Simulated Minimum Distance Estimation of Dynamic Models with Errors-in-Variables

Nikolay Gospodinov* Ivana Komunjer† Serena Ng‡

April 24, 2015

Abstract

Empirical analysis often involves using inexact measures of the predictors suggested by economic theory. The bias created by the correlation between the mismeasured regressors and the error term motivates the need for instrumental variable estimation. This paper considers a class of estimators that can be used in dynamic models with measurement errors 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 the latter are not analytically tractable, a special algorithm is designed to simulate the model without completely specifying the processes that generate the latent predictors. The proposed estimators perform well in simulations of the autoregressive distributed lag model. The methodology is used to estimate the long run risk model.

JEL Classification: C1, C3

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

*Federal Reserve Bank of Atlanta, 1000 Peachtree Street, N.E., Atlanta, GA 30309. 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 would like to thank the Editor, Tom Wansbeek, and two anonymous referees for their suggestions and comments which greatly improved an earlier version of the paper. Financial support from the National Science Foundation grants SES-0962473 and SES-0962431 is gratefully acknowledged. The views expressed here are the authors' and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

1 Introduction

Empirical analysis often involves using incorrectly measured data which complicates identification of the behavioral parameters and testing of economic hypothesis.[1] The problem is acute in cross-section and survey data where errors in data collection and reporting are inevitable, and this is still an active area of research. 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 use 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 series shows a steeper downturn in 2007-2009 than the GDP series. 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 parameters of an accelerator model is affected by how the data are reported.

This paper is concerned with estimation of autoregressive distributed lag models (hereafter, ADL(p,q)) when the regressors are measured with errors that are possibly serially correlated, making it difficult to find valid instruments.[2] An early account of the problem can be found in Grether and Maddala (1973); Buonaccorsi (2012) provides a recent survey of the literature. As is well known, identification in distributed lag models is impossible without further assumptions when the predictors are serially uncorrelated or normally distributed.[3] But as Goldberger (1972, p.996) pointed out, identification is still possible in the presence of measurement errors. The instrumental variable (IV) approach uses additional information from two mismeasured indicators of the latent regressor: one to replace the latent regressor and a second to instrument the first. The case of many mismeasured indicators is studied in Bai and Ng (2010). A second approach is to drop the

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

[2] The potential of instrumental variable estimators in time series regressions with serially correlated measurement errors is studied in Biørn (2014).

[3] See Maravall (1979), Wansbeek and Meijer (2000), and Aigner, Hsiao, Kapteyn, and Wansbeek (1984) for identification conditions in measurement error models. Gillard (2010) present an overview of approaches to handle the errors-in-variables (EIV) problem from different fields.

normality assumption; see, e.g., Reiersøl (1950). Pal (1980), Dagenais and Dagenais (1997), Lewbel (1997) and Meijer, Spierdijk, and Wansbeek (2012) exploit heteroskedasticity, skewness and excess kurtosis for identification without relying on instruments. Our approach falls in the third category along the lines of Grilliches and Hausman (1986) and Biørn (1992) for panel data: we combine the information from several biased estimators to identify and estimate the parameters of the model. A novelty of our approach is the use of simulations to map out the possibly non-tractable relation between the unknown parameters and the biases induced.

We consider the case when no external instruments are available. Our point of departure is that provided the regressors are serially correlated, the ordinary least squares (OLS) residuals will be serially correlated. There is in general enough information in the OLS estimator and the least squares residuals to permit identification of the parameters of interest. In a way, our approach is to combine information in these sample estimates, or auxiliary statistics, whose bias is magnified by the persistence of the regressors. Identifying the model parameters is then possible provided the probability limit of the auxiliary statistics, or binding function, is invertible.

In simple models, where the binding function can be derived analytically, the classical minimum distance (CMD) estimator has standard properties. This 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.[4] In more complex models where the binding function is not analytically tractable, we use Monte-Carlo methods to approximate this mapping. However, our simulated minimum distance (SMD) estimator differs from the ones considered in Smith (1993), Gourieroux, Monfort, and Renault (1993), and Gallant and Tauchen (1996). These estimators treat the predictors as exogenous and hold them 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 distribution of the dependent variable given the covariates are not variation free in the sense of Engle, Hendry, and Richard (1983). Thus, even though the correctly measured predictors can be held fixed, the mismeasured ones cannot.

The construction of SMD estimators in models with endogenous variables is far from being trivial. For instance, Gourieroux, Monfort, and Renault (1993) point out that "models in which non-strongly exogenous variables appear have the serious drawback of not being simulable." While this is true in general, we capitalize on the fact that in linear models, processes with identical covariance

[4] 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.

structures lead to observational equivalence. Thus, to guarantee consistency of the SMD estimator, it will be sufficient to simulate endogenous regressors with an appropriate autocovariance structure, even if the exact data generation process of those regressors is unknown. We propose a simulation algorithm for the endogenous regressor that guarantees consistency of the SMD estimator. While not specifying the complete measurement error structure may be less efficient, our simulator is less sensitive to misspecification.

The paper proceeds as follows. Section 2 introduces the time series econometric setup and uses a simple regression model to explain our identification and estimation strategies. Section 3 formally discusses identification and estimation in general autoregressive distributed lag models. Section 4 presents Monte Carlo simulation evidence and an application to the long-run risks model. The last section concludes. Technical proofs are relegated to an Appendix.

As a matter of notation, we use $\Gamma_z(j) \equiv E(z_t z'_{t-j})$ to denote the autocovariance of order j of a generic covariance (or weakly) stationary mean-zero vector-valued time series $\{z_t\}$. We use $\Gamma_{zw}(j,k) \equiv E(z_{t-j} w'_{t-k})$ to denote the cross-covariance between two mean-zero covariance stationary processes $\{z_t\}$ and $\{w_t\}$. If $E(z_t) = 0$, $E(z_t z'_t) = \Gamma_z(0)$, and $\Gamma_z(j) = 0$ for $j \geq 1$, then $\{z_t\}$ is a white noise (WN). In this case, we write $z_t \sim WN(0, \Gamma_z(0))$.

2 The Econometric Setup

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

$$\alpha(L) y_t = \beta(L) x_t + u_t, \tag{1}$$

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:

$$X_t = x_t + \epsilon_t.$$

On the other hand, y_t is observed without error.[5] Additional regressors can be accommodated provided they are correctly observed. In that case, y_t and X_t above can be interpreted as the residuals from projections of the dependent variable and the mismeasured regressor on all other regressors. The ADL(p,q) model expressed in terms of the observables is then given by:

$$\alpha(L) y_t = \beta(L) X_t + V_t, \quad \text{where} \quad V_t = u_t - \beta(L)\epsilon_t. \tag{2}$$

Assumptions on the latent variables of the model, u_t, ϵ_t, and x_t, are as follows.

Assumption A

[5] ARMA models when y_t is observed with error are studied in Komunjer and Ng (2014).

(a) $u_t \sim WN(0, \sigma_u^2)$. *For every* (t, τ), $E(u_t x_\tau) = 0$ *and* $E(u_t \epsilon_\tau) = 0$.

(b) $\{(x_t, \epsilon_t)'\}$ *is covariance stationary with* $E(x_t) = 0$, $E(\epsilon_t) = 0$, $E(x_t \epsilon_\tau) = 0$ *for every* (t, τ).

(c) The roots of $\alpha(z) = 0$, $z \in \mathbb{C}$ *are all strictly outside the unit circle.*

(d) The covariance matrix of $(y_{t-1}, \ldots, y_{t-p}, x_t, x_{t-1}, \ldots, x_{t-q})'$ *is nonsingular.*

We assume in (a) that the model is dynamically correctly specified and that all the relevant regressors have been included in (1). Hence, u_t is serially uncorrelated. The white noise assumption on u_t can accommodate disturbances that are conditionally heteroskedastic.[6] Though latent, the regressor x_t is assumed exogenous, and its measurement error ϵ_t orthogonal to u_t. (a) and (b) combined ensure that all the latent variables of the model are covariance stationary. This, together with the stability condition (c) then guarantees covariance stationarity of the observables $\{(y_t, X_t)'\}$. Since x_t and ϵ_t are mean zero, the intercept is suppressed in (1). The regressor x_t is observed with error ϵ_t whenever $\Gamma_\epsilon(0) \neq 0$. The measurement error is classical, i.e. orthogonal to x_t at all leads and lags, but is allowed to be serially correlated. We only need ϵ_t to be covariance stationary. Correct specification of its dynamic structure is however not necessary. Moreover, ϵ_t like u_t is allowed to be conditionally heteroskedastic. Assumption (d) is standard for least squares analysis, except for the fact that it involves the latent variables $x_t, \ldots, x_{t-q}$.

From (2), we see that V_t is generally serially correlated. As first documented in Grether and Maddala (1973), measurement errors in the exogenous variables may lead to the appearance of spurious long lags in adjustments: even if ϵ_t is white noise, V_t is a q-order moving average (MA(q)) process. Thus, the order q of the ADL(p,q) model affects the identification of α and β.

The model defined by (2) can be rewritten as:

$$\begin{aligned} y_t &= W_t'\gamma + V_t \\ \gamma &\equiv (\alpha_1, \ldots, \alpha_p, \beta_0, \ldots, \beta_q)' \\ W_t &\equiv (y_{t-1}, \ldots, y_{t-p}, X_t, \ldots, X_{t-q})'. \end{aligned} \tag{3}$$

As is well known, the OLS estimator for γ is generally biased when $E(W_t V_t) \neq 0$. Instrumental variable estimation requires X_{t-j} to be correlated with ϵ_{t-j} for $j > q$, and instruments that are both strong and valid may not be available. In these cases, identification of γ is not possible without further information. We propose to use the information contained in the autocovariance structure of V_t. Because the autocovariances of V_t also depend on the autocovariances of ϵ_t which

[6] Anticipating the estimation results to follow, it is worth pointing out that while the potential presence of conditional heteroskedasticity does not affect the consistency of our estimator, it would affect its efficiency.

are not of direct interest, the problem is to find a balance between the information that they bring, and the additional parameters that characterize them. Once identification is established, we show how consistent estimates can be obtained. The precise implementation again depends on the complexity of the model as given by p and q. In simple models, classical minimum distance estimation is possible. In the next subsections, we study the $(p, q) = (0, 0)$ case. The choice of auxiliary statistics, identification and estimation will be discussed. Section 3 then analyzes the general ADL(p,q) model where the simulated minimum distance estimation is useful.

2.1 ADL(0,0) Model

Consider the regression model:

$$y_t = x_t\beta + u_t, \tag{4}$$

with a latent regressor x_t, and a mismeasured observed regressor $X_t = x_t + \epsilon_t$. In terms of the observables (y_t, X_t), the model becomes:

$$y_t = X_t\beta + V_t, \quad \text{where} \quad V_t = u_t - \beta\epsilon_t. \tag{5}$$

Because of the measurement error, the regressor X_t is endogenous, $E(X_tV_t) \neq 0$, which causes problems in estimating β. Several solutions to the problem have been proposed in the literature. When both the latent regressor x_t and measurement error ϵ_t are known to be serially uncorrelated, identification and estimation can proceed by exploiting certain features of the data (heteroskedasticity, skewness and excess kurtosis) or external instruments as discussed in the introduction. In a time series context when x_t is serially correlated, different estimation strategies are possible under different assumptions regarding the measurement error. In the case where ϵ_t is uncorrelated (white noise) or it has a finite-order MA structure, lags of X_t can be used as instruments (see Biørn (2014)). The practical interest of our approach is in situations when X_{t-k} $(k \geq 1)$ may not be valid instruments. This occurs when the measurement error follows a process with an autoregressive (AR) component. Thus, we focus on the case where both x_t and ϵ_t are serially correlated, with unknown autocorrelation structures.

Serial correlation in the measurement error has the important implication that X_{t-1} is no longer a valid instrument in (5). Though longer lags could be valid, they may have weak correlation with X_t. To begin, consider estimating β using OLS. The OLS estimator $\widehat{\beta}$ has an attenuation bias given by:

$$\text{plim}_{\,T\to\infty}(\widehat{\beta} - \beta) = -\beta\frac{\Gamma_\epsilon(0)}{\Gamma_X(0)} \equiv [\beta].$$

Since the bias $[\beta]$ is a function of two unknown parameters, β and $\Gamma_\epsilon(0)$, estimating β using the OLS estimator alone is impossible.

Our point of departure is the simple observation that the bias $[\beta]$ also affects the time series properties of the least squares residuals $\widehat{V}_t \equiv y_t - X_t\widehat{\beta} = V_t - X_t(\widehat{\beta} - \beta)$. Consider the autocovariances $\widehat{\Gamma}_{\widehat{V}}(j) \equiv \frac{1}{T}\sum_{t=j+1}^{T}\widehat{V}_t\widehat{V}_{t-j}$, and cross-covariances $\widehat{\Gamma}_{\widehat{V}X}(j,0) \equiv \frac{1}{T}\sum_{t=j+1}^{T}\widehat{V}_{t-j}X_t$ $(j \geq 0)$. Then, as $T \to \infty$, we have

$$\begin{aligned}
\text{plim}_{\,T\to\infty}\widehat{\Gamma}_{\widehat{V}}(0) &= \sigma_u^2 + \beta^2\Gamma_\epsilon(0) + [\beta]\beta\Gamma_\epsilon(0) \\
\text{plim}_{\,T\to\infty}\widehat{\Gamma}_{\widehat{V}}(1) &= \beta^2\Gamma_\epsilon(1) + 2\beta\Gamma_\epsilon(1)[\beta] + [\beta]^2\Gamma_X(1) \\
\text{plim}_{\,T\to\infty}\widehat{\Gamma}_{\widehat{V}X}(1,0) &= -\beta\Gamma_\epsilon(1) - [\beta]\Gamma_X(1).
\end{aligned}$$

Observe that these moments use $\widehat{V}_t$ instead of V_t. Hence, the autocovariances and cross-covariances of the least squares residuals are functions of the least squares bias $[\beta]$, and thus contain useful information regarding the parameters of the model (4).

2.2 Identification

The parameters of the ADL(0,0) model in the presence of measurement errors are given by:

$$\theta = (\beta, \sigma_u^2, \Gamma_\epsilon(0), \Gamma_\epsilon(1))'. \tag{6}$$

Note that β is the parameter of direct interest in (4), while σ_u^2 , $\Gamma_\epsilon(0)$ and $\Gamma_\epsilon(1)$ are nuisance parameters. We propose to identify θ from the auxiliary statistics:

$$\widehat{\psi} = \left(\widehat{\beta}, \widehat{\Gamma}_{\widehat{V}}(0), \widehat{\Gamma}_{\widehat{V}}(1), \widehat{\Gamma}_{VX}(1,0)\right)' \tag{7}$$

whose probability limit as T goes to infinity (or binding function) is given by:

$$\psi(\theta) = \begin{pmatrix} \beta + [\beta] \\ \sigma_u^2 + \beta^2\Gamma_\epsilon(0) + [\beta]\beta\Gamma_\epsilon(0) \\ \beta^2\Gamma_\epsilon(1) + 2\beta\Gamma_\epsilon(1)[\beta] + [\beta]^2\Gamma_X(1) \\ -\beta\Gamma_\epsilon(1) - [\beta]\Gamma_X(1) \end{pmatrix}. \tag{8}$$

To show that θ is (globally) identifiable from $\psi(\theta)$, we need to establish that the mapping $\theta \mapsto \psi(\theta)$ is invertible. The following result summarizes conditions under which identification obtains.

Lemma 1 *Consider the ADL(0,0) model* (4)*. Under Assumptions A(a)-(d):*

(a) $(\beta = 0, \sigma_u^2)'$ *is globally identified;*

(b) $(\beta \neq 0, \sigma_u^2)'$ *is globally identified if* $\Gamma_x(1) \neq 0$*;*

(c) θ *is globally identified if (i)* $\Gamma_x(1) \neq 0$ *and (ii)* $\beta \neq 0$.

The proof, given in the Appendix, involves inverting the binding function in (8) and showing that a unique solution to $\psi(\theta) = \psi$ exists. Serial correlation in the latent regressor x_t is needed.[7] It is worth pointing out that β and σ_u^2 are identifiable irrespective of whether or not $\beta = 0$.[8] However, $\Gamma_\epsilon(0)$ and $\Gamma_\epsilon(1)$ can only be identified if $\beta \neq 0$. This is because the regression residuals have no information about β if X_t has no role in the regression model. Thus, if $\beta = 0$, the only way we can learn about the measurement error is by looking at the regressor X_t.

The required condition $\Gamma_x(1) \neq 0$ is not directly testable. However, when $\beta \neq 0$, we can use the fact that $\Gamma_X(1) = \Gamma_x(1) + \Gamma_\epsilon(1)$ in order to learn about the serial correlation of the latent regressor. When $\beta \neq 0$, an IV estimator with X_{t-1} as instrument has probability limit

$$\text{plim}_{\,T\to\infty}\widehat{\beta}_{IV} = \beta\left(1 - \frac{\Gamma_\epsilon(1)}{\Gamma_X(1)}\right) = \beta\frac{\Gamma_x(1)}{\Gamma_X(1)},$$

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 the latter is biased for β.

The identification results of Lemma 1 continue to hold when the measurement error is uncorrelated (or white noise). In this case, it may be of interest to determine whether the nuisance parameter $\Gamma_\epsilon(0)$ can be identified when $\beta = 0$. 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 $\Gamma_X(0) = \Gamma_x(0) + \Gamma_\epsilon(0)$, we have

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

We can use this expression for $\Gamma_\epsilon(0)$ to assess the severity of measurement error prior to any regression analysis. The result is, however, specific to white noise processes.

2.3 Estimation

We now turn to the problem of estimating the ADL(0,0) model. The CMD estimator is defined as:

$$\widehat{\theta} = \text{argmin}_{\,\theta}\|\widehat{\psi} - \psi(\theta)\|_W,$$

[7] Note that when ϵ_t is white noise, then $\Gamma_x(1) = \Gamma_X(1)$ and the requirement is that $\Gamma_X(1) \neq 0$ which is easy to test.

[8] This 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 $\theta = (\beta, \sigma_u^2, \Gamma_\epsilon(0))'$, while parts (a) and (b) of our result apply to $(\beta, \sigma_u^2)'$ alone.

where $\|v\|_W \equiv v'Wv$ and W is a positive definite weighting matrix. For the ADL(0,0) model, θ, $\widehat{\psi}$ and $\psi(\theta)$ are defined in (6), (7) and (8), respectively, and W is the identity matrix. Given the invertibility of the binding function, the CMD estimator equals $\widehat{\theta} = \psi^{-1}(\widehat{\psi})$. While a closed-form expression of the binding function $\psi(\theta)$ is possible to derive for the ADL(0,0) model, this is often not feasible in more complex models. However, we can use Monte-Carlo methods to compute the mapping from θ to ψ. We now present such an estimator for the ADL(0,0) model. This is useful for understanding the estimator in the general case.

Simulating the data according to the model in (4) is not straightforward because the model contains no information regarding the data generating process for the latent regressor x_t nor its measurement error ϵ_t. To deal with this model incompleteness, we exploit the following simple principle: since the auxiliary statistics only depend on the first and second order moments of the observed data, it is sufficient that the simulated data have correct first and second order moments. Put differently, the dynamic specifications used to simulate x_t or ϵ_t need not be correct, provided they lead to the correct values of the first and second order moment properties of the observables.

To formalize the argument, say that S sets of simulated data $(\mathbf{y}^S(\theta), \mathbf{X}^S(\theta))$ have been obtained given an assumed value for θ, and consider

$$\psi^S(\theta) \equiv \frac{1}{S}\sum_{s=1}^{S} \widehat{\psi}(\mathbf{y}^s(\theta), \mathbf{X}^s(\theta)).$$

This allows us to define the SMD estimator $\widehat{\theta}^S$ as:

$$\widehat{\theta}^S = \text{argmin}_{\,\theta} \|\widehat{\psi} - \psi^S(\theta)\|_W. \tag{9}$$

As in the classical minimum distance estimation, consistency of $\widehat{\theta}^S$ requires that the mapping $\psi(\theta)$ be invertible. The new additional requirement is that the simulated mapping $\psi^S(\theta)$ "approximates" $\psi(\theta)$ as the number of simulated samples S gets large, in a sense that

$$E_{(\mathbf{y}^S(\theta), \mathbf{X}^S(\theta))}[\widehat{\psi}(\mathbf{y}^S(\theta), \mathbf{X}^S(\theta))] = \psi(\theta). \tag{10}$$

The above "consistent simulation" property ensures that the auxiliary statistics computed using the simulated data provide a consistent functional estimator of the binding function.

The consistent simulation condition (10) is automatically satisfied in most of the traditional work on simulation estimation where it follows directly from the assumed exogeneity and correct specification of the dynamics of the model variables. It holds, for example, when $\widehat{\psi}$ is a vector of unconditional moments, and $\widehat{\theta}^S$ is the simulated method of moments estimator of Duffie and Singleton (1993); or else when $\widehat{\psi}$ is the score of the likelihood and $\widehat{\theta}^S$ the efficient methods of moments estimator of Gallant and Tauchen (1996). Finally, the property also holds in the indirect inference estimator of Gourieroux, Monfort, and Renault (1993) where $\widehat{\psi}$ are the parameters of an auxiliary regression (see, for example, p.S89 in Gourieroux, Monfort, and Renault (1993)).

In incomplete models such as (4), the consistent simulation condition (10) is not trivial to obtain. As far as we are aware, the only reference to simulation estimation of measurement error models is Jiang and Turnbull (2004). Their method relies on the existence of "validation" data that can be used to estimate the nuisance parameters of the model. Without validation data, simulation estimation cannot be implemented in the standard way. This is due to the endogeneity of the observed regressor X_t. More specifically, there are two issues that need to be dealt with.

First, there is the issue of simulating the measurement error ϵ_t. We exploit the fact that the nuisance parameters in θ are the autocovariances $\Gamma_\epsilon(0)$ and $\Gamma_\epsilon(1)$. Hence, simulation of ϵ_t can be based on any dynamic specification that respects those moments. For instance, we can simulate ϵ_t as an AR(1) process, $\epsilon_t = \rho\epsilon_{t-1} + \xi_t$ with $\xi_t \sim iidN(0, \sigma_\xi^2)$ where ρ and σ_ξ are chosen so that

$$\rho = \frac{\Gamma_\epsilon(1)}{\Gamma_\epsilon(0)} \quad \text{and} \quad \sigma_\xi^2 = (1-\rho^2)\Gamma_\epsilon(0).$$

It is important to emphasize that the true data generating process for ϵ_t, which is unknown, need not be an AR(1). All that is needed is that the parameters ρ and σ_ξ of the AR(1) model used for simulation be chosen so that the simulated ϵ_t's have correct variance $\Gamma_\epsilon(0)$ and autocovariance $\Gamma_\epsilon(1)$.

Second, there is the issue of how to simulate the latent regressor x_t. A naive approach would be to set the simulated x_t^s so that $x_t^s + \epsilon_t^s = X_t$. This would correspond to the classical indirect inference approach in which the regressor X_t is held fixed in simulations, and only the disturbance u_t and the measurement error ϵ_t are simulated. Though appealing in its simplicity, this naive approach would lead to incorrect inference. To see why, let u_t^s and ϵ_t^s denote the simulated values of u_t and ϵ_t, respectively. Then, for any given value of θ, the simulated value y_t^s of y_t is obtained as:

$$y_t^s = \beta X_t + V_t^s, \quad \text{where} \quad V_t^s = u_t^s - \beta\epsilon_t^s. \tag{11}$$

Despite being correctly specified, the simulated regression in (11) has one fundamental difference with the observed regression in (5): in simulations, $E(X_t V_t^s) = 0$, while $E(X_t V_t) \neq 0$ in the data. The generic problem is that when X_t is not exogenous, it 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 the dependence structure found in the data. The simulated X_t^s will need to be endogenous, i.e. correlated with V_t^s, with the dependence structure that matches that of the observed regressor X_t.

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 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. When x_t is 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 $\varphi = [1 - \Gamma_\epsilon(0)/\Gamma_X(0)]^{1/2}$. By construction, the mean and variance of X_t^s are equal to the mean and variance of X_t. But with serially correlated latent regressors, we will need the simulated regressors to preserve not only the variance, but also the autocovariance structure in the data. For this reason, we propose the following simulation method:

Algorithm SMD for the ADL(0,0) model

1. Compute the auxiliary statistics $\widehat{\psi}$ from the observed data.
2. Given θ, for $s = 1, \dots S$ and $t = 1, \dots T$:
   (i) simulate $u_t^s \sim iidN(0, \sigma_u^2)$;
   (ii) simulate $\epsilon_t^s = \rho\epsilon_{t-1}^s + \xi_t^s$, $\xi_t^s \sim iidN(0, \sigma_\xi^2)$, with $\rho = \Gamma_\epsilon(1)/\Gamma_\epsilon(0)$, $\sigma_\xi^2 = (1-\rho^2)\Gamma_\epsilon(0)$;
   (iii) let $x_t^s = \varphi_1 X_t + \varphi_2 X_{t-1}$;
   (iv) let $X_t^s = x_t^s + \epsilon_t^s$;
   (v) let $y_t^s = \beta x_t^s + u_t^s$;
   (vi) compute $\widehat{\psi}$ from the simulated data $(\mathbf{y}^s(\theta), \mathbf{X}^s(\theta))$.
3. Minimize $\|\widehat{\psi} - \frac{1}{S}\sum_{s=1}^S \widehat{\psi}(\mathbf{y}^s(\theta), \mathbf{X}^s(\theta))\|_W$ over θ.

The key to our simulation method is Step 2(iii) 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:

$$\begin{aligned}\Gamma_X(0) - \Gamma_\epsilon(0) &= (\varphi_1^2 + \varphi_2^2)\Gamma_X(0) + 2\varphi_1\varphi_2\Gamma_X(1) \\ \Gamma_X(1) - \Gamma_\epsilon(1) &= (\varphi_1^2 + \varphi_2^2)\Gamma_X(1) + \varphi_1\varphi_2\Gamma_X(0) + \varphi_1\varphi_2\Gamma_X(2), \end{aligned} \tag{12}$$

where $\Gamma_\epsilon(1) = 0$ if the measurement error is white noise. Step 2(iii) 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, $\Gamma_x^s(0) = \Gamma_x(0)$ and $\Gamma_x^s(1) = \Gamma_x(1)$. Given the assumed values for $\Gamma_\epsilon(0)$ and $\Gamma_\epsilon(1)$, and given the estimates of $\Gamma_X(0)$ and $\Gamma_X(1)$ obtained from the observed data, (12) is a system of two equations in two unknowns. A unique solution for φ_1 and φ_2 can be obtained by noting that the system in (12) is linear in $r_0 \equiv (\varphi_1^2 + \varphi_2^2, \varphi_1\varphi_2)'$,

$$\underbrace{\begin{pmatrix} \Gamma_X(0) & 2\Gamma_X(1) \\ \Gamma_X(1) & \Gamma_X(0)+\Gamma_X(2) \end{pmatrix}}_{R_0} \underbrace{\begin{pmatrix} \varphi_1^2+\varphi_2^2 \\ \varphi_1\varphi_2 \end{pmatrix}}_{r_0} = \underbrace{\begin{pmatrix} \Gamma_X(0)-\Gamma_\epsilon(0) \\ \Gamma_X(1)-\Gamma_\epsilon(1) \end{pmatrix}}_{Q_0}.$$

Assuming that R_0 is invertible, $r_0 = R_0^{-1}Q_0 = (r_{01}, r_{02})'$.[9] Then, φ_1 and φ_2 can be computed as:

$$\varphi_1 = \frac{1}{2}\left[\sqrt{r_{01}+2r_{02}} + \sqrt{r_{01}-2r_{02}}\right], \quad \varphi_2 = \frac{1}{2}\left[\sqrt{r_{01}+2r_{02}} - \sqrt{r_{01}-2r_{02}}\right].$$

Combining Steps 2(ii) and 2(iii), the simulated regressor X_t^s has the same autocovariances as the observed regressor, i.e.

$$\Gamma_X^s(0) = \Gamma_X(0) \quad \text{and} \quad \Gamma_X^s(1) = \Gamma_X(1).$$

Moreover, the simulated regressor is endogenous, and $E(X_t^s V_t^s) = E(X_t V_t) \neq 0$. This comes from the fact that the simulated X_t^s respects the measurement error equation $X_t^s = x_t^s + \epsilon_t^s$, and that the simulated latent regressor x_t^s is truly exogenous. Since the simulations respect all the moments that appear in the binding function $\psi(\theta)$, the consistent simulation property (10) is satisfied. Section 3 extends this result to more general ADL models.

3 Identification and Estimation of ADL(p,q) Models

This section considers the general ADL(p,q) model. It will be shown that in the presence of measurement errors, the model has $(p+3q+4)$ parameters

$$\theta \equiv \left(\gamma', \sigma_u^2, \Gamma_\epsilon(0), \ldots, \Gamma_\epsilon(2q+1)\right)'. \tag{13}$$

These parameters are to be identified from the probability limits of $(p+3q+4)$ auxiliary statistics:

$$\widehat{\psi} \equiv \left(\widehat{\gamma}', \widehat{\Gamma}_{\widehat{V}X}(1,0), \ldots, \widehat{\Gamma}_{\widehat{V}X}(q+1,0), \widehat{\Gamma}_{\widehat{V}}(0), \ldots, \widehat{\Gamma}_{\widehat{V}}(q+1)\right)'. \tag{14}$$

We then turn to estimation of the parameters θ.

[9] Note that we have:

$$r_{01} = 1 - \frac{(\Gamma_X(0)+\Gamma_X(2))\Gamma_\epsilon(0) - 2\Gamma_X(1)\Gamma_\epsilon(1)}{(\Gamma_X(0)+\Gamma_X(2))\Gamma_X(0) - 2\Gamma_X(1)^2},$$

from which it is straightforward to show that $r_{01} \geq 0$ if and only if

$$\frac{\Gamma_X(0)+\Gamma_X(2)}{2\Gamma_X(1)} - \frac{\Gamma_X(1)}{\Gamma_X(0)} \text{ has the same sign as } \frac{\Gamma_X(0)+\Gamma_X(2)}{2\Gamma_X(1)} - \frac{\Gamma_x(1)}{\Gamma_x(0)}.$$

Intuitively, the persistence of the latent regressor $\Gamma_x(1)/\Gamma_x(0)$ should not be too different from the persistence of the observed regressor $\Gamma_X(1)/\Gamma_X(0)$, so that the above signs remain the same. When $r_{01} \geq 0$, the two solutions φ_1, φ_2 are guaranteed to be real.

3.1 Identification and Choice of Auxiliary Statistics

To understand (13) and (14), note first that the OLS estimator has asymptotic bias:

$$\text{plim}_{\,T\to\infty}\,(\widehat{\gamma}-\gamma)=\Gamma_W(0)^{-1}\Gamma_{VW}(0,0)\equiv[\gamma],$$

with γ and W_t as defined in (3). The parameters entering $[\gamma]$ are those appearing in the cross-covariance $\Gamma_{VW}(0,0)$. Since $X_t=x_t+\epsilon_t$ and $V_t=u_t-\beta(L)\epsilon_t$, the OLS bias $[\gamma]$ now also depends on the measurement error autocovariances $\Gamma_\epsilon(i)$ with $0\leqslant i\leqslant q$. This implies that in addition to the $(p+q+2)$ parameters $(\gamma',\sigma_u^2)'$ of the ADL(p,q) model, there are now $(q+1)$ nuisance parameters $(\Gamma_\epsilon(0),\ldots,\Gamma_\epsilon(q))'$. Thus, at least $(p+q+2)+(q+1)$ auxiliary statistics are needed to identify all the parameters.

The OLS estimator provides $(p+q+1)$ statistics; the variance of the least squares residuals $\widehat{\Gamma}_{\widehat{V}}(0)$ provides another. But we still need another $(q+1)$ auxiliary statistics. By orthogonality of the least squares residuals, $\widehat{\Gamma}_{\widehat{V}X}(0,i)=0$, $0\leqslant i\leqslant q$. We are left to to consider the moments $\widehat{\Gamma}_{\widehat{V}X}(k,0)$ and $\widehat{\Gamma}_{\widehat{V}}(k)$ for $k\geq 1$ whose probability limits are:[10]

$$\begin{aligned}\Gamma_{\widehat{V}}(k) &= \Gamma_V(k)-\Big(\Gamma_{VW}(k,0)+\Gamma_{VW}(0,k)\Big)'[\gamma]+[\gamma]'\Gamma_W(k)[\gamma] \qquad (15)\\ \Gamma_{\widehat{V}X}(k,0) &= \Gamma_{VX}(k,0)-\Gamma_{WX}(k,0)'[\gamma]. \qquad (16)\end{aligned}$$

It is not hard to see that for any $k\geq 1$, $\Gamma_{\widehat{V}}(1),\ldots,\Gamma_{\widehat{V}}(k)$ and $\Gamma_{\widehat{V}X}(1,0),\ldots,\Gamma_{\widehat{V}X}(k,0)$ depend on k new nuisance parameters $\Gamma_\epsilon(q{+}1),\ldots,\Gamma_\epsilon(q{+}k)$. Take for example $k=1$. Evidently, $\widehat{\Gamma}_{\widehat{V}X}(1,0)$ and $\widehat{\Gamma}_{\widehat{V}}(1)$ depend on: (i) the parameters of the ADL(p,q) model, (γ,σ_u^2), (ii) the nuisance parameters $(\Gamma_\epsilon(0),\ldots,\Gamma_\epsilon(q))$ already appearing in the OLS bias $[\gamma]$, and (iii) a new nuisance parameter $\Gamma_\epsilon(q{+}1)$. Thus, when $k=1$, the inclusion of two auxiliary statistics $\widehat{\Gamma}_{\widehat{V}X}(1,0)$ and $\widehat{\Gamma}_{\widehat{V}}(1)$ increases the number of nuisance parameters by one.

In general, there are $(p+q+1)+1+2k$ auxiliary statistics

$$\widehat{\psi}_k=\Big(\widehat{\gamma}',\widehat{\Gamma}_{\widehat{V}X}(1,0),\ldots,\widehat{\Gamma}_{\widehat{V}X}(k,0),\widehat{\Gamma}_{\widehat{V}}(0),\widehat{\Gamma}_{\widehat{V}}(1),\ldots,\widehat{\Gamma}_{\widehat{V}}(k)\Big)' \qquad (17)$$

to determine $(p+q+1)+1+(q+1+k)$ parameters

$$\theta_k=\big(\gamma',\sigma_u^2,\Gamma_\epsilon(0),\ldots,\Gamma_\epsilon(q),\Gamma_\epsilon(q+1),\ldots,\Gamma_\epsilon(q+k)\big)', \qquad (18)$$

with the order condition given by

$$k\geq q+1. \qquad (19)$$

[10] In principle, we can also consider $\widehat{\Gamma}_{\widehat{V}X}(0,q+k)$ for $k\geq 1$, but it is straightforward to see that these cross-covariances are informative only if X_t is strongly persistent.

Setting $k = q + 1$ satisfies the rule which leads to (13) and (14). When $q = 0$, we have $k = 1$, $\theta = (\beta, \sigma_u^2, \Gamma_\epsilon(0), \Gamma_\epsilon(1))'$, and $\widehat{\psi} = \left(\widehat{\beta}, \widehat{\Gamma}_{\widehat{V}}(0), \widehat{\Gamma}_{\widehat{V}}(1), \widehat{\Gamma}_{VX}(1,0)\right)'$, which agrees with the earlier analysis for the ADL(0,0) model. For the ADL(1,1) model, for example, $k = 2$. We need to identify 8 parameters $\theta = (\alpha, \beta_0, \beta_1, \sigma_u^2, \Gamma_\epsilon(0), \Gamma_\epsilon(1), \Gamma_\epsilon(2), \Gamma_\epsilon(3))'$ from 8 auxiliary statistics $\widehat{\psi} = (\widehat{\alpha}, \widehat{\beta}_0, \widehat{\beta}_1, \widehat{\Gamma}_{\widehat{V}}(0), \widehat{\Gamma}_{\widehat{V}}(1), \widehat{\Gamma}_{\widehat{V}}(2), \widehat{\Gamma}_{\widehat{V}X}(1,0), \widehat{\Gamma}_{\widehat{V}X}(2,0))'$.

We now turn to the question of identifiability of θ. The ADL(1,0) model is simple enough that this condition can be analytically verified. The 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, $\gamma = (\alpha, \beta)'$, $W_t = (y_{t-1}, X_t)'$, and $\theta = (\alpha, \beta, \sigma_u^2, \Gamma_\epsilon(0), \Gamma_\epsilon(1))'$. Assuming that $\Gamma_W(0)$ is nonsingular, the least squares bias is given by

$$\text{plim}_{\,T\to\infty}\,(\widehat{\gamma} - \gamma) = \begin{pmatrix} \beta \frac{\Gamma_\epsilon(0)\Gamma_{yX}(1,0)}{\Gamma_y(0)\Gamma_X(0) - \Gamma_{yX}(1,0)^2} \\ -\beta \frac{\Gamma_\epsilon(0)\Gamma_y(0)}{\Gamma_y(0)\Gamma_X(0) - \Gamma_{yX}(1,0)^2} \end{pmatrix} \equiv [\gamma].$$

The auxiliary statistic is $\widehat{\psi} = \left(\widehat{\alpha}, \widehat{\beta}, \widehat{\Gamma}_{\widehat{V}}(0), \widehat{\Gamma}_{\widehat{V}}(1), \widehat{\Gamma}_{\widehat{V}X}(1,0)\right)'$. To (globally) identify θ requires inverting the binding function $\psi(\theta) = \text{plim}_{\,T\to\infty}\widehat{\psi}$.

Lemma 2 *Consider the ADL(1,0) model* (2)*. Under Assumptions A(a)-(d):*

(a) $(\alpha, \beta = 0, \sigma_u^2)'$ *is globally identified;*

(b) $(\alpha, \beta \neq 0, \sigma_u^2)'$ *is globally identified if* $\Gamma_x(1) \neq 0$*;*

(c) θ *is globally identified if: (i)* $\Gamma_x(1) \neq 0$*, and (ii)* $\beta \neq 0$.

The required restrictions are the same as those of Lemma 1. This is not surprising given that the only difference between the ADL(0,0) and ADL(1,0) models comes from the presence of an additional regressor y_{t-1}. With the lagged dependent variable being correctly measured, it is not surprising that identification requires the same conditions as when this regressor is absent. We conjecture, however, that higher order ADL(p,q) models with $q \geq 1$ require more restrictions on the serial correlation of the latent regressor.

Checking invertibility of the binding function is difficult for ADL(p,q) models with $q \geq 1$. In the ADL(1,1) model, for example, there are 8 nonlinear equations in 8 unknowns to be solved. As is often the case in complex non-linear models, global identification is difficult, if not impossible, to analytically verify. In the next section, we describe how to approximate the ADL(p,q) binding function using simulations. Once such approximations are available, one can check for invertibility using numerical methods.

3.2 Simulated Minimum Distance Estimation

Once the auxiliary statistic is defined, we can use the analytic expression of its probability limit or binding function to establish identification and construct the CMD estimator. Such an analysis is possible for small order models such as ADL(0,0) or ADL(1,0). However, this task proves to be impossible for the ADL(1,1) model despite serious efforts. For this reason, we use the SMD estimator defined in (9).

The auxiliary statistics for the ADL(p,q) model defined in (14) depends on the first $2q+2$ autocovariances of X_t. Hence, it is necessary to simulate an exogenous process for x_t which preserves those $2q+2$ autocovariances, i.e. a process $\{x_t^s\}$ such that:

$$\Gamma_x^s(k) = \Gamma_X(k) - \Gamma_\epsilon(k), \quad k = 0, \ldots, 2q+2. \tag{20}$$

Exogeneity of the simulated latent process $\{x_t^s\}$ means that it must be independent of $\{u_t^s\}$. To do so, we extend the simulation procedure presented earlier for the ADL(0,0) model to the general ADL(p,q) model. Let

$$x_t^s = \varphi_0 X_t + \varphi_1 X_{t-1} + \ldots + \varphi_{2q+1} X_{t-(2q+1)}, \tag{21}$$

where the $2q+2$ parameters $(\varphi_0, \ldots, \varphi_{2q+1})'$ are to be determined to satisfy (20). Following a reasoning similar to that for the ADL(0,0) model analyzed in Section 2.3, the restrictions can be written in a matrix form:

$$\underbrace{\begin{pmatrix} \Gamma_X(0) - \Gamma_\epsilon(0) \\ \vdots \\ \Gamma_X(2q+1) - \Gamma_\epsilon(2q+1) \end{pmatrix}}_{Q_q} = \underbrace{\begin{pmatrix} \Gamma_X(0) & \ldots & 2\Gamma_X(2q+1) \\ \vdots & & \vdots \\ \Gamma_X(2q+1) & \ldots & \Gamma_X(0) + \Gamma_X(4q+2) \end{pmatrix}}_{R_q} \underbrace{\begin{pmatrix} \varphi_0^2 + \varphi_1^2 + \ldots + \varphi_{2q+1}^2 \\ \vdots \\ \varphi_0 \varphi_{2q+1} \end{pmatrix}}_{r_q}.$$

Assuming R_q is invertible, $r_q = R_q^{-1} Q_q = (r_{q0}, \ldots, r_{q,2q+1})'$. The coefficients $(\varphi_0, \ldots, \varphi_{2q+1})$ in (21) are then obtained from the solution $(r_{q0}, \ldots, r_{q,2q+1})$ by solving the nonlinear system of $(2q+2)$ equations in $(2q+2)$ unknowns:

$$\begin{aligned} \varphi_0^2 + \varphi_1^2 + \ldots + \varphi_{2q+1}^2 &= r_{q0} \\ &\vdots \\ \varphi_0 \varphi_{2q+1} &= r_{q,2q+1}. \end{aligned} \tag{22}$$

The dimension of the system (22) only depends on the lag-length q in the ADL(p,q) model and is relatively easy to solve. In the $q = 0$ case, the solution was given in Section 2.3. In the general case, the solution can be obtained using a numerical solver.

Our general simulation algorithm can now be described as follows:

Algorithm SMD for ADL(p,q) model with parameters $\theta = (\gamma', \sigma_u^2, \Gamma_\epsilon(0), \ldots, \Gamma_\epsilon(2q+1))'$:

1. Compute the auxiliary statistics $\widehat{\psi}$ from the observed data.

2. Given θ, for $s = 1, \ldots S$ and $t = 1, \ldots T$:

   (i) simulate $u_t^s \sim iidN(0, \sigma_u^2)$

   (ii) simulate $\epsilon_t^s = \rho_1 \epsilon_{t-1}^s + \ldots + \rho_{2q+1} \epsilon_{t-(2q+1)}^s + \xi_t$, $\xi_t \sim iidN(0, \sigma_\xi^2)$, where $\rho = (\rho_1, \ldots, \rho_{2q+1})'$ and σ_ξ^2 solve the Yule-Walker equations:

   $$\Gamma_\epsilon \rho = \gamma_{2q+1} \quad \text{and} \quad \sigma_\xi^2 = \Gamma_\epsilon(0) - \rho' \gamma_{2q+1},$$

   where $\gamma_{2q+1} = (\Gamma_\epsilon(1), \ldots, \Gamma_\epsilon(2q+1))'$ and Γ_ϵ is the covariance matrix $[\Gamma_\epsilon(i-j)]_{i,j=1}^{2q+1}$;

   (iii) let $x_t^s = \varphi_0 X_t + \ldots + \varphi_{2q+1} X_{t-(2q+1)}$ where $(\varphi_0, \ldots, \varphi_{2q+1})$ solve the system in (22);

   (iv) let $X_t^s = x_t^s + \epsilon_t^s$;

   (v) let $y_t^s = \alpha_1 y_{t-1}^s + \ldots + \alpha_p y_{t-p}^s + \beta_0 x_t^s + \ldots + \beta_q x_{t-q}^s + u_t^s$;

   (vi) compute the auxiliary statistics $\widehat{\psi}$ in (14) from the simulated data $(\mathbf{y}^s(\theta), \mathbf{X}^s(\theta))$.

3. Minimize $\|\widehat{\psi} - \frac{1}{S} \sum_{s=1}^S \widehat{\psi}(\mathbf{y}^s(\theta), \mathbf{X}^s(\theta))\|_W$ over θ.

As before, the measurement errors are simulated as an AR$(2q+1)$ process, but this need not be the true data generating process. The AR model only needs to provide correct first $2q+2$ autocovariances of the measurement error. The parameters of this AR model in step (ii) are calibrated using the Yule-Walker equations. In step (iii), the simulated latent regressor x_t^s is postulated to be a linear function of the observed regressors $(X_t, \ldots, X_{t-(2q+1)})$. The parameters $(\varphi_0, \ldots, \varphi_{2q+1})'$ are chosen so as to preserve the autocovariances of the observed regressors. Since these in turn depend on the assumed model parameters $(\Gamma_\epsilon(0), \ldots, \Gamma_\epsilon(2q+1))'$, x_t^s will need to be re-calculated in each simulation. The simulator produces latent regressors $\{x_t^s\}$ that are independent from $\{u_t^s\}$. The exogeneity guarantees the validity of all cross-covariances between y_t^s and X_t^s.

To derive the asymptotic properties of our SMD estimator, we impose the following assumptions.

Assumption B

(a) The $(2q+2) \times (2q+2)$ *autocovariance matrix* R_q *is non-singular.*

(b) $\{(y_t, X_t)'\}$ *is* α*-mixing of size* $-r/(r-1)$, $r > 1$, *and for some* $\delta > 0$ *and all* t*:* $E(|X_t|^{2r+\delta}) \leqslant \Delta < \infty$, $E(|y_t|^{2r+\delta}) \leqslant \Delta < \infty$.

The nonsingularity condition in Assumption B(a) ensures the R_q matrix used to compute the coefficients $\varphi_0, \ldots, \varphi_{2q+1}$ is invertible. This is needed in the simulation of x_t^s in Step 2(iii) of the SMD algorithm. The mixing and bounded moment conditions in B(b) are used to establish the almost sure convergence of the auxiliary statistics. The properties of our algorithm are formally stated below.

Lemma 3 *Let Assumptions A and B hold. Assume in addition that the binding function $\psi : \theta \mapsto \psi(\theta)$ is invertible. Then, under the SMD algorithm described above, the SMD estimator $\widehat{\theta}^S$ is a consistent estimator of θ.*

Lemma 3 is stated for estimation of θ from $\widehat{\psi}$ as defined in (13) and (14), respectively. Of note is that there are $2q+2$ nuisance parameters in θ. The parameters pertaining to the persistence of ϵ_t are needed for identification of α and β, and hence are regarded as nuisance. However, the magnitude of $\widehat{\Gamma}_\epsilon(0)$ relative to $\widehat{\Gamma}_X(0)$ can be used to gauge the severity of measurement error. Furthermore, the persistence of the latent process can be recovered from the relation $\widehat{\Gamma}_x(j) = \widehat{\Gamma}_X(j) - \widehat{\Gamma}_\epsilon(j)$. This sheds light on whether the assumptions of our analysis are satisfied.

Nonetheless, for large q, the nuisance parameters can increase the dimension of θ substantially. To avoid the proliferation of nuisance parameters, we can impose an additional restriction that would require $\Gamma_\epsilon(1), \ldots, \Gamma_\epsilon(2q+1)$ to be well approximated by a parameter vector $\phi = (\phi_1, \ldots, \phi_m)'$ with

$$m \leqslant 2q + 1. \tag{23}$$

Since the ADL(p,q) model has $p + q + 1$ parameters and $m + 2$ nuisance parameters, (23) is a necessary order condition for identification under the ϕ parameterization. Such a parametrization is not necessary for our estimation method to work. Its role is to help solve numerical optimization issues if the lag q of the ADL(p,q) model happens to be large. For instance, in the ADL(1,1) model, the condition would require that $\Gamma_\epsilon(1)$, $\Gamma_\epsilon(2)$ and $\Gamma_\epsilon(3)$ be well approximated by a $m \leqslant 3$ dimensional parameter vector ϕ. The parameterization has no effect on estimation. It is only in larger models that a smaller m may be desirable.

We should reiterate that the order conditions (19) and (23) are not 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), Meijer, Spierdijk, and Wansbeek (2012), and Erickson and Whited (2000, 2002), higher order moments of ϵ_t or of X_t can also be used to achieve identification.

4 Monte Carlo Simulations and Application

4.1 Simulations

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

$$\begin{aligned} y_t &= \alpha y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + u_t, \quad u_t \sim iidN(0, \sigma_u^2), \\ x_t &= \rho_x x_{t-1} + u_{xt}, \quad u_{xt} \sim iidN(0, \sigma_{ux}^2), \\ X_t &= x_t + \epsilon_t, \qquad \epsilon_t = e_t + \theta e_{t-1}, \quad e_t \sim iidN(0, \sigma_e^2). \end{aligned}$$

The parameters are $\rho_x = (0.2, 0.5, 0.8)$, $\theta = 0$ (case 'ϵ_t WN' in the tables) or 0.4 (case 'ϵ_t MA(1)' in the tables), $\alpha = 0$ (ADL(0,0) model) or 0.6 (ADL(1,0) and ADL(1,1) models), $\beta_1 = 0$ (ADL(0,0) and ADL(1,0) models) or 0.5 (ADL(1,1) model), and $\beta_0 = 1$. The measurement error process is calibrated such that the signal-to-noise ratio is $R^2 = \frac{\text{var }(x_t)}{\text{var }(X_t)} = 0.7$. 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 $\sigma_u^2 = \sigma_{ux}^2 = 1$. In practice, we do not know if ϵ_t is serially correlated or not. Thus, we always estimate a model that allows 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 begin with the simple regression model when $\alpha = \beta_1 = 0$. As these parameters are not estimated, $\theta = (\beta_0, \sigma_u^2, \Gamma_\epsilon(0), \phi)'$ and $\widehat{\psi} = (\widehat{\beta}, \widehat{\Gamma}_{\widehat{V}}(0), \widehat{\Gamma}_{\widehat{V}}(1), \widehat{\Gamma}_{\widehat{V}X}(1,0))'$. 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 $\theta = (\alpha,\ \beta_0,\ \beta_1,\ \sigma_u^2,\ \Gamma_\epsilon(0),\ \Gamma_\epsilon(1),\ \Gamma_\epsilon(2),\ \Gamma_\epsilon(3))'$ with $\alpha = 0.6$ and $\beta_1 = 0$ or 0.5. The auxiliary statistics are $\widehat{\psi}(\theta) = (\widehat{\alpha},\ \widehat{\beta}_0,\ \widehat{\beta}_1,\ \widehat{\Gamma}_{\widehat{V}}(0),\ \widehat{\Gamma}_{\widehat{V}}(1),\ \widehat{\Gamma}_{\widehat{V}}(2),\ \widehat{\Gamma}_{\widehat{V}X}(1,0),\ \widehat{\Gamma}_{\widehat{V}X}(2,0))'$. We report the estimated short- and long-run response of y_t to x_t as given by $\widehat{\beta}_0$ and $\widehat{\beta}(1) = \widehat{\beta}_0 + \widehat{\beta}_1$. 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 are reasonably precise and exhibit some downward biases that tend to increase with the degree of persistence in x_t. Table 3 shows results for the ADL(1,1) model. While the CMD estimator reduces substantially the large bias of the OLS estimator, the SMD estimator provides further bias corrections.

4.2 Long-Run Risks Model

The risks that affect consumption and their role in explaining the equity premium puzzle have been a focus of extensive research effort. Bansal and Yaron (2004) propose a model where consumption growth contains a small long-run persistent predictive component. Their basic constant-volatility specification can be cast as an ADL(0,0) model with uncorrelated measurement errors:

$$\begin{aligned} y_{t+1} &= \mu_y + \beta x_t + \sigma_u u_{t+1} \\ X_{t+1} &= x_t + \sigma_\epsilon \epsilon_{t+1}, \end{aligned}$$

where $y_{t+1} = \triangle d_{t+1}$ is the dividend growth rate, $X_{t+1} = \triangle c_{t+1}$ is the consumption growth rate, x_t is a latent AR(1) process with autoregressive coefficient ρ_x, and u_{t+1} and ϵ_{t+1} are mutually independent, $iidN(0, \sigma_u^2)$ and $iidN(0, \sigma_\epsilon^2)$ errors, respectively.[11] In order to calibrate the dividend growth volatility, the model requires that β, which can be interpreted as the leverage ratio on expected consumption growth (Bansal and Yaron (2004)), and σ_u/σ_ϵ are both greater than one. Also, high persistence of the latent component x_t, measured by a value of ρ near one, is critical for the potential resolution of the equity premium puzzle. Below, we will evaluate the plausibility of these parameter values and restrictions using our proposed method. We note that our approach is similar in spirit to the one used by Contanstinides and Ghosh (2011) but it is based on a different set of moment conditions.

Before we proceed with the estimation results, we make several remarks. First, the OLS estimator of β from a regression of y_{t+1} on the observed X_{t+1} (instead of the latent x_t) is downward biased. The IV estimator that uses X_t as an instrument is asymptotically valid. But both of these estimators do not provide information about the multitude of the measurement error, the implied value of the persistence parameter ρ_x and the variability of the long-run risks component. Since the moments employed in estimation can be computed analytically, we use both the classical and simulated method of moments. The results from the CMD estimation are very similar to those from the SMD but we report only the SMD estimates due to their bias-correction properties.

We report results for quarterly (1952:Q2–2012:Q4) and annual (1931–2009) data. The consumption growth is the percentage growth rate of real per-capita personal consumption expenditures on

[11] This specification of the model assumes that $E(X_t) = E(x_t) = \mu$. Alternatively, one could assume $X_{t+1} = \mu + x_t + \sigma_\epsilon \epsilon_{t+1}$ and $E(x_t) = 0$ (as in Bansal and Yaron (2004)) and identify μ from the mean of the observed X_t.

nondurable goods and services from the Bureau of Economic Analysis. The dividend growth is the percentage growth rate of real dividends on the Center for Research in Security Prices (CRSP) value-weighted stock market portfolio. For the SMD estimation, N is set equal to 100. The OLS, IV and SMD estimates of the parameters μ_y, β, σ_u^2 and σ_ϵ^2 are presented in Table 4.

The first interesting observation from Table 4 is that SMD estimates of β are larger, both economically and statistically, than the IV and, especially, OLS estimates. This lends support to the "levered" nature of dividends and the larger values of β used for calibrating the model in Bansal and Yaron (2004). For annual data, that includes the Great Depression, the ratio σ_u/σ_ϵ is 2.64. This is lower than the value of 4.5 used in Bansal and Yaron (2004). For the post-war quarterly data, this ratio is even lower. We attribute this to the larger variance of the measurement error (or transitory component) estimated by SMD. To put this in perspective, note that $\Gamma_X(0)$ for quarterly data is 0.547 and for annual data is 7.899 so that the variance of the measurement error is 69% and 61% of the variance of the observed consumption growth, respectively. Recall from Section 2.2 that a quick estimate of the measurement error variance can be backed out directly from the data, i.e. $\sigma_\epsilon^2 = \Gamma_X(0) - \Gamma_X(1)^2/\Gamma_X(2)$. Using the sample values of $\Gamma_X(0)$, $\Gamma_X(1)$ and $\Gamma_X(2)$ these back-of-the-envelope calculations yield an estimate of 0.369 for σ_ϵ^2 for quarterly data. Furthermore, using our SMD estimate of σ_ϵ^2, we can compute the implied estimate of ρ_x as $\rho_x = \Gamma_X(1)/\Gamma_x(0)$, where $\Gamma_x(0) = \Gamma_X(0) - \sigma_\epsilon^2$. This gives estimates for ρ_x of 0.630 for quarterly data and 0.480 for annual data. Although these values are far from unity, they do seem to suggest a presence of a persistent, long-run component in consumption growth.

5 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 exogenous and thus cannot be held fixed in simulations. The algorithm can be extended to dynamic panels and can accommodate additional regressors. The proposed methodology can be useful when external instruments are either unavailable or are weak.

A Appendix: Proofs

Proof of Lemma 1 Write the binding function as:

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

First, consider the case $\beta = 0$. Note that $\psi_1 = \beta\frac{\Gamma_x(0)}{\Gamma_X(0)} = 0$. But $\Gamma_X(0) - \Gamma_\epsilon(0) = \Gamma_x(0) \neq 0$. Hence, $\beta = 0$ if and only if $\psi_1 = 0$, and $\beta = 0$ is directly identifiable from ψ_1. For σ_u^2, we have $\sigma_u^2 = \psi_2$, so $(\beta = 0, \sigma_u^2)'$ is identified from ψ. Next, we consider the case $\beta \neq 0$. In this case, $\psi_1 \neq 0$ and we can solve for β by considering

$$A \equiv \Gamma_X(1)\psi_1^2 + 2\psi_4\psi_1 + \psi_3.$$

Using the definition of ψ, this quantity can be computed in two ways: $A = \beta^2(\Gamma_X(1) - \Gamma_\epsilon(1)) = \beta^2\Gamma_x(1)$ and $A = \beta(\psi_4 + \Gamma_X(1)\psi_1)$. So if $\Gamma_x(1) \neq 0$, then $A \neq 0$ and we use the two expressions for A to obtain:

$$\beta = \frac{A}{\psi_4 + \Gamma_X(1)\psi_1}.$$

For σ_u^2, consider

$$D \equiv \psi_2\psi_4 - \Gamma_X(0)\psi_1\psi_3 + \Gamma_X(1)\psi_1\psi_2 - \Gamma_X(0)\psi_1^2\psi_4.$$

Then, $D = \sigma_u^2(\psi_4 + \Gamma_X(1)\psi_1)$. Dividing both sides by $\psi_4 + \Gamma_X(1)\psi_1 \neq 0$ gives

$$\sigma_u^2 = \frac{D}{\psi_4 + \Gamma_X(1)\psi_1}.$$

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

$$B \equiv \Gamma_X(0)\left(\psi_3 + \psi_1\psi_4\right),$$

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

$$\Gamma_\epsilon(0) = \frac{B}{A} = \Gamma_X(0)\frac{\psi_3 + \psi_1\psi_4}{\Gamma_X(1)\psi_1^2 + 2\psi_4\psi_1 + \psi_3}.$$

Finally, for $\Gamma_\epsilon(1)$, let $C \equiv -\psi_4^2 + \Gamma_X(1)\psi_3$, and note that $C = \Gamma_\epsilon(1)A$. Under our assumptions, $A \neq 0$ and $\Gamma_\epsilon(1)$ is identified as

$$\Gamma_\epsilon(1) = \frac{C}{A} = \frac{-\psi_4^2 + \Gamma_X(1)\psi_3}{\Gamma_X(1)\psi_1^2 + 2\psi_4\psi_1 + \psi_3}.$$

Proof of Lemma 2 The analysis can be simplified by noting that σ_u^2 will identified from $\Gamma_{\widehat{V}}(0)$. Thus, we only need to consider identification of

$$\theta = (\alpha, \beta, \Gamma_\epsilon(0), \Gamma_\epsilon(1))'$$

from

$$\widehat{\psi} = (\widehat{\alpha}, \widehat{\beta}, \widehat{\Gamma}_{\widehat{V}}(1), \widehat{\Gamma}_{\widehat{V}X}(1,0))'.$$

As before, θ is globally identified from $\psi = (\psi_1, \psi_2, \psi_3, \psi_4)'$ if the binding function $\psi(\theta)$ is invertible. Consider then the system of equations $\psi(\theta) = \psi$ 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:

$$\begin{aligned} \Gamma_{Wy}(0,0) &= \Gamma_W(0)\begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix} \\ \Gamma_{Wy}(0,0)\Gamma_W(0)^{-1}\Gamma_{Wy}(1,0) &= \begin{pmatrix}\psi_1 & \psi_2\end{pmatrix}\Gamma_W(1)\begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix} - \psi_3 \\ \Gamma_{yX}(1,0) &= \psi_4 + \begin{pmatrix}\psi_1 & \psi_2\end{pmatrix}\Gamma_{WX}(1,0). \end{aligned} \tag{24}$$

The system of 4 equations in 4 unknowns in (24) has the important feature that only the left-hand side of (24) depends on θ. The right hand side consists either of $(\psi_1, \ldots, \psi_4)$ or the elements in $\Gamma_W(0)$, $\Gamma_W(1)$ and $\Gamma_{WX}(1,0)$ for which sample estimates are available. Global identifiability of θ from ψ holds if it can be established that the system (24) has a unique solution in θ.

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, and the above quantity is strictly positive. Thus, $\psi_2 = 0$ if and only if $\beta = 0$. In this case, $\alpha = \psi_1$ and $(\alpha, \beta = 0)$ 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 $E(X_t y_{t-j}) = \sum_{i=0}^{\infty} \alpha^i(\beta\Gamma_x(j+i))$ and

$$\Gamma_{yX}(1,0) = \beta\left[\Gamma_x(1) + \sum_{i=1}^{\infty} \alpha^i \Gamma_x(1+i)\right], \tag{25}$$

so $\Gamma_{yX}(1,0) = 0$ occurs, for example, whenever x is white noise. In this case, α can be directly identified from ψ_1, $\alpha = \psi_1$. As for β, notice that the components $\psi_2(\theta), \psi_3(\theta), \psi_4(\theta)$ are as in the ADL(0,0) case and identification can proceed as in Lemma 1 provided $\Gamma_x(1) \neq 0$.

It remains to consider the case $\beta \neq 0$, $\Gamma_{yX}(1,0) \neq 0$. For this, we further write the elements on the left-hand side of (24) in terms of $(\alpha, \beta, \Gamma_\epsilon(0), \Gamma_\epsilon(1))$.

$$\alpha = \psi_1 + \frac{\Gamma_{yX}(1,0)(\psi_3 + \psi_2\psi_4)}{(\Gamma_y(0)\Gamma_{yX}(2,0) - \Gamma_y(1)\Gamma_{yX}(1,0))\psi_1 - \Gamma_{yX}(1,0)(\Gamma_{yX}(0,0) - \Gamma_{yX}(2,0))\psi_2}.$$

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:

$$\alpha - \psi_1 = \frac{N}{D},$$

with

$$\begin{aligned}
N &= \Gamma_{yX}(1,0)(\psi_3 + \psi_2\psi_4) \\
D &= (\Gamma_y(0)\Gamma_{yX}(2,0) - \Gamma_y(1)\Gamma_{yX}(1,0))\psi_1 - \Gamma_{yX}(1,0)(\Gamma_{yX}(0,0) - \Gamma_{yX}(2,0))\psi_2.
\end{aligned}$$

Note that

$$\alpha - \psi_1 = -\beta\Gamma_\epsilon(0)\frac{\Gamma_{yX}(1,0)}{\Gamma_y(0)\Gamma_X(0) - \Gamma_{yX}(1,0)^2} \neq 0.$$

Thus, $D = 0$ if and only if $\psi_3 + \psi_2\psi_4 = 0$. Moreover, it also holds that:

$$\Gamma_\epsilon(0)\beta = \frac{(\psi_3 + \psi_2\psi_4)(\Gamma_{yX}(1,0)^2 - \Gamma_X(0)\Gamma_y(0))}{D},$$

so $\psi_3 + \psi_2\psi_4 = 0$ if and only if $\Gamma_\epsilon(0)\beta = 0$ which we excluded. Thus, both $N \neq 0$ and $D \neq 0$.

For β, the solution is:

$$\begin{aligned}
\beta &= \psi_2 - \frac{\Gamma_y(0)(\psi_3 + \psi_2\psi_4)}{(\Gamma_y(0)\Gamma_{yX}(2,0) - \Gamma_y(1)\Gamma_{yX}(1,0))\psi_1 - \Gamma_{yX}(1,0)(\Gamma_{yX}(0,0) - \Gamma_{yX}(2,0))\psi_2} \\
&= \psi_2 - \frac{\Gamma_y(0)}{\Gamma_{yX}(1,0)}(\alpha - \psi_1).
\end{aligned}$$

Thus $(\alpha, \beta \neq 0)'$ are identified from ψ. Finally, for $\Gamma_\epsilon(0)$ we have:

$$\Gamma_\epsilon(0)\beta = \frac{(\psi_3 + \psi_2\psi_4)(\Gamma_{yX}(1,0)^2 - \Gamma_X(0)\Gamma_y(0))}{D},$$

so if in addition $\beta \neq 0$, $\Gamma_\epsilon(0)$ is identified. Similarly, $\Gamma_\epsilon(1)$ is then also identified from ψ_4.

Proof of Lemma 3 To establish the consistency of $\widehat{\theta}^S$, we need to check the following high-level conditions:

1. $\widehat{\psi} \xrightarrow{a.s.} \psi$ as $T \to \infty$;

2. $\frac{1}{S}\sum_{s=1}^{S} \widehat{\psi}(\mathbf{y}^s(\theta), \mathbf{X}^s(\theta)) \xrightarrow{a.s.} \psi(\theta)$ as $S \to \infty$;

3. $\psi : \theta \mapsto \psi(\theta)$ is invertible.

Using the same reasoning as in Gourieroux, Monfort, and Renault (1993) (see their proof of Proposition 1), under 1 and 2, the limit of the optimization problem:

$$\min_{\theta} \left\| \widehat{\psi} - \frac{1}{S} \sum_{s=1}^{S} \widehat{\psi}(\mathbf{y}^s(\theta), \mathbf{X}^s(\theta)) \right\|_{\widehat{W}}$$

with $\widehat{W} \xrightarrow{a.s.} W$, is

$$\min_{\theta} \|\psi - \psi(\theta)\|_W = \theta.$$

Then, the consistency of the SMD estimator $\widehat{\theta}^S$ follows. We now check the high-level conditions 1 and 2. Condition 3 is assumed.

Condition 1. The auxiliary statistics $\widehat{\psi}$ in (14) is a continuous function of $\widehat{\Gamma}_y(j)$, $0 \leqslant j \leqslant p+q+1$, $\widehat{\Gamma}_X(k)$, $0 \leqslant k \leqslant 2q+1$, and $\widehat{\Gamma}_{yX}(l,m)$, $0 \leqslant l \leqslant p+q+1$, $0 \leqslant m \leqslant 2q+1$. Moreover, by Assumption B(b) and Cauchy-Schwartz inequality, there exists $\delta_1 = \delta/2 > 0$ such that for all t: $E(|y_t y_{t-j}|^{r+\delta_1}) \leqslant \left[E(|y_t|^{2r+\delta}) E(|y_{t-j}|^{2r+\delta})\right]^{1/2} \leqslant \Delta < \infty$ for all $0 \leqslant j \leqslant p+q+1$, with a similar result for all the other covariances and cross-covariances. Then, by Theorem 3.47 in White (1984), $\widehat{\Gamma}_y(j)$, $\widehat{\Gamma}_X(k)$, and $\widehat{\Gamma}_{yX}(l,m)$ converge almost surely to $\Gamma_y(j)$, $\Gamma_X(k)$, and $\Gamma_{yX}(l,m)$, respectively. Thus, by the continuous mapping theorem, $\widehat{\psi} \xrightarrow{a.s.} \psi$.

Condition 2. First, note that under the full rank assumption B(a), the SMD algorithm is implementable. We next discuss the mixing properties of the simulated variables. If $\{X_t\}$ is α-mixing of size $-a$, then by Theorem 3.49 in White (1984), $\{x_t^s\}$ is α-mixing of size $-a$. Being a Gaussian AR($2q+1$) process, $\{\epsilon_t^s\}$ is α-mixing of size $-a$ for any $a \in \mathbb{R}$ since the mixing coefficients $\alpha(m)$ decay exponentially with m (see, e.g., Example 3.46 in White (1984)). In addition, $\{x_t^s\}$, $\{u_t^s\}$ and $\{\epsilon_t^s\}$ are independent. Thus, $\{(y_t^s, X_t^s)'\}$ is α-mixing of size $-a$. Under Assumption B(b), $a = r/(r-1)$ with $r > 1$. We now check that the simulated data satisfies the required moment conditions. For this, note that for some constant $1 < C < +\infty$ (that depends on r and δ), we have:

$$E^s(|X_t^s|^{2r+\delta}) = E^s(|x_t^s + \epsilon_t^s|^{2r+\delta}) \leqslant C\left[E^s(|x_t^s|^{2r+\delta}) + E^s(|\epsilon_t^s|^{2r+\delta})\right].$$

Now, there exists Δ_1 such that $E^s(|x_t^s|^{2r+\delta}) \leqslant \Delta_1 < \infty$ because under Step 2(iii) of the SMD algorithm, x_t^s is a linear function of $(X_t, \ldots, X_{2q+1})$, which all satisfy $E(|X_t|^{2r+\delta}) \leqslant \Delta < \infty$. Next, under Step 2(ii), ϵ_t^s is an AR(2q+1) Gaussian process so there exists Δ_2 such that: $E^s(|\epsilon_t^s|^{2r+\delta}) \leqslant \Delta_2 < \infty$. Thus, there exists $\bar{\Delta}$ such that for all t: $E^s(|X_t^s|^{2r+\delta}) \leqslant \bar{\Delta} < \infty$. Using a similar reasoning, under Step 2(v), since $E^s(|x_t^s|^{2r+\delta}) \leqslant \Delta_1 < \infty$, $E^s(|u_t^s|^{2r+\delta}) \leqslant \Delta_3 < \infty$ (since u_t is Gaussian), there exists $\widetilde{\Delta}$ such that: $E(|y_t|^{2r+\delta}) \leqslant \widetilde{\Delta} < \infty$. This means that the simulated data satisfy the same mixing and bounded moment conditions B(b) as the true data. Using the same reasoning as in the proof of Condition 1, it follows that the auxiliary statistics computed over simulated data converge almost surely to their limit $E^s[\widehat{\psi}(\mathbf{y}^S(\theta), \mathbf{X}^S(\theta))]$. It remains to show that this limit equals $\psi(\theta)$. For this, recall that the auxiliary statistics $\widehat{\psi}$ in (14) computed over the simulated data depends on $\widehat{\Gamma}_y^s(j)$, $0 \leqslant j \leqslant p+q+1$, $\widehat{\Gamma}_X^s(k)$, $0 \leqslant k \leqslant 2q+1$, and $\widehat{\Gamma}_{yX}^s(l,m)$, $0 \leqslant l \leqslant p+q+1$, $0 \leqslant m \leqslant 2q+1$. The proposed SMD algorithm ensures that:

$$\begin{aligned} \Gamma_y^s(j) &= \Gamma_y(j), \quad 0 \leqslant j \leqslant p+q+1 \\ \Gamma_X^s(k) &= \Gamma_X(k), \quad 0 \leqslant k \leqslant 2q+1 \\ \Gamma_{yX}^s(l,m) &= \Gamma_{yX}(l,m), \quad 0 \leqslant l \leqslant p+q+1, 0 \leqslant m \leqslant 2q+1. \end{aligned}$$

Thus,

$$E^s[\widehat{\psi}(\mathbf{y}^S(\theta), \mathbf{X}^S(\theta))] = E[\widehat{\psi}(\mathbf{y}, \mathbf{X})] = \psi(\theta),$$

which establishes Condition 2.

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.

BANSAL, R., AND A. YARON (2004): "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59, 1481–1509.

BIØRN, E. (1992): "Panel Data with Measurement Errors," in *The Econometrics of Panel Data. Handbook of the Theory and Applications*, ed. by L. Matyas, and P. Sevestre, Dordrecht. Kluwer.

——— (2014): "Serially Correlated Measurement Errors in Time Series Regression: The Potential of Instrumental Variablees Estimators," University of Oslo.

BUONACCORSI, J. (2012): "Mesaurement Error in Dynamic Models," in *Lecture Notes in Statistics 211: Longitudinal Data Analysis Subject to Measurement Errors, Missing Values, and Outliers*, ed. by B. Sutradhar, vol. 211, pp. 53–76. Springer.

CONTANSTINIDES, G., AND A. GHOSH (2011): "Asset Pricing Tests with Long Run Risks in Consumption Growth," *Review of Asset Pricing Studies*, 1, 93–136.

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.

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.

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 Friendma's Permanent Income Hypothesis," *Journal of Monetary Economics*, 26:2, 267–283.

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," *REVSTAT Statistical Journal*, 8(1), 57–80.

GOLDBERGER, A. (1972): "Structural Equation Methods in the Social Sciences," *Econometrica*, 40(6), 979–1001.

GOURIEROUX, C., A. MONFORT, AND E. RENAULT (1993): "Indirect Inference," *Journal of Applied Econometrics*, 85, 85–118.

GRETHER, D., AND G. MADDALA (1973): "Errors in Variables and Serially Correlated Disturbances in Distributed Lag Models," *Econometrica*, 41:2, 255–262.

GRILLICHES, Z., AND J. HAUSMAN (1986): "Errors-in-Variables in Panel Data," *Journal of Econometrics*, 32:3, 93–118.

JIANG, W., AND B. TURNBULL (2004): "The Indirect Method: Inference Based on Intermediate Statistics: A Synthesis and Examples," *Statistical Science*, 19:2, 239–263.

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 Avaialbe, 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.

MARAVALL, A. (1979): *Identification in Dynamic Shock-Error Models*. Springer-Verlag, New York.

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.

ORPHANIDES, A., AND S. VAN NORDEN (2002): "The Unrelaibility 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.

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.

Wansbeek, T., and E. Meijer (2000): *Measurement Error and Latent Variables in Econometrics.* Elseiver, Amsterdam.

White, H. (1984): *Asymptotic Theory for Econometricians.* Academic Press, New York.

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.

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

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

		Estimates of $\beta_0 = 1$					Standard Deviations				
T	ρ_x	OLS	IDEAL	IV	CMD	SMD	OLS	IDEAL	IV	CMD	SMD
ϵ_t WN											
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):

$$\begin{aligned} \theta &= (\alpha, \beta_0, \beta_1, \sigma_u^2, \Gamma_\epsilon(0), \Gamma_\epsilon(1), \Gamma_\epsilon(2), \Gamma_\epsilon(3))' \\ \widehat{\psi}(\theta) &= (\widehat{\alpha}, \widehat{\beta}_0, \widehat{\beta}_1, \widehat{\Gamma}_{\widehat{V}}(0), \widehat{\Gamma}_{\widehat{V}}(1), \widehat{\Gamma}_{\widehat{V}}(2), \widehat{\Gamma}_{\widehat{V}X}(1,0), \widehat{\Gamma}_{\widehat{V}X}(2,0))'. \end{aligned}$$

		Estimates of $\beta_0 = 1$ and $\beta(1) = \beta_0 + \beta_1 = 1$						Standard Deviations					
T	ρ_x	OLS		CMD		SMD		OLS		CMD		SMD	
ϵ_t WN		β_0	$\beta(1)$	β_0	$\beta(1)$	β_0	$\beta(1)$	β_0	$\beta(1)$	β_0	$\beta(1)$	β_0	$\beta(1)$
200	0.2	0.711	0.657	1.051	1.030	1.078	1.067	0.070	0.095	0.229	0.242	0.130	0.211
200	0.5	0.679	0.669	0.946	0.907	1.035	1.031	0.069	0.087	0.209	0.189	0.143	0.192
200	0.8	0.545	0.578	0.896	0.814	0.925	0.948	0.063	0.076	0.172	0.167	0.160	0.188
500	0.2	0.709	0.648	1.000	0.973	1.073	1.073	0.044	0.060	0.180	0.176	0.085	0.136
500	0.5	0.677	0.660	0.913	0.872	1.035	1.036	0.044	0.055	0.142	0.121	0.094	0.120
500	0.8	0.543	0.569	0.894	0.800	0.934	0.937	0.041	0.049	0.116	0.105	0.114	0.132
1000	0.2	0.709	0.646	0.961	0.931	1.071	1.068	0.031	0.042	0.132	0.126	0.056	0.092
1000	0.5	0.677	0.658	0.901	0.858	1.033	1.033	0.030	0.038	0.098	0.085	0.061	0.081
1000	0.8	0.543	0.567	0.897	0.797	0.937	0.928	0.028	0.033	0.082	0.073	0.096	0.108
ϵ_t MA(1)													
200	0.2	0.696	0.741	1.061	1.060	1.080	1.076	0.068	0.099	0.254	0.254	0.137	0.223
200	0.5	0.656	0.758	0.990	0.971	1.033	1.047	0.066	0.090	0.211	0.186	0.142	0.199
200	0.8	0.522	0.675	0.923	0.865	0.898	1.003	0.058	0.079	0.158	0.168	0.154	0.206
500	0.2	0.693	0.733	1.030	1.026	1.077	1.076	0.043	0.062	0.217	0.197	0.092	0.148
500	0.5	0.653	0.750	0.968	0.949	1.039	1.044	0.042	0.057	0.147	0.120	0.098	0.132
500	0.8	0.520	0.668	0.925	0.857	0.941	1.004	0.037	0.050	0.113	0.111	0.108	0.146
1000	0.2	0.693	0.731	1.009	1.001	1.075	1.072	0.030	0.043	0.174	0.153	0.062	0.100
1000	0.5	0.653	0.748	0.960	0.938	1.035	1.038	0.029	0.039	0.103	0.086	0.067	0.091
1000	0.8	0.520	0.667	0.930	0.852	0.968	1.001	0.026	0.034	0.081	0.081	0.078	0.106

Table 3: ADL(1,1):

$$\begin{aligned} \theta &= (\alpha, \beta_0, \beta_1, \sigma_u^2, \Gamma_\epsilon(0), \Gamma_\epsilon(1), \Gamma_\epsilon(2), \Gamma_\epsilon(3))' \\ \widehat{\psi}(\theta) &= (\widehat{\alpha}, \widehat{\beta}_0, \widehat{\beta}_1, \widehat{\Gamma}_{\widehat{V}}(0), \widehat{\Gamma}_{\widehat{V}}(1), \widehat{\Gamma}_{\widehat{V}}(2), \widehat{\Gamma}_{\widehat{V}X}(1,0), \widehat{\Gamma}_{\widehat{V}X}(2,0))'. \end{aligned}$$

		Estimates of $\beta_0 = 1$ and $\beta(1) = \beta_0 + \beta_1 = 1.5$						Standard Deviations					
T	ρ_x	OLS		CMD		SMD		OLS		CMD		SMD	
ϵ_t WN		β_0	$\beta(1)$	β_0	$\beta(1)$	β_0	$\beta(1)$	β_0	$\beta(1)$	β_0	$\beta(1)$	β_0	$\beta(1)$
200	0.2	0.697	0.966	0.958	1.378	1.041	1.518	0.074	0.104	0.187	0.240	0.151	0.227
200	0.5	0.697	0.973	0.965	1.358	0.999	1.441	0.074	0.097	0.167	0.196	0.158	0.219
200	0.8	0.601	0.826	0.994	1.316	0.957	1.348	0.071	0.091	0.182	0.198	0.166	0.214
500	0.2	0.695	0.958	0.929	1.344	1.032	1.526	0.046	0.066	0.119	0.155	0.097	0.150
500	0.5	0.695	0.965	0.952	1.348	1.002	1.462	0.047	0.062	0.107	0.124	0.106	0.156
500	0.8	0.600	0.818	0.990	1.314	0.972	1.382	0.046	0.059	0.119	0.120	0.112	0.143
1000	0.2	0.694	0.956	0.919	1.331	1.026	1.520	0.032	0.046	0.082	0.108	0.066	0.104
1000	0.5	0.695	0.962	0.949	1.344	1.006	1.473	0.033	0.043	0.074	0.086	0.076	0.117
1000	0.8	0.600	0.816	0.993	1.313	0.985	1.403	0.032	0.040	0.083	0.082	0.076	0.090
ϵ_t MA(1)													
200	0.2	0.715	1.070	0.970	1.414	1.036	1.523	0.071	0.104	0.181	0.228	0.155	0.226
200	0.5	0.706	1.080	0.974	1.394	0.987	1.457	0.069	0.096	0.154	0.183	0.153	0.203
200	0.8	0.605	0.947	0.988	1.353	0.929	1.400	0.063	0.090	0.161	0.185	0.154	0.213
500	0.2	0.713	1.063	0.945	1.391	1.031	1.523	0.045	0.065	0.119	0.149	0.101	0.150
500	0.5	0.703	1.074	0.962	1.392	0.994	1.462	0.044	0.061	0.100	0.116	0.101	0.138
500	0.8	0.603	0.941	0.985	1.357	0.949	1.420	0.041	0.058	0.109	0.113	0.108	0.157
1000	0.2	0.713	1.061	0.939	1.385	1.026	1.517	0.031	0.046	0.082	0.105	0.067	0.101
1000	0.5	0.703	1.072	0.960	1.390	0.996	1.458	0.030	0.042	0.069	0.081	0.073	0.103
1000	0.8	0.603	0.939	0.988	1.357	0.962	1.437	0.028	0.040	0.076	0.078	0.080	0.123

Table 4: Estimation results for the long-run risks model.

	OLS	IV	SMD
quarterly data			
μ_y	0.258	-0.298	-0.822
β	0.339	2.025	4.736
σ_u^2	2.760	4.326	0.148
σ_ϵ^2	-	-	0.379
annual data			
μ_y	-3.335	-3.955	-9.938
β	2.089	2.365	5.015
σ_u^2	83.28	83.90	33.56
σ_ϵ^2	-	-	4.818