)/2.$$

Now q^* depends on the choice of auxiliary statistics. Since the residual autocovariances and cross-covariances $\hat{\Gamma}_{\hat{V}}(0), \dots, \hat{\Gamma}_{\hat{V}}(k), \hat{\Gamma}_{\hat{V}_X}(1,0), \dots, \hat{\Gamma}_{\hat{V}_X}(k,0)$ depend on $\Gamma_\epsilon(1), \dots, \Gamma_\epsilon(k+q)$ in addition to $(\gamma, \sigma_u^2, \sigma_\epsilon^2)$, we need $q^* = k+q$.

For example, in the ADL(1,1) model, if we consider $k=2$ and include $\hat{\Gamma}_{\hat{V}_X}(2,0)$ among others in our auxiliary statistics, we have $q^* = 3$ autocovariances of ϵ_t to consider. However, if we use the same auxiliary statistics as in the simple regression model with correlated measurement errors,

$k = 1$ so $q^* = 2$. Assumption A(d) then requires that $\Gamma_\epsilon(1)$ and $\Gamma_\epsilon(2)$ are well approximated by $m = 2k - 1 = 1$ parameter, ϕ . This would be consistent with ϵ_t being an AR(1) or an MA(1) process. This suggests the following general rule: choose an order k in (10). Set

$$m = 2k - 1 \quad (11)$$

and require that $k + q$ autocovariances of ϵ , $\Gamma_\epsilon(1), \dots, \Gamma_\epsilon(k + q)$, be well approximated by m parameters ϕ_1, \dots, ϕ_m .

We should point out that neither of the above order conditions are strictly necessary for identification since additional information relating to heteroskedasticity and skewness of ϵ_t can also be exploited. In the spirit of Pal (1980), Dagenais and Dagenais (1997), Lewbel (1997), Spierdijk and Wansbeek (2011), and Erickson and Whited (2000, 2002), higher order moments of ϵ_t or of X_t can also be used to achieve identification. For example, the third moment of x_t is related to that of X_t by $s_X = s_x + s_\epsilon$. From $V_t = u_t - \beta\epsilon_t$, we have $s_V = \sigma_u^3 s_u - \beta^3 \sigma_\epsilon^3 s_\epsilon$ and $\widehat{\Gamma}_{\widehat{V}X^2}(0, 0) = \frac{1}{T} \sum_{t=1}^T \widehat{V}_t X_t^2 \xrightarrow{p} s_X \cdot \text{bias}(\theta) - \beta s_\epsilon$.

Once the order condition is established, we can proceed with a local identification analysis by evaluating the rank of the $\frac{\partial \psi_G(\theta)}{\partial \theta}$. Numerical evaluation of the Jacobian matrix reveals the not-so-surprising result that θ_D is not identified when β_0 or $\beta_1 = 0$, and x_t is not serially correlated. In the ADL(1,1) model, θ_D can be identified even when x_t is not serially correlated provided that $\beta_0 \neq 0$ or $\beta_1 \neq 0$. The reason is that serial correlation in \widehat{V}_t can also come from autocorrelation in y_t and V_t .

4 Dynamic Panel Models

The least squares bias in the ADL(p,q) model is due solely to the measurement error in X_t because $E(V_t y_{t-j}) = 0$ for $j = 1, \dots, p$. This will no longer be the case in the dynamic panel model. Let N and T be the number of units and time periods. Consider a panel AR(1) model (or equivalently a PADL(1,0) model) with a scalar exogenous regressor and additive fixed effects:

$$y_{it} = \eta_i + \alpha y_{i,t-1} + \beta x_{it} + u_{it}, \quad (12)$$

for $i = 1, \dots, N$ and $t = 1, \dots, T$. Nickell (1981) showed that $\widehat{\theta}_{N,T}$ is inconsistent as $N \rightarrow \infty$ with T fixed. Estimates can be obtained by either by using instruments or by explicitly correcting for the small T bias.⁷ Kiviet (1995) showed that these bias-corrected estimators tend to have smaller root mean-squared error (RMSE) than the IV estimators. However, the bias corrections are model specific and are invalid when there are additional covariates or lagged independent variables.

⁷Kiviet (1995), Bun and Carree (2005), and Phillips and Sul (2007) approximate the biases using large N asymptotics. Large N and T approximations has been used by Hahn and Kuersteiner (2002) to obtain $b_{\infty, \infty}(\cdot)$.

As is well documented, measurement errors in survey and longitudinal data are prevalent.⁸ Existing work identifies the parameters from the covariance between the errors in the equations and the measurement errors, or higher order moments of the data.⁹ Erickson and Whited (2012) compare high-order moment estimators with dynamic panel estimators and estimators that use lags of the mismeasured regressor. They find that all estimators perform well under correct specification but are biased under misspecification. We suggest a new approach of combining biased estimators.

Let the ideal regressor x_{it} be possibly contaminated by serially uncorrelated classical measurement errors:

$$X_{it} = x_{it} + \epsilon_{it}.$$

In terms of observables, the dynamic panel model becomes:

$$y_{it} = \eta_i + \alpha y_{i,t-1} + \beta' X_{it} + V_{it}, \quad V_{it} = u_{it} - \beta \epsilon_{it} \quad (13)$$

so the composite error V_{it} is correlated with X_{it} , and $E(V_{it}X_{it}) = -\beta\sigma_\epsilon^2$.

Assumption P For all $s, t = 1, \dots, T$ and for all $i = 1, \dots, N$:

- (a) u_{it} is iid across i and t with $E(u_{it}) = 0$, $E(u_{it}^2) = \sigma_u^2$, $E(u_{it}y_{i0}) = 0$, $E(u_{it}x_{is}) = 0$, and $E(u_{it}\eta_i) = 0$.
- (b) ϵ_{it} is iid across i and t with $E(\epsilon_{it}) = 0$, $E(\epsilon_{it}^2) = \sigma_\epsilon^2$, and $E(\epsilon_{it}x_{is}) = 0$, and $E(\epsilon_{it}u_{is}) = 0$.

The disturbances u_{it} are iid across both time and individuals, and uncorrelated with the initial observations y_{i0} . While u_{it} is orthogonal to the individual fixed effect η_i as well as to all values of the explanatory variable x_{it} , the correlation between η_i and x_{it} is left unspecified. The measurement error is iid and classical, and uncorrelated with u_{it} . Thus, $V_{it} = u_{it} - \beta\epsilon_{it}$ and $\Gamma_V(0) = \sigma_u^2 + \beta^2\sigma_\epsilon^2$ and $\Gamma_V(j) = 0$ for $j \geq 1$.

As is standard in panel data model with fixed effects, the within-transformation is used to eliminate the unobserved individual effects η_i in (13). For a generic variable z , let $\tilde{z}_{it} = z_{it} - \frac{1}{T} \sum_{s=1}^T z_{is}$ denote its demeaned value. We use the notation y_{-1} to denote the lagged dependent variable and $\tilde{y}_{i,t-1} = y_{i,t-1} - \frac{1}{T} \sum_{s=1}^T y_{i,s-1}$ its demeaned version. The sample autocovariance of order j of variable z is defined as $\Gamma_{\tilde{z}}(j) = \text{plim}_{N \rightarrow \infty} \hat{\Gamma}_{\tilde{z}}(j)$ where $\hat{\Gamma}_{\tilde{z}}(j) = \frac{1}{N(T-1)} \sum_{i=1}^N \sum_{t=1+j}^T \tilde{z}_{it} \tilde{z}'_{i,t-j}$. The cross

⁸The problem appears to be particularly important in investment data. See Erickson and Whited (2000) and Eberly, Rebelo, and Vincent (2009) among others.

⁹Erickson and Whited (2002), Erickson, Jiang, and Whited (2014) exploit non-normality of the data to identify the parameters of the model. See also Biorn (2008), Wansbeek and Koning (1991), Meijer, Spierdijk, and Wansbeek (2012), and Spierdijk and Wansbeek (2011) for discussion of choice of moments.

covariance between generic variables z_1 and z_2 is $\widehat{\Gamma}_{\tilde{z}_1\tilde{z}_2}(u, v) = \frac{1}{N(T-1)} \sum_{i=1}^N \sum_{t=\max\{u,v\}}^T \tilde{z}_{1,i,t-u} \tilde{z}_{2,i,t-v}'$ and $\Gamma_{\tilde{z}_1\tilde{z}_2}(u, v)$ is its probability limit as N tends to ∞ .

The PADL(1,0) model written in demeaned data form becomes

$$\tilde{y}_{it} = \alpha \tilde{y}_{i,t-1} + \beta \tilde{X}_{it} + \tilde{V}_{it}, \quad \tilde{V}_{it} = \tilde{u}_{it} - \beta \tilde{\epsilon}_{it}.$$

Though its structure is similar to the ADL(1,0) model, there is one fundamental difference: the within-transformation induces serial correlation in variables that are initially uncorrelated. This, in particular, will be the case for \tilde{V}_{it} .

Now let $\tilde{W}_{it} = (\tilde{y}_{i,t-1}, \tilde{X}_{it})'$ and assume the covariance matrix $\Gamma_{\tilde{W}}(0)$ is full rank. The least squares dummy variable (LSDV or fixed effect) estimator for $\gamma = (\alpha, \beta)'$ is the pooled OLS estimator in a regression of \tilde{y}_{it} on \tilde{X}_{it} . It has the property that

$$\text{plim}_{N \rightarrow \infty} \hat{\gamma} - \gamma = \Gamma_{\tilde{W}}(0)^{-1} \Gamma_{\tilde{W}\tilde{V}}(0, 0) \equiv \text{bias}_P(\theta_P), \quad (14)$$

where

$$\Gamma_{\tilde{W}\tilde{V}}(0, 0) = \begin{pmatrix} -\sigma_u^2 h_T(\alpha) \\ -\beta \sigma_\epsilon^2 \end{pmatrix}, \quad h_T(\alpha) = \frac{T(1-\alpha) - 1 + \alpha^T}{T(T-1)(1-\alpha)^2}.$$

The fixed- T -Nickell bias is represented by $h_T(\alpha)$ and is present even in the absence of measurement error.¹⁰ This is in contrast with the ADL(1,0) model considered before where $E(y_{t-1}V_t) = 0$. Measurement error introduces a new source of bias to the dynamic panel model because of $E(X_t V_t) \neq 0$. However, the measurement error bias and the small- T bias are opposite in sign. Thus, as noted in Biorn (1992), measurement error can alleviate the effect of heterogeneity; estimators that are sensitive to heterogeneity can be more robust to measurement errors.

The LSDV residuals $\hat{V}_{it} = \tilde{y}_{it} - \hat{\gamma}' \tilde{W}_{it}$ are serially correlated even if \tilde{V}_{it} is uncorrelated over time. The error \tilde{V}_{it} is correlated with the demeaned regressors \tilde{y}_{it} , \tilde{X}_{it} , and their lags in a nontrivial manner because of the interaction between the fixed effect bias and measurement error bias. Direct computations show that the autocovariances of \hat{V}_{it} have probability limits (as $N \rightarrow \infty$) given by

$$\Gamma_{\hat{V}}(0) = \Gamma_{\tilde{V}}(0) - \text{bias}_P(\theta_P)' \Gamma_{\tilde{W}}(0) \text{bias}_P(\theta_P) \quad (15)$$

$$\Gamma_{\hat{V}}(1) = \Gamma_{\tilde{V}}(1) + \text{bias}_P(\theta_P)' \Gamma_{\tilde{W}}(1) \text{bias}_P(\theta_P) - (\Gamma_{\tilde{W}\tilde{V}}(1, 0) + \Gamma_{\tilde{W}\tilde{V}}(0, 1))' \text{bias}_P(\theta_P) \quad (16)$$

with

$$\begin{aligned} \Gamma_{\tilde{V}}(0) &= \sigma_u^2 + \beta^2 \sigma_\epsilon^2, & \Gamma_{\tilde{V}}(1) &= -\frac{1}{T} [\sigma_u^2 + \beta^2 \sigma_\epsilon^2], \\ \Gamma_{\tilde{W}\tilde{V}}(1, 0) &= \begin{pmatrix} -\sigma_u^2 [(T+1)h_T(\alpha) - 1] \\ \frac{1}{T} \beta \sigma_\epsilon^2 \end{pmatrix}, & \Gamma_{\tilde{W}\tilde{V}}(0, 1) &= \begin{pmatrix} -\sigma_u^2 [(T-1)h_T(\alpha) + \frac{T-2}{T} h_{T-1}(\alpha)] \\ \frac{1}{T} \beta \sigma_\epsilon^2 \end{pmatrix}. \end{aligned}$$

¹⁰The above expressions then reduce to equation (12) in Bun and Carree (2005).

The expression for $\Gamma_{\hat{V}}(0)$ reduces to that derived in Bun and Carree (2005) when no measurement errors are present. The lagged autocovariance reflects the cumulated fixed effect bias arising from the correlation between $\tilde{y}_{i,t-1}$ and \tilde{V}_{it} as well as $\tilde{V}_{i,t-1}$. Note that $\Gamma_{\tilde{W}\tilde{V}}(0,1)$ does not vanish even as $T \rightarrow \infty$ which we can exploit for identification.

The parameters of the PADL(1,0) model are

$$\theta_P = (\alpha, \beta, \sigma_u^2, \sigma_\epsilon^2)',$$

and we consider the auxiliary statistics

$$\hat{\psi}_P(\theta_P) = \begin{pmatrix} \hat{\alpha} & \hat{\beta} & \hat{\Gamma}_{\hat{V}}(0) & \hat{\Gamma}_{\hat{V}}(1) \end{pmatrix}',$$

whose binding function $\psi_P(\theta_P) = \text{plim}_{N \rightarrow \infty} \hat{\psi}_P(\theta_P)$ is given by equations (14)-(16). As before, θ_P is identified from ψ_P is the mapping $\theta_P \mapsto \psi_P(\theta_P)$ is invertible. To examine invertibility, we first transform the system of equations $\psi_P(\theta_P) = \psi_P$ into:

$$\begin{aligned} \Gamma_{\tilde{W}\tilde{y}}(0,0) &= \Gamma_{\tilde{W}}(0) \begin{pmatrix} \psi_{P1} \\ \psi_{P2} \end{pmatrix} \\ \Gamma_{\tilde{y}}(0) &= \psi_{P3} + (\psi_{P1} \quad \psi_{P2}) \Gamma_{\tilde{W}}(0) \begin{pmatrix} \psi_{P1} \\ \psi_{P2} \end{pmatrix} \\ \Gamma_{\tilde{W}\tilde{y}}(0,0)' \Gamma_{\tilde{W}}(0)^{-1} \Gamma_{\tilde{W}\tilde{y}}(1,0) &= -\psi_{P4} + (\psi_{P1} \quad \psi_{P2}) \Gamma_{\tilde{W}}(1) \begin{pmatrix} \psi_{P1} \\ \psi_{P2} \end{pmatrix} \end{aligned} \quad (17)$$

Note that the system (17) is equivalent to the original set of equations $\psi_P(\theta_P) = \psi_P$ since it only involves pre-multiplying the first two equations by $\Gamma_{\tilde{W}}(0)$ which is nonsingular, and plugging them into the last two equations. As was the case in our analysis of ARX(1,0) in the previous section, only the left hand side terms of (17) depend on θ_P . They can be expanded using:

$$\begin{aligned} \Gamma_{\tilde{W}\tilde{y}}(0,0) &= \begin{pmatrix} \alpha \Gamma_{\tilde{y}_{-1}}(0) + \beta \Gamma_{\tilde{X}\tilde{y}_{-1}}(0,0) - \sigma_u^2 h_T(\alpha) \\ \alpha \Gamma_{\tilde{X}\tilde{y}_{-1}}(0,0) + \beta \Gamma_{\tilde{X}}(0) - \beta \sigma_\epsilon^2 \end{pmatrix} \\ \Gamma_{\tilde{y}}(0) &= \alpha^2 \Gamma_{\tilde{y}_{-1}}(0) + 2\alpha\beta \Gamma_{\tilde{X}\tilde{y}_{-1}}(0,0) + \beta^2 \Gamma_{\tilde{X}}(0) + \sigma_u^2 + \beta^2 \sigma_\epsilon^2 - 2\sigma_u^2 h_T(\alpha) - 2\beta \sigma_\epsilon^2 \\ \Gamma_{\tilde{W}\tilde{y}}(1,0) &= \begin{pmatrix} \alpha \Gamma_{\tilde{y}_{-1}}(1) + \beta \Gamma_{\tilde{X}\tilde{y}_{-1}}(0,1) - \sigma_u^2 [(T+1)h_T(\alpha) - 1] \\ \alpha \Gamma_{\tilde{X}\tilde{y}_{-1}}(1,0) + \beta \Gamma_{\tilde{X}}(1) + \beta \frac{1}{T} \sigma_\epsilon^2 \end{pmatrix} \end{aligned}$$

Now define the mapping $\theta_P \mapsto \Psi_P(\theta_P)$ by letting

$$\Psi_P(\theta_P) = \begin{pmatrix} \Gamma_{\tilde{W}\tilde{y}}(0,0) \\ \Gamma_{\tilde{y}}(0) \\ \Gamma_{\tilde{W}\tilde{y}}(0,0)' \Gamma_{\tilde{W}}(0)^{-1} \Gamma_{\tilde{W}\tilde{y}}(1,0) \end{pmatrix}$$

Identification of θ_P can now be analyzed in terms of invertibility of Ψ_P .

Lemma 4 *Suppose that Assumption P holds. Then θ_P is globally identifiable from ψ_P if and only if the system of equations $\Psi_P(\theta_P) = \Psi_P$ has a unique solution in θ_P . If $\text{rank} \left(\frac{\partial \Psi_P(\theta_P)}{\partial \theta_P} \right) = 4$, then θ_P is locally identifiable from ψ_P .*

Due to the presence of $h_T(\alpha)$, which is a highly nonlinear function in α , $\psi_P(\theta_P)$ is difficult to invert analytically. Lemma 4 is obtained by first isolating the terms in $\psi_P(\theta)$ that do not depend on θ_P , and then collecting the terms that do depend on θ_P into $\Psi_P(\theta)$. Local injectivity of $\psi_P(\theta_P)$ is equivalent to local injectivity of $\Psi_P(\theta_P)$. Hence local identification of θ_P amounts to checking the rank of the Jacobian matrix, $\frac{\partial \Psi_P(\theta_P)}{\partial \theta_P}$.

5 Classical and Simulated Minimum Distance Estimation

The above analysis provides conditions under which θ can be identified from appropriately defined $\psi(\theta)$. This leads naturally to classical minimum distance (CMD) estimation that identifies θ from $\hat{\psi}(\theta)$. The CMD estimator is generically defined by

$$\hat{\theta} = \underset{\theta}{\text{argmin}} J(\theta), \quad J(\theta) = \|\hat{\psi}(\theta_0) - \psi(\theta)\|_{W_T}, \quad (18)$$

where W_T is a weighting matrix. Provided that the conditions stated in Newey and McFadden (1994) hold, $\hat{\theta}$ is instrument-free estimator with classical property that

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N\left(0, [\psi_\theta(\theta_0) \text{Avar}(\hat{\psi})^{-1} \psi_\theta(\theta_0)']^{-1}\right) \equiv N(0, \text{Avar}(\hat{\theta}))$$

where $\psi_\theta(\theta_0)$ is the matrix of derivatives of $\psi(\theta)$ with respect to θ evaluated at θ_0 .

In simple linear models, the probability limit of $\hat{\psi}(\theta)$ is easily derived. As shown above for the ADL(0,0) model, the binding function can be inverted to obtain closed-form expressions for $\psi(\theta)$ in the case of exact identification. But the probability limit of the terms in $\hat{\psi}(\theta)$ can be tedious and sometimes impossible to calculate. While we can still invert the binding function for the ADL(1,0) model, the exercise was impossible for the ADL (1,1) model despite serious efforts. For the dynamic panel model, Ψ_P is a complicated function of θ_P even in the absence of measurement error. A closed-form solution for $\theta_P(\theta)$ is out of the question.

Simulated minimum distance (SMD) estimators use Monte-Carlo methods to compute the mapping from θ to ψ and is therefore feasible even when $\psi(\theta)$ is not analytically tractable. Specifically, let $(\mathbf{y}^S(\theta), \mathbf{X}^S(\theta))$ denote S sets of data simulated under an assumed value for θ , where each set $\mathbf{y}^s(\theta)$ is $T \times 1$ and $\mathbf{X}^s(\theta)$ is $T \times K$. The estimator is

$$\hat{\theta}^S = \underset{\theta}{\text{argmin}} J^S(\theta), \quad J^S(\theta) = \|\hat{\psi}(\theta_0) - \psi^S(\theta)\|_{W_T}, \quad (19)$$

where

$$\psi^S(\theta) = \frac{1}{S} \sum_{s=1}^S \hat{\psi}(\mathbf{y}^s(\theta), \mathbf{X}^s(\theta)) \approx E_{(\mathbf{y}^S(\theta), \mathbf{X}^S(\theta))}[\hat{\psi}(\mathbf{y}^S(\theta), \mathbf{X}^S(\theta))].$$

As in classical minimum distance estimation, the mapping $\psi(\theta)$ must be injective; simulations merely provide an approximation to $\hat{\theta} = \psi^{-1}(\hat{\psi})$. When ψ is a vector of unconditional moments, $\hat{\theta}$ is the simulated method of moments estimator of Duffie and Singleton (1993). When ψ are the scores of the likelihood, the efficient methods of moments estimator of Gallant and Tauchen (1996) obtains. The indirect inference estimator of Gourieroux, Monfort, and Renault (1993) defines ψ to be the parameters of an auxiliary regression.¹¹ Simulation estimators automatically provides second order bias correction if the auxiliary model admits an Edgeworth expansion, though the bias reduction comes at the cost of efficiency, see Gourieroux, Monfort, and Renault (1993). Gourieroux, Phillips, and Yu (2010) showed that indirect inference provides a substantial reduction in bias when the linear dynamic panel model is also the auxiliary model.

As far as we are aware, the only reference to simulation estimation of measurement error models is Jiang and Turnbull (2004) who used indirect inference as a way to adjust the bias in the auxiliary parameters with the help of validation data. Without validation data, simulation estimation cannot be implemented in the standard way. To see why, suppose that the auxiliary model coincides with the regression model. If $E(V_t X_t) = 0$, the auxiliary model can be estimated by OLS to yield $\hat{\psi}$ and $\text{plim } \hat{\psi} = \psi$. But when $E(V_t X_t) \neq 0$, $\text{plim } \hat{\psi} = \psi + \text{bias} \neq \psi$. If X_t is fixed in simulations and V_t^s drawn such that $X_t \perp V_t^s$, then by construction, $\text{plim } \hat{\psi}^s(\theta) = \psi(\theta)$. The binding function $\psi(\theta)$ will not be consistently estimated because the bias term is omitted.

The generic problem is that when X_t is not weakly exogenous for the parameters, they can not be held fixed in simulations. In the measurement error model, the parameters in the marginal distribution of X_t and those of the conditional distribution of y_t given X_t are not variation free. For the simulation estimation to work, it is necessary that the simulated X_t^s preserves these relations. One can approximate x_t by a parametric model such as an AR(2) model $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t$. The Yule-Walker equations then yield

$$\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = \begin{pmatrix} \Gamma_X(1) & \Gamma_X(2) \\ \Gamma_X(0) - \sigma_\epsilon^2 & \Gamma_X(1) \end{pmatrix}^{-1} \begin{pmatrix} \Gamma_X(0) - \sigma_\epsilon^2 - \sigma_w^2 \\ \Gamma_X(1) \end{pmatrix}.$$

The drawback of this approach is that θ has to be further augmented to include σ_w^2 , and from a computational point of view, the presence of σ_ϵ^2 makes it difficult to control stability of x_t since ϕ_1, ϕ_2 can take on values arbitrarily large.

The question then is how to simulate X_t with the desired properties without fully specifying its dynamic properties. We make use of the fact that covariance stationary processes with identical

¹¹The estimators of Gallant and Tauchen (1996) and Gourieroux, Monfort, and Renault (1993) consider pseudo-maximum likelihood estimator of ψ .

second moments are observationally equivalent. Thus, it is only necessary for the simulated data to match the first and second moment properties the observed data. To motivate our simulator, note that if x_t were serially uncorrelated, the mean and variance of the simulated data can be preserved by letting $X_t^s = x_t^s + \epsilon_t^s$ with $x_t^s = \varphi X_t$ and $\varphi = (1 - \frac{\sigma_\epsilon^2}{\Gamma_X(0)})^{1/2}$. By construction, the mean and variance of X_t^s are equal to the mean and variance of X_t . But with serially correlated regressors, we will need the simulated regressors to not only preserve the variance, but also the autocovariance structure in the data. For this reason, we propose the following simulation method:

Algorithm SMD

1. Compute the auxiliary statistics $\hat{\psi}$ from the data.
2. For $s = 1, \dots, S$ and $t = 1, \dots, T$ and given $\{\epsilon_t^s\}$ and $\{u_t^s\}$:
 - (i) let $x_t^s = \varphi_1 X_t + \varphi_2 X_{t-1}$;
 - (ii) simulate $X_t^s = x_t^s + \epsilon_t^s$;
 - (iii) $y_t^s = \alpha + \beta x_t^s + u_t^s$;
 - (iv) compute $\hat{\psi}(\theta^s)$ from the simulated data (\mathbf{y}^s, X^s) .
- 3 Minimize $\|\hat{\psi} - \frac{1}{S} \sum_s \hat{\psi}^s(\theta)\|_{W_T}$ over θ .

The key to Algorithm SMD is Step 2(i) in which we postulate that x_t^s is linear in X_t and X_{t-1} . The constants φ_1 and φ_2 are chosen to satisfy the pair of equations:

$$\begin{aligned}\Gamma_X(0) - \Gamma_\epsilon(0) &= (\varphi_1^2 + \varphi_2^2)\Gamma_X(0) + 2\varphi_1\varphi_2\Gamma_X(1) \\ \Gamma_X(1) - \Gamma_\epsilon(1) &= (\varphi_1^2 + \varphi_2^2)\Gamma_X(1) + \varphi_1\varphi_2\Gamma_X(0) + \varphi_2\Gamma_X(2).\end{aligned}\tag{20}$$

Step 2(i) thus models x_t^s as a rescaled but deterministic function of the data X_t . This method does not directly model the dynamics of x_t (or of x_t^s), but by construction, $\Gamma_x^s(0) = \Gamma_X(0) - \Gamma_\epsilon(0)$ and $\Gamma_x^s(1) = \Gamma_X(1) - \Gamma_\epsilon(1)$. Given estimates of $\Gamma_X(0)$ and $\Gamma_X(1)$, there are two equations in two unknowns. A unique solution for φ_1 and φ_2 can be obtained by noting that the system in (20) is linear in r, ϕ , where $r = \varphi_1^2 + \varphi_2^2$ and $r_2 = \varphi_1\varphi_2$:

$$\underbrace{\begin{pmatrix} \Gamma_X(0) & 2\Gamma_X(1) \\ \Gamma_X(1) & \Gamma_X(0) + \Gamma_X(2) \end{pmatrix}}_A \underbrace{\begin{pmatrix} r_1 \\ r_2 \end{pmatrix}}_B = \underbrace{\begin{pmatrix} \Gamma_X(0) - \Gamma_\epsilon(0) \\ \Gamma_X(1) - \Gamma_\epsilon(1) \end{pmatrix}}_B.$$

Assuming that A is invertible, $(r_1 \ r_2)' = A^{-1}B$. Then φ_1 and φ_2 can be taken as

$$\varphi_1 = \frac{1}{2} [\sqrt{r_1 + 2r_2} + \sqrt{r_1 - 2r_2}], \quad \varphi_2 = \frac{1}{2} [\sqrt{r_1 + 2r_2} - \sqrt{r_1 - 2r_2}].$$

The simulation procedure can be easily adapted to allow for skewed errors or errors with fat tails. This is used in Gospodinov and Ng (2014) to estimate moving-average models without imposing invertibility. The procedure can also be adapted to use in dynamic panel regressions.

One we can simulate the model to obtain draws of X_t , the arguments in Gourieroux, Monfort, and Renault (1993) can be used to analyze the properties of the SMD estimator. The order condition requires that the dimension of ψ exceeds the dimension of θ . The necessary and sufficient condition for local identification is that $\frac{\partial \psi(\theta)}{\partial \theta'}$ has full column rank. Once identification is granted, the SMD estimator has standard properties. If the auxiliary parameters $\hat{\psi}$ are \sqrt{T} -consistent and asymptotically normal, the simulation estimator is \sqrt{T} -consistent and asymptotically normal:

$$\sqrt{T}(\hat{\theta}^S - \theta_0) \xrightarrow{d} N\left(0, \left(1 + \frac{1}{S}\right)[\psi_\theta(\theta_0) \text{Avar}(\hat{\psi})^{-1} \psi_\theta(\theta_0)']^{-1}\right) = N(0, \text{Avar}(\hat{\theta}^S)).$$

The asymptotic variance is of the double sandwich form. It depends on the asymptotic variance of the auxiliary statistics, whether the auxiliary model is correctly specified, and is smaller the larger is the number of simulations, S . Also, $T \times J(\hat{\theta}^S) \xrightarrow{d} \chi_{\dim(\psi) - \dim(\theta)}^2$ as $T \rightarrow \infty$ and W_T is the optimal weighting matrix. The distance metric (DM) statistic for testing the null hypothesis that a constraint is true is $DM = \frac{TS}{S+1}(J(\hat{\theta}_C^S) - J(\hat{\theta}^S))$ where $\hat{\theta}_C^S$ is the constrained SMD estimate.

The variance of $\hat{\theta}^S$ is based on the delta method and can be inaccurate in finite samples. We construct confidence intervals for an element of θ , θ_i , by inverting the DM statistics under a sequence of constraints $H_0 : g(\theta) = 0$. This approach is also considered in Czellar and Zivot (2008). If η is the significance level of the test and $q_{1-\eta, df}$ is the $(1 - \eta)$ -th quantile of the chi-squared distribution with df degrees of freedom, the $100(1 - \eta)\%$ confidence interval for θ_i is given by the set of values of θ satisfying $DM \leq q_{1-\eta, df}$.

5.1 Simulations

We use 5000 replications to illustrate the properties of the CMD and SMD estimators. For $t = 1, \dots, T = (200, 500, 1000)$, the data are generated from the ADL(1,1) model

$$\begin{aligned} y_t &= \alpha y_{t-1} \beta_0 x_t + \beta_1 x_{t-1} + u_{yt}, & u_{yt} &\stackrel{iid}{\sim} \mathcal{N}(0, \sigma_{uy}^2), \\ x_t &= \rho_x x_{t-1} + u_{xt}, & u_{xt} &\stackrel{iid}{\sim} \mathcal{N}(0, \sigma_{ux}^2), \\ X_t &= x_t + \epsilon_t, & \epsilon_t &= e_t + \theta e_{t-1}, & e_t &\stackrel{iid}{\sim} \mathcal{N}(0, \sigma_e^2). \end{aligned}$$

We assume $(\alpha, \beta) = (0, 1)$ with $\rho_x = (0.2, 0.5, 0.8)$. The measurement error process is calibrated such that the signal-to-noise ratio $R^2 = \frac{\text{var}(x_t)}{\text{var}(X_t)} = 0.7$. This is achieved by solving σ_e^2 from

$$\sigma_e^2(1 + \theta^2) = \frac{1 - R^2}{R^2} \frac{\sigma_{ux}^2}{1 - \rho_x^2}.$$

In the simulations, we let $\sigma_{uy}^2 = \sigma_{ux}^2 = 1$. In practice, we do not know if ϵ_t is serially correlated or not. We always estimate the model to allow for serial correlation in ϵ_t even when ϵ_t is white noise. The SMD simulates ϵ_t as an AR(1) process even though the true process is MA(1). We only consider the case of exact identification.

We begin with the simple regression model when $\alpha = \beta_1 = 0$. As these parameters are not estimated, $\theta = (\beta_0, \sigma_u^2, \sigma_\epsilon^2, \phi)$ and $\hat{\psi} = (\hat{\beta}, \hat{\Gamma}_{\hat{V}}(0), \hat{\Gamma}_{\hat{V}}(1), \hat{\Gamma}_{\hat{V}X}(1, 0))'$. The results are reported in Table 1. In the top panel where ϵ_t is white noise, X_{t-1} is a valid instrument. The estimator is denoted by IV. For comparison purposes, Table 1 also reports the estimates from the infeasible estimator (IDEAL) based on the true (latent) regressor x_t . As expected, the average of the IDEAL estimates is well centered around the true value of β . The OLS estimates are significantly downward biased when X_t is used as regressor instead of x_t . The bias is larger the less persistent is x_t . The IV estimator gives highly variable estimates when $\rho_x = 0.2$. The CMD is more stable than IV. The SMD estimator matches up well with the CMD, showing that simulation estimation of the mapping from θ to ψ did not induce much efficiency loss. The bottom panel shows that when ϵ_t is serially correlated, the IV estimates are highly unreliable. The CMD and SMD estimates are similar to the case of white noise measurement error.

The parameters of the ADL(1,1) model are $\theta = (\alpha, \beta_0, \beta_1, \sigma_u^2, \sigma_\epsilon^2, \phi)$. The auxiliary statistics are $\hat{\psi}(\theta) = (\hat{\alpha}, \hat{\beta}_0, \hat{\beta}_1, \hat{\Gamma}_{\hat{V}}(0), \hat{\Gamma}_{\hat{V}}(1), \hat{\Gamma}_{\hat{V}X}(1, 0))'$. We report the estimated short- and long-run response of y_t to x_t as given by $\hat{\beta}_0$ and $\frac{\hat{\beta}_0 + \hat{\beta}_1}{1 - \hat{\alpha}}$. Table 2 reports results for ADL(1,0). This is a special ADL(1,1) model with $\beta_1 = 0$, but this constraint is not imposed in the estimation. The estimates exhibit some downward biases that tend to increase with the degree of persistence in x_t . The estimates are reasonably precise, taking into account that estimates of models with an autoregressive structure tend to be downward biased. Table 3 shows results for the ADL(1,1) model. The estimates are somewhat more precise than those in Table 2 for the ADL(1,0) model.

Results for the panel data model are reported in Table 4 based on $N = 200$. Two models are considered: PADL(1,0) and PADL(1,1). We assume that for all i , $\eta_i \sim N(0, 1)$, and x_{it} is an AR(1) process with autocorrelation of 0.5. The parameters are again chosen such that σ_ϵ^2 is about 0.7 of $\Gamma_X(0)$. The first column is the infeasible estimator based on the latent regressor, x , and is denoted LSDV- x . The estimates are biased even when x is correctly observed because of the fixed effect bias. The biases are even larger when X is used in place of x . The SMD estimates are closer to the true values.

5.2 Application to the Traditional Phillips Curve

Despite the advances in the development and empirical assessment of the New Keynesian Phillips curve, the traditional Phillips curve continues to be widely used for inflation forecasting and pol-

icy analysis, and it remains a topic of active academic research. The traditional Phillips curve considered in Gordon (2013) and Watson (2014) is an autoregressive distributed lag model with exogenous variables z , which we denote by ADLZ(p,q,s). Specifically,

$$\alpha(L)(1-L)\pi_t = \beta(L)x_t + \zeta(L)z_t + u_t,$$

where π_t is inflation, z_t is a vector of supply shock variables, $\alpha(L)$, $\beta(L)$ and $\zeta(L)$ are lag polynomials of orders p , q and s respectively, and u_t is a serially uncorrelated error term. The long-run effect of unemployment on the change in inflation $\Delta\pi_t$ is

$$\bar{\beta} = \frac{\beta(1)}{\alpha(1)}.$$

An important variable in the traditional Phillips curve is the ‘demand gap’ variable x_t which is meant to capture the deviation of real activity from its equilibrium. There is no unique measure of real activity. It is quite common in empirical work to use the unemployment rate as a proxy. Feng and Hu (2013) present evidence that the official U.S. unemployment rate underestimates the true unemployment rate due to misclassification of labor force status in the Current Population Survey. Even if real activity is correctly measured, the natural rate of unemployment is not observed. Any constructed ‘gap’ variable used in estimation is potentially subject to measurement error. In what follows, we explicitly account for possible measurement error in the unemployment gap $x_t = U_t - U_t^*$ and consider different proxies for U_t^* .

The data, same as in Watson (2014) and available from the author’s website, covers the period 1959:Q3–2013:Q3. The inflation rate π_t is the core inflation (PCE less food and energy) and U_t is the civilian unemployment rate for all workers ages 16+. The control variables are $z_t = (Z_{t1}, Z_{t-1,1}, Z_{t-1,2}, Z'_{t,3})'$, where Z_{t1} is the relative price of nonpetroleum imports, $Z_{t,2}$ is change in the relative price of food and energy inflation, and $Z_{t,3}$ denotes two dummy variables for the 1971–1974 Nixon-era price controls (see Gordon (2013) and Watson (2014)).¹² All series are seasonally adjusted.

We consider six measures of U_t^* , giving $X_t(k) = U_t - U_t^*(k)$ for $k = 1, \dots, 6$.

$$\begin{aligned} U_t^*(1) &= \min(U_t, \dots, U_{t-11}); & U_t^*(2) &= U_{t-4}; & U_t^*(3) &= BP(U_t, \text{one-sided}); \\ U_t^*(4) &= HP(U_t, 1600); & U_t^*(5) &= HP(U_t, 6400). & U_t^*(6) &= \text{constant}. \end{aligned}$$

The first three are ‘real-time gaps’, considered in Stock and Watson (2010), where BP denotes a band-pass filter. These are one-sided filters that are appropriate for out-of-sample forecasting inflation. In contrast, the last two are based on the Hodrick-Prescott (HP) filter which is a two-sided

¹²More specifically, Z_{t1} is the difference in the rates of inflation of the GDP deflator for non-petroleum imports and the overall GDP deflator, and Z_{t2} is the difference between PCE (all-items) inflation and PCE (less food and energy) with different effects pre- and post-1984:Q1.

filter with smoothing parameter λ . For quarterly data, a smoothing parameter of 1600 is quite standard. The larger smoothing parameter of 6400 is used in Gordon (2013). The last gap measure, U_t^* is assumed constant which is commonly done in practice. All six versions of X_t are persistent and the condition $\Gamma_X(1) \neq 0$ is easily satisfied. We allow the measurement error $\epsilon_t = X_t - x_t$ to be possibly serially correlated. Interestingly, the six gap variables are not particularly strongly correlated with each other. Though $X_t(4)$ and $X_t(5)$ have a correlation of 0.98, $X_t(2)$ and $X_t(6)$ only have a correlation of 0.34. Most of the remaining correlations are between 0.5 and 0.8.

The CMD estimator is feasible if the binding function can be explicitly derived. The current ADLZ(p,q,s) model has additional predictors z_t , making the CMD estimator difficult to implement. This is precisely the situation when the SMD estimator is appealing because it can easily accommodate changes to the model. Here, we treat z_t as exogenous in Algorithm SMD-EIV with $S = 100$. Most authors choose $q \geq 2$. Because measurement errors can induce serial correlation in the residuals, more lags than necessary might have been included. When we explicitly allow for measurement error, we find $q = 1$ to be appropriate. The traditional Phillips curve is thus estimating using the ADLZ(1,1,1) model and $\bar{\beta} = \frac{\beta_0 + \beta_1}{1 - \alpha_1}$. Confidence intervals for $\bar{\beta}$ are obtained by inverting the distance metric test as discussed above. The $100(1 - \eta)\%$ confidence interval for $\bar{\beta}$ is given by the set of values satisfying $DM \leq q_{1-\eta,1}$. The endpoints of the confidence interval (CI) are obtained as

$$\begin{aligned}\bar{\beta}_L &= \inf\{\bar{\beta} \in \Theta^C : \Pr(DM \leq q_{1-\eta,1} \mid H_0) \geq 1 - \eta\}, \\ \bar{\beta}_U &= \sup\{\bar{\beta} \in \Theta^C : \Pr(DM \leq q_{1-\eta,1} \mid H_0) \geq 1 - \eta\}.\end{aligned}$$

Table 5 reports the OLS and SMD estimates of the long-run effect of unemployment gap on inflation, $\hat{\bar{\beta}}$ and $\hat{\sigma}_\epsilon^2$. The point estimates of $\bar{\beta}$ are quite similar and for most models and estimators, the long-run estimate is statistically insignificant. The main differences between the OLS and SMD estimators of $\bar{\beta}$ lie in their assessment of the sampling uncertainty around $\hat{\bar{\beta}}$. Consistent with our simulation results, explicitly recognizing the presence of measurement error produces more imprecise estimates of $\bar{\beta}$ and wider confidence intervals. For example, while the OLS estimator in the model with $X_t(6)$ indicates that $\hat{\bar{\beta}}$ is statistically different than zero at 10% significance level, allowing for measurement error via our proposed approach renders the SMD estimate insignificant. The measurement error ranges from 3-4% for $X_t(6)$ and $X_t(1)$ to 37% for $X_t(2)$ suggesting that some specifications of the Phillips curve are characterized by a signal-to-noise ratio which is large enough to provide reliable empirical guidance for policy analysis. Similar results are obtained using output gap instead of unemployment gap but the estimated magnitude of the measurement error in these models tends to be larger.

6 Conclusion

This paper makes two contributions. First, we show that several biased estimates can jointly identify a model with mismeasured regressors without the need for external instruments. The key is to exploit persistence in the data. Second, we develop a simulation algorithm for situations where the regressors are not weakly exogenous for the parameters of interest. The algorithm can be used in time series and panel data regressions and can accommodate additional regressors. The proposed methodology is especially useful when external instruments are either unavailable or are weak. This can be potentially useful in endogeneity problems.

Table 1: ADL(0,0): $(\alpha, \beta_0, \beta_1) = (0, 1, 0)$

$$\begin{aligned}\theta &= (\beta_0, \sigma_u^2, \sigma_\epsilon^2, \phi) \\ \hat{\psi} &= (\hat{\beta}, \hat{\Gamma}_{\hat{V}}(0), \hat{\Gamma}_{\hat{V}}(1), \hat{\Gamma}_{\hat{V}_X}(1, 0))' .\end{aligned}$$

		$\hat{\beta}_0$					SE($\hat{\beta}_0$)				
T	ρ_x	OLS	IDEAL	IV	CMD	SMD	OLS	IDEAL	IV	CMD	SMD
ϵ_t White Noise											
200	0.200	0.702	1.006	1.078	1.058	1.089	0.068	0.086	1.151	0.401	0.375
200	0.500	0.699	1.006	1.030	0.960	0.982	0.064	0.079	0.215	0.223	0.220
200	0.800	0.690	1.004	1.013	0.972	0.983	0.062	0.066	0.108	0.129	0.132
500	0.200	0.700	1.000	1.054	1.010	1.037	0.043	0.053	0.451	0.302	0.293
500	0.500	0.699	1.000	1.008	0.967	0.978	0.041	0.049	0.122	0.167	0.166
500	0.800	0.695	1.000	1.004	0.988	0.992	0.040	0.041	0.063	0.075	0.079
1000	0.200	0.699	1.000	1.016	0.975	0.993	0.030	0.037	0.259	0.242	0.238
1000	0.500	0.699	1.000	1.002	0.973	0.978	0.028	0.034	0.083	0.128	0.129
1000	0.800	0.698	1.000	1.001	0.993	0.994	0.028	0.028	0.043	0.050	0.052
ϵ_t MA(1)											
200	0.200	0.702	1.006	-3.044	1.122	1.123	0.069	0.087	246.882	0.416	0.397
200	0.500	0.700	1.007	1.093	1.004	1.003	0.065	0.083	9.395	0.187	0.189
200	0.800	0.690	1.006	1.031	1.002	1.001	0.064	0.075	0.171	0.100	0.102
500	0.200	0.700	1.000	0.815	1.048	1.042	0.043	0.054	49.865	0.292	0.282
500	0.500	0.699	1.001	1.062	0.998	0.994	0.042	0.052	1.902	0.119	0.120
500	0.800	0.695	1.001	1.009	1.000	0.997	0.042	0.046	0.093	0.061	0.061
1000	0.200	0.700	1.000	-1.57	1.008	0.997	0.030	0.038	154.439	0.206	0.206
1000	0.500	0.699	1.000	1.018	0.996	0.990	0.029	0.036	0.199	0.083	0.085
1000	0.800	0.698	1.000	1.004	0.999	0.996	0.028	0.032	0.061	0.042	0.043

Table 2: ADL(1,0):

$$\begin{aligned}\theta &= (\alpha, \beta_0, \beta_1, \sigma_u^2, \sigma_\epsilon^2, \phi) \\ \hat{\psi}(\theta) &= (\hat{\alpha}, \hat{\beta}_0, \hat{\beta}_1, \hat{\Gamma}_{\hat{v}}(0), \hat{\Gamma}_{\hat{v}}(1), \hat{\Gamma}_{\hat{v}_X}(1, 0))'\end{aligned}$$

		Estimates of $\beta_0 = 1$ and $\frac{\beta_0 + \beta_1}{1 - \alpha} = 2.5$						Standard Errors					
T	ρ_x	OLS		CMD		SMD		OLS		CMD		SMD	
ϵ_t white noise													
200	0.2	0.696	1.858	1.125	2.838	1.137	2.803	0.068	0.261	0.454	5.755	0.438	1.026
200	0.5	0.656	1.996	0.989	2.434	1.018	2.547	0.066	0.224	0.263	1.523	0.264	0.456
200	0.8	0.522	2.129	0.931	2.422	0.936	2.516	0.058	0.179	0.216	0.247	0.170	0.248
500	0.2	0.693	1.858	1.053	2.588	1.079	2.652	0.043	0.166	0.358	0.765	0.366	0.777
500	0.5	0.653	2.000	0.961	2.421	0.988	2.487	0.042	0.143	0.162	0.262	0.177	0.292
500	0.8	0.520	2.144	0.928	2.433	0.928	2.481	0.037	0.113	0.136	0.144	0.111	0.172
1000	0.2	0.693	1.857	1.004	2.486	1.039	2.561	0.030	0.116	0.255	0.531	0.292	0.603
1000	0.5	0.653	2.002	0.956	2.418	0.980	2.471	0.029	0.100	0.108	0.178	0.129	0.217
1000	0.8	0.520	2.151	0.933	2.440	0.931	2.481	0.026	0.077	0.090	0.097	0.078	0.105
ϵ_t MA(1)													
200	0.2	0.711	1.790	1.105	2.741	1.138	2.803	0.070	0.275	0.466	1.459	0.434	1.115
200	0.5	0.679	1.945	0.941	2.367	1.000	2.504	0.069	0.243	0.251	0.440	0.270	0.501
200	0.8	0.544	2.079	0.907	2.400	0.918	2.504	0.063	0.203	0.199	0.252	0.185	0.289
500	0.2	0.709	1.787	0.981	2.428	1.078	2.640	0.044	0.174	0.355	0.787	0.354	0.783
500	0.5	0.677	1.947	0.901	2.317	0.970	2.455	0.044	0.154	0.156	0.272	0.183	0.317
500	0.8	0.543	2.094	0.899	2.407	0.913	2.482	0.041	0.127	0.125	0.152	0.119	0.168
1000	0.2	0.709	1.785	0.893	2.237	1.030	2.537	0.031	0.122	0.228	0.485	0.276	0.583
1000	0.5	0.677	1.948	0.891	2.308	0.964	2.447	0.030	0.107	0.108	0.191	0.132	0.228
1000	0.8	0.543	2.102	0.900	2.412	0.915	2.482	0.028	0.087	0.085	0.105	0.083	0.114

Table 3: ADL(1,1)

$$\begin{aligned}\theta &= (\alpha, \beta_0, \beta_1, \sigma_u^2, \sigma_\epsilon^2, \phi) \\ \hat{\psi}(\theta) &= (\hat{\alpha}, \hat{\beta}_0, \hat{\beta}_1, \hat{\Gamma}_{\hat{v}}(0), \hat{\Gamma}_{\hat{v}}(1), \hat{\Gamma}_{\hat{v}_X}(1, 0))'\end{aligned}$$

		Estimates of $\beta_0 = 1$ and $\frac{\beta_0 + \beta_1}{1 - \alpha} = 3.75$						Standard Errors					
T	ρ_x	OLS		CMD		SMD		OLS		CMD		SMD	
ϵ_t white noise													
200	0.2	0.715	3.014	0.963	3.629	1.023	3.810	0.071	0.352	0.192	0.564	0.223	0.631
200	0.5	0.706	3.226	0.972	3.685	1.014	3.802	0.069	0.301	0.156	0.385	0.172	0.422
200	0.8	0.605	3.392	0.997	3.741	1.008	3.931	0.063	0.238	0.173	0.267	0.168	1.259
500	0.2	0.713	3.011	0.941	3.590	1.003	3.756	0.045	0.222	0.124	0.355	0.146	0.394
500	0.5	0.703	3.228	0.961	3.675	1.002	3.768	0.044	0.190	0.101	0.239	0.109	0.255
500	0.8	0.603	3.407	0.989	3.736	0.987	3.843	0.041	0.149	0.112	0.164	0.108	0.250
1000	0.2	0.713	3.010	0.936	3.583	1.001	3.747	0.031	0.155	0.085	0.246	0.100	0.273
1000	0.5	0.703	3.230	0.960	3.674	1.000	3.761	0.030	0.132	0.069	0.166	0.075	0.178
1000	0.8	0.603	3.415	0.989	3.735	0.984	3.812	0.028	0.101	0.078	0.113	0.075	0.168
ϵ_t MA(1)													
200	0.2	0.697	2.892	0.950	3.596	1.018	3.777	0.074	0.375	0.203	0.623	0.234	0.705
200	0.5	0.697	3.146	0.964	3.672	1.010	3.796	0.074	0.334	0.170	0.442	0.185	0.490
200	0.8	0.601	3.334	1.009	3.752	0.997	3.943	0.071	0.280	0.194	0.324	0.187	0.803
500	0.2	0.695	2.884	0.920	3.529	0.992	3.709	0.046	0.236	0.127	0.387	0.154	0.436
500	0.5	0.695	3.144	0.951	3.652	0.995	3.760	0.047	0.210	0.108	0.274	0.116	0.293
500	0.8	0.600	3.349	0.999	3.740	0.985	3.855	0.046	0.175	0.122	0.199	0.114	0.341
1000	0.2	0.694	2.883	0.912	3.513	0.987	3.700	0.032	0.164	0.088	0.269	0.108	0.303
1000	0.5	0.695	3.147	0.948	3.648	0.996	3.753	0.033	0.145	0.074	0.190	0.080	0.203
1000	0.8	0.600	3.357	0.999	3.738	0.985	3.810	0.032	0.119	0.084	0.137	0.080	0.198

Table 4: Panel Data, $N = 200$.

$$\begin{aligned} y_{it} &= \eta_i + \alpha y_{it-1} + \beta_0 X_{it} + \beta_1 X_{it-1} + u_{it} \\ X_{it} &= x_{it} + \epsilon_{it}. \end{aligned}$$

		LSDV-x		LSDV-X		SMD	
		Est.	S.E.	Est.	S.E.	Est.	S.E.
$T = 10, \text{PADL}(1,0)$							
β_0	1.000	1.005	0.022	0.658	0.023	1.063	0.273
α_1	0.800	0.722	0.059	0.720	0.066	0.750	0.114
σ_u^2	1.000	1.172	0.511	1.196	0.406	0.863	0.337
σ_ϵ^2	0.756	-	-	-	-	0.534	0.318
$T = 20, \text{PADL}(1,0)$							
β_0	1.000	1.013	0.019	0.684	0.017	1.055	0.236
α_1	0.800	0.765	0.034	0.784	0.034	0.770	0.059
σ_u^2	1.000	1.327	0.392	1.376	0.311	0.937	0.248
σ_ϵ^2	0.756	-	-	-	-	0.555	0.268
$T = 10, \text{PADL}(1,1)$							
β_0	1.000	0.965	0.038	0.679	0.029	0.994	0.241
β_1	0.600	0.669	0.062	0.514	0.040	0.568	0.113
α_1	0.800	0.744	0.046	0.751	0.047	0.802	0.067
σ_u^2	1.000	0.945	0.428	1.643	0.402	0.931	0.391
σ_ϵ^2	0.756	-	-	-	-	0.466	0.333
$T = 20, \text{PADL}(1,1)$							
β_0	1.000	0.990	0.019	0.714	0.017	1.037	0.189
β_1	0.600	0.637	0.039	0.500	0.026	0.600	0.104
α_1	0.800	0.778	0.021	0.800	0.020	0.797	0.028
σ_u^2	1.000	1.232	0.373	2.621	0.442	0.973	0.274
σ_ϵ^2	0.756	-	-	-	-	0.579	0.283

Table 5: OLS and SMD Estimates of the traditional Phillips Curve.

	OLS		SMD			$\widehat{\Gamma}_X(0)$
	$\widehat{\bar{\beta}}$	90% CI of $\bar{\beta}$	$\widehat{\bar{\beta}}$	90% CI of $\bar{\beta}$	$\widehat{\sigma}_\epsilon^2$	
$X_t(1)$	-0.028	[-0.077, 0.022]	-0.025	[-0.225, 0.139]	0.071	1.727
$X_t(2)$	-0.049	[-0.129, 0.032]	-0.052	[-0.270, 0.072]	0.421	1.128
$X_t(3)$	-0.067	[-0.140, 0.006]	-0.081	[-0.217, 0.056]	0.104	0.802
$X_t(4)$	-0.017	[-0.111, 0.078]	-0.021	[-0.252, 0.184]	0.081	0.587
$X_t(5)$	-0.024	[-0.098, 0.051]	-0.032	[-0.239, 0.145]	0.214	0.857
$X_t(6)$	-0.040	[-0.078, -0.002]	-0.044	[-0.199, 0.079]	0.073	2.583

A Appendix: Proofs

Proof of Lemma 1 Note first that since $\Gamma_X(0) - \sigma_\epsilon^2 = \Gamma_x(0) > 0$, the equation for ψ_{A1} implies that $\beta = 0$ if and only if $\psi_{A1} = 0$. Thus, $\beta = 0$ is directly identified off of ψ_{A1} . Moreover, we then have $\sigma_u^2 = \psi_{A2}$. Thus $(\beta = 0, \sigma_u^2)$ is identified. For $\beta \neq 0$, we have $\text{sgn}(\beta) = \text{sgn}(\psi_{A1})$. If $\Gamma_X(1) \neq 0$, then from the equation for ψ_{A3} ,

$$\beta \frac{\sigma_\epsilon^2}{\Gamma_X(0)} = \text{sgn}(\psi_{A1}) \sqrt{\frac{\psi_{A3}}{\Gamma_X(1)}}. \quad (21)$$

Plugging into the solution for ψ_{A1} and noticing that $\psi_{A1} = \text{sgn}(\psi_{A1}) |\psi_{A1}|$ then gives

$$\beta = \text{sgn}(\psi_{A1}) \left[|\psi_{A1}| + \sqrt{\frac{\psi_{A3}}{\Gamma_X(1)}} \right].$$

For σ_u^2 , combine (21) with $\psi_{A2} = \sigma_u^2 + \beta \sigma_\epsilon^2 \psi_{A1}$ to obtain

$$\sigma_u^2 = \psi_{A2} - |\psi_{A1}| \Gamma_X(0) \sqrt{\frac{\psi_{A3}}{\Gamma_X(1)}}.$$

Thus, serial correlation in the regressor X is sufficient to globally identify $(\beta \neq 0, \sigma_u^2)$. Finally, when $\beta \neq 0$,

$$\sigma_\epsilon^2 = \frac{\Gamma_X(0)}{\beta} \text{sgn}(\psi_{A1}) \sqrt{\frac{\psi_{A3}}{\Gamma_X(1)}}$$

where β is defined above. The expression for σ_u^2 follows directly from ψ_{A2} . Thus, $\Gamma_X(1) \neq 0$ and $\beta \neq 0$ are sufficient for identification of θ_A . To show necessity, suppose $\Gamma_X(1) = 0$. For $\kappa \in (1 - \delta, 1)$ and $\delta > 0$, let

$$\begin{aligned} \bar{\beta} &\equiv \kappa \beta \\ \bar{\sigma}_\epsilon^2 &\equiv \frac{1}{\kappa} \sigma_\epsilon^2 + \left(1 - \frac{1}{\kappa}\right) \Gamma_X(0) \\ \bar{\sigma}_u^2 &\equiv \sigma_u^2 + \beta^2 (1 - \kappa) (\Gamma_X(0) - \sigma_\epsilon^2). \end{aligned}$$

Clearly, $\theta_A = (\beta, \sigma_\epsilon^2, \sigma_u^2)$ and $\bar{\theta}_A = (\bar{\beta}, \bar{\sigma}_\epsilon^2, \bar{\sigma}_u^2)$ are observationally equivalent since $\psi_A(\theta_A) = \psi_A(\bar{\theta}_A)$. It remains to show that $\beta \neq 0$ is also necessary for identification. This is immediate because if $\beta = 0$, $\psi_A(\theta_A) = \psi_A(\tilde{\theta}_A)$ with $\theta_A = (0, \sigma_\epsilon^2, \sigma_u^2)$ and $\tilde{\theta}_A = (0, \tilde{\sigma}_\epsilon^2, \sigma_u^2)$ for any values of σ_ϵ^2 and $\tilde{\sigma}_\epsilon^2$.

Proof of Lemma 2 Write the binding function as:

$$\psi_B(\theta_B) = \begin{pmatrix} \beta \left(1 - \frac{\Gamma_\epsilon(0)}{\Gamma_X(0)}\right) \\ \beta^2 \Gamma_\epsilon(0) \left(1 - \frac{\Gamma_\epsilon(0)}{\Gamma_X(0)}\right) + \sigma_u^2 \\ \beta^2 \left[\Gamma_\epsilon(1) - 2\Gamma_\epsilon(1) \frac{\Gamma_\epsilon(0)}{\Gamma_X(0)} + \left(\frac{\Gamma_\epsilon(0)}{\Gamma_X(0)}\right)^2 \Gamma_X(1) \right] \\ -\beta \left(\Gamma_\epsilon(1) - \Gamma_\epsilon(0) \frac{\Gamma_X(1)}{\Gamma_X(0)} \right) \end{pmatrix}.$$

First, consider the case $\beta = 0$. Note that $\psi_{B1} = \beta \frac{\Gamma_x(0)}{\Gamma_X(0)} = 0$. But $\Gamma_X(0) - \Gamma_\epsilon(0) = \Gamma_x(0) \neq 0$. Hence, $\beta = 0$ if and only if $\psi_{B1} = 0$, and $\beta = 0$ is directly identified off of ψ_{B1} . For σ_u^2 , we have $\sigma_u^2 = \psi_{B2}$, so $(\beta = 0, \sigma_u^2)$ is identified from ψ_B . Next, we consider the case $\beta \neq 0$. In this case $\psi_{B1} \neq 0$, and we can solve for β by considering

$$A \equiv \Gamma_X(1)\psi_{B1}^2 + 2\psi_{B4}\psi_{B1} + \psi_{B3}.$$

Using the definition of ψ_B , this quantity can be computed in two ways: $A = \beta^2(\Gamma_X(1) - \Gamma_\epsilon(1)) = \beta^2\Gamma_x(1)$ and $A = \beta(\psi_{B4} + \Gamma_X(1)\psi_{B1})$. So if $\Gamma_x(1) \neq 0$, then $A \neq 0$ and we use the two expressions for A to obtain:

$$\beta = \frac{A}{\psi_{B4} + \Gamma_X(1)\psi_{B1}}.$$

For σ_u^2 , consider

$$D \equiv \psi_{B2}\psi_{B4} - \Gamma_X(0)\psi_{B1}\psi_{B3} + \Gamma_X(1)\psi_{B1}\psi_{B2} - \Gamma_X(0)\psi_{B1}^2\psi_{B4}.$$

Then, $D = \sigma_u^2(\psi_{B4} + \Gamma_X(1)\psi_{B1})$. Dividing both sides by $\psi_{B4} + \Gamma_X(1)\psi_{B1} \neq 0$ gives

$$\sigma_u^2 = \frac{D}{\psi_{B4} + \Gamma_X(1)\psi_{B1}}.$$

Thus $\Gamma_x(1) \neq 0$ is sufficient to globally identify $(\beta \neq 0, \sigma_u^2)$. Finally, to identify $\Gamma_\epsilon(0)$, assume $\Gamma_x(1) \neq 0$, and $\beta \neq 0$. Consider

$$B \equiv \Gamma_X(0)(\psi_{B3} + \psi_{B1}\psi_{B4}),$$

and note that $B = A\Gamma_\epsilon(0)$. Since $A \neq 0$ under our assumptions,

$$\Gamma_\epsilon(0) = \frac{B}{A} = \Gamma_X(0) \frac{\psi_{B3} + \psi_{B1}\psi_{B4}}{\Gamma_X(1)\psi_{B1}^2 + 2\psi_{B4}\psi_{B1} + \psi_{B3}}.$$

Finally, for $\Gamma_\epsilon(1)$, let $C \equiv -\psi_{B4}^2 + \Gamma_X(1)\psi_{B3}$, and note that $C = \Gamma_\epsilon(1)A$. Under our assumptions, $A \neq 0$ and $\Gamma_\epsilon(1)$ is identified as

$$\Gamma_\epsilon(1) = \frac{C}{A} = \frac{-\psi_{B4}^2 + \Gamma_X(1)\psi_{B3}}{\Gamma_X(1)\psi_{B1}^2 + 2\psi_{B4}\psi_{B1} + \psi_{B3}}.$$

Proof of Lemma 3 The proof is as follows. First, we consider the case when $\beta = 0$. Note that

$$1 - \frac{\Gamma_\epsilon(0)\Gamma_y(0)}{\Gamma_y(0)\Gamma_X(0) - \Gamma_{yX}(1,0)^2} = \frac{\Gamma_y(0)\Gamma_x(0) - \Gamma_{yX}(1,0)^2}{\Gamma_y(0)\Gamma_X(0) - \Gamma_{yX}(1,0)^2}$$

and since $\Gamma_{yX}(1,0) = \Gamma_{yx}(1,0)$ both the numerator and the denominator are determinants of positive definite covariance matrices, thus the above quantity is strictly positive. Thus, $\psi_2 = 0$ if and only if $\beta = 0$. In this case $\alpha = \psi_1$, so $(\alpha, \beta = 0)$ is identified.

Next, consider the case when $\beta \neq 0$. There are again two cases to consider: $\Gamma_{yX}(1, 0) = 0$ and $\Gamma_{yX}(1, 0) \neq 0$. Consider $\Gamma_{yX}(1, 0) = 0$ first. Since $X_t = x_t + \epsilon_t$ and $y_t = \sum_{i=0}^{\infty} \alpha^i (\beta x_{t-i} + u_{t-i})$, we have $E[X_t y_{t-j}] = \sum_{i=0}^{\infty} \alpha^i (\beta \Gamma_x(j+i))$. In particular,

$$\Gamma_{yX}(1, 0) = \beta \left[\Gamma_x(1) + \sum_{i=1}^{\infty} \alpha^i \Gamma_x(1+i) \right], \quad (22)$$

so $\Gamma_{yX}(1, 0)$ occurs, for example, whenever x is white noise. In this case, α is directly identified off ψ_1 , $\alpha = \psi_1$. As for β , notice that the components $\psi_{C2}^-(\theta_C^-), \psi_{C3}^-(\theta_C^-), \psi_{C4}^-(\theta_C^-)$ are as in the ADL(0,0) case and identification can proceed as in Lemma 2.

It remains to consider the case $\beta \neq 0$, $\Gamma_{yX}(1, 0) \neq 0$. For this, we further expand the elements on the left-hand side of (9). We then obtain that $\psi_{\bar{C}}^-(\theta_{\bar{C}}^-) = \psi_{\bar{C}}^-$ has a unique solution in $\theta_{\bar{C}}^-$ if and only if the system:

$$\begin{aligned} \alpha \Gamma_y(0) + \beta \Gamma_{yX}(1, 0) &= \pi_1 \\ \alpha \Gamma_{yX}(1, 0) + \beta (\Gamma_X(0) - \Gamma_{\epsilon}(0)) &= \pi_2 \\ \left(\begin{array}{c} \alpha \Gamma_y(0) + \beta \Gamma_{yX}(1, 0) \\ \alpha \Gamma_{yX}(1, 0) + \beta (\Gamma_X(0) - \Gamma_{\epsilon}(0)) \end{array} \right)' \Gamma_W(0)^{-1} \left(\begin{array}{c} \alpha \Gamma_y(1) + \beta \Gamma_{yX}(2, 0) \\ \alpha \Gamma_{yX}(0, 0) + \beta (\Gamma_X(1) - \Gamma_{\epsilon}(1)) \end{array} \right) &= \pi_3 \\ \alpha \Gamma_{yX}(2, 0) + \beta (\Gamma_X(1) - \Gamma_{\epsilon}(1)) &= \pi_4 \end{aligned}$$

has a unique solution in θ , for given $\pi = (\pi_1, \dots, \pi_4)$ and $\psi_{\bar{C}}^-$. Of course, the π 's in the above equations are functions of the original $\psi_{\bar{C}}^-$'s: they correspond to the right-hand side of (9). The above equations have a unique solution in $\theta_{\bar{C}}^-$ as a function of π . Replacing π 's with the values on the right-hand side of (9) shows that the solution for α is:

$$\alpha = \psi_{C1}^- + \frac{\Gamma_{yX}(1, 0)(\psi_{C3}^- + \psi_{C2}^- \psi_{C4}^-)}{(\Gamma_y(0)\Gamma_{yX}(2, 0) - \Gamma_y(1)\Gamma_{yX}(1, 0))\psi_{C1}^- - \Gamma_{yX}(1, 0)(\Gamma_{yX}(0, 0) - \Gamma_{yX}(2, 0))\psi_{C2}^-}.$$

Of course, for the solution to be valid we need to check that the denominator is not zero. For this, write the equality above as:

$$\alpha - \psi_{C1}^- = \frac{N}{D},$$

with

$$\begin{aligned} N &= \Gamma_{yX}(1, 0)(\psi_{C3}^- + \psi_{C2}^- \psi_{C4}^-) \\ D &= (\Gamma_y(0)\Gamma_{yX}(2, 0) - \Gamma_y(1)\Gamma_{yX}(1, 0))\psi_{C1}^- - \Gamma_{yX}(1, 0)(\Gamma_{yX}(0, 0) - \Gamma_{yX}(2, 0))\psi_{C2}^-. \end{aligned}$$

Note that

$$\alpha - \psi_{C1}^- = -\beta \Gamma_{\epsilon}(0) \frac{\Gamma_{yX}(1, 0)}{\Gamma_y(0)\Gamma_X(0) - \Gamma_{yX}(1, 0)^2} \neq 0.$$

Thus, $D = 0$ if and only if $\psi_{C3}^- + \psi_{C2}^- \psi_{C4}^- = 0$. For β , the solution is:

$$\begin{aligned}\beta &= \psi_{C2}^- - \frac{\Gamma_y(0)(\psi_{C3}^- + \psi_{C2}^- \psi_{C4}^-)}{(\Gamma_y(0)\Gamma_{yX}(2,0) - \Gamma_y(1)\Gamma_{yX}(1,0))\psi_{C1}^- - \Gamma_{yX}(1,0)(\Gamma_{yX}(0,0) - \Gamma_{yX}(2,0))\psi_{C2}^-} \\ &= \psi_{C2}^- - \frac{\Gamma_y(0)}{\Gamma_{yX}(1,0)}(\alpha - \psi_{C1}^-).\end{aligned}$$

Thus $(\alpha, \beta \neq 0)$ are identified from ψ_C^- if $\Gamma_x(1) \neq 0$ and $\psi_{C3}^- + \psi_{C2}^- \psi_{C4}^- \neq 0$. Finally, for $\Gamma_\epsilon(0)$ we have:

$$\Gamma_\epsilon(0)\beta = \frac{(\psi_{C3}^- + \psi_{C2}^- \psi_{C4}^-)(\Gamma_{yX}(1,0)^2 - \Gamma_X(0)\Gamma_y(0))}{D},$$

so if in addition $\beta \neq 0$, $\Gamma_\epsilon(0)$ is identified. Similarly, $\Gamma_\epsilon(1)$ is then also identified from ψ_{C4}^- .

Proof of Lemma 4 We have

$$\Gamma_{\widetilde{W}}(0) = \begin{pmatrix} \Gamma_{\widetilde{y}_{-1}}(0) & \Gamma_{\widetilde{X}\widetilde{y}_{-1}}(0,0) \\ \Gamma_{\widetilde{X}\widetilde{y}_{-1}}(0,0) & \Gamma_{\widetilde{X}}(0) \end{pmatrix},$$

which is assumed to be non-singular. Since $\Gamma_{\widetilde{W}}(0)$ and $\Gamma_{\widetilde{W}}(1)$ do not depend on θ_P , the system of equations $\psi_P = \psi_P(\theta_P)$ is equivalent to $\Psi_P = \Psi_P(\theta_P)$, where

$$\begin{aligned}\Psi_{P1}(\theta) &= \Gamma_{\widetilde{W}\widetilde{y}}(0,0) = \Gamma_{\widetilde{W}}(0) \begin{pmatrix} \psi_{P1}(\theta) \\ \psi_{P2}(\theta) \end{pmatrix} = \begin{pmatrix} \alpha\Gamma_{\widetilde{y}_{-1}}(0) + \beta\Gamma_{\widetilde{X}\widetilde{y}_{-1}}(0,0) - \sigma_u^2 h_T(\alpha) \\ \alpha\Gamma_{\widetilde{X}\widetilde{y}_{-1}}(0,0) + \beta\Gamma_{\widetilde{X}}(0) - \beta\sigma_\epsilon^2 \end{pmatrix}, \\ \Psi_{P2}(\theta) &= \Gamma_{\widetilde{y}}(0) = \psi_{P3}(\theta) + (\psi_{P1}(\theta) \quad \psi_{P2}(\theta)) \Gamma_{\widetilde{W}}(0) \begin{pmatrix} \psi_{P1}(\theta) \\ \psi_{P2}(\theta) \end{pmatrix} \\ &= \alpha^2\Gamma_{\widetilde{y}_{-1}}(0) + 2\alpha\beta\Gamma_{\widetilde{X}\widetilde{y}_{-1}}(0,0) + \beta^2\Gamma_{\widetilde{X}}(0) + \sigma_u^2 + \beta^2\sigma_\epsilon^2 - 2\sigma_u^2 h_T(\alpha) - 2\beta\sigma_\epsilon^2,\end{aligned}$$

and

$$\begin{aligned}\Psi_{P3}(\theta) &= \Gamma_{\widetilde{W}\widetilde{y}}(0,0)' \Gamma_{\widetilde{W}}(0)^{-1} \Gamma_{\widetilde{W}\widetilde{y}}(1,0) = -\psi_{P4} + (\psi_{P1}(\theta) \quad \psi_{P2}(\theta)) \Gamma_{\widetilde{W}}(1) \begin{pmatrix} \psi_{P1}(\theta) \\ \psi_{P2}(\theta) \end{pmatrix} \\ &= \begin{pmatrix} \alpha\Gamma_{\widetilde{y}_{-1}}(1) + \beta\Gamma_{\widetilde{X}\widetilde{y}_{-1}}(0,1) - \sigma_u^2 [(T+1)h_T(\alpha) - 1] \\ \alpha\Gamma_{\widetilde{X}\widetilde{y}_{-1}}(1,0) + \beta\Gamma_{\widetilde{X}}(1) + \beta\frac{1}{T}\sigma_\epsilon^2 \end{pmatrix}.\end{aligned}$$

Only the right-hand side terms depend on θ_P .

References

- AIGNER, D., C. HSIAO, A. KAPTEYN, AND T. WANSBEEK (1984): “Latent Variable Models in Econometrics,” in *Handbook of Econometrics*, vol. 2. North Holland.
- ARUOBA, B., F. DIEBOLD, J. NALEWAIK, F. SCHORFHEIDE, AND D. SONG (2013): “Improving GDP Measurement: A Measurement Error Perspective,” FRB Philadelphia Working Paper 13-16.
- BAI, J., AND S. NG (2010): “Instrumental Variables Estimation in a Data Rich Environment,” *Econometric Theory*, 26:6, 1607–1637.
- BIORN, E. (1992): “The Bias of Some Estimators for Panel Models with Measurement Errors,” *Empirical Economics*, 17, 51–66.
- (2008): “Measurement Error in a Dynamic Panel Data Analysis: A Synthesis on Modeling and GMM Estimation,” University of Oslo, mimeo.
- BUN, M. J., AND M. A. CARREE (2005): “Bias-Corrected Estimation in Dynamic Panel Data Models,” *Journal of Business and Economic Statistics*, 23, 200–210.
- CZELLAR, V., AND E. ZIVOT (2008): “Improved Small Sample Inference for Efficient Method of Moments and Indirect Inference Estimators,” University of Washington.
- DAGENAIS, M., AND D. DAGENAIS (1997): “Higher Moment Estimators for Linear Regression Models with Errors in Variables,” *Journal of Econometrics*, 76(1-2), 193–221.
- DUFFIE, D., AND K. SINGLETON (1993): “SIMulated Moments Estimation of Markov Models of Asset Prices,” *Econometrica*, 61, 929–952.
- EBERLY, J., S. REBELO, AND N. VINCENT (2009): “Investment and Value: a Neoclassical Benchmark,” CIRPEE Working paper 09-08.
- ENGLE, R., D. HENDRY, AND J. RICHARD (1983): “Exogeneity,” *Econometrica*, 51:2, 277–304.
- ERICKSON, T. (2001): “Constructing Instruments for Regressions with Measurement Error when no Additional Data are Available,” *Econometrica*, 69(1), 221–222.
- ERICKSON, T., C. JIANG, AND T. WHITED (2014): “Minimum Distance Estimation of the Errors using Linear Cumulant Equations,” Working Paper.
- ERICKSON, T., AND T. WHITED (2000): “Measurement Error and the Relationship between Investment and q ,” *Journal of Political Economy*, 108, 1027–57.
- (2002): “Two-Step GMM Estimation of the Errors-in-Variables Model using Higher Order Moments,” *Econometric Theory*, 18, 776–799.
- (2012): “Treating Measurement Error in Tobin’s q ,” *Review of Financial Studies*, 25(4), 1286–1329, SSRN working paper.
- ERMINI, L. (1993): “Effects of Transitory Consumption and Temporal Aggregation on the Permanent Income Hypothesis,” *Review of Economics and Statistics*, 75:5, 736–74.
- FALK, B., AND B. LEE (1990): “Time Series Implications of Friedman’s Permanent Income Hypothesis,” *Journal of Monetary Economics*, 26:2, 267–283.

- FENG, S., AND Y. HU (2013): “Misclassification Errors and the Underestimation of U.S. Unemployment Rates,” *American Economic Review*, 103(2), 1054–70.
- GALLANT, R., AND G. TAUCHEN (1996): “Which Moments to Match,” *Econometric Theory*, 12, 657–681.
- GILLARD, J. (2010): “An Overview of Linear Structural Models in Errors in Variable Regression,” *Statistical Journal*, 8(1), 57–80.
- GOLDBERGER, A. (1972): “Structural Equation Methods in the Social Sciences,” *Econometrica*, 40(6), 979–1001.
- GORDON, R. J. (2013): “The Phillips Curve is Alive and Well: Inflation and the NAIRU During the Slow Recovery,” NBER Working Paper 19390.
- GOSPODINOV, N., AND S. NG (2014): “Minimum Distance Estimation of Possibly Non-Invertible Moving-Average Models,” Working Paper.
- GOURIEROUX, C., A. MONFORT, AND E. RENAULT (1993): “Indirect Inference,” *Journal of Applied Econometrics*, 85, 85–118.
- GOURIEROUX, G., P. PHILLIPS, AND Y. YU (2010): “Indirect Inference of Dynamic Panel Models,” *Journal of Econometrics*, 157(1), 68–77.
- HAHN, J., AND G. KUERSTEINER (2002): “Asymptotically Unbiased Inference for a Dynamic Model with Fixed Effects when both n and T are large,” *Econometrica*, 70, 1639–1657.
- HSIAO, C. (1977): “Identification For a Linear Dynamic SIMultaneous Error-Shock Model,” *International Economic Review*, 18(1), 181–194.
- JIANG, W., AND B. TURNBULL (2004): “The Indirect Method: Inference Based on Intermediate Statistics: A Synthesis and Examples,” *Statistical Science*, 19:2, 239–263.
- KIVIET, J. (1995): “On Bias, Inconsistency and Efficiency of Various Estimators in Dynamic Panel Data Models,” *Journal of Econometrics*, 68, 53–78.
- KOMUNJER, I., AND S. NG (2014): “Measurement Errors in Dynamic Models,” *Econometric Theory*, 30, 150–175.
- LEWBEL, A. (1997): “Constructing Instruments for Regression with Measurement Error When No Additional Data are Available, with an Application to Patents and R & D,” *Econometrica*, 65, 1201–1213.
- LEWBEL, A. (2012): “Using Heteroskedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models,” *Journal of Business and Economic Statistics*, 30, 67–80.
- MEIJER, E., L. SPIERDIJK, AND T. WANSBEEK (2012): “Consistent Estimation of Linear Panel Data Models with Measurement Error,” Working Paper.
- NALEWALK, J. (2010): “The Income- and Expenditure-Side Measures of Output Growth,” *Brookings Papers for Economic Activity*, 1, 71–106.
- NEWKEY, W., AND D. MCFADDEN (1994): “Large Sample Estimation and Hypothesis Testing,” in *Handbook of Econometrics*, vol. 36:4, pp. 2111–2234. North Holland.
- NICKELL, S. (1981): “Biases in Dynamic Models with Fixed Effects,” *Econometrica*, 49, 1417–1426.

- ORPHANIDES, A., AND S. VAN NORDEN (2002): “The Unreliability of Output Gap Estimates in Real Time,” *Review of Economics and Statistics*, 84(4), 569–83.
- ORPHANIDES, A., AND J. C. WILLIAMS (2002): “Robust Monetary Rules with Unknown Natural Rates,” *Brookings Paper on Economic Activity*, 2, 63–118.
- PAL, M. (1980): “Consistent Moment Estimators of Regression Coefficients in the Presence of Errors in Variables,” *Journal of Econometrics*, 14, 349–364.
- PHILLIPS, P., AND D. SUL (2007): “Bias in Dynamic Panel Estimation with Fixed Effects, Incidental Trends and Cross-Section Dependence,” *Journal of Econometrics*, 137, 162–188.
- REIERSØL, O. (1950): “Identifiability of a Linear Relation Between Variables Which Are Subject to Error,” *Econometrica*, 23, 375–389.
- SARGENT, T. (1989): “Two Models of Measurements and the Investment Accelerator,” *Journal of Political Economy*, 97:2.
- SCHENNACH, S., AND Y. HU (2013): “Nonparametric Identification and Semiparametric Estimation of Classical Measurement Error Models Without Using Side Information,” *Journal of American Statistical Association*, 108(501), 177–186.
- SMITH, A. (1993): “Estimating Nonlinear Time Series Models Using Simulated Vector Autoregressions,” *Journal of Applied Econometrics*, 8:S1, S63–S84.
- SPIERDIJK, L., AND T. WANSBEEK (2011): “A Comprehensive Approach to Measurement Error in Panel Data Models,” University of Groningen, mimeo.
- STOCK, J., AND M. W. WATSON (2010): “Modeling Inflation After the Crisis,” NBER Working Paper 16488.
- WANSBEEK, T., AND R. KONING (1991): “Measurement Error and Panel Data,” *Statistica Neerlandica*, 45, 85–92.
- WATSON, M. W. (2014): “Inflation Persistence, the NAIRU, and the Great Recession,” *American Economic Review*, 104(5), 31–36.
- WILCOX, D. (1992): “The Construction of U.S. Consumption Data: Some Facts and Their Implications for Empirical Work,” *American Economic Review*, 82:4, 992–941.