MEMORANDUM

No 28/2014

Serially Correlated Measurement Errors in Time Series Regression: The Potential of Instrumental Variable Estimators

Erik Biørn

ISSN: 0809-8786

Department of Economics
University of Oslo

This series is published by the
University of Oslo
Department of Economics

P. O.Box 1095 Blindern
N-0317 OSLO Norway
Telephone: + 47 22855127
Fax: + 47 22855035
Internet: http://www.sv.uio.no/econ
e-mail: econdep@econ.uio.no

In co-operation with
The Frisch Centre for Economic Research

Gaustadalleén 21
N-0371 OSLO Norway
Telephone: +47 22 95 88 20
Fax: +47 22 95 88 25
Internet: http://www.frisch.uio.no
e-mail: frisch@frisch.uio.no

Last 10 Memoranda

No 27/14	Erik Biørn *The Price-Quantity Decomposition of Capital Values Revisited: Framework and Examples*
No 26/14	Olav Bjerkholt *Econometric Society 1930: How it Got Founded*
No 25/14	Nils Chr. Framstad *The Effect of Small Intervention Costs on the Optimal Extraction of Dividends and Renewable Resources in a Jump-Diffusion Model*
No 24/14	Leif Andreassen, Maria Laura Di Tommaso and Steinar Strøm *Wages Anatomy: Labor Supply of Nurses and a Comparison with Physicians*
No 23/14	Derek J. Clark, Tore Nilssen and Jan Yngve Sand *Keep on Fighting: Dynamic Win Effects in an All-Pay Auction*
No 22/14	John K. Dagsvik and Zhiyang Jia *Labor Supply as a Choice among Latent Jobs: Unobserved Heterogeneity and Identification*
No 21/14	Simen Gaure *Practical Correlation Bias Correction in Two-way Fixed Effects Linear Regression*
No 20/14	Rolf Aaberge, Tarjei Havnes and Magne Mogstad *A Theory for Ranking Distribution Functions*
No 19/14	Alice Ciccone *Is It All About CO2 Emissions? The Environmental Effects of Tax Reform for New Vehicles in Norway*
No 18/14	Mikolaj Czajkowski, Nick Hanley and Karine Nyborg *Social Norms, Morals and Self-interest as Determinants of Pro-environment Behaviours*

Previous issues of the memo-series are available in a PDF® format at:
http://www.sv.uio.no/econ/english/research/unpublished-works/working-papers/

Serially Correlated Measurement Errors in Time Series Regression: The Potential of Instrumental Variable Estimators

Erik Biørn

Department of Economics, University of Oslo,
P.O. Box 1095 Blindern, 0317 Oslo, Norway

E-mail: erik.biorn@econ.uio.no

Memo 28/2014-v1

(This version December 2014)

Abstract: The measurement error problem in linear time series regression, with focus on the impact of error memory, modeled as finite-order MA processes, is considered. Three prototype models, two bivariate and one univariate ARMA, and ways of handling the problem by using instrumental variables (IVs) are discussed as examples. One has a bivariate regression equation that is static, although with dynamics, entering via the memory of its latent variables. The examples illustrate how 'structural dynamics' interacting with measurement error memory create bias in Ordinary Least Squares (OLS) and illustrate the potential of IV estimation procedures. Supplementary Monte Carlo simulations are provided for two of the example models.

Keywords: Errors in variables, ARMA, Error memory, Simultaneity bias, Attenuation, Monte Carlo.

JEL classification: C22, C26, C32, C36, C52, C53

Introduction

Estimation of coefficients in time-series regression models where autoregression and errors in variables (EIV) jointly occur, is interesting in several contexts. Motivating examples that involve such variables are: a stock of finished goods or of fixed capital constructed from cumulated flows, in which case improper measurements may produce serially correlated errors, and a flow variable, *e.g.*, income and sales, for which improper periodization of transactions may create serial correlation between errors which are close in time. Grether and Maddala (1973), Pagano (1974), and Staudenmayer and Buonaccorsi (2005) consider distributed lag models where errors in variables and serially correlated disturbances interact.[1] In the present paper, the bias of Ordinary Least Squares (OLS) estimators for dynamic equations is considered in cases where both disturbances and errors may have memory. Further we examine the potential inconsistency of Instrumental Variables (IV) estimators when lagged values of the regressors are used as IVs. Like the studies mentioned, the present paper is concerned with strict time series data, although not with Maximum Likelihood estimation. The main attention is given to applications of IV procedures. Our approach may therefore be called an ARMA-EIV-IV approach.

It is well known that related problems arise in panel data, and explorations of IV estimation in pure time-series contexts may give insights relevant for panel data situations. Relative to time series situations, panel data create additional possibilities for ensuring consistency in estimation and raise some new problems. For example, designing IV procedures when there is a 'two-dimensional' (unit-time) variation, across unit and time, requires that unobserved unit-specific heterogeneity, in the behaviour of the units or in the data measurement process, be handled. The latter may call for procedures that combine variables in levels and in differences, using IVs in levels for variables in differences, or doing the opposite, as exemplified in Biørn (1996, 2015) and Biørn and Han (2013).

IV procedures valid for the standard situation in time series regression with memory-free errors can be modified to handle finite memory, *e.g.* formalized as moving average (MA) processes, by reducing the IV set. The essence of this reduction is to ensure that all remaining IVs 'get clear of' the memory of the error process so that the IVs are uncorrelated with the errors/disturbances (*the orthogonality condition*), while being correlated with the variables for which they serve (*the rank condition*). This double claim restricts the admissible signal and noise memories, which will be exemplified throughout the paper.

To illustrate the potential of the ARMA-EIV-IV approach we will consider three models, mimicking prototype cases. All models have 'dynamic elements', although with different 'focus'. Only two have more than one 'structural variable' with observable counterparts. The models can be labeled as follows: The first contains a bivariate static equation with memory in both error and disturbance (Section 1), the second is a univariate ARMA model with white noise or MA(1) error (Section 2),

[1] Maravall and Aigner (1977), Maravall (1979) and Nowak (1993) discuss identification problems for such models.

and the third is an ARMAX model[2] with an MA(1) error (Section 3). Results from supplementary simulation experiments will be presented along with the models and methods.

1 Bivariate EIV model with memory in error and disturbance

Model 1 describes a static time series regression equation with finite memory not only of the latent regressor ξ_t, but also of the measurement error in the regressor ϵ_t and of the disturbance v_t (which may also include a measurement error in the regressand), equal to N_ξ, N_ϵ and N_v, respectively. All variables are assumed (covariance) stationary. The model is:

$$
(1.1)\qquad \begin{array}{llll}
y_t = \beta\xi_t + v_t & \mathsf{E}(\xi_t) = 0, & \mathsf{E}(\xi_t\xi_{t-s}) = \sigma_{\xi\xi(s)}, & s = 0, 1, \ldots, N_\xi, \\
x_t = \xi_t + \epsilon_t, & \mathsf{E}(v_t) = 0, & \mathsf{E}(v_t v_{t-s}) = \sigma_{vv(s)}, & s = 0, 1, \ldots, N_v, \\
\xi_t \perp \epsilon_t \perp v_t, & \mathsf{E}(\epsilon_t) = 0, & \mathsf{E}(\epsilon_t\epsilon_{t-s}) = \sigma_{\epsilon\epsilon(s)}, & s = 0, 1, \ldots, N_\epsilon,
\end{array}
$$

where $\perp$ denotes 'orthogonal to'.[3] For simplicity, ξ_t is assumed to have zero expectation and hence, that the equation's intercept is zero. At the moment, we do not impose any further restrictions on these sets of autocovariances. It follows that

$$
(1.2)\qquad y_t = \beta x_t + v_t - \beta\epsilon_t,
$$

$$
(1.3)\qquad \begin{array}{ll}
\mathsf{E}(y_t y_{t-s}) \equiv \sigma_{yy(s)} = \beta^2\sigma_{\xi\xi(s)} + \sigma_{vv(s)}, & \\
\mathsf{E}(y_t x_{t-s}) \equiv \sigma_{yx(s)} = \beta\sigma_{\xi\xi(s)}, & \\
\mathsf{E}(x_t x_{t-s}) \equiv \sigma_{xx(s)} = \sigma_{\xi\xi(s)} + \sigma_{\epsilon\epsilon(s)}, & s = 0, 1, 2, \ldots ,
\end{array}
$$

where $\equiv$ denotes 'equal by definition'. For this model, we consider OLS and IV estimation.

The plims of the direct and the reverse OLS estimators based on (1.2),

$$
(1.4)\qquad \widehat{\beta}_x^{OLS} = \frac{\sum_t y_t x_t}{\sum_t x_t^2},
$$

$$
(1.5)\qquad \widehat{\beta}_y^{OLS} = \frac{\sum_t y_t^2}{\sum_t x_t y_t},
$$

are, when relying on the usual convergence in moments assumptions, see Fuller (1987, p. 11), respectively:

$$
(1.6)\qquad \bar{\beta}_{x(0)} = \text{plim}(\widehat{\beta}_x^{OLS}) = \frac{\mathsf{E}(y_t x_t)}{\mathsf{E}(x_t^2)} = \frac{\sigma_{yx(0)}}{\sigma_{xx(0)}} = \beta k_{x(0)}^{-1},
$$

$$
(1.7)\qquad \bar{\beta}_{y(0)} = \text{plim}(\widehat{\beta}_y^{OLS}) = \frac{\mathsf{E}(y_t^2)}{\mathsf{E}(x_t y_t)} = \frac{\sigma_{yy(0)}}{\sigma_{xy(0)}} = \beta k_{y(0)},
$$

where

[2] As usual, ARMAX is a shorthand for ARMA augmented with exogenous variables.

[3] It is assumed that $\sigma_{\xi\xi(s)} = 0$ for $s > N_\xi$, $\sigma_{\epsilon\epsilon(s)} = 0$ for $s > N_\epsilon$, and $\sigma_{vv(s)} = 0$ for $s > N_v$.

$$k_{x(0)} = \frac{\sigma_{xx(0)}}{\sigma_{\xi\xi(0)}} = 1+\frac{\sigma_{\epsilon\epsilon(0)}}{\sigma_{\xi\xi(0)}},$$
$$k_{y(0)} = \frac{\sigma_{yy(0)}}{\beta^2\sigma_{\xi\xi(0)}} = 1+\frac{\sigma_{vv(0)}}{\beta^2\sigma_{\xi\xi(0)}}.$$

The factors $k_{x(0)}^{-1}$ and $k_{y(0)}$ represent the measurement error bias (simultaneity bias) of, respectively, the direct and the reverse OLS estimators. In most textbook expositions of OLS applied to EIV models memory-free errors and disturbances, $N_v = N_\epsilon = 0$, are assumed.

Allowing for $N_\nu > 0$, $N_\epsilon > 0$, we next consider estimators using, respectively, x_{t-s} and y_{t-s} $(s>0)$ as IVs for x_t in (1.2):

$$\widehat{\beta}^{IV}_{x(s)} = \frac{\sum_t y_t x_{t-s}}{\sum_t x_t x_{t-s}}, \tag{1.8}$$
$$\widehat{\beta}^{IV}_{y(s)} = \frac{\sum_t y_t y_{t-s}}{\sum_t x_t y_{t-s}}. \tag{1.9}$$

From (1.3), again relying on usual convergence in moments assumptions, we obtain

$$\bar{\beta}_{x(s)} = \text{plim}[\widehat{\beta}^{IV}_{x(s)}] = \frac{\mathsf{E}(y_t x_{t-s})}{\mathsf{E}(x_t x_{t-s})} = \frac{\sigma_{yx(s)}}{\sigma_{xx(s)}} = \beta k_{x(s)}^{-1}, \qquad s = 0,1,2,\dots\ , \tag{1.10}$$
$$\bar{\beta}_{y(s)} = \text{plim}[\widehat{\beta}^{IV}_{y(s)}] = \frac{\mathsf{E}(y_t y_{t-s})}{\mathsf{E}(x_t y_{t-s})} = \frac{\sigma_{yy(s)}}{\sigma_{xy(s)}} = \beta k_{y(s)}, \qquad s = 0,1,2,\dots\ , \tag{1.11}$$

where

$$k_{x(s)} = \frac{\sigma_{xx(s)}}{\sigma_{\xi\xi(s)}} = 1 + \frac{\sigma_{\epsilon\epsilon(s)}}{\sigma_{\xi\xi(s)}}, \qquad \sigma_{\xi\xi(s)} \neq 0 \Longrightarrow s < N_\xi, \tag{1.12}$$
$$k_{y(s)} = \frac{\sigma_{yy(s)}}{\beta^2} = 1 + \frac{\sigma_{vv(s)}}{\beta^2\sigma_{\xi\xi(s)}}, \qquad \sigma_{\xi\xi(s)} \neq 0 \Longrightarrow s < N_\xi. \tag{1.13}$$

The factors $k_{x(s)}^{-1}$ and $k_{y(s)}$, which do not exist unless $\sigma_{\xi\xi(s)} \neq 0$, represent the measurement error bias (simultaneity bias) of the respective estimators. Therefore,

$x_{t-s}\,(s>0)$ is a valid IV for x_t means : $\{k_{x(s)}=1, \bar{\beta}_{x(s)}=\beta\} \Longleftrightarrow \sigma_{\epsilon\epsilon(s)}=0$ & $\sigma_{\xi\xi(s)} \neq 0$,
$y_{t-s}\,(s>0)$ is a valid IV for x_t means : $\{k_{y(s)}=1, \bar{\beta}_{y(s)}=\beta\} \Longleftrightarrow \sigma_{vv(s)}=0$ & $\sigma_{\xi\xi(s)} \neq 0$,

so that for (1.2)

$\widehat{\beta}^{IV}_{x(s)}$, *with* $N_\xi \geq s > N_\epsilon$, *is consistent for* β.
$\widehat{\beta}^{IV}_{y(s)}$, *with* $N_\xi \geq s > N_v$, *is consistent for* β.

Hence, the memory configuration of the signal, noise and disturbance is essential for the existence of consistent estimators, and hence for identifiability of β.

Let $\widehat{\beta}$ be *any* estimator of β and let $\bar{\beta}$ denote plim$(\widehat{\beta})$. The corresponding ordinary and asymptotic residuals can be written as

$$\widehat{e}_t \equiv y_t - \widehat{\beta}x_t = \beta\xi_t + v_t - \widehat{\beta}(\xi_t+\epsilon_t) = (\beta-\widehat{\beta})\xi_t - \widehat{\beta}\epsilon_t + v_t, \tag{1.14}$$
$$\bar{e}_t \equiv y_t - \bar{\beta}x_t = \beta\xi_t + v_t - \bar{\beta}(\xi_t+\epsilon_t) = (\beta-\bar{\beta})\xi_t - \bar{\beta}\epsilon_t + v_t. \tag{1.15}$$

From (1.10) and (1.11) it follows, in particular, that $\widehat{\beta}=\widehat{\beta}^{IV}_{x(s)}$ and $\widehat{\beta}=\widehat{\beta}^{IV}_{y(s)}$ have asymptotic residuals that can be written as:

$$e_{[xs]t} \equiv y_t - \bar{\beta}_{x(s)}x_t = \beta[(1-k^{-1}_{x(s)})\xi_t - k^{-1}_{x(s)}\epsilon_t] + v_t, \tag{1.16}$$
$$e_{[ys]t} \equiv y_t - \bar{\beta}_{y(s)}x_t = \beta[(1-k_{y(s)})\xi_t - k_{y(s)}\epsilon_t] + v_t,\ s=0,1,2,\dots\ , \tag{1.17}$$

and hence,

$$e_{[xs]t} - v_t = \beta[(1-k^{-1}_{x(s)})\xi_t - k^{-1}_{x(s)}\epsilon_t],$$
$$e_{[ys]t} - v_t = \beta[(1-k_{y(s)})\xi_t - k_{y(s)}\epsilon_t].$$

Combining the latter equations with (1.1), we find that the asymptotic residuals have autocovariances of order τ given by, respectively,

$$\mathrm{cov}(e_{[xs]t}, e_{[xs]t-\tau}) = \beta^2[(1-k^{-1}_{x(s)})^2\sigma_{\xi\xi(\tau)} + k^{-2}_{x(s)}\sigma_{\epsilon\epsilon(\tau)}] + \sigma_{vv(\tau)}, \tag{1.18}$$
$$\mathrm{cov}(e_{[ys]t}, e_{[ys]t-\tau}) = \beta^2[(1-k_{y(s)})^2\sigma_{\xi\xi(\tau)} + k^{2}_{y(s)}\sigma_{\epsilon\epsilon(\tau)}] + \sigma_{vv(\tau)}. \tag{1.19}$$

Therefore, *if $\widehat{\beta}^{IV}_{x(s)}$ and $\widehat{\beta}^{IV}_{y(s)}$ are inconsistent for β, as consequences of, respectively, $s \le N_\epsilon$ and $s \le N_v$, they produce serially correlated residuals when ξ_t is autocorrelated.* This holds even if ϵ_t and v_t are white noise ($N_v = N_\epsilon = 0$), since then

$$\mathrm{cov}(e_{[xs]t}, e_{[xs]t-\tau}) = \beta^2(1-k^{-1}_{x(s)})^2\sigma_{\xi\xi(\tau)},$$
$$\mathrm{cov}(e_{[ys]t}, e_{[ys]t-\tau}) = \beta^2(1-k_{y(s)})^2\sigma_{\xi\xi(\tau)},$$

so that the covariances have the same sign as $\sigma_{\xi\xi(\tau)}$. Grether and Maddala (1973) pointed out this implication of autocorrelated signals for asymptotic OLS residuals, $e_{[x0]t}$ (in our notation), in a static measurement error model. Equations (1.18)–(1.19) generalize this result to hold for inconsistent IV estimators as well.

Monte Carlo simulation results, with $R=100$ replications, are given below.[4] *Table 1* illustrates the effect of changes in the distribution of the disturbance v_t (also including possible measurement error in the regressand) on the distribution of $\widehat{\beta}^{IV}_{x(s)}$ and $\widehat{\beta}^{IV}_{y(s)}$. The table shows the mean (`mean`), (empirical) standard deviation (`stdev`), maximum, minimum (`max, min`) and the relative root mean square error (`relmse`). The range spanned by `max` and `min`, the `stdev` and the `relmse` of the estimates are larger the more strongly backdated the IV is and the larger the variance of the disturbance (error in the regressand). Under the assumptions made in Examples 1.1 and 1.2, including zero memory of the error in the regressor ($N_v = 0$), all IV estimates using x-IVs, *i.e.*, $\widehat{\beta}^{IV}_{x(s)}$ for $s = 1,2,3,4,5$, are consistent. Their `mean` are very close to the input value in the simulations, $\beta = 0.8$, *i.e.*, the bias is small, but $\widehat{\beta}^{IV}_{x(5)}$ have large `relmse`, respectively, 20% (Example 1.1) and 28% (Example 1.2) about twice the `relmse` for $\widehat{\beta}^{IV}_{x(4)}$. Turning to the estimates based on y-IVs, we find notable changes in the results. In Example 1.1, $\widehat{\beta}^{IV}_{y(1)}$ is inconsistent, with `mean` 0.8343, while $\widehat{\beta}^{IV}_{y(s)}$ for $s = 2,3,4,5$ are consistent and have `mean` close to 0.8 (between 0.790 and 0.802). When in Example 1.2 the memory of the error in y_t is increased from $N_\epsilon = 1$

[4] The simulations are based on program modules constructed in the Gauss software code. The time series length used is 200, of which the last $T=100$ are used as estimation sample. The author is grateful to Xuehui Han for her excellent job with the programming of the routines for the Monte Carlo simulations.

to $N_\epsilon = 4$, only $\widehat{\beta}^{IV}_{y(5)}$ is consistent, but its `relmse` is as large as 53%. The point estimate, 0.8512, however, is not markedly different from those of the inconsistent estimators, which have much smaller `relmse` and `max-min` range.

Table 2 illustrates how a changed distribution of the measurement error in the regressor, ϵ_t, notably its spread, impacts the distribution of the two sets of estimators. Relative to Examples 1.1 and 1.2 the memory of the error in the regressor, N_ϵ, is increased from 0 to 4, while the error in the regressand (including the disturbance), N_v, is set to 0 throughout. In Example 1.3, $\text{var}(\epsilon_t)$ is 1.5, and in Examples 1.4 and 1.5 it is raised to 3.75 and 7.5, respectively. All $\widehat{\beta}^{IV}_{y(s)}$ (with $s = 1, \ldots, 5$) are now consistent, because of the assumed zero memory of v_t, while among $\widehat{\beta}^{IV}_{x(s)}$ only the one with $s = 5$ has this property. However, the `stdev` and `relmse` of the latter are substantial, even in the case with the lowest error spread, the latter is as large as 59%, making its point estimate, 0.88, practically 'insignificant'. That a 'weak IV problem' arises in this case is confirmed from the last column of Table 3. The inconsistent estimators $\widehat{\beta}^{IV}_{x(s)}$ for $s = 1, 2, 3, 4$ have all much smaller `relmse`. The same is true for $\widehat{\beta}^{IV}_{y(s)}$ for $s = 1, 2, 3, 4$, all of which are consistent. Regarding the *sign of the bias*, *i.e.*, the difference between the mean estimates of β and the input value $\beta = 0.8$, an interesting result is that for $s = 1, 2, 3, 4$, all $\widehat{\beta}^{IV}_{x(s)}$ have a negative bias and all $\widehat{\beta}^{IV}_{y(s)}$ have a positive bias. However, in all cases the bias of $\widehat{\beta}^{IV}_{y(s)}$ changes its sign when the IV-lag, s, is increased from 4 to 5. For $\widehat{\beta}^{IV}_{x(s)}$ the bias changes sign when s is increased from 4 to 5 in Examples 1.3 and 1.5, although not in the intermediate case, Example 1.4.

Table 1:
Model 1: Simulated IV estimates. Impact of changed v_t distribution.
$\beta = 0.8,\ N_\xi = 8,\ \sigma_{\xi\xi(s)} = N_\xi + 1 - s\ (s = 0, 1, \ldots, N_\xi).\ T = 100, R = 100$

		$s=1$	$s=2$	$s=3$	$s=4$	$s=5$
Example 1.1: $N_v = 1$, $N_\epsilon = 0$, $\text{var}(\xi_t) = 45$, $\text{var}(v_t) = 0.65$, $\text{var}(\epsilon_t) = 0.1$. $(\sigma_{vv(0)}, \sigma_{vv(1)}) = (0.625, 0.25)$. $\sigma_{\epsilon\epsilon(0)} = 0.1$						
$\widehat{\beta}^{IV}_{x(s)}$	mean	0.7900	0.7920	0.7898	0.7964	0.8103
	stdev	0.0417	0.0508	0.0639	0.0799	0.1579
	max	0.9417	0.9688	1.0691	1.1320	1.8715
	min	0.6916	0.6758	0.6524	0.6531	0.5343
	relmse	0.0537	0.0643	0.0809	0.0999	0.1978
$\widehat{\beta}^{IV}_{y(s)}$	mean	0.8343	0.7904	0.7907	0.7966	0.8021
	stdev	0.0461	0.0506	0.0640	0.0847	0.1269
	max	1.0309	0.9318	0.9969	1.1345	1.2730
	min	0.7317	0.6805	0.6472	0.6058	0.4541
	relmse	0.0718	0.0644	0.0809	0.1059	0.1586
Example 1.2: $N_v = 4$, $N_\epsilon = 0$, $\text{var}(\xi_t) = 45$, $\text{var}(v_t) = 1.5$, $\text{var}(\epsilon_t) = 0.1$. $(\sigma_{vv(0)}, \sigma_{vv(1)}, \sigma_{vv(2)}, \sigma_{vv(3)}, \sigma_{vv(4)}) = (0.5, 0.4, 0.3, 0.2, 0.1)$. $\sigma_{\epsilon\epsilon(0)} = 0.1$						
$\widehat{\beta}^{IV}_{x(s)}$	mean	0.8041	0.8039	0.8035	0.8045	0.8194
	stdev	0.0548	0.0637	0.0810	0.1094	0.2252
	max	0.9474	0.9695	1.0028	1.1367	2.3284
	min	0.6513	0.6145	0.5452	0.4000	0.0372
	relmse	0.0687	0.0798	0.1014	0.1369	0.2825
$\widehat{\beta}^{IV}_{y(s)}$	mean	0.8699	0.8599	0.8486	0.8336	0.8512
	stdev	0.0561	0.0638	0.0819	0.1136	0.4241
	max	0.9974	1.0073	1.0408	1.2138	4.4434
	min	0.7359	0.7116	0.6253	0.4598	0.2810
	relmse	0.1121	0.1094	0.1191	0.1480	0.5340

Table 2:

Model 1: Simulated IV estimates. Impact of changed ϵ_t spread.

$\beta = 0.8,\ N_\xi = 8,\ \sigma_{\xi\xi(s)} = N_\xi + 1 - s,\ (s = 0, 1, \dots, N_\xi).\ T = 100, R = 100$

		$s=1$	$s=2$	$s=3$	$s=4$	$s=5$
Example 1.3: $N_v = 0$, $N_\epsilon = 4$, $\mathrm{var}(\xi_t) = 45$, $\mathrm{var}(v_t) = 0.5$, $\mathrm{var}(\epsilon_t) = 1.5$. $\sigma_{vv(0)} = 0.5$. $(\sigma_{\epsilon\epsilon(0)}, \sigma_{\epsilon\epsilon(1)}, \sigma_{\epsilon\epsilon(2)}, \sigma_{\epsilon\epsilon(3)}, \sigma_{\epsilon\epsilon(4)}) = (0.5, 0.4, 0.3, 0.2, 0.1)$						
$\widehat{\beta}^{IV}_{x(s)}$	mean	0.7572	0.7614	0.7705	0.7821	0.8829
	stdev	0.0442	0.0547	0.0720	0.1311	0.4674
	relmse	0.0769	0.0837	0.0972	0.1654	0.5933
$\widehat{\beta}^{IV}_{y(s)}$	mean	0.8012	0.8023	0.8070	0.8085	0.7861
	stdev	0.0488	0.0581	0.0789	0.0996	0.5123
	relmse	0.0610	0.0726	0.0990	0.1249	0.6406
Example 1.4: $N_v = 0$, $N_\epsilon = 4$, $\mathrm{var}(\xi_t) = 45$, $\mathrm{var}(v_t) = 0.5$, $\mathrm{var}(\epsilon_t) = 3.75$. $\sigma_{vv(0)} = 0.5$. $(\sigma_{\epsilon\epsilon(0)}, \sigma_{\epsilon\epsilon(1)}, \sigma_{\epsilon\epsilon(2)}, \sigma_{\epsilon\epsilon(3)}, \sigma_{\epsilon\epsilon(4)}) = (1.25, 1.00, 0.75, 0.50, 0.25)$						
$\widehat{\beta}^{IV}_{x(s)}$	mean	0.7051	0.7124	0.7277	0.7714	0.7882
	stdev	0.0755	0.0956	0.1170	0.1904	0.4237
	relmse	0.1515	0.1621	0.1720	0.2407	0.5298
$\widehat{\beta}^{IV}_{y(s)}$	mean	0.8049	0.8027	0.8029	0.8239	0.7175
	stdev	0.0778	0.1017	0.1273	0.2435	0.8519
	relmse	0.0974	0.1272	0.1592	0.3058	1.0698
Example 1.5: $N_v = 0$, $N_\epsilon = 4$, $\mathrm{var}(\xi_t) = 45$, $\mathrm{var}(v_t) = 0.5$, $\mathrm{var}(\epsilon_t) = 7.5$. $\sigma_{vv(0)} = 0.5$. $(\sigma_{\epsilon\epsilon(0)}, \sigma_{\epsilon\epsilon(1)}, \sigma_{\epsilon\epsilon(2)}, \sigma_{\epsilon\epsilon(3)}, \sigma_{\epsilon\epsilon(4)}) = (2.5, 2.0, 1.5, 1.0, 0.5)$						
$\widehat{\beta}^{IV}_{x(s)}$	mean	0.6360	0.6577	0.6923	0.7523	1.0230
	stdev	0.0903	0.1115	0.1439	0.2022	1.6571
	relmse	0.2340	0.2259	0.2246	0.2597	2.0901
$\widehat{\beta}^{IV}_{y(s)}$	mean	0.8160	0.8254	0.8433	0.9130	0.4770
	stdev	0.1592	0.1999	0.2620	0.5341	6.5339
	relmse	0.2000	0.2519	0.3319	0.6824	8.1774

Table 3:

Model 1: Examples 1.3–1.5, autocovariances: x vs. x_{-s} and y_{-s}. $T = 100, R = 100$

		$s=1$	$s=2$	$s=3$	$s=4$	$s=5$
Example 1.3:	mean corr(x, x_{-s})	0.8711	0.7449	0.6181	0.4978	0.3822
	mean corr(x, y_{-s})	0.8093	0.6971	0.5869	0.4802	0.3806
Example 1.4:	mean corr(x, y_{-s})	0.8701	0.7380	0.6093	0.4832	0.3593
	mean corr(x, y_{-s})	0.7888	0.6839	0.5807	0.4796	0.3786
Example 1.5:	mean corr(x, x_{-s})	0.8569	0.7158	0.5793	0.4404	0.3018
	mean corr(x, y_{-s})	0.7348	0.6326	0.5330	0.4319	0.3321

2 ARMA model with white noise or MA measurement error

Model 2, unlike Model 1, is explicitly dynamic and has no exogenous variable. It is a univariate ARMA(1,1)-model for a latent variable μ_t, observed through y_t with, in its simplest version, a white noise error δ_t:

$$\begin{aligned} &\mu_t = \gamma\mu_{t-1} + v_t + \lambda v_{t-1}, \qquad |\gamma| < 1, \\ &y_t = \mu_t + \delta_t, \\ &\mathsf{E}(\mu_0) = 0, \quad v_t \sim \mathsf{IID}(0, \sigma_v^2), \quad \delta_t \sim \mathsf{IID}(0, \sigma_\delta^2), \quad v_t \perp \delta_t \perp \mu_\tau, \quad \forall\tau. \end{aligned} \tag{2.1}$$

The observed and latent 'structural' variables satisfy, respectively,

$$(1-\gamma\mathsf{L})y_t = (1-\gamma\mathsf{L})\delta_t + (1+\lambda\mathsf{L})v_t, \tag{2.2}$$

$$\mu_t = \frac{1+\lambda\mathsf{L}}{1-\gamma\mathsf{L}}v_t = v_t + \frac{\gamma+\lambda}{1-\gamma\mathsf{L}}v_{t-1}, \tag{2.3}$$

where L denotes the backshift operator. Hence, μ_t follows a geometric lag distribution whose first term is 'free' and whose remaining terms have a variance that is $(\gamma+\lambda)^2$ times the variance of the first. Since (2.1) and (2.3) imply

$$y_t = \delta_t+v_t+(\gamma+\lambda)\sum_{i=1}^{\infty}\gamma^i v_{t-i},$$

y_t follows a geometric lag distribution whose first term is 'free' and whose remaining terms have a variances that are $(\gamma+\lambda)^2\sigma_v^2/(\sigma_\delta^2+\sigma_v^2)$ times the variance of the first. Letting $v_t^* = v_t+\lambda v_{t-1}$ denote the model's MA(1) part, it also follows that

$$\text{(2.4)} \qquad \operatorname{cov}(y_{t-s}, \delta_t+v_t^*) \begin{cases} \neq 0, & s=0, \\ = 0, & s=1,2,\dots\ , \end{cases}$$

while the variance and autocovariances of μ_t become

$$\text{(2.5)} \qquad \mathsf{E}(\mu_t\mu_{t-s}) = \begin{cases} \sigma_{\mu\mu(0)} = (1+\chi^2)\sigma_v^2, & s=0, \\ \sigma_{\mu\mu(1)} = [\gamma(1+\chi^2)+\lambda]\sigma_v^2, & s=1, \\ \sigma_{\mu\mu(s)} = \gamma^{s-1}[\gamma(1+\chi^2)+\lambda]\sigma_v^2, & s\geq 2, \end{cases}$$

where

$$\chi^2 = \operatorname{var}\left[\frac{\gamma+\lambda}{1-\gamma\mathsf{L}}\left(\frac{v_t}{\sigma_v}\right)\right] = \frac{(\gamma+\lambda)^2}{1-\gamma^2}.$$

By combining (2.1) and (2.5) it follows that

$$\text{(2.6)} \qquad \mathsf{E}(y_t y_{t-s}) \equiv \sigma_{yy(s)} = \begin{cases} \sigma_{\mu\mu(0)} + \sigma_\delta^2 = (1+\chi^2)\sigma_v^2 + \sigma_\delta^2, & s=0, \\ \sigma_{\mu\mu(1)} = [\gamma(1+\chi^2)+\lambda]\sigma_v^2, & s=1, \\ \sigma_{\mu\mu(s)} = \gamma^{s-1}[\gamma(1+\chi^2)+\lambda]\sigma_v^2, & s=2,3,\dots\ . \end{cases}$$

The OLS estimator of γ based on (2.2),

$$\widehat{\gamma}^{OLS} = \frac{\sum_t y_t y_{t-1}}{\sum_t y_{t-1}^2},$$

is inconsistent and from (2.6) we get

$$\text{(2.7)} \qquad \bar{\gamma}_{(1)} = \operatorname{plim}[\widehat{\gamma}^{OLS}] = \frac{\sigma_{yy(1)}}{\sigma_{yy(0)}} = \frac{\sigma_{\mu\mu(1)}}{\sigma_{\mu\mu(0)}+\sigma_\delta^2} = \frac{\gamma(1+\chi^2)+\lambda}{1+\chi^2+\frac{\sigma_\delta^2}{\sigma_v^2}}.$$

The sign of the inconsistency depends on the AR parameter γ, the MA parameter λ and the error variance σ_δ^2. Therefore, the attenuation[5] of $\widehat{\gamma}^{OLS}$ may be off-set by the memory of v_t^*, provided that λ and γ have equal sign, since we have

$$\text{(2.8)} \qquad \bar{\gamma}_{(1)} \gtreqless \gamma \iff \frac{\lambda}{\gamma} \gtreqless \frac{\sigma_\delta^2}{\sigma_v^2}.$$

A special case is the familiar result $\lambda=\sigma_\delta^2=0 \implies \bar{\gamma}_{(1)}=\gamma$: strict AR(1) with no measurement error and white noise disturbance gives consistency of OLS. On the other hand,

$$\lambda = -\gamma \implies \mu_t = v_t,\ y_t = v_t+\delta_t \implies \chi^2 = 0,\ \bar{\gamma}_{(1)} = 0$$

[5]Term denoting the tendency of an OLS estimator in a one-regressor EIV model to be biased towards zero.

is a boundary case with respect to attenuation, in the sense that the OLS estimator is not only *biased towards* zero, its plim *is* zero.

Using in (2.2) $y_{t-s}\,(s\geq 2)$ as IV for y_{t-1}, to obtain

$$\widehat{\gamma}^{IV}_{(s)} = \frac{\sum_t y_t y_{t-s}}{\sum_t y_{t-1} y_{t-s}},$$

consistency is ensured, as (2.6) implies

$$\text{(2.9)} \qquad \bar{\gamma}_{(s)} = \text{plim}[\widehat{\gamma}^{IV}_{(s)}] = \frac{\sigma_{yy(s)}}{\sigma_{yy(s-1)}} = \frac{\sigma_{\mu\mu(s)}}{\sigma_{\mu\mu(s-1)}} = \gamma, \quad s = 2, 3, \dots .$$

This way of using IVs exploits that y_{t-s} $(s\geq 2)$ is correlated (more or less strongly depending on γ, λ and s) with y_{t-1}, and is uncorrelated with $(1-\gamma\mathsf{L})\delta_t+(1+\lambda\mathsf{L})v_t$ regardless of λ. It follows from (2.6) that $\text{corr}(y_{t-1}, y_{t-s})$ is smaller the smaller is γ and the larger is s. This signalizes that the weaker is the autocorrelation and the further a y is backdated, the weaker will it be as IV for the current y.

We can generalize (and often increase the model's realism) by allowing for memory in the measurement error. Extending δ_t in (2.1) to an MA(1) process gives

$$\text{(2.10)} \qquad \begin{array}{ll} \mu_t & = \gamma\mu_{t-1} + v_t + \lambda v_{t-1}, \quad |\gamma| < 1, \\ y_t & = \mu_t + \delta_t + \psi\delta_{t-1}, \\ \mathsf{E}(\mu_0) & = 0, \quad v_t \sim \mathsf{IID}(0, \sigma_v^2), \quad \delta_t \sim \mathsf{IID}(0, \sigma_\delta^2), \quad v_t \perp \delta_t. \end{array}$$

It follows that (2.2) is generalized to

$$\text{(2.11)} \qquad (1-\gamma\mathsf{L})y_t = (1-\gamma\mathsf{L})(1+\psi\mathsf{L})\delta_t + (1+\lambda\mathsf{L})v_t.$$

Since then

$$y_t = \delta_t+\psi\delta_{t-1}+v_t+(\gamma+\lambda)\textstyle\sum_{i=1}^{\infty}\gamma^i v_{t-i},$$

we obtain, letting $v_t^* = v_t + \lambda v_{t-1}$ and $\delta_t^* = \delta_t + \psi\delta_{t-1}$ be the model's two MA(1) processes,

$$\text{(2.12)} \qquad \text{cov}(y_{t-s}, \delta_t^*+v_t^*) \begin{cases} \neq 0, & s = 0, 1, \\ = 0, & s = 2, 3, \dots . \end{cases}$$

Still (2.5) holds, while (2.6) is generalized to

$$\text{(2.13)} \qquad \mathsf{E}(y_t y_{t-s}) \equiv \sigma_{yy(s)} = \begin{cases} \sigma_{\mu\mu(0)} + (1+\psi^2)\sigma_\delta^2, & s = 0, \\ \sigma_{\mu\mu(1)} + \psi\sigma_\delta^2, & s = 1, \\ \sigma_{\mu\mu(s)}, & s = 2, 3, \dots . \end{cases}$$

The resulting generalization of (2.7) and (2.9) is

$$\text{(2.14)} \; \bar{\gamma}_{(s)} = \frac{\sigma_{yy(s)}}{\sigma_{yy(s-1)}} = \begin{cases} \dfrac{\sigma_{\mu\mu(1)} + \psi\sigma_\delta^2}{\sigma_{\mu\mu(0)} + (1+\psi^2)\sigma_\delta^2} = \dfrac{\gamma(1+\chi^2)+\lambda+\psi\frac{\sigma_\delta^2}{\sigma_v^2}}{(1+\chi^2)+(1+\psi^2)\frac{\sigma_\delta^2}{\sigma_v^2}}, & s=1, \\ \dfrac{\gamma\sigma_{\mu\mu(1)}}{\sigma_{\mu\mu(1)} + \psi\sigma_\delta^2} = \dfrac{\gamma[\gamma(1+\chi^2)+\lambda]}{\gamma(1+\chi^2)+\lambda+\psi\frac{\sigma_\delta^2}{\sigma_v^2}}, & s=2, \\ \dfrac{\sigma_{\mu\mu(s)}}{\sigma_{\mu\mu(s-1)}} = \gamma, & s=3, 4, \dots . \end{cases}$$

This model, with a one-period memory of the disturbance and the measurement error, ensures consistent IV estimation for $s \geq 3$, because y_{t-s} is correlated with y_{t-1} and is uncorrelated with the composite MA(2) process $(1-\gamma\mathsf{L})(1+\psi\mathsf{L})\delta_t + (1+\lambda\mathsf{L})v_t$ – it gets clear of the memory of the latter process. If $s = 2$ and $\gamma > 0$, $\lambda > 0$, a negative (positive) asymptotic bias occurs when ψ is positive (negative), since

$$\bar{\gamma}_{(2)} - \gamma = \frac{-\psi\frac{\sigma_\delta^2}{\sigma_v^2}}{\gamma(1+\chi^2)+\lambda+\psi\frac{\sigma_\delta^2}{\sigma_v^2}}.$$

Therefore $\widehat{\gamma}_{(2)}^{IV}$ exemplifies an attenuated IV estimator. For the OLS estimator $\widehat{\gamma}^{OLS}$, however, the sign of the bias depends on both λ and ψ. If $\gamma > 0$, $\lambda > 0$ we have the following generalization of (2.8):[6]

$$\text{(2.15)} \qquad \bar{\gamma}_{(1)} \gtreqless \gamma \iff \frac{\lambda}{\gamma} \gtreqless \left(1 + \psi^2 - \frac{\psi}{\gamma}\right)\frac{\sigma_\delta^2}{\sigma_v^2}.$$

Again, attenuation (of OLS) may be counteracted by the MA part of the error processes. Now, however,

$$\lambda = -\gamma \Longrightarrow \mu_t = v_t,\ y_t = v_t + \delta_t + \psi\delta_{t-1} \Longrightarrow \chi^2 = 0,\ \bar{\gamma}_{(1)} = \frac{\psi\sigma_\delta^2}{\sigma_v^2+\sigma_\delta^2(1+\psi^2)},$$

so that $\lambda = -\gamma$ represents a boundary case where the strength (and sign) of the attenuation depends on the MA coefficient of the measurement error, ψ.

Tables 4 and 5 contain the expressions for $\bar{\gamma}_{(1)}$ and $\bar{\gamma}_{(2)}$ in the boundary cases

$$\begin{array}{lll} \lambda = 0 & \Longrightarrow & \mu_t = \gamma\mu_{t-1} + v_t, \\ \gamma = 0 & \Longrightarrow & \mu_t = v_t + \lambda v_{t-1}, \\ \lambda = -\gamma & \Longrightarrow & \mu_t = v_t, \end{array}$$

for three combinations of σ_δ^2 and ψ: $\sigma_\delta^2 = 0$ represents the no measurement error case and $\psi = 0$ represents the white noise measurement error case.

Table 6 contains numerical examples based on synthetic data for $s = 2, 3, 4, 5$ for six selected parameter constellations. As only *one replication* is performed in each case ($R = 1$), these examples are strictly not Monte Carlo simulations. They invite a few comments: The estimator in Example 2.a, where y_t is free from measurement error ($\sigma_\delta^2 = 0$) and v_t is white noise ($\lambda = 0$), is consistent even for $s = 2$, while in Examples 2.b-2.f, $\widehat{\gamma}_{(s)}^{IV}$ is consistent for $s = 3, 4, 5$ and inconsistent for $s = 2$; see (2.14) and Table 5. This concurs with 'level shift' we notice in the estimate sequence when s increases from 2 to 3.

Model 2 can be generalized further, specifying μ_t as an ARMA(n, m) process and extending the measurement error from MA(1) to MA(k). This gives the model

[6] The boundary case $\gamma = 0$, which makes y_t the sum of two MA(1) processes, $y_t = v_t^* + \delta_t^*$, and makes $\bar{\gamma}_{(1)}$ a variance-weighted average of the autocorrelation coefficients of v_t^* and δ_t^*, $\rho_v^* = \lambda/(1+\lambda^2)$ and $\rho_\delta^* = \psi/(1+\psi^2)$:

$$\bar{\gamma}_{(1)} = \frac{\text{var}(v_t^*)\rho_v^* + \text{var}(\delta_t^*)\rho_\delta^*}{\text{var}(v_t^*) + \text{var}(\delta_t^*)} \qquad (\gamma = 0).$$

Table 4:

Model 2: OLS estimator plim: $\bar{\gamma}_{(1)} = \text{plim}(\widehat{\gamma}^{OLS})$ *in boundary cases*

	$\sigma_\delta^2=0$	$\psi=0$	$\sigma_\delta^2>0, \psi\neq 0$
$\lambda=0$, $[\chi^2=\gamma^2/(1-\gamma^2)]$	γ	$\dfrac{\gamma\sigma_v^2}{\sigma_v^2+\sigma_\delta^2(1-\gamma^2)}$	$\dfrac{\gamma\sigma_v^2+\psi\sigma_\delta^2(1-\gamma^2)}{\sigma_v^2+\sigma_\delta^2(1+\psi^2)(1-\gamma^2)}$
$\gamma=0$, $(\chi=\lambda)$	$\dfrac{\lambda}{1+\lambda^2}$	$\dfrac{\lambda\sigma_v^2}{\sigma_v^2(1+\lambda^2)+\sigma_\delta^2}$	$\dfrac{\lambda\sigma_v^2+\psi\sigma_\delta^2}{\sigma_v^2(1+\lambda^2)+\sigma_\delta^2(1+\psi^2)}$
$\lambda=-\gamma$, $(\chi=0)$	0	0	$\dfrac{\psi\sigma_\delta^2}{\sigma_v^2+\sigma_\delta^2(1+\psi^2)}$

Table 5:

Model 2: IV estimator plim for $s=2$*:* $\bar{\gamma}_{(2)} = \text{plim}(\widehat{\gamma}_{(2)}^{IV})$ *in boundary cases*

	$\sigma_\delta^2=0$	$\psi=0$	$\sigma_\delta^2>0, \psi\neq 0$
$\lambda=0$, $[\chi^2=\gamma^2/(1-\gamma^2)]$	γ	γ	$\dfrac{\gamma\sigma_v^2}{\sigma_v^2+\psi\sigma_\delta^2(1-\gamma^2)}$
$\gamma=0$, $(\chi=\lambda)$	0	0	0
$\lambda=-\gamma$, $(\chi=0)$	0	0	0

Table 6:

Model 2: Impact of changed v_t *and* δ_t *distributions.*
Artificial data. $\gamma=0.8$. $T=100$. $R=1$

Example	λ	ψ	σ_v^2	σ_δ^2	s	Estimate
2.a	0.0	0.0	0.10	0.00	2	0.8357
					3	0.8852
					4	0.8961
					5	0.8308
2.b	0.3	0.3	0.10	0.10	2	0.8401
					3	0.9429
					4	1.0021
					5	0.9472
2.c	0.5	0.5	0.10	0.10	2	0.6634
					3	0.9000
					4	0.8890
					5	0.7476
2.d	0.8	0.8	0.10	0.10	2	0.5056
					3	0.5668
					4	0.5738
					5	0.6274
2.e	0.0	0.5	0.10	0.10	2	0.6232
					3	1.0457
					4	0.9307
					5	0.6341
2.f	0.5	0.0	0.10	0.10	2	0.8770
					3	0.8116
					4	0.7865
					5	0.7579

$$\begin{aligned}
\mu_t &= \gamma(\mathsf{L})\mu_t + v_t^*, \quad |\gamma| < 1, \\
y_t &= \mu_t + \delta_t^*, \\
\mathsf{E}(\mu_0) &= 0, \quad \mu(\mathsf{L}) = \gamma_1\mathsf{L}+\cdots+\gamma_n\mathsf{L}^n, \\
v_t^* = \lambda(\mathsf{L})v_t, &\quad \lambda(\mathsf{L}) = 1+\lambda_1\mathsf{L}+\cdots+\lambda_m\mathsf{L}^m, \quad v_t \sim \mathsf{IID}(0,\sigma_v^2), \\
\delta_t^* = \psi(\mathsf{L})\delta_t, &\quad \psi(\mathsf{L}) = 1+\psi_1\mathsf{L}+\cdots+\psi_k\mathsf{L}^k, \quad \delta_t \sim \mathsf{IID}(0,\sigma_\delta^2).
\end{aligned} \tag{2.16}$$

It follows that (2.11) and (2.12) are extended to

$$[1-\gamma(\mathsf{L})]y_t = [1-\gamma(\mathsf{L})]\psi(\mathsf{L})\delta_t + \lambda(\mathsf{L})v_t, \tag{2.17}$$

$$\text{cov}(y_{t-s}, \delta_t^*+v_t^*) \begin{cases} \neq 0, & s = 0,1,\dots,k+n, \\ = 0, & s = k+n+1, k+n+2, \dots \ . \end{cases} \tag{2.18}$$

Since y_{t-s} is uncorrelated with $[1-\gamma(\mathsf{L})]\psi(\mathsf{L})\delta_t+\lambda(\mathsf{L})v_t$ when $s \geq k+n+1$, consistency is ensured for s in this region. Choosing an IV for y_{t-1} among $y_{t-2},\dots,y_{t-k-n}$ violates orthogonality.

3 ARMAX model with MA(1) measurement error

Model 3 has elements from both Model 1 and Model 2, although contains neither as special cases. It augments the ARMA(1,1) part of the latent variable μ_t in (2.10) by an exogenous, error-ridden regressor with latent part ξ_t, to give an ARMAX mechanism for μ_t. The latent regressor ξ_t is MA(N_ξ), while the measurement errors in the regressand (including a disturbance in the equation) and the regressor are both MA(1). Overall, this gives a model with three uncorrelated MA(1) processes:

$$\begin{aligned}
\mu_t &= \gamma\mu_{t-1} + \beta\xi_t + v_t + \lambda v_{t-1}, \ \mathsf{E}(\mu_0)=0, \quad v_t \sim \mathsf{IID}(0,\sigma_v^2), \ |\gamma|<1, \\
y_t &= \mu_t + \delta_t + \psi\delta_{t-1}, \qquad \delta_t \sim \mathsf{IID}(0,\sigma_\delta^2), \\
x_t &= \xi_t + \epsilon_t + \phi\epsilon_{t-1}, \qquad \epsilon_t \sim \mathsf{IID}(0,\sigma_\epsilon^2), \\
\mathsf{E}(\mu_0) &= 0, \\
\mathsf{E}(\xi_t) &= 0, \quad \mathsf{E}(\xi_t\xi_{t-s}) = \begin{cases} \sigma_{\xi\xi(s)}, & s \leq N_\xi, \\ 0, & s > N_\xi, \end{cases} \\
&\xi_t \perp v_t \perp \delta_t \perp \epsilon_t.
\end{aligned} \tag{3.1}$$

From (3.1), after elimination first of μ_t, next of ξ_t, we obtain, respectively,

$$(1-\gamma\mathsf{L})y_t = \beta\xi_t + (1+\lambda\mathsf{L})v_t + (1-\gamma\mathsf{L})(1+\psi\mathsf{L})\delta_t, \tag{3.2}$$

$$(1-\gamma\mathsf{L})y_t = \beta x_t + w_t, \tag{3.3}$$

where

$$w_t = (1+\lambda\mathsf{L})v_t - \beta(1+\phi\mathsf{L})\epsilon_t + (1-\gamma\mathsf{L})(1+\psi\mathsf{L})\delta_t. \tag{3.4}$$

Introducing

$$\tau_t = \frac{1}{1-\gamma\mathsf{L}}\xi_t, \tag{3.5}$$

which is an ARMA($1, N_\xi$) process, we obtain a generalization of (2.3) that can be written as

$$(3.6)\qquad \mu_t = \beta\tau_t + \frac{1+\lambda\mathsf{L}}{1-\gamma\mathsf{L}}v_t = \beta\tau_t + v_t + \frac{\gamma+\lambda}{1-\gamma\mathsf{L}}v_{t-1}.$$

Since (3.1) and (3.5) imply

$$(3.7)\quad \mathsf{E}(\tau_t\tau_{t-s}) \equiv \sigma_{\tau\tau(s)} = \frac{\gamma^{|s|}\sigma_{\xi\xi(0)}}{1-\gamma^2}, \qquad s=0,\pm1,\pm2,\dots,$$

$$(3.8)\quad \mathsf{E}(\mu_t\xi_{t-s}) \equiv \sigma_{\mu\xi(s)} = \beta\mathsf{E}(\tau_t\xi_{t-s}) = \beta\sum_{i:|i-s|\le N_\xi}\gamma^i\sigma_{\xi\xi(|i-s|)}, \quad s=0,\pm1,\pm2,\dots,$$

it follows that (2.5) is generalized to:

$$(3.9)\qquad \mathsf{E}(\mu_t\mu_{t-s}) \equiv \sigma_{\mu\mu(s)} = \begin{cases} \beta^2\sigma_{\tau\tau(0)} + [1+\chi^2]\sigma_v^2, & s=0,\\ \gamma\beta^2\sigma_{\tau\tau(0)} + [\gamma(1+\chi^2)+\lambda]\sigma_v^2, & s=1,\\ \gamma^s\beta^2\sigma_{\tau\tau(0)} + \gamma^{s-1}[\gamma(1+\chi^2)+\lambda]\sigma_v^2, & s\ge 2.\end{cases}$$

We further find that the observed variables in (3.3) have autocovariances and cross-autocovarances given by

$$(3.10)\qquad \mathsf{E}(y_ty_{t-s}) = \begin{cases} \sigma_{\mu\mu(0)}+(1+\psi^2)\sigma_\delta^2, & s=0,\\ \sigma_{\mu\mu(1)}+\psi\sigma_\delta^2, & s=1,\\ \sigma_{\mu\mu(s)}, & s=2,3,\dots\,,\end{cases}$$

$$(3.11)\qquad \mathsf{E}(x_tx_{t-s}) = \begin{cases} \sigma_{\xi\xi(0)}+(1+\phi^2)\sigma_\epsilon^2, & s=0,\\ \sigma_{\xi\xi(1)}+\phi\sigma_\epsilon^2, & s=1,\\ \sigma_{\xi\xi(s)}, & s=2,3,\dots,N_\xi,\\ 0, & s>N_\xi,\end{cases}$$

$$(3.12)\qquad \mathsf{E}(y_tx_{t-s}) = \sigma_{\mu\xi(s)} = \beta\sum_{i:|i-s|\le N_\xi}\gamma^i\sigma_{\xi\xi(|i-s|)}, \quad s=0,\pm1,\pm2,\dots\,,$$

and that the (non)orthogonality properties of this equation are

$$(3.13)\qquad \mathsf{E}(y_{t-s}w_t) = \begin{cases} \lambda\sigma_v^2+(\psi-\gamma)\sigma_\delta^2, & s=1,\\ -\psi\gamma\sigma_\delta^2, & s=2,\\ 0, & s=3,4,\dots\,,\end{cases}$$

$$(3.14)\qquad \mathsf{E}(x_{t-z}w_t) = \begin{cases} -\beta(1+\phi^2)\sigma_\epsilon^2, & z=0,\\ -\beta\phi\sigma_\epsilon^2, & z=1,\\ 0, & z=2,3,\dots\,.\end{cases}$$

Since in (3.3) $\mathsf{E}(y_{t-1}w_t)\neq 0$ $(s=1)$ and $\mathsf{E}(x_tw_t)\neq 0$ $(z=0)$, we once again have a measurement error bias (simultaneity bias) problem for OLS. Potential IVs for (y_{t-1},x_t) to handle this are: $y_{t-3},y_{t-4},\dots$ and $x_{t-2},x_{t-3},\dots,x_{t-N_\xi}$. As in Models 1 and 2, strongly backdated y-IVs – and, depending on $\sigma_{\xi\xi(s)}$, also strongly backdated x-IVs – may be weak IVs.

An illustration: Relying on the usual convergence in moments assumptions, we find that the estimators of γ and β obtained from (3.3) when instrumenting[7] (y_{t-1},x_t) by (y_{t-s},x_{t-z}) have plims

[7] We in this illustration confine attention to 'exact identification' cases, with only one IV allocated to each instrumented variable.

$$\begin{bmatrix} \bar{\gamma}_{(sz)} \\ \bar{\beta}_{(sz)} \end{bmatrix} = \begin{bmatrix} \mathsf{E}(y_{t-s}y_{t-1}) & \mathsf{E}(y_{t-s}x_t) \\ \mathsf{E}(x_{t-z}y_{t-1}) & \mathsf{E}(x_{t-z}x_t) \end{bmatrix}^{-1} \begin{bmatrix} \mathsf{E}(y_{t-s}y_t) \\ \mathsf{E}(x_{t-z}y_t) \end{bmatrix}$$

Since this can be written as

$$\begin{bmatrix} \bar{\gamma}_{(sz)} \\ \bar{\beta}_{(sz)} \end{bmatrix} = \begin{bmatrix} \gamma \\ \beta \end{bmatrix} + \begin{bmatrix} \mathsf{E}(y_{t-s}y_{t-1}) & \mathsf{E}(y_{t-s}x_t) \\ \mathsf{E}(x_{t-z}y_{t-1}) & \mathsf{E}(x_{t-z}x_t) \end{bmatrix}^{-1} \begin{bmatrix} \mathsf{E}(y_{t-s}w_t) \\ \mathsf{E}(x_{t-z}w_t) \end{bmatrix},$$

it follows in view of (3.13) that

$$\begin{bmatrix} \bar{\gamma}_{(sz)} \\ \bar{\beta}_{(sz)} \end{bmatrix} = \begin{bmatrix} \gamma \\ \beta \end{bmatrix}, \quad s \geq 3 \quad N_\xi \geq z \geq 2.$$

It is essential that $\sigma_{\xi\xi(s)} \neq 0$. For $(sz) = (21), (22), (23)$ we for example find that the inconsistencies of the IV estimators can be expressed as

$$\begin{bmatrix} \bar{\gamma}_{(21)} - \gamma \\ \bar{\beta}_{(21)} - \beta \end{bmatrix} = \begin{bmatrix} \sigma_{\mu\mu(1)} & \sigma_{\mu\xi(2)} \\ \sigma_{\mu\xi(0)} & \sigma_{\xi\xi(1)} + \phi\sigma_\epsilon^2 \end{bmatrix}^{-1} \begin{bmatrix} -\psi\gamma\sigma_\delta^2 \\ -\beta\phi\sigma_\epsilon^2 \end{bmatrix},$$

$$\begin{bmatrix} \bar{\gamma}_{(22)} - \gamma \\ \bar{\beta}_{(22)} - \beta \end{bmatrix} = \begin{bmatrix} \sigma_{\mu\mu(1)} & \sigma_{\mu\xi(2)} \\ \sigma_{\mu\xi(1)} & \sigma_{\xi\xi(2)} \end{bmatrix}^{-1} \begin{bmatrix} -\psi\gamma\sigma_\delta^2 \\ 0 \end{bmatrix},$$

$$\begin{bmatrix} \bar{\gamma}_{(23)} - \gamma \\ \bar{\beta}_{(23)} - \beta \end{bmatrix} = \begin{bmatrix} \sigma_{\mu\mu(2)} & \sigma_{\mu\xi(2)} \\ \sigma_{\mu\xi(2)} & \sigma_{\xi\xi(3)} \end{bmatrix}^{-1} \begin{bmatrix} -\psi\gamma\sigma_\delta^2 \\ 0 \end{bmatrix},$$

where $\sigma_{\mu\xi(s)}$ and $\sigma_{\mu\mu(s)}$ are given by (3.8) and (3.9), respectively. Obviously, the sign and size of ψ, the MA-coefficient of the measurement error of y_t affects the inconsistency of the estimators of (β, γ).

References

Biørn, E. (1996): Panel Data with Measurement Errors. Chapter 10 in: Mátyás L. and Sevestre, P. (eds.): *The Econometrics of Panel Data. A Handbook of the Theory with Applications.* Dordrecht: Kluwer.

Biørn, E. (2015): Panel Data Dynamics with Mis-measured Variables: Modeling and GMM Estimation. *Empirical Economics*, forthcoming.

Biørn, E. and Han, X. (2013): Panel Data Models With Dynamics and Measurement Error: GMM Performance When Autocorrelation and Error Memory Interact — Some Small Sample Evidence. Paper presented at the 19th International Conference on Panel Data, London, July 2013.

Fuller, W.A. (1987): *Measurement Error Models.* New York: Wiley.

Grether, D.M., and Maddala, G.S. (1973): Errors in Variables and Serially Correlated Disturbances in Distributed Lag Models. *Econometrica* **41**, 255–262.

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

Maravall, A., and Aigner, D.J. (1977): Identification of the Dynamic Shock-Error Model: The Case of Dynamic Regression. In *Latent Variables in Socio-Economic Models*, ed. by D.J. Aigner and A.S. Goldberger. Amsterdam: North-Holland.

Nowak, E. (1993): The Identification of Multivariate Linear Dynamic Errors-in-Variables Models. *Journal of Econometrics* **59**, 213–227.

Pagano, M. (1974): Estimation of Models of Autoregressive Signal Plus White Noise. *Annals of Statistics* **2**, 99–108.

Staudenmayer, J., and Buonaccorsi, J.P. (2005): Measurement Error in Linear Autoregressive Models. *Journal of the American Statistical Association* **100**, 841–852.