MEASUREMENT ERRORS IN DYNAMIC PANEL DATA ANALYSIS: A SYNTHESIS ON MODELING AND GMM ESTIMATION

ERIK BIØRN
Department of Economics,
University of Oslo,
P.O. Box 1095 Blindern,
0317 Oslo, Norway
E-mail: erik.biorn@econ.uio.no

Preliminary version, May 25, 2009
Comments are welcome. Please do not quote

ABSTRACT: A panel data model which combines an autoregressive fixed effects panel data equation and a static equation with measurement errors is considered. Examples of models for pure time series data with errors in variables are the starting point. Two versions of the Generalized Method of Moments (GMM) for this panel data model are considered, with fixed effects accounted for by, respectively, (i) transforming the equation to differences and keeping the instruments in levels, (ii) keeping the equation in levels and transforming the instruments to differences. In specifying the set of valid moment conditions and instrument sets, patterns with finite memory of disturbances, of latent regressors and of the measurement errors are discussed. Focus is on examining how the potential instrument sets which satisfy both the rank conditions for the instrumented variables and the orthogonality conditions for the composite errors and disturbances, change when the model's memory pattern changes. Sometimes the joint occurrence of measurement errors and long memories implies that the potential instrument sets becomes too small to make consistent estimation possible. An application based on panel data for capital stock and output from Norwegian manufacturing firms is presented.

KEYWORDS: Panel data, Measurement error, Dynamic modeling, ARMA model, GMM

JEL CLASSIFICATION: C21, C23, C31, C33, C51, E21

ACKNOWLEDGEMENTS: Earlier versions of the paper were presented at the Conference on Factor Structures for Panel and Multivariate Time Series Data, Maastricht, September 2008, and at seminars at the University of Oslo, May 2008, and Statistics Norway, October 2008. I am grateful to Xuehui Han for efficient programming of the estimation and testing procedures in the Gauss software code and to Terje Skjerpen and conference/seminar participants for comments and suggestions.

1 Introduction

For more than six decades it has been common knowledge that the occurrence of endogenous right-hand side (RHS) variables correlated with the disturbance in a static regression equation biases Ordinary Least Squares (OLS) estimators – the 'simultaneity problem'. Similar simultaneity biases arise if regressors are mis-measured or if, in a dynamic equation, lagged endogenous regressors and a disturbance with memory occur concurrently. Specifically, inconsistency of OLS follows if (a) in a static equation random measurement errors affect the regressors in a *static* equation and if (b) in a dynamic equation *lagged endogenous variables* occur jointly with *autocorrelated disturbances.* In the (a) case – unless extraneous information, say parameter restrictions or valid instrument variables (IVs) for the error-ridden regressors, exists – coefficients cannot be identified from uni-dimensional data (cross-section or time series data). In the (b) case, consistent estimation can be ensured by using lagged values of the exogenous, and sometimes also values of the lagged endogenous variables, as IVs.

If panel data are available the situation may be more favourable. This basically is due to the fact that its two-dimensional variation can serve to alleviate the identification and simultaneity problems, because linear data transformations can be performed along one dimension, for instance to eliminate heterogeneity, and still one data dimension is left, making the construction of valid IVs possible. Such ideas have been explored for panel data models in the (a) case in Griliches and Hausman (1986), Wansbeek and Koning (1991), Biørn (1992, 1996, 2000), Wansbeek and Meijer (2000, section 6.9), Wansbeek (2001) and Biørn and Krishnakumar (2008, Section 10.2). For the (b) case, ways of ensuring consistency when using panel with finite time series length are discussed in Balestra and Nerlove (1966), Anderson and Hsiao (1981, 1982), Sevestre and Trognon (1985, 1996), Holtz-Eakin *et al.* (1988), Arellano and Bond (1991), Ahn and Schmidt (1995), Blundell and Bond (1998), and Arellano (2003, Chapters 7 and 8). For *uni-dimensional data*, consistent IV estimation of static errors in variables models is discussed in Fuller (1987, Sections 1.4 and 2.4), among others. Grether and Maddala (1973), Pagano (1974), and Staudenmayer and Buonaccorsi (2005) discuss distributed lag models for pure time series combining errors in variables and serially correlated disturbances. Maravall and Aigner (1977), Maravall (1979) and Nowak (1993) discuss identification problems for such models.

This paper sets out to elaborate Generalized Method of Moments (GMM) to handle the joint occurrence of random measurement errors and autoregressive mechanisms in panel data. Simultaneity in panel data is discussed in Biørn and Krishnakumar (2008, Section 10.3) and will not be dealt with here. Motivating examples easily come to mind. One is a partial adjustment, or an equilibrium-correcting mechanism for a firm's capital-labour ratio motivated by an error-inflicted labour/capital service price ratio. Another is an autoregressive household consumption function or savings function with poorly measured income and net wealth. A third example is an autoregressive equation explaining individual wage rates by noisy measures of ability, education achievement or work experience.

Our specific attention will be given to the estimation of linear, autoregressive equations from balanced panel data with additive, random measurement errors (errors in variables, EIV) by the GMM. We show that it is possible to estimate such equations consistently while eliminating fixed and unstructured individual heterogeneity in two different ways: (A) Keeping the equation in its level form and using *x-values (or y-values) in differences as IVs* for the level x-values. (B) Transforming the equation to differences and using *x-values (or y-values) in levels as IVs* for the differenced x-values. The errors do not need to be white noise, which – although quite often assumed – may be unrealistic. (In the following 'error' will be used as synonymous with 'random measurement error'.)

A notable point is that GMM estimators valid for non-memory situations can be modified to account for *finite memory of errors or disturbances* by modifying the IV set. This memory pattern will be represented by moving average (MA) processes. The essence of this modification is to *reduce the IV set* in such a way that all IVs 'get clear of' the memory of the error process, in order to ensure that the IVs and the errors/disturbances are still uncorrelated (*the orthogonality condition*), while ensuring fulfilment of the claim that the IVs be correlated with the variables for which they serve as IVs (*the rank condition*). Certain extensions of static EIV models to finite memory cases are discussed in Biørn (2000, 2003). The present paper sets out to extend these ideas more thoroughly from static to autoregressive models.

The paper proceeds as follows. In Section 2 five related time series models with measurement errors and their OLS biases are considered to fix ideas, before in Section 3 the panel data model is presented and examples related to the literature are described. The GMM principle, without particular reference to the present panel data setting, is described in Section 4, for later reference. In Section 5 focus is on the orthogonality and rank conditions and the valid IV sets. Section 6 elaborates GMM estimation for an ARMAX-EIV equation in levels and for its first-differenced version. Similarities and dissimilarities between implementing the GMM principle in an ARX panel data model without errors and in a static EIV model for panel data are synthesized in Section 7. An application based on panel data for capital stock and output from Norwegian manufacturing firms is presented in Section 8. It illustrates the on the one hand the effect of the transformation from levels to differences and the implied changes in the instrument set and how the instrument sent and the coefficient estimates change when memory pattern of the latent regressor and the error elements changes. Section 9 concludes.

2 Measurement error biases in time series models

Before considering the measurement error problem in a dynamic panel data context and possible solutions to the estimation problem, we illustrate the corresponding problem for pure time series models by five examples. One is a static model, the other are dynamic. Three examples relate to univariate models, the others contain exogenous variables. The models differ with respect to the memory pattern and serve to illustrate how dynamics and measurement errors may interact in creating bias when using OLS regression.

Example 1: A static EIV equation with memory in error and disturbance:
The first example is a benchmark case where a static regression equation has a memory in either the latent regressor ξ_t, the measurement error ϵ_t or the disturbance v_t:

$$
\begin{aligned}
&y_t &&= \beta\xi_t + v_t,\\
&x_t &&= \xi_t + \epsilon_t,\\
&\mathsf{E}(\xi_t) &&= \mathsf{E}(v_t) = \mathsf{E}(\epsilon_t) = 0,\\
&\mathsf{E}(\xi_t\xi_{t-s}) &&= \sigma_{\xi\xi(S)}, \quad \mathsf{E}(v_tv_{t-s}) = \sigma_{uu(s)}, \quad \mathsf{E}(\epsilon_t\epsilon_{t-s}) = \sigma_{\epsilon\epsilon(s)}, \quad \xi_t \perp \epsilon_t \perp v_t,
\end{aligned}
\tag{2.1}
$$

$\perp$ symbolizing 'orthogonal to'. For simplicity we assume that the latent 'structural' variable, μ_t, has zero expectation, and omit the intercept from the equation. It is well known that if neither of ϵ_t and v_t has a memory [$\sigma_{uu(s)}=\sigma_{\epsilon\epsilon(s)}=0$, $s=1,2,\dots$], then the plim of the OLS estimator of β is [see, *e.g.*, Fuller (1987, p. 11)]

$$\bar{\beta} = \frac{\sigma_{yx}}{\sigma_{xx}} = \beta\, k \tag{2.2}$$

where k, often denoted as attenuation, is

$$k = \frac{\sigma_{\xi\xi(0)}}{\sigma_{\xi\xi(0)}+\sigma_{\epsilon\epsilon(0)}}. \tag{2.3}$$

If there is memory in (v_t, ϵ_t), (2.2)–(2.3) still holds and the OLS residuals converge to

$$e_t = y_t - \bar{\beta}x_t = \beta[(1-k)\xi_t - k\epsilon_t] + v_t, \tag{2.4}$$

which shows that the discrepancy between the asymptotic OLS residual and the disturbance equals $e_t - v_t = \beta[(1-k)\xi_t - k\epsilon_t]$. The autocovariance of this asymptotic residual is

$$\operatorname{cov}(e_t, e_{t-s}) = \beta^2[(1-k)^2\sigma_{\xi\xi(s)} + k^2\sigma_{\epsilon\epsilon(s)}] + \sigma_{uu(s)}, \qquad s = 0, 1, 2, \dots \tag{2.5}$$

This gives the following conclusion: Even if the disturbance v_t is white noise, OLS induces not only attenuation [$k \in (0,1)$], but also spurious serial correlation in residuals as long as the latent regressor or the error is autocorrelated. This was pointed out for a related example by Grether and Maddala (1973).

Example 2: An autoregressive model with white noise error:
The second example is a simple time series model without any covariate:

$$
\begin{aligned}
&\mu_t &&= \phi(\mathsf{L})\mu_t + \delta_t \iff \mu_t = \tfrac{1}{1-\phi(\mathsf{L})}\delta_t\\
&y_t &&= \mu_t + \epsilon_t,\\
&\delta_t &&\sim \mathsf{IID}(0, \sigma_\delta^2), \qquad \epsilon_t \sim \mathsf{IID}(0, \sigma_\epsilon^2), \qquad \mathsf{E}(\mu_0) = 0, \qquad \delta_t \perp \epsilon_t,
\end{aligned}
\tag{2.6}
$$

where $\phi(\mathsf{L}) = \phi_1\mathsf{L} + \phi_2\mathsf{L}^2 + \dots \phi_P\mathsf{L}^P$, L being the backshift operator. All values of z satisfying $1-\phi(z)=0$ are assumed to lie outside the unit circle, so that μ_t and y_t are stationary variables. For simplicity we assume that the latent variable μ_t has zero expectation and omit the intercept from the equation. It follows that the observed, error-contaminated variable y_t follows the ARMA(P,P)-process:

$$y_t = \phi(\mathsf{L})y_t + \epsilon_t + \delta_t - \phi(\mathsf{L})\epsilon_t. \tag{2.7}$$

Obviously, OLS estimation of this equation is inconsistent for the parameters of $\Phi(\mathsf{L})$. This kind of model and procedures for consistent estimation are discussed, mostly under the additional assumption of normality of the errors, in Pagano (1974) and Staudenmayer and Buonaccorsi (2005).

Example 3: An ARMA(1,1) model with a white noise error:
The next example is an ARMA$(1,1)$-process for a latent variable μ_t observed with a white noise error δ_t:

$$\begin{aligned} \mu_t &= \gamma\mu_{t-1} + v_t + \lambda v_{t-1} \iff \mu_t = \tfrac{1+\lambda\mathsf{L}}{1-\gamma\mathsf{L}}v_t = \left[1 + \tfrac{(\gamma+\lambda)\mathsf{L}}{1-\gamma\mathsf{L}}\right]v_t, \quad |\gamma|<1,\\ y_t &= \mu_t + \delta_t,\\ \mathsf{E}(\mu_0) &= 0, \quad v_t \sim \mathsf{IID}(0,\sigma_v^2), \quad \delta_t \sim \mathsf{IID}(0,\sigma_\delta^2), \quad v_t \perp \delta_t. \end{aligned} \tag{2.8}$$

It follows that the observed variable y_t is generated by

$$y_t = \frac{1+\lambda\mathsf{L}}{1-\gamma\mathsf{L}}v_t + \delta_t. \tag{2.9}$$

Hence,

$$\mathsf{E}(\mu_t^2) = \sigma_{\mu\mu} = (1+\chi^2)\sigma_v^2, \qquad \mathsf{E}(\mu_t\mu_{t-1}) = \sigma_{\mu\mu(1)} = [\gamma(1+\chi^2)+\lambda]\sigma_v^2,$$

where

$$\chi = \frac{\gamma+\lambda}{\sqrt{1-\gamma^2}}.$$

Taking the plim of the OLS estimator of γ, $\widehat{\gamma} = [\sum_t y_t y_{t-1}]/[\sum_t y_{t-1}^2]$, gives

$$\bar{\gamma} = \frac{\sigma_{yy(1)}}{\sigma_{yy}} = \frac{\sigma_{\mu\mu(1)}}{\sigma_{\mu\mu}+\sigma_\delta^2} = \frac{\gamma[1+\chi^2]+\lambda}{[1+\chi^2]+\frac{\sigma_\delta^2}{\sigma_v^2}}, \tag{2.10}$$

where $\sigma_{yy} = \text{var}(y_t)$, $\sigma_{yy(1)} = \text{cov}(y_t, y_{t-1})$. Hence, the bias in $\widehat{\gamma}$ is affected by both the model's MA parameter ($\lambda \neq 0$) and its error variance ($\sigma_\delta^2 > 0$). Since

$$\bar{\gamma} \gtreqless \gamma \iff \frac{\lambda}{\gamma} \gtreqless \frac{\sigma_\delta^2}{\sigma_v^2},$$

assuming $\lambda > 0$, the attenuation of OLS may (partly) be counteracted by the memory of the MA part of the equation's disturbance.

Example 4: An ARMA(1,1) model with an MA(1) error:

The next example also is an ARMA(1,1)-process, now with the error process extended to an MA(1) process, which increases the complexity of the memory structure:

$$(2.11)\qquad \begin{aligned} \mu_t &= \gamma\mu_{t-1} + v_t + \lambda v_{t-1} \iff \mu_t = \tfrac{1+\lambda \mathsf{L}}{1-\gamma \mathsf{L}} v_t, \quad |\gamma| < 1, \\ y_t &= \mu_t + \delta_t + \psi\delta_{t-1}, \\ \mathsf{E}(\mu_0) &= 0, \quad v_t \sim \mathsf{IID}(0,\sigma_v^2), \quad \delta_t \sim \mathsf{IID}(0,\sigma_\delta^2), \quad v_t \perp \delta_t. \end{aligned}$$

It follows that

$$(2.12)\qquad y_t = \frac{1+\lambda \mathsf{L}}{1-\gamma \mathsf{L}} v_t + (1+\psi \mathsf{L})\delta_t,$$

showing that y_t is generated by an ARMA(1,2) process, $1-\gamma\mathsf{L}$ being a common factor in the lag polynomials for y_t and δ_t. We further obtain

$$\mathsf{E}(\mu_t^2) = \sigma_{\mu\mu} = (1+\chi^2)\sigma_v^2, \qquad \mathsf{E}(\mu_t\mu_{t-1}) = \sigma_{\mu\mu(1)} = [\gamma(1+\chi^2)+\lambda]\sigma_v^2$$

Taking the plim of the OLS estimator of γ, $\widehat{\gamma} = [\sum_t y_t y_{t-1}]/[\sum_t y_{t-1}^2]$, gives the following generalization of (2.10)

$$(2.13)\qquad \bar{\gamma} = \frac{\sigma_{yy(1)}}{\sigma_{yy}} = \frac{\sigma_{\mu\mu(1)} + \psi\sigma_\delta^2}{\sigma_{\mu\mu} + (1+\psi^2)\sigma_\delta^2} = \frac{\gamma(1+\chi^2)+\lambda+\psi\frac{\sigma_\delta^2}{\sigma_v^2}}{(1+\chi^2)+(1+\psi^2)\frac{\sigma_\delta^2}{\sigma_v^2}}.$$

Table 1: Example 4: Values of $\bar{\gamma} = \text{plim}(\widehat{\gamma})$ in boundary cases

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

Now, the bias of OLS depends on both λ and ψ, the first-order parameters in the MA processes for the equation's disturbance and error, respectively. Table 1 summarizes the expressions for $\bar{\gamma}$ in 15 boundary cases, $\psi=0$ corresponding to Example 3.

If, in particular, $\gamma=0$, when writing the equation as

$$y_t = v_t^* + \delta_t^*, \qquad v_t^* = v_t+\lambda v_{t-1}, \qquad \delta_t^* = \delta_t+\psi\delta_{t-1},$$

$\bar{\gamma}$ becomes a variance-weighted average of the autocorrelation coefficients of v_t^* and δ_t^*, which are, respectively, $\rho_v^* = \lambda/(1+\lambda^2)$ and $\rho_\delta^* = \psi/(1+\psi^2)$:

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

EXAMPLE 5: AN ARMA(1,1) MODEL WITH EXOGENOUS VARIABLE AND WITH MA(1) ERRORS: The final time series example extends Example 4 by including in the ARMA(1,1)-process an exogenous, error-ridden regressor, givng an ARMAX model. Both errors are MA(1) processes. The model is:

$$\begin{aligned}
&\mu_t &&= \gamma\mu_{t-1} + \beta\xi_t + v_t + \lambda v_{t-1}, \qquad |\gamma| < 1,\\
&y_t &&= \mu_t + \delta_t + \psi\delta_{t-1},\\
&x_t &&= \xi_t + \epsilon_t + \phi\epsilon_{t-1},\\
&\mathsf{E}(\xi_t) &&= 0, \quad \mathsf{E}(\xi_t\xi_{t-s}) = \begin{cases} \sigma_{\xi\xi(s)}, & s = 0,1,2;\\ 0, & s \geq 3,\end{cases}\\
&\mathsf{E}(\mu_0) &&= 0, \quad v_t \sim \mathsf{IID}(0,\sigma_v^2), \quad \delta_t \sim \mathsf{IID}(0,\sigma_\delta^2), \quad \epsilon_t \sim \mathsf{IID}(0,\sigma_\epsilon^2),\\
&\xi_t &&\perp v_t \perp \delta_t \perp \epsilon_t.
\end{aligned} \tag{2.14}$$

These assumptions imply

$$\mathsf{E}(y_t y_{t-s}) = \begin{cases} \mathsf{E}(\mu_t^2)+(1+\psi^2)\sigma_\delta^2, & s = 0,\\ \mathsf{E}(\mu_t\mu_{t-1})+\psi\sigma_\delta^2, & s = 1,\\ \mathsf{E}(\mu_t\mu_{t-2}), & s = 2,\end{cases} \tag{2.15}$$

$$\mathsf{E}(x_t x_{t-s}) = \begin{cases} \sigma_{\xi\xi(0)} + (1+\phi^2)\sigma_\epsilon^2, & s = 0,\\ \sigma_{\xi\xi(1)}+\phi\sigma_\epsilon^2, & s = 1,\\ \sigma_{\xi\xi(2)}, & s = 2.\end{cases} \tag{2.16}$$

$$\mathsf{E}(x_{t-\theta}y_{t-\tau}) = \mathsf{E}(\xi_{t-\theta}\xi_{t-\tau}), \qquad \theta = 0,1;\ \tau = 0,1,2. \tag{2.17}$$

From (2.14) it follows that

$$\mu_t = \frac{\beta}{1-\gamma\mathsf{L}}\xi_t + \frac{1+\lambda\mathsf{L}}{1-\gamma\mathsf{L}}v_t, \tag{2.18}$$

$$y_t = \frac{\beta}{1-\gamma\mathsf{L}}x_t + \frac{1+\lambda\mathsf{L}}{1-\gamma\mathsf{L}}v_t - \frac{\beta(1+\phi\mathsf{L})}{1-\gamma\mathsf{L}}\epsilon_t + (1+\psi\mathsf{L})\delta_t. \tag{2.19}$$

Since $|\gamma| < 1$ ensures invertibility, the last equation is equivalent to

$$y_t = \gamma y_{t-1} + \beta x_t + w_t, \tag{2.20}$$

$$w_t = (1+\lambda\mathsf{L})v_t - \beta(1+\phi\mathsf{L})\epsilon_t + (1-\gamma\mathsf{L})(1+\psi\mathsf{L})\delta_t. \tag{2.21}$$

The plims of the OLS estimator of (β,γ) and of the 'IV estimators' obtained by using, respectively, [1] x_{t-1} as IV for y_{t-1}, [2] x_{t-2} as IV for y_{t-1}, and [3] x_{t-1} as IV for x_t, are given by

$$[OLS]: \quad \begin{bmatrix} \bar{\gamma}^{OLS}\\ \bar{\beta}^{OLS}\end{bmatrix} = \begin{bmatrix} \mathsf{E}(y_{t-1}^2) & \mathsf{E}(x_t y_{t-1})\\ \mathsf{E}(y_{t-1}x_t) & \mathsf{E}(x_t^2)\end{bmatrix}^{-1} \begin{bmatrix} \mathsf{E}(y_{t-1}y_t)\\ \mathsf{E}(x_t y_t).\end{bmatrix},$$

$$[IV1]: \quad \begin{bmatrix} \bar{\gamma}^{IV1}\\ \bar{\beta}^{IV1}\end{bmatrix} = \begin{bmatrix} \mathsf{E}(x_{t-1}y_{t-1}) & \mathsf{E}(x_t y_{t-1})\\ \mathsf{E}(x_{t-1}x_t) & \mathsf{E}(x_t^2)\end{bmatrix}^{-1} \begin{bmatrix} \mathsf{E}(x_{t-1}y_t)\\ \mathsf{E}(x_t y_t)\end{bmatrix},$$

$$[IV2]: \quad \begin{bmatrix} \bar{\gamma}^{IV2}\\ \bar{\beta}^{IV2}\end{bmatrix} = \begin{bmatrix} \mathsf{E}(x_{t-2}y_{t-1}) & \mathsf{E}(x_t y_{t-1})\\ \mathsf{E}(x_{t-2}x_t) & \mathsf{E}(x_t^2)\end{bmatrix}^{-1} \begin{bmatrix} \mathsf{E}(x_{t-2}y_t)\\ \mathsf{E}(x_t y_t)\end{bmatrix},$$

$$[IV3]: \quad \begin{bmatrix} \bar{\gamma}^{IV3}\\ \bar{\beta}^{IV3}\end{bmatrix} = \begin{bmatrix} \mathsf{E}(y_{t-1}^2) & \mathsf{E}(x_{t-1}y_{t-1})\\ \mathsf{E}(y_{t-1}x_t) & \mathsf{E}(x_{t-1}x_t)\end{bmatrix}^{-1} \begin{bmatrix} \mathsf{E}(y_{t-1}y_t)\\ \mathsf{E}(x_{t-1}y_t)\end{bmatrix}.$$

However, in cases [1], [2] and [3] the IVs may be invalid. In the Appendix it is shown that

$$(2.22)\qquad \begin{aligned} \mathsf{E}(\xi_t\mu_t) &= \beta\Lambda(\gamma), \\ \mathsf{E}(\xi_t\mu_{t-1}) &= \beta(\sigma_{\xi\xi(1)}+\gamma\sigma_{\xi\xi(2)}), \\ \mathsf{E}(\xi_t\mu_{t-2}) &= \beta\sigma_{\xi\xi(2)}, \end{aligned}$$

$$(2.23)\qquad \begin{aligned} \mathsf{E}(\xi_{t-1}\mu_t) &= \beta[\gamma\Lambda(\gamma)+\sigma_{\xi\xi(1)}], \\ \mathsf{E}(\xi_{t-2}\mu_t) &= \beta[\gamma^2\Lambda(\gamma)+\gamma\sigma_{\xi\xi(1)}+\sigma_{\xi\xi(2)}], \end{aligned}$$

$$(2.24)\qquad \begin{aligned} \mathsf{E}(\mu_t^2) &= \tfrac{\beta^2}{1-\gamma^2}[\Lambda(\gamma)+\gamma\sigma_{\xi\xi(1)}+\gamma^2\sigma_{\xi\xi(2)}]+(1+\chi^2)\sigma_v^2, \\ \mathsf{E}(\mu_t\mu_{t-1}) &= \tfrac{\beta^2}{1-\gamma^2}[\gamma\Lambda(\gamma)+\sigma_{\xi\xi(1)}+\gamma\sigma_{\xi\xi(2)}]+[\gamma(1+\chi^2)+\lambda]\sigma_v^2, \\ \mathsf{E}(\mu_t\mu_{t-2}) &= \tfrac{\beta^2}{1-\gamma^2}[\gamma^2\Lambda(\gamma)+\gamma\sigma_{\xi\xi(1)}+\sigma_{\xi\xi(2)}]+\gamma[\gamma(1+\chi^2)+\lambda]\sigma_v^2. \end{aligned}$$

where

$$\Lambda(\gamma) = \sigma_{\xi\xi(0)}+\gamma\sigma_{\xi\xi(1)}+\gamma^2\sigma_{\xi\xi(2)},$$

Combining (2.15)–(2.17) with (2.22)–(2.24) the plims of the four sets of estimators can be derived. We find that not only [OLS], but also [IV1], [IV2] and [IV3] differ from $[\gamma,\beta]'$.

Consider, to fix ideas, as a simpler *example* where *the latent exogenous variable* ξ_t *has no memory, i.e.*, $\sigma_{\xi\xi(1)}=\sigma_{\xi\xi(2)}=0$. Then the regressors are orthogonal and we have

$$\mathsf{E}(x_t y_{t-1}) = \mathsf{E}(\xi_t\mu_{t-1}) = 0, \quad \mathsf{E}(x_t y_{t-2}) = \mathsf{E}(\xi_t\mu_{t-2}) = 0, \quad \Lambda(\gamma) = \sigma_{\xi\xi(0)}.$$

The plims of the four sets of estimators are

$$\bar{\gamma}^{OLS} = \frac{\mathsf{E}(y_{t-1}y_t)}{\mathsf{E}(y_{t-1}^2)} = \frac{\frac{\beta^2}{1-\gamma^2}\gamma\sigma_{\xi\xi(0)}+[\gamma(1+\chi^2)+\lambda]\sigma_v^2+\psi\sigma_\delta^2}{\frac{\beta^2}{1-\gamma^2}\sigma_{\xi\xi(0)}(1+\chi^2)\sigma_v^2+(1+\psi^2)\sigma_\delta^2}, \qquad \bar{\beta}^{OLS} = \frac{\mathsf{E}(x_t y_t)}{\mathsf{E}(x_t^2)} = \frac{\beta\sigma_{\xi\xi(0)}}{\sigma_{\xi\xi(0)}+(1+\phi^2)\sigma_\epsilon^2},$$

$$\bar{\gamma}^{IV1} = \frac{\mathsf{E}(x_{t-1}y_t)}{\mathsf{E}(x_{t-1}y_{t-1})} = \frac{\beta\gamma\sigma_{\xi\xi(0)}}{\beta\sigma_{\xi\xi(0)}} = \gamma, \qquad \bar{\beta}^{IV1} = \frac{\mathsf{E}(x_t y_t)}{\mathsf{E}(x_t^2)} = \frac{\beta\sigma_{\xi\xi(0)}}{\sigma_{\xi\xi(0)}+(1+\phi^2)\sigma_\epsilon^2},$$

$$\bar{\gamma}^{IV2} = \frac{\mathsf{E}(x_{t-2}y_t)}{\mathsf{E}(x_{t-2}y_{t-1})} = \frac{\beta\gamma^2\sigma_{\xi\xi(0)}}{\beta\gamma\sigma_{\xi\xi(0)}} = \gamma, \qquad \bar{\beta}^{IV2} = \frac{\mathsf{E}(x_t y_t)}{\mathsf{E}(x_t^2)} = \frac{\beta\sigma_{\xi\xi(0)}}{\sigma_{\xi\xi(0)}+(1+\phi^2)\sigma_\epsilon^2},$$

$$\bar{\gamma}^{IV3} = \frac{\mathsf{E}(y_{t-1}y_t)}{\mathsf{E}(y_{t-1}^2)} = \frac{\frac{\beta^2}{1-\gamma^2}\gamma\sigma_{\xi\xi(0)}+[\gamma(1+\chi^2)+\lambda]\sigma_v^2+\psi\sigma_\delta^2}{\frac{\beta^2}{1-\gamma^2}\sigma_{\xi\xi(0)}(1+\chi^2)\sigma_v^2+(1+\psi^2)\sigma_\delta^2}, \qquad \bar{\beta}^{IV3} = \frac{\mathsf{E}(x_{t-1}y_t)}{\mathsf{E}(x_{t-1}x_t)} = \frac{\beta\gamma\sigma_{\xi\xi(0)}}{\phi\sigma_\epsilon^2}.$$

All estimators of β are inconsistent. For γ, the estimators [IV1] and [IV2] ensure consistency. This reflects on the one hand that not all 'IV sets' are valid since they are not orthogonal to w_t, confer (2.20)–(2.21). On the other hand that the orthogonality of x_t and y_{t-1} implies that an inconsistency of the β estimator will not infect the γ estimator and *vice versa*.

Examples 1 through 5 serve to demonstrate that for dynamic time series models the bias in OLS estimation – and frequently also in IV estimation – induced by the occurrence of (measurement) errors and the bias induced by its memory are 'intermingled'. At the one extreme, for a static regression equation with error-ridden regressor [confer Grether and Maddala (1973) and Example 1] a memory of the errors may induce spurious serial correlation in OLS residuals. At the other extreme, OLS estimation of an error-ridden AR equation where both the disturbance and error are white noise processes, is biased as well [confer Example 3 with $\lambda = 0$].

Since panel data have a cross-sectional dimension added to the time dimension, and since it is known that this added dimension may improve the possibility for identification and consistent estimation [see Biørn (2000)], the following question arises: How should we proceed to estimate regression coefficients in AR, ARMA or ARMAX models with EIV consistently and efficiently taking advantage of the cross-sectional dimension added to the time dimension of the data? The focus of the rest of the paper will be on this problem.

3 An ARMAX-EIV model for panel data

We now consider a panel data model with individual heterogeneity which is, formally, an extension of Examples 2–5 in Section 2. First, the scalar regressors (ξ, μ) and error in the regressor (δ) in the time series models are extended to vectors $(\boldsymbol{\xi}, \boldsymbol{\mu})$ and $\boldsymbol{\delta}$, respectively. Second, the orders of the MA parts, *i.e.*, the memory of the errors and disturbances as well as the memory of the latent exogenous variables can take any *finite* values.

Assume that $N\ (\geq 2)$ individuals, indexed by i, are observed in $T\ (\geq 2)$ periods, indexed by t, and consider the following dynamic equation system for panel data with errors:

$$\mu_{it} = \alpha_i + \boldsymbol{\xi}_{it}\boldsymbol{\beta} + \mu_{i,t-1}\lambda + u_{it}, \qquad |\lambda| < 1, \tag{3.1}$$

$$y_{it} = \mu_{it} + \nu_{it}, \tag{3.2}$$

$$\boldsymbol{q}_{it} = \boldsymbol{\xi}_{it} + \boldsymbol{\eta}_{it}, \tag{3.3}$$

$$\boldsymbol{\xi}_{it} \perp u_{it} \perp \nu_{it} \perp \boldsymbol{\eta}_{it}, \qquad i = 1, \ldots, N;\ t = 1, \ldots, T. \tag{3.4}$$

where α_i is a *fixed* effect, specific to individual i, $(\mu_{it}, \boldsymbol{\xi}_{it})$ are latent variables, $(y_{it}, \boldsymbol{q}_{it})$ are their observable counterparts, $(\nu_{it}, \boldsymbol{\eta}_{it})$ are errors, u_{it} is a disturbance, and $(\lambda, \boldsymbol{\beta}', k)$ are constants. Boldface letters denote vectors, *rows for variables, columns for coefficients.* Unlike standard errors-in-variables models, where the errors are contemporaneously uncorrelated white noise, we allow $(u_{it}, \nu_{it}, \boldsymbol{\eta}_{it})$ to have finite memory equal to (N_u, N_ν, N_η), respectively. This would for instance follow if

$$\begin{aligned}
u_{it} &= v_{it} + \textstyle\sum_{s=1}^{N_u} v_{i,t-s}\theta_s, && v_{it} \sim \mathsf{IID}(0, \sigma_v^2), \\
\nu_{it} &= \delta_{it} + \textstyle\sum_{s=1}^{N_\nu} \delta_{i,t-s}\psi_s, && \delta_{it} \sim \mathsf{IID}(0, \sigma_\delta^2), \\
\boldsymbol{\eta}_{it} &= \boldsymbol{\epsilon}_{it} + \textstyle\sum_{s=1}^{N_\eta} \boldsymbol{\epsilon}_{i,t-s}\boldsymbol{\phi}_s, && \boldsymbol{\epsilon}_{it} \sim \mathsf{IID}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}), \qquad i = 1, \ldots, N;\ t = 1, \ldots, T.
\end{aligned}$$

Finally, we let $\boldsymbol{\xi}_{it}$ have an N_ξ-period memory, which would for instance follow if

$$\boldsymbol{\xi}_{it} = \boldsymbol{\zeta}_{it} + \textstyle\sum_{s=1}^{N_\xi} \boldsymbol{\zeta}_{i,t-s}\boldsymbol{\Lambda}_s, \qquad \boldsymbol{\zeta}_{1t}, \ldots, \boldsymbol{\zeta}_{N_\xi t} \sim \mathsf{IID}(\boldsymbol{\mu}_{\boldsymbol{\zeta}_t}, \boldsymbol{\Sigma}_{\boldsymbol{\zeta}_t}).$$

Allowing for memory of errors is an interesting extension of a standard AR-EIV model in several practical situations. One motivation may be that y or $\boldsymbol{q}$ include a stock variable, *i.e.*, of finished goods in a manufacturing firm or of fixed production capital constructed from on cumulated flows, in which case (measurement) errors tend to vary cyclically.

For a flow variable the motivation may come from improper periodization of economic transactions, for example sales, investment, cash-flow and income, which may create serial correlation – often negative – between errors which are close in time. Another motivation may be that the level variables rather than being stationary may be integrated of order P (say) with white noise measurement errors. When differencing P times to make the observed variables stationary, the resulting errors become MA(P) processes.

As illustrated in Biørn (2000, Section 2.b) [see also Biørn and Krishnakumar (2008, Section 10.2.2)], an IID property for latent regressors would have been detrimental to identification of slope coefficients in a static panel data model. It is important to allow for some memory of $\boldsymbol{\xi}$: "$\cdots$ in order to ensure identification of the slope coefficient vector from panel data, there should not be 'too much structure' on the second order moments of the latent exogenous regressors along the time dimension, and not 'too little structure' on the second order moments of the errors and disturbances along the time dimension." [Biørn (2000, p. 398)]. Therefore $N_\xi \geq 1$ is an important assumption of the model.

By eliminating μ_{it} and $\boldsymbol{\xi}_{it}$ from (3.1)–(3.3) we obtain

$$y_{it} = \alpha_i + \boldsymbol{q}_{it}\boldsymbol{\beta} + y_{i,t-1}\lambda + w_{it}, \tag{3.5}$$

$$w_{it} = u_{it} + \nu_{it} - \nu_{i,t-1}\lambda - \boldsymbol{\eta}_{it}\boldsymbol{\beta}. \tag{3.6}$$

A one-period differencing which eliminates the fixed individual effect, yields

$$\Delta y_{it} = \Delta\boldsymbol{q}_{it}\boldsymbol{\beta} + \Delta y_{i,t-1}\lambda + \Delta w_{it}, \tag{3.7}$$

$$\Delta w_{it} = \Delta u_{it} + \Delta\nu_{it} - \Delta\nu_{i,t-1}\lambda - \Delta\boldsymbol{\eta}_{it}\boldsymbol{\beta}. \tag{3.8}$$

Solving (3.5) and (3.7) by an infinite number of backward substitutions, we get

$$y_{it} = \frac{\alpha_i}{1-\lambda} + \textstyle\sum_{s=0}^{\infty} \lambda^s[\boldsymbol{q}_{i,t-s}\boldsymbol{\beta} + w_{i,t-s}], \tag{3.9}$$

$$\Delta y_{it} = \textstyle\sum_{s=0}^{\infty} \lambda^s[\Delta\boldsymbol{q}_{i,t-s}\boldsymbol{\beta} + \Delta w_{i,t-s}]. \tag{3.10}$$

It is evident that $(y_{it}, \boldsymbol{q}_{i,t+\tau}, w_{i,t+\theta})$ and $(\Delta y_{it}, \Delta\boldsymbol{q}_{i,t+\tau}, \Delta w_{i,t+\theta})$ are correlated for some (τ, θ) and uncorrelated for others. Examining this is needed to delimit orthogonality conditions and valid IVs for GMM estimation of $(\boldsymbol{\beta}, \lambda)$. This is the topic of Section 5.

Special cases of this model, some familiar and some not much discussed in the literature – ARX and ARMAX being acronyms for AR and ARMA models with exogenous variables, respectively – are:

(i) Static EIV model with memory in errors & disturbances: $\lambda=0$.
(ii) Static EIV model, no memory in errors & disturbances: $\lambda=0; \theta_s=\psi_s=0; \boldsymbol{\phi}_s=\mathbf{0}$, $s=1,2,\ldots$.
(iii) ARMAX model with no errors: $\sigma_\delta^2=0$, $\boldsymbol{\Sigma_\epsilon}=\mathbf{0}$.
(iv) ARX(1) model with no errors: $\sigma_\delta^2=0$, $\boldsymbol{\Sigma_\epsilon}=\mathbf{0}$, $\theta_s=\psi_s=0$, $\boldsymbol{\phi}_s=\mathbf{0}$, $s=1,2,\ldots$.
(v) EIV-ARX(1) model, no memory in errors: $\theta_s=\psi_s=0$, $\boldsymbol{\phi}_s=\mathbf{0}$, $s=1,2,\ldots$.

Cases (i) and (v) include Case (ii), and Case (iii) includes Case (iv). Case (v) will be discussed as an example in Section 5.

How does this model *relate to the literature*? Case (ii), and to some extent the more general Case (i), is discussed in Griliches and Hausman (1986), Wansbeek and Koning (1991), Biørn (1992, 1996, 2000), Wansbeek (2002) and Biørn and Krishnakumar (2008, Section 10.2). Case (iv) resembles cases discussed in Balestra and Nerlove (1966), Anderson and Hsiao (1981, 1982), Sevestre and Trognon (1985, 1996), Arellano and Bond (1991), and Arellano and Bover (1995). Holtz-Eakin *et al.* (1988) consider case (iii) as well as a generalization of Case (iii) which allows for higher-order autocorrelation, although in a bivariate context. The extension of this higher-order model to include errors with no memory is discussed in Holtz-Eakin *et al.* (1988, pp. 1376–1377). Finally, by modifying the specification of the dynamic behaviour of the latent regressor vector, given in (**??**), allowing for cross-sectional dependence this model could also include elements from the literature on *factor models* for panel data, see, *e.g.*, Pesaran (2006, 2007).

Throughout we will assume that the *time series length T is finite.* If $T \to \infty$ were allowed, we could take advantage of the repeated measurement property of panel data to construct consistent estimators, since our assumptions in conjunction with the law of large moments imply $\text{plim}_{T\to\infty}(\bar{\boldsymbol{q}}_{\cdot t} - \bar{\boldsymbol{\xi}}_{\cdot t}) = \mathbf{0}$, $\text{plim}_{T\to\infty}(\bar{y}_{\cdot t} - \bar{\mu}_{\cdot t}) = \text{plim}_{T\to\infty}(\bar{u}_{\cdot t}) = \text{plim}_{T\to\infty}(\bar{w}_{\cdot t}) = 0$, etc., subscript $\cdot t$ denoting averages specific to period t. This solution to the measurement error problem – which may be of limited practical interest since the transformation eliminates all between-individual variation in the data, which often dominates – is elaborated for static measurement errors models in Biørn (2000, Section 3) [see also Biørn and Krishnakumar (2008, Section 10.2.4)], and will not be expanded for the more general model considered here.

4 Generalities on GMM estimation in a linear model

Before elaborating estimation of $(\boldsymbol{\beta}, \gamma)$ by using GMM on (3.5) or (3.7), we recapitulate for later reference, the essentials of the method when applied to a *linear* equation.

Assume that we want to estimate the $(K \times 1)$ coefficient vector $\boldsymbol{\beta}$ in the equation

$$y = \boldsymbol{x}\boldsymbol{\beta} + \epsilon, \tag{4.1}$$

where y and ϵ are scalars and $\boldsymbol{x}$ is a $(1 \times K)$ regressor vector (which may well include lagged values of y). There exists an IV vector $\boldsymbol{z}$, of dimension $(1 \times G)$, for $\boldsymbol{x}$ $(G \geq K)$, satisfying the orthogonality conditions

$$\mathsf{E}(\boldsymbol{z}'\epsilon) = \mathsf{E}[\boldsymbol{z}'(y - \boldsymbol{x}\boldsymbol{\beta})] = \mathbf{0}_{G,1}. \tag{4.2}$$

From n observations on $(y, \boldsymbol{x}, \boldsymbol{z})$, denoted as $\{y_j, \boldsymbol{x}_j, \boldsymbol{z}_j\}_{j=1}^{j=n}$, we define the vector valued $(G \times 1)$ function of corresponding empirical means,

$$\boldsymbol{g}_n(y, \boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\beta}) = \tfrac{1}{n}\textstyle\sum_{j=1}^{n} \boldsymbol{z}_j'(y_j - \boldsymbol{x}_j\boldsymbol{\beta}), \tag{4.3}$$

to be considered the empirical counterpart to $\mathsf{E}[\boldsymbol{z}'(y - \boldsymbol{x}\boldsymbol{\beta})]$ based on the sample. The *essence of the GMM* [see, *e.g.*, Davidson and MacKinnon (2004, Section 9.2)] is to choose as an estimator for $\boldsymbol{\beta}$ the value which brings $\boldsymbol{g}_n(y, \boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\beta})$ as close to the zero vector $\mathbf{0}_{G,1}$, as possible. GMM solves the estimation problem by *minimizing a distance measure* *represented by a quadratic form in $g_n(y, \boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\beta})$ for a suitably chosen positive definit $(G{\times}G)$ weighting matrix $\boldsymbol{W}_n$, i.e.,*

$$(4.4) \qquad \boldsymbol{\beta}^*_{GMM} = \boldsymbol{\beta}^*_{GMM}(\boldsymbol{W}_n) = \text{argmin}_{\boldsymbol{\beta}}[\boldsymbol{g}_n(y, \boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\beta})' \boldsymbol{W}_n \boldsymbol{g}_n(y, \boldsymbol{x}, \boldsymbol{z}; \boldsymbol{\beta})].$$

All estimators obtained in this way are consistent, and the choice of $\boldsymbol{W}_n$ determines the estimator's efficiency. A choice which leads to an asymptotically efficient estimator of $\boldsymbol{\beta}$, is to set $\boldsymbol{W}_n$ equal (or proportional) to the inverse of (an estimate of) the (asymptotic) covariance matrix of $\frac{1}{n}\sum_{j=1}^{n} \boldsymbol{z}'_j \epsilon_j$; see, *e.g.*, Davidson and MacKinnon (1993, Theorem 17.3) and Harris and Mátyás (1999, section 1.3.3).

If ϵ is homoskedastic, with variance σ^2_ϵ, the appropriate choice is $\boldsymbol{W}_n = \frac{n^2}{\sigma^2_\epsilon}[\sum_j \boldsymbol{z}'_j \boldsymbol{z}_j]^{-1}$. The resulting estimator implied by (4.4) is

$$(4.5) \quad \widehat{\boldsymbol{\beta}}_{GMM} = \boldsymbol{\beta}^*_{GMM}[\tfrac{n^2}{\sigma^2_\epsilon}(\textstyle\sum_j \boldsymbol{z}'_j \boldsymbol{z}_j)^{-1}]$$
$$= [(\textstyle\sum_j \boldsymbol{x}'_j \boldsymbol{z}_j)(\sum_j \boldsymbol{z}'_j \boldsymbol{z}_j)^{-1}(\sum_j \boldsymbol{z}'_j \boldsymbol{x}_j)]^{-1}[(\sum_j \boldsymbol{x}'_j \boldsymbol{z}_j)(\sum_j \boldsymbol{z}'_j \boldsymbol{z}_j)^{-1}(\sum_j \boldsymbol{z}'_j y_j)].$$

If ϵ_j has an unspecified heteroskedasticity, we first construct residuals $\widehat{\epsilon}_j$ from (4.5), considered a *first-step GMM estimator*, and replace $\boldsymbol{W}_n$ by $[n^{-2}\sum_j \boldsymbol{z}'_j \widehat{\epsilon}^2_j \boldsymbol{z}_j]^{-1}$; see White (1984, sections IV.3 and VI.2). Inserting this into (4.4) gives

$$(4.6) \quad \widetilde{\boldsymbol{\beta}}_{GMM} = \boldsymbol{\beta}^*_{GMM}[n^2(\textstyle\sum_j \boldsymbol{z}'_j \widehat{\epsilon}^2_j \boldsymbol{z}_j)^{-1}]$$
$$= [(\textstyle\sum_j \boldsymbol{x}'_j \boldsymbol{z}_j)(\sum_j \boldsymbol{z}'_j \widehat{\epsilon}^2_j \boldsymbol{z}_j)^{-1}(\sum_j \boldsymbol{z}'_j \boldsymbol{x}_j)]^{-1}[(\sum_j \boldsymbol{x}'_j \boldsymbol{z}_j)(\sum_j \boldsymbol{z}'_j \widehat{\epsilon}^2_j \boldsymbol{z}_j)^{-1}(\sum_j \boldsymbol{z}'_j y_j)].$$

The latter, *second-step GMM estimator*, is in a sense an optimal GMM estimator in the presence of unspecified heteroskedasticity of ϵ_{it}.

The validity of the orthogonality condition (4.2) can be tested by the Sargan-Hansen statistic [confer Hansen (1982), Newey (1985), and Arellano and Bond (1991)], corresponding to the asymptotically efficient estimator $\widetilde{\boldsymbol{\beta}}_{GMM}$:

$$(4.7) \qquad \mathcal{J} = (\textstyle\sum_j \widehat{\epsilon}'_j \boldsymbol{z}_j)(\sum_j \boldsymbol{z}'_j \widehat{\epsilon}^2_j \boldsymbol{z}_j)^{-1}(\sum_j \boldsymbol{z}'_j \widehat{\epsilon}_j).$$

Under the null, $\mathcal{J}$ is asymptotically distributed as χ^2 with a number of degrees of freedom equal to the number of overidentifying restrictions, *i.e.*, the number of orthogonality conditions less the number of coefficients estimated under the null.

5 Orthogonality conditions, rank conditions and instruments

When delimiting the set of potential IVs for GMM estimation for our panel data model, we need to examine which variables or variable transformations satisfy jointly the rank conditions for the IVs – the requirement of being correlated with the vector of variables needing IVs – and the orthogonality between the IVs and the composite errors/disturbances of the equation. The problem is how to implement this general idea and how to figure out a principle for IV selection which is useful for computer programming.

A question of particular interest is: how will the potential IV set change when the model's memory pattern changes? The solution adopted here is to operate on *index sets* to keep track of the lag and leads of the actual variables.

Index sets for the lead/lag indicators

We first define index sets for the lag/lead indicator τ. Let in general, $\boldsymbol{Z}_{\alpha\bullet\beta}$ denotes the index set for τ which ensures the variables α and β to be *orthogonal* and $\boldsymbol{S}_{\gamma\bullet\delta}$ the index set for τ which ensures the variables γ and δ to be *correlated*. For our problem the following index sets are of particular interest [where the instruments are specified before $\bullet$, and, respectively, the composite disturbance and the instrumented variable after.]:

$$
\begin{array}{lrcl}
(a) & \boldsymbol{Z}_{\Delta q\bullet w} & \equiv & \{\tau : \text{cov}(\Delta \boldsymbol{q}_{i,t+\tau}, w_{it}) = \boldsymbol{0}\}, \\
(b) & \boldsymbol{Z}_{\Delta y\bullet w} & \equiv & \{\tau : \text{cov}(\Delta y_{i,t+\tau}, w_{it}) = 0\}, \\
(c) & \boldsymbol{Z}_{q\bullet \Delta w} & \equiv & \{\tau : \text{cov}(\boldsymbol{q}_{i,t+\tau}, \Delta w_{it}) = \boldsymbol{0}\}, \\
(d) & \boldsymbol{Z}_{y\bullet \Delta w} & \equiv & \{\tau : \text{cov}(y_{i,t+\tau}, \Delta w_{it}) = 0\},
\end{array}
$$

$$
\begin{array}{lrcl}
(e) & \boldsymbol{S}_{\Delta q\bullet q} & \equiv & \{\tau : \text{cov}(\Delta \boldsymbol{q}_{i,t+\tau}, \boldsymbol{q}_{it}) \neq \boldsymbol{0}\}, \\
(f) & \boldsymbol{S}_{\Delta y\bullet q} & \equiv & \{\tau : \text{cov}(\Delta y_{i,t+\tau}, \boldsymbol{q}_{it}) \neq \boldsymbol{0}\}, \\
(g) & \boldsymbol{S}_{\Delta q\bullet y(-1)} & \equiv & \{\tau : \text{cov}(\Delta \boldsymbol{q}_{i,t+\tau}, y_{i,t-1}) \neq \boldsymbol{0}\}, \\
(h) & \boldsymbol{S}_{\Delta y\bullet y(-1)} & \equiv & \{\tau : \text{cov}(\Delta y_{i,t+\tau}, y_{i,t-1}) \neq \boldsymbol{0}\}, \\
(i) & \boldsymbol{S}_{q\bullet \Delta q} & \equiv & \{\tau : \text{cov}(\boldsymbol{q}_{i,t+\tau}, \Delta \boldsymbol{q}_{it}) \neq \boldsymbol{0}\}, \\
(j) & \boldsymbol{S}_{y\bullet \Delta q} & \equiv & \{\tau : \text{cov}(y_{i,t+\tau}, \Delta \boldsymbol{q}_{it}) \neq \boldsymbol{0}\}, \\
(k) & \boldsymbol{S}_{q\bullet \Delta y(-1)} & \equiv & \{\tau : \text{cov}(\boldsymbol{q}_{i,t+\tau}, \Delta y_{i,t-1}) \neq \boldsymbol{0}\}, \\
(l) & \boldsymbol{S}_{y\bullet \Delta y(-1)} & \equiv & \{\tau : \text{cov}(y_{i,t+\tau}, \Delta y_{i,t-1}) \neq \boldsymbol{0}\}.
\end{array}
$$

The sets (a-b) and (e-h) relate to, respectively, the orthogonality conditions and the rank conditions for IVs valid for the RHS variables of (3.5). The sets (c-d) and (i-l) relate to, respectively, the orthogonality conditions and the rank conditions for IVs valid for the RHS variables of (3.7). These index sets can be derived from five index sets involving only (y, q, w) variables in levels:

$$
(5.1) \qquad
\begin{array}{rcl}
\boldsymbol{Z}_{q\bullet w} & \equiv & \{\tau : \text{cov}(\boldsymbol{q}_{i,t+\tau}, w_{it}) = \boldsymbol{0}\}, \\
\boldsymbol{Z}_{y\bullet w} & \equiv & \{\tau : \text{cov}(y_{i,t+\tau}, w_{it}) = 0\},
\end{array}
$$

$$
(5.2) \qquad
\begin{array}{rcl}
\boldsymbol{S}_{q\bullet q} & \equiv & \{\tau : \text{cov}(\boldsymbol{q}_{i,t+\tau}, \boldsymbol{q}_{it}) \neq \boldsymbol{0}\}, \\
\boldsymbol{S}_{y\bullet q} & \equiv & \{\tau : \text{cov}(y_{i,t+\tau}, \boldsymbol{q}_{it}) \neq \boldsymbol{0}\}, \\
\boldsymbol{S}_{q\bullet y} & \equiv & \{\tau : \text{cov}(\boldsymbol{q}_{i,t+\tau}, y_{it}) \neq \boldsymbol{0}\}.
\end{array}
$$

The relationships are as follows:

$$
(5.3) \qquad
\begin{array}{rcl}
\boldsymbol{Z}_{\Delta q\bullet w} & = & \boldsymbol{Z}_{q\bullet w} \cup \boldsymbol{Z}_{q(-1)\bullet w}, \\
\boldsymbol{Z}_{\Delta y(-1)\bullet w} & = & \boldsymbol{Z}_{y(-1)\bullet w} \cup \boldsymbol{Z}_{y(-1)\bullet w}, \\
\boldsymbol{Z}_{q\bullet \Delta w} & = & \boldsymbol{Z}_{q\bullet w} \cup \boldsymbol{Z}_{q\bullet w(-1)}, \\
\boldsymbol{Z}_{y(-1)\bullet \Delta w} & = & \boldsymbol{Z}_{y(-1)\bullet w} \cup \boldsymbol{Z}_{y\bullet w(-1)},
\end{array}
$$

$$
(5.4) \qquad
\begin{array}{rcl}
\boldsymbol{S}_{\Delta q\bullet q} & = & \boldsymbol{S}_{q\bullet q} \cup \boldsymbol{S}_{q(-1)\bullet q}, \\
\boldsymbol{S}_{\Delta y\bullet q} & = & \boldsymbol{S}_{y\bullet q} \cup \boldsymbol{S}_{y(-1)\bullet q}, \\
\boldsymbol{S}_{\Delta q\bullet y(-1)} & = & \boldsymbol{S}_{q\bullet y(-1)} \cup \boldsymbol{S}_{q(-1)\bullet y(-1)}, \\
\boldsymbol{S}_{\Delta y\bullet y(-1)} & = & \boldsymbol{S}_{y\bullet y(-1)} \cup \boldsymbol{S}_{y(-1)\bullet y(-1)}, \\
\boldsymbol{S}_{q\bullet \Delta q} & = & \boldsymbol{S}_{q\bullet q} \cup \boldsymbol{S}_{q\bullet q(-1)}, \\
\boldsymbol{S}_{y\bullet \Delta q} & = & \boldsymbol{S}_{y\bullet q} \cup \boldsymbol{S}_{y\bullet q(-1)}, \\
\boldsymbol{S}_{q\bullet \Delta y(-1)} & = & \boldsymbol{S}_{q\bullet y(-1)} \cup \boldsymbol{S}_{q\bullet y(-1)}, \\
\boldsymbol{S}_{y\bullet \Delta y(-1)} & = & \boldsymbol{S}_{y\bullet y(-1)} \cup \boldsymbol{S}_{y\bullet y(-1)}.
\end{array}
$$

The crucial point in delimiting valid IV sets for $\boldsymbol{x}_{it} = (\boldsymbol{q}_{it}, y_{i,t-1})$ and $\Delta\boldsymbol{x}_{it} = (\Delta\boldsymbol{q}_{it}, \Delta y_{i,t-1})$ will be to construct intersections of the relevant pairs of $\boldsymbol{Z}$ and $\boldsymbol{S}$ index sets from (5.3) and (5.4), which we simplify as follows:

$$\begin{array}{rcl} \boldsymbol{Z}_{\Delta x \bullet w} &=& \boldsymbol{Z}_{x\bullet w} \cup \boldsymbol{Z}_{x(-1)\bullet w}, \\ \boldsymbol{Z}_{x\bullet \Delta w} &=& \boldsymbol{Z}_{x\bullet w} \cup \boldsymbol{Z}_{x\bullet w(-1)}, \end{array} \tag{5.5}$$

$$\begin{array}{rcl} \boldsymbol{S}_{\Delta q \bullet x} &=& \boldsymbol{S}_{q\bullet x} \cup \boldsymbol{S}_{q(-1)\bullet x}, \\ \boldsymbol{S}_{\Delta y \bullet x} &=& \boldsymbol{S}_{y\bullet x} \cup \boldsymbol{S}_{y(-1)\bullet x}, \\ \boldsymbol{S}_{q\bullet \Delta x} &=& \boldsymbol{S}_{q\bullet x} \cup \boldsymbol{S}_{q\bullet x(-1)}, \\ \boldsymbol{S}_{y\bullet \Delta x} &=& \boldsymbol{S}_{y\bullet x} \cup \boldsymbol{S}_{y\bullet x(-1)}. \end{array} \tag{5.6}$$

We have:

> GENERAL PRESCRIPTION:
>
> *1. The index sets $\boldsymbol{Z}_{\Delta x\bullet w} \cap \boldsymbol{S}_{\Delta q\bullet x}$ and $\boldsymbol{Z}_{\Delta y\bullet w} \cap \boldsymbol{S}_{\Delta y\bullet x}$ delimit the τ-values for $\Delta\boldsymbol{x}_{i,t+\tau}$ which give valid IVs for the RHS variables in (3.5).*
>
> *2. The index sets $\boldsymbol{Z}_{x\bullet \Delta w} \cap \boldsymbol{S}_{q\bullet \Delta x}$ and $\boldsymbol{Z}_{y\bullet \Delta w} \cap \boldsymbol{S}_{y\bullet \Delta x}$ delimit the τ-values for $\boldsymbol{x}_{i,t+\tau}$ which give valid IVs for the RHS variables in (3.7).*

The index sets thus obtained depend on the length of memory parameters $(N_\xi, N_\eta, N_u, N_\nu)$. Below we consider first the general case and next the more transparent case without memory of errors and disturbances.

Potential IVs in the general case

It follows from the model setup that

$$\begin{array}{c} (1-\lambda \mathsf{L})y_{it} = \alpha_i + \boldsymbol{\xi}_{it}\boldsymbol{\beta} + \omega_{it}, \\ \boldsymbol{q}_{it} = \boldsymbol{\xi}_{it} + \boldsymbol{\eta}_{it}, \\ w_{it} = \omega_{it} - \boldsymbol{\eta}_{it}\boldsymbol{\beta}. \end{array}$$

where

$$\omega_{it} = u_{it} + \nu_{it} - \nu_{i,t-1}\lambda.$$

Let (N_q, N_w, N_ω) and $(N_{\Delta q}, N_{\Delta w}, N_{\Delta\omega})$ be the memory of, respectively, $(\boldsymbol{q}_{it}, w_{it}, \omega_{it})$ and $(\Delta\boldsymbol{q}_{it}, \Delta w_{it}, \Delta\omega_{it})$. They can be expressed by means of $(N_\xi, N_u, N_\nu, N_\eta)$ as follows:

$$\begin{array}{rcl} N_q &=& \max[N_\xi, N_\eta], \\ N_\omega &=& \max[N_u, N_\nu+1], \\ N_w &=& \max[N_u, N_\nu+1, N_\eta], \\ N_{\Delta q} = N_q+1 &=& \max[N_\xi+1, N_\eta+1], \\ N_{\Delta\omega} = N_\omega+1 &=& \max[N_u+1, N_\nu+2], \\ N_{\Delta w} = N_w+1 &=& \max[N_u+1, N_\nu+2,\ N_\eta+1]. \end{array} \tag{5.7}$$

From (5.1)–(5.2) we obtain

$$\begin{aligned}
\boldsymbol{Z}_{q\bullet w} &= \{\tau : |\tau| \ge N_\eta{+}1\}, \\
\boldsymbol{Z}_{y\bullet w} &= \{\tau : \tau \le -(N_\omega{+}1)\},
\end{aligned} \tag{5.8}$$

$$\begin{aligned}
\boldsymbol{S}_{q\bullet q} &= \{\tau : |\tau| \le N_q\}, \\
\boldsymbol{S}_{y\bullet q} &= \{\tau : \tau \ge -N_\xi\}, \\
\boldsymbol{S}_{q\bullet y} &= \{\tau : \tau \le N_\xi\}.
\end{aligned} \tag{5.9}$$

These equations can act as a generator of potential sets of instrumental variables for GMM estimation of (3.5) and (3.7). They can be used for constructing

$\boldsymbol{q}$-IVs for $\Delta\boldsymbol{q}$ and Δy,
y-IVs for $\Delta\boldsymbol{q}$ and Δy,
$\Delta\boldsymbol{q}$-IVs for $\boldsymbol{q}$ and y,
Δy-IVs for $\boldsymbol{q}$ and y.

We shall not discuss all cases in detail in the following, but confine attention to the following example, derived from (5.8)–(5.9), in combination with (5.3)–(5.4) lead to

$$\begin{aligned}
\boldsymbol{Z}_{\Delta q\bullet w} &= \{\tau : \tau \notin [-N_\eta, N_\eta{+}1]\}, \\
\boldsymbol{Z}_{\Delta y\bullet w} &= \{\tau : \tau \le -(N_\omega{+}1)\}, \\
\boldsymbol{Z}_{\Delta y(-1)\bullet w} &= \{\tau : \tau \le -N_\omega\}, \\
\boldsymbol{Z}_{q\bullet \Delta w} &= \{\tau : \tau \notin [-(N_\eta{+}1), N_\eta]\}, \\
\boldsymbol{Z}_{y\bullet \Delta w} &= \{\tau : \tau \le -(N_\omega{+}2)\}, \\
\boldsymbol{Z}_{y(-1)\bullet \Delta w} &= \{\tau : \tau \le -(N_\omega{+}1)\},
\end{aligned} \tag{5.10}$$

$$\begin{aligned}
\boldsymbol{S}_{\Delta q\bullet q} &= \{\tau : -N_q \le \tau \le N_q{+}1\}, \\
\boldsymbol{S}_{\Delta y\bullet q} &= \{\tau : \tau \ge -N_\xi\}, \\
\boldsymbol{S}_{\Delta y(-1)\bullet q} &= \{\tau : \tau \ge -(N_\xi{-}1)\}, \\
\boldsymbol{S}_{q\bullet \Delta q} &= \{\tau : -(N_q{+}1) \le \tau \le N_q\}, \\
\boldsymbol{S}_{y\bullet \Delta q} &= \{\tau : \tau \ge -(N_\xi{+}1)\}, \\
\boldsymbol{S}_{y(-1)\bullet \Delta q} &= \{\tau : \tau \ge -N_\xi\}.
\end{aligned} \tag{5.11}$$

We find:

1. Eq. in levels, IVs in differences:

Potential IVs for $\boldsymbol{q}_{it}$*:*

$\Delta\boldsymbol{q}_{i,t+\tau}$ for $\tau \in (\boldsymbol{Z}_{\Delta q\bullet w} \cap \boldsymbol{S}_{\Delta q\bullet q}) \Longrightarrow \tau \notin [-N_\eta, N_\eta{+}1]$ and $\tau \in [-N_q, N_q+1]$
(Index set empty if $N_\xi \le N_\eta$)

$\Delta y_{i,t+\tau}$ for $\tau \in (\boldsymbol{Z}_{\Delta y\bullet w} \cap \boldsymbol{S}_{\Delta y\bullet q}) \Longrightarrow \tau \in [-N_\xi, -(N_\omega{+}1)]$
(Index set empty if $N_\xi < N_\omega{+}1$).

2. Eq. in differences, IVs in levels:

Potential IVs for $\Delta\boldsymbol{q}_{it}$*:*

$\boldsymbol{q}_{i,t+\tau}$ for $\tau \in (\boldsymbol{Z}_{q\bullet\Delta w} \cap \boldsymbol{S}_{q\bullet\Delta q}) \Longrightarrow \tau \notin [-(N_\eta{+}1), N_\eta]$, $\tau \in [-(N_q{+}1), N_q]$
(Index set empty if $N_\xi \le N_\eta$)

$y_{i,t+\tau}$ for $\tau \in (\boldsymbol{Z}_{y\bullet\Delta w} \cap \boldsymbol{S}_{y\bullet\Delta q}) \Longrightarrow \tau \in [-(N_\xi{+}1), -(N_\omega{+}2)]$
(Index set empty if $N_\xi < N_\omega{+}1$).

We see that

(a) Increasing $N_\xi \Longrightarrow$ *Potential* IV set is extended.

(b) Increasing $N_u \Longrightarrow$ *Potential* IV set is diminished.

(c) Increasing $N_\nu \Longrightarrow$ *Potential* IV set is diminished.

(d) Increasing $N_\eta \Longrightarrow$ Effect on *potential* IV set indeterminate.

A sufficiently large set of *potential* IVs exists only if the latent regressors have a sufficiently long memory, while the disturbances and errors in the endogenous variable have sufficiently short memories. The reason why an increase in the memory of the errors in the exogenous variables has in indeterminate effect is due to the fact that an increase in this memory in general increases the memory of both $\boldsymbol{q}$ and w. Some of these potential IV sets may be empty, depending on the $(N_\xi, N_\eta, N_u, N_\nu)$ configuration.

The box above, intending to indicate *potential* IVs for the two versions of the equation, do by no means suggest that all prescribed IVs should be used in practice. Their number and the hence the number of orthogonality conditions may be unduly large, depending on the values of N and T. 'Weak instruments' – in particular when level values are taken to serve as IVs for differenced ones – is a problem often addressed. Recently, Roodman (2009) has characterized 'instrument proliferation' as 'an underappreciated problem .... in the application of difference and system GMM'. Aspects of the problem, inter alia related to small-sample bias and to estimation efficiency, are also discussed by Altonji and Segal (1996) and Ziliak (1997).

The case with no memory of errors and disturbances

Let us examine the simpler case where the latent regressor $\boldsymbol{\xi}_{it}$ has a finite memory equal to N_ξ, while the disturbances and errors have no memory. This implies $N_\eta = N_u = N_\nu = 0$ and hence $N_q = N_\xi, N_\omega = N_w = 1$. This is the case denoted as case (v) in Section 3.

Using (5.1)–(5.2), it follows that

$$\begin{aligned} \boldsymbol{Z}_{q\bullet w} &= \{\tau : \tau \neq 0\}, \\ \boldsymbol{Z}_{y\bullet w} &= \{\tau : \tau \leq -2\}, \\ \boldsymbol{S}_{q\bullet q} &= \{\tau : |\tau| \leq N_\xi\}, \\ \boldsymbol{S}_{y\bullet q} &= \{\tau : \tau \geq -N_\xi\}, \end{aligned}$$

so that (5.10)–(5.11) are simplified to

$$\begin{aligned} \boldsymbol{Z}_{\Delta q\bullet w} &= \{\tau : \tau \neq 0, 1\}, \\ \boldsymbol{Z}_{\Delta y\bullet w} &= \{\tau : \tau \leq -2\}, \\ \boldsymbol{Z}_{q\bullet \Delta w} &= \{\tau : \tau \neq -1, 0\}, \\ \boldsymbol{Z}_{y\bullet \Delta w} &= \{\tau : \tau \leq -3\}, \end{aligned} \tag{5.12}$$

$$\begin{aligned} \boldsymbol{S}_{\Delta q\bullet q} &= \{\tau : -N_\xi \leq \tau \leq N_\xi + 1\}, \\ \boldsymbol{S}_{\Delta y\bullet q} &= \{\tau : \tau \geq -N_\xi\}, \\ \boldsymbol{S}_{q\bullet \Delta q} &= \{\tau : -(N_\xi + 1) \leq \tau \leq N_\xi\}, \\ \boldsymbol{S}_{y\bullet \Delta q} &= \{\tau : \tau \geq -(N_\xi + 1)\}. \end{aligned} \tag{5.13}$$

and hence

1. Eq. in levels, IVs in differences:

Potential IVs for $\boldsymbol{q}_{it}$*:*

$\Delta \boldsymbol{q}_{i,t+\tau}$ for $\tau \in (\boldsymbol{Z}_{\Delta q \bullet w} \cap \boldsymbol{S}_{\Delta q \bullet q}) \Longrightarrow \tau = -N_\xi, \dots, -2, -1$ & $2, 3, \dots, N_\xi + 1$,
$\Delta y_{i,t+\tau}$ for $\tau \in (\boldsymbol{Z}_{\Delta y \bullet w} \cap \boldsymbol{S}_{\Delta y \bullet q}) \Longrightarrow -N_\xi \leq \tau \leq -2$.

2. Eq. in differences, IVs in levels:

Potential IVs for $\Delta \boldsymbol{q}_{it}$*:*

$\boldsymbol{q}_{i,t+\tau}$ for $\tau \in (\boldsymbol{Z}_{q \bullet \Delta w} \cap \boldsymbol{S}_{q \bullet \Delta q}) \Longrightarrow \tau = -(N_\xi + 1), \dots, -3, -2$ & $1, 2, \dots, N_\xi$,
$y_{i,t+\tau}$ for $\tau \in (\boldsymbol{Z}_{y \bullet \Delta w} \cap \boldsymbol{S}_{y \bullet \Delta q}) \Longrightarrow -(N_\xi + 1) \leq \tau \leq -3$.

6 GMM estimation of equations in levels and in differences

In this section, relying on the exposition in Section 4, we elaborate on the GMM estimation of equations in levels and in differences. Let $\boldsymbol{x}_{it} = (\boldsymbol{q}_{it}, y_{i,t-1})$ and $\boldsymbol{\gamma} = (\boldsymbol{\beta}', \lambda)'$ and write Equations (3.5) and (3.7) as, respectively,

$$y_{it} = \alpha_i + \boldsymbol{x}_{it}\boldsymbol{\gamma} + w_{it}, \tag{6.1}$$

$$\Delta y_{it} = \Delta \boldsymbol{x}_{it}\boldsymbol{\gamma} + \Delta w_{it}. \tag{6.2}$$

After stacking by periods,

$$\begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{bmatrix} = \begin{bmatrix} \alpha_i \\ \alpha_i \\ \vdots \\ \alpha_i \end{bmatrix} + \begin{bmatrix} \boldsymbol{x}_{i1} \\ \boldsymbol{x}_{i2} \\ \vdots \\ \boldsymbol{x}_{iT} \end{bmatrix} \boldsymbol{\gamma} + \begin{bmatrix} w_{i1} \\ w_{i2} \\ \vdots \\ w_{iT} \end{bmatrix},$$

$$\begin{bmatrix} \Delta y_{i2} \\ \Delta y_{i3} \\ \vdots \\ \Delta y_{iT} \end{bmatrix} = \begin{bmatrix} \Delta \boldsymbol{x}_{i2} \\ \Delta \boldsymbol{x}_{i3} \\ \vdots \\ \Delta \boldsymbol{x}_{iT} \end{bmatrix} \boldsymbol{\gamma} + \begin{bmatrix} \Delta w_{i2} \\ \Delta w_{i3} \\ \vdots \\ \Delta w_{iT} \end{bmatrix},$$

the equation systems read, in compact notation,

$$\boldsymbol{y}_{Li} = \boldsymbol{\alpha}_i + \boldsymbol{X}_{Li}\boldsymbol{\gamma} + \boldsymbol{w}_{Li}, \tag{6.3}$$

$$\boldsymbol{y}_{Di} = \boldsymbol{X}_{Di}\boldsymbol{\gamma} + \boldsymbol{w}_{Di}. \tag{6.4}$$

We use as IV for $\boldsymbol{x}_{it}$ in the *equation in levels*, (6.1), in general, a vector *in differences* written as

$$\boldsymbol{z}_{Di(t)} = \boldsymbol{Q}_t \Delta \boldsymbol{x}_{it}, \tag{6.5}$$

where $\boldsymbol{Q}_t$ is a suitable selection matrix and use as IV for $\Delta\boldsymbol{x}_{it}$ in the *equation in differences*, (6.2), in general, a vector *in levels* written as

$$\boldsymbol{z}_{Li(t,t-1)} = \boldsymbol{P}_{t,t-1}\boldsymbol{x}_{it}, \tag{6.6}$$

where $\boldsymbol{P}_{t,t-1}$ is a suitable selection matrix. The IV matrix for the matrix $\boldsymbol{X}_{Li}$ in (6.3) therefore becomes

$$\boldsymbol{Z}_{Di} = \begin{bmatrix} \boldsymbol{x}'_{Di(1)} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{x}'_{D(2)} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{x}'_{D(T)} \end{bmatrix}. \tag{6.7}$$

Likewise, the IV matrix for the matrix $\boldsymbol{X}_{Di}$ in (6.4) becomes

$$\boldsymbol{Z}_{Li} = \begin{bmatrix} \boldsymbol{x}'_{Li(2,1)} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{x}'_{L(3,2)} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{x}'_{L(T,T-1)} \end{bmatrix}. \tag{6.8}$$

The GMM-estimator corresponding to the first-step GMM estimator (4.5), for the equation in levels and for the equation in differences then become, respectively

$$\begin{aligned}\widehat{\gamma}_L = \{&[\textstyle\sum_i \boldsymbol{X}'_{Li}\boldsymbol{Z}_{Di}][\sum_i \boldsymbol{Z}'_{Di}\boldsymbol{Z}_{Di}]^{-1}[\sum_i \boldsymbol{Z}'_{Di}\boldsymbol{X}_{Li}]\}^{-1} \\ &\times \{[\textstyle\sum_i \boldsymbol{X}'_{Li}\boldsymbol{Z}_{Di}][\sum_i \boldsymbol{Z}'_{Di}\boldsymbol{Z}_{Di}]^{-1}[\sum_i \boldsymbol{Z}'_{Di}\boldsymbol{y}_{Li}]\},\end{aligned} \tag{6.9}$$

$$\begin{aligned}\widehat{\gamma}_D = \{&[\textstyle\sum_i \boldsymbol{X}'_{Di}\boldsymbol{Z}_{Li}][\sum_i \boldsymbol{Z}'_{Li}\boldsymbol{Z}_{Li}]^{-1}[\sum_i \boldsymbol{Z}'_{Li}\boldsymbol{X}_{Di}]\}^{-1} \\ &\times \{[\textstyle\sum_i \boldsymbol{X}'_{Di}\boldsymbol{Z}_{Di}][\sum_i \boldsymbol{Z}'_{Li}\boldsymbol{Z}_{Li}]^{-1}[\sum_i \boldsymbol{Z}'_{Li}\boldsymbol{y}_{Di}]\},\end{aligned} \tag{6.10}$$

They are obtained by minimizing, respectively, the quadratic forms

$$[\tfrac{1}{N}\textstyle\sum_i \boldsymbol{w}'_{Li}\boldsymbol{Z}_{Di}][\frac{1}{N^2}\sum_i \boldsymbol{Z}'_{Di}\boldsymbol{Z}_{Di}]^{-1}[\frac{1}{N}\boldsymbol{Z}'_{Di}\Delta\boldsymbol{w}_{Li}], \tag{6.11}$$

$$[\tfrac{1}{N}\textstyle\sum_i \boldsymbol{w}'_{Di}\boldsymbol{Z}_{Li}][\frac{1}{N^2}\sum_i \boldsymbol{Z}'_{Li}\boldsymbol{Z}_{Li}]^{-1}[\frac{1}{N}\boldsymbol{Z}'_{Li}\Delta\boldsymbol{w}_{Di}]. \tag{6.12}$$

If w_{it} has an unspecified heteroskedasticity, which can be due to heteroskedasticity of at least one of $(u_{it}, \nu_{it}, \boldsymbol{\eta}_{it})$, we obtain (asymptotically) more efficient GMM-estimators by the following procedure: First, form residuals vectors from (6.3) and (6.4), respectively, denoted as $\widehat{\boldsymbol{w}}_{Li}$ and $\widehat{\boldsymbol{w}}_{Di}$, respectively. Next, replace in (6.9) $\sum_i \boldsymbol{Z}'_{Di}\boldsymbol{Z}_{Di}$ with $\sum_i \boldsymbol{Z}'_{Di}\widehat{\boldsymbol{w}}_{Li}\widehat{\boldsymbol{w}}'_{Li}\boldsymbol{Z}_{Di}$, and replace in (6.10) $\sum_i \boldsymbol{Z}'_{Li}\boldsymbol{Z}_{Li}$ with $\sum_i \boldsymbol{Z}'_{Li}\widehat{\boldsymbol{w}}_{Di}\widehat{\boldsymbol{w}}'_{Di}\boldsymbol{Z}_{Li}$, The resulting estimators, which exemplify the second-step estimator (4.6), are, respectively

$$\begin{aligned}\widetilde{\gamma}_L = \{&[\textstyle\sum_i \boldsymbol{X}'_{Li}\boldsymbol{Z}_{Di}][\sum_i \boldsymbol{Z}'_{Di}\widehat{\boldsymbol{w}}_{Li}\widehat{\boldsymbol{w}}'_{Li}\boldsymbol{Z}_{Di}]^{-1}[\sum_i \boldsymbol{Z}'_{Di}\boldsymbol{X}_{Li}]\}^{-1} \\ &\times \{[\textstyle\sum_i \boldsymbol{X}'_{Li}\boldsymbol{Z}_{Di}][\sum_i \boldsymbol{Z}'_{Di}\widehat{\boldsymbol{w}}_{Li}\widehat{\boldsymbol{w}}'_{Li}\boldsymbol{Z}_{Di}]^{-1}[\sum_i \boldsymbol{Z}'_{Di}\boldsymbol{y}_{Li}]\},\end{aligned} \tag{6.13}$$

$$\begin{aligned}\widetilde{\gamma}_D = \{&[\textstyle\sum_i \boldsymbol{X}'_{Di}\boldsymbol{Z}_{Li}][\sum_i \boldsymbol{Z}'_{Li}\widehat{\boldsymbol{w}}_{Di}\widehat{\boldsymbol{w}}'_{Di}\boldsymbol{Z}_{Li}]^{-1}[\sum_i \boldsymbol{Z}'_{Li}\boldsymbol{X}_{Di}]\}^{-1} \\ &\times \{[\textstyle\sum_i \boldsymbol{X}'_{Di}\boldsymbol{Z}_{Di}][\sum_i \boldsymbol{Z}'_{Li}\widehat{\boldsymbol{w}}_{Di}\widehat{\boldsymbol{w}}'_{Di}\boldsymbol{Z}_{Li}]^{-1}[\sum_i \boldsymbol{Z}'_{Li}\boldsymbol{y}_{Di}]\},\end{aligned} \tag{6.14}$$

Testing of orthogonality conditions

The testing of the orthogonality conditions can, by using (4.7), be carried out for, respectively, the level equation and the equation in differences, in the following way:

$$\mathcal{J}_L = [\textstyle\sum_i \boldsymbol{w}'_{Li}\boldsymbol{Z}_{Di}][\sum_i \boldsymbol{Z}'_{Di}\widehat{\boldsymbol{w}}_{Li}\widehat{\boldsymbol{w}}'_{Li}\boldsymbol{Z}_{Di}]^{-1}[\sum_i \boldsymbol{Z}'_{Di}\boldsymbol{w}_{Li}], \tag{6.15}$$

$$\mathcal{J}_D = [\textstyle\sum_i \boldsymbol{w}'_{Di}\boldsymbol{Z}_{Li}][\sum_i \boldsymbol{Z}'_{Li}\widehat{\boldsymbol{w}}_{Di}\widehat{\boldsymbol{w}}'_{Di}\boldsymbol{Z}_{Li}]^{-1}[\sum_i \boldsymbol{Z}'_{Li}\boldsymbol{w}_{Di}]. \tag{6.16}$$

Under the respective null hypotheses, $\mathcal{J}_L$ and $\mathcal{J}_D$ are asymptotically distributed as χ^2 with a number of degrees of freedom equal to the number of overidentifying restrictions, *i.e.*, the number of orthogonality conditions less the number of coefficients estimated under the null.

Non-stationarity and the intercept

The treatment of the *level equation's intercept* when using the orthogonality conditions of the form (4.2) with $\boldsymbol{z}$ constructed from differences as in (6.5) and (6.7), needs a comment. When $\boldsymbol{q}_{it}$ and y_{it} are *stationary in the mean*, then $\mathsf{E}(\Delta\boldsymbol{x}_{it}) = \mathbf{0}$, so that the equation's intercept will be annihilated in the moment equations used in the GMM. If, however, mean stationarity is relaxed so that $\mathsf{E}(\Delta\boldsymbol{x}_{it}) \neq \mathbf{0}$, which may be reasonable in may practical situations, we get

$$\mathsf{E}[(\Delta\boldsymbol{x}_{it})'w_{it}] = \mathsf{E}[(\Delta\boldsymbol{x}_{it})'y_{it}] - \mathsf{E}[(\Delta\boldsymbol{x}_{it})']c - \mathsf{E}[(\Delta\boldsymbol{x}_{it})'\boldsymbol{x}_{it}]\boldsymbol{\gamma} = \mathbf{0},$$

denoting the intercept in the level equation as c. Utilizing $\mathsf{E}(w_{it}) = \mathsf{E}(y_{it}) - c - \mathsf{E}(\boldsymbol{x}_{it})\boldsymbol{\gamma} = 0$ to eliminate c, we obtain:

$$\mathsf{E}\left[(\Delta\boldsymbol{x}_{it})'[y_{it} - \mathsf{E}(y_{it})]\right] - \mathsf{E}\left[(\Delta\boldsymbol{x}_{it})'[\boldsymbol{x}_{it} - \mathsf{E}(\boldsymbol{x}_{it})]\right]\boldsymbol{\gamma} = \mathbf{0}.$$

To implement the latter orthogonality conditions in the GMM procedure for the equation in levels, IVs in differences, we could therefore replace $\mathsf{E}(y_{it})$ and $\mathsf{E}(\boldsymbol{x}_{it})$ by zeros if mean stationarity can be assumed and by corresponding global or period specific sample means otherwise. Illustrations will be given in Section 8.

7 GMM in AR models versus in EIV models: A synthesis

We can now survey **[A]** the common properties and **[B]** the discrepancies in the way the GMM is applied for pure autoregressive panel data models and for static panel data models with regressors measured with errors:

[A] The *common features* are basically six:

First, GMM exploits *orthogonality conditions* and *rank conditions*, which jointly delimit of potential IVs.

Second, the equation when expressed in observed variables can be transformed to differences in order to eliminate *heterogeneity*, either fixed or random assumed to be correlated with regressors.

Third, GMM estimation can be performed on the equation transformed to *differences* with IVs in *levels*, or on the equation in *levels*, using IVs in *differences*. It is important that at least one of the two are in differences to get clear of the nuisance created by the individual effects. In the equation is kept in levels and there is reason to assume that the latent regressor is non-stationary in the mean, the level variables should be measured from their (global or period specific) means to compensate for this.

Fourth, in choosing level variables to serve as IVs for an equation transformed to differences, we select values for periods *other than those occurring in defining the differences*. In choosing variables in differences for inclusion in an IV set for an equation in levels, the periods represented by the levels are excluded. GMM estimation can proceed in *one step or in two steps*, the second step serving to improve the estimator's efficiency by adjusting for disturbance or error heteroskedasticity of unspecified form.

Fifth, when memory of the errors or disturbances are allowed for, the valid IV set should be reduced. Its essence is to *reduce the IV set* such that all IVs 'get clear of' the memory of the error process, in order to ensure that the IVs and the errors/disturbances are still uncorrelated (*the orthogonality condition*), while ensuring fulfilment of the claim that the IVs be correlated with the variables for which they serve as IVs (*the rank condition*).

Sixth, the validity of the orthogonality conditions can be tested by the *Sargan-Hansen* test. The quality of the instruments entering the GMM can be assessed by R^2 criteria.

[B] The *discrepancies* are basically three:

First, for the pure *autoregressive* panel data model, *only lagged* endogenous and exogenous variables can serve as IVs. Their validity can be inferred from the model's memory pattern. For the pure static *measurement error* panel data model, *lagged and leaded* values of endogenous and exogenous variables can serve jointly as IVs, provided that certain additional conditions are satisfied.

Second, the pure *autoregressive* model has a memory pattern which rationalizes both the orthogonality condition and the the rank condition for the IV set. The static *measurement error* panel data model has no such memory pattern for its structural variables. Therefore, the validity of the rank condition must be postulated as a supplement, imposing mild restrictions on the temporal properties of the latent exogenous variables. The *mixed model* considered in the present paper can be view as taking an intermediate position.

Three, for the pure static *measurement error* panel data model, (leaded or lagged) y-values can serve as IVs for other ys. For the pure *autoregressive* panel data model, no y-values can serve as IVs for other ys.

8 An application

In this section, we present preliminary results from a simple application of the approach described in Sections 5 and 6 to data from the Pulp and Paper industry in Norway: $N = 61$ firms are observed in $T = 10$ consecutive years (1984–1993). We consider a relationship between an assumed endogenous measure of capital stock in machinery (k) and an assumed exogenous measure of output (x) – both log-transformed. The *capital-output relationship* may be considered a *prima facie* example in the present context, for two reasons: First, it exemplifies very well dynamic modelling with panel data, in a case where it may be disputed whether to express the equation in levels or to transform it to changes – confer the acceleration principle in business-cycle theory versus the constant capital coefficient (elasticity) assumption which occurs in several theories of economic growth. Second, both capital stocks and output flows are notoriously difficult variables to measure properly.

Results for a *static logged capital-logged output equation in levels* with IVs in differences are given in Table 2. In panel A only the latent regressor ξ is assumed to have a memory, in panel B both ξ and the error elements have a memory. Four versions of the results are reported, arranged as a 2×2 classification: (a) based on the untransformed level data (columns 1, 2, 5, 6) versus based on level data measured from their annual means to compensate for possible non-stationarity in the expectations (columns 3, 4, 7, 8): and (b) using only xs as instruments (columns 1, 3, 5, 7) versus using both xs and ys as instruments (columns 2, 4, 6, 8). Four memory configurations are considered: $(N_\nu, N_u, N_\eta, N_\xi) = (0,0,0,2), (0,0,0,4), (1,1,1,4)$ and $(2,2,2,4)$, respectively. Results for a corresponding *AR(1) equation in levels* with IVs in differences for a similar 2×2 classification are given in Table 3.

Sargan-Hansen $\mathcal{J}$-test-statistics for IV-error orthogonality, with corresponding p-values, are reported for each set of results. Finally, 'quality indexes' based on R^2s for the IVs are given: Traditional R^2s for the IVs in differences. In addition are given modified R^2s obtained from the Pesaran-Smith (1994) criterion, denoted as $\texttt{PS}\,R^2$.

From the results in Table 2 we note that for the static equation, using level observations measured from year means, and hence relaxing mean stationarity of the regressor, we get estimates of the capital elasticity β exceeding one, while using non-demeaned observations, and hence imposing mean stationarity, the estimates become smaller than one. Allowing for increased memory of the latent regressor, but no memory of the errors and disturbances reduces the standard errors and improves overall IV quality, as indicated by the R^2 measures. Introducing a one period memory in the errors-disturbances, retaining a four-period memory of the latent regressor, has a small impact on the coefficient estimates, but reduces the IV quality somewhat.

The results for the AR(1) equation in Table 3 depart substantially from those for the corresponding static equations. The λ estimates are below, but close to, one, and since the estimated $\beta + \lambda$ are close to one in all alternatives, the results suggest substantial sluggishness in capital adjustment to output, cumulating to a long-run elasticity approximately equal to one. Relaxing stationarity of the latent regressor, while including both x-differences and y-differences among the IVs tends to increase the estimated short-run elasticity β to significantly positive values.

The results reported in Tables 4 and 5 for similar capital-output equations written in logged first-differences are somewhat mixed. Again, in panel A only the latent regressor ξ (in level form) is assumed to have a memory, in panel B both ξ and the error elements have a memory, and again, four memory configurations are considered: $(N_\nu, N_u, N_\eta, N_\xi) = (0,0,0,3), (0,0,0,4), (1,1,1,4)$ and $(2,2,2,5)$, respectively. Since, in view of (5.7), we must have $N_\xi – N_\nu \geq 3$ to ensure the IV set being sufficiently large for making GMM estimation of the difference version of the equation feasible, N_ξ is increased by one in the first and third case relative to the values assumed in Tables 2 and 3. For the static equation in differences, confer Table 4, no estimate of β exceed 0.02 and none turn out significantly positive. The results for the differenced AR(1) equation in Table 5 depart substantially from those for the level equation in Table 4, with λ estimates in the range 0.3–0.5, departing significantly from both 0 and 1.

9 Concluding remarks

In this paper, we have considered several GMM estimators that can handle jointly the heterogeneity problem and the measurement error problem in static and dynamic panel data models. These problems may be intractable when only pure (single or repeated) cross section data or pure time series data are available. Estimators using either equations in differences with level values as IVs, or equations in levels with differenced values as IVs are useful. In both cases, the differences may be taken over one period or more.

GMM exploits certain orthogonality conditions and certain rank conditions, which jointly delimit a class of potential IVs. We transform to differences in order to eliminate fixed heterogeneity (or random heterogeneity which is correlated with regressors). GMM estimation can be implemented in two ways: (a) on the equation transformed to differences with IVs in levels, (b) on the equation in levels using IVs in differences. It is important that at least one of the two are in differences to get clear of the nuisance created by the individual effects. We can perform GMM estimation in one step or in two steps, the second step serving to improve the estimator's efficiency by adjusting for disturbance or error heteroskedasticity of unspecified form.

For the pure ARMA model, only lagged endogenous and exogenous variables can be IVs. For the static EIV model, lagged and leaded endogenous and exogenous variables can both be valid IVs. For the mixed model considered here it is convenient to formalize by letting index sets represent the leaded and lagged values of the observed endogenous and exogenous variables which satisfy the orthogonality and rank conditions. The AR, ARX, ARMA, and ARMAX models have a memory pattern which rationalizes both the orthogonality condition and the rank condition for the IVs. For the static EIV model, on the other hand, a memory pattern which satisfies the rank condition must be postulated as a supplement. The mixed model takes an intermediate position, showing that there is a trade-off between the moving average orders of the latent regressors, the errors and the disturbances. Long memory of errors or disturbances may make the number of valid IVs insufficient. There is, however, an asymmetry between the memory of the errors of the endogenous and of the exogenous variables in this respect.

Using levels as IVs for differences or *vice versa* as a general estimation strategy within a GMM framework may raise problems related to 'weak instruments'. Finding operational ways of identifying such IVs among those utilizing essential orthogonality conditions to reduce their potential damage in terms of inefficiency, are challenges for future research. For the capital-output relationship example considered, the level version of the equation with instruments in differences seems preferable to choosing the opposite strategy by a wide margin.

Table 2: Static equation in levels. GMM with IVs in differences

$N = 61, T = 10$. Standard errors in parenthesis.

R^2 = Standard R^2. PS R^2= Pesaran-Smith-corrected R^2. J-test=Hansen-Sargan orthogonality test

A. Memory in latent regressor only

	Memory: $(N_\nu, N_u, N_\eta, N_\xi) = (0,0,0,2)$				Memory: $(N_\nu, N_u, N_\eta, N_\xi) = (0,0,0,4)$			
	ξ stationary		ξ non-stationary		ξ stationary		ξ non-stationary	
IVs in diff.	xs	xs & ys	xs	xs & ys	xs	xs & ys	xs	xs & ys
β= coef. of x	1.1750	1.1535	0.8700	0.7634	1.1627	1.1556	0.7815	0.7464
	(0.0336)	(0.0215)	(0.0983)	(0.0641)	(0.0311)	(0.0206)	(0.0770)	(0.0559)
IV quality statistics:								
R^2	0.0427	0.1539	0.0703	0.1252	0.1080	0.2445	0.1369	0.2165
PS R^2	0.0423	0.1532	0.0573	0.1059	0.1069	0.2430	0.1119	0.1854
J-test,p=	0.0919	0.2433	0.2532	0.3415	0.0839	0.9601	0.0946	0.9677

B. Memory in both latent regressor and errors

	Memory: $(N_\nu, N_u, N_\eta, N_\xi) = (1,1,1,4)$				Memory: $(N_\nu, N_u, N_\eta, N_\xi) = (2,2,2,4)$			
	ξ stationary		ξ non-stationary		ξ stationary		ξ non-stationary	
IVs in differences	xs	xs & ys	xs	xs & ys	xs	xs & ys	xs	xs & ys
β= coef. of x	1.1670	1.1538	0.7628	0.7222	1.1788	1.1547	0.7273	0.6988
	(0.0301)	(0.0207)	(0.0657)	(0.0501)	(0.0292)	(0.0212)	(0.0610)	(0.0462)
IV quality statistics:								
R^2	0.1075	0.2327	0.1293	0.2031	0.0700	0.1971	0.0809	0.1543
PS R^2	0.1065	0.2311	0.1088	0.1752	0.0692	0.1958	0.0695	0.1357
J-test,p=	0.1222	0.9437	0.1933	0.9537	0.4557	0.7138	0.5823	0.7444

Table 3: AR(1) equation in levels. GMM with IVs in differences

$N = 61, T = 10$. Standard errors in parenthesis.

R^2 = Standard R^2. PS R^2= Pesaran-Smith-corrected R^2. J-test=Hansen-Sargan orthogonality test

A. Memory in latent regressor only

	Memory: $(N_\nu, N_u, N_\eta, N_\xi) = (0,0,0,2)$				Memory: $(N_\nu, N_u, N_\eta, N_\xi) = (0,0,0,4)$			
	ξ stationary		ξ non-stationary		ξ stationary		ξ non-stationary	
IVs in diff.	xs	xs & ys	xs	xs & ys	xs	xs & ys	xs	xs & ys
β= coef. of x	0.0216	0.0275	0.0317	0.0589	0.0074	0.0134	0.0219	0.0383
	(0.0263)	(0.0164)	(0.0289)	(0.0166)	(0.0101)	(0.0090)	(0.0176)	(0.0145)
λ= coef. of k_{-1}	0.9857	0.9824	0.9668	0.9495	0.9952	0.9925	0.9711	0.9610
	(0.0200)	(0.0140)	(0.0277)	(0.0219)	(0.0081)	(0.0077)	(0.0225)	(0.0197)
IV quality statistics:								
R^2	0.0317	0.1472	0.0769	0.1256	0.0757	0.2321	0.1507	0.2155
PS R^2	0.0317	0.1472	0.0768	0.1251	0.0757	0.2321	0.1505	0.2149
J-test,p=	0.2352	0.4218	0.2249	0.3662	0.3386	0.7428	0.3281	0.7718

B. Memory in both latent regressor and errors

	Memory: $(N_\nu, N_u, N_\eta, N_\xi) = (1,1,1,4)$				Memory: $(N_\nu, N_u, N_\eta, N_\xi) = (2,2,2,4)$			
	ξ stationary		ξ non-stationary		ξ stationary		ξ non-stationary	
IVs in diff.	xs	xs & ys	xs	xs & ys	xs	xs & ys	xs	xs & ys
β= coef. of x	0.0024	0.0200	0.0138	0.0470	0.0066	0.0315	0.0024	0.0521
	(0.0073)	(0.0075)	(0.0143)	(0.0165)	(0.0072)	(0.0077)	(0.0140)	(0.0175)
λ= coef. of k_{-1}	0.9981	0.9867	0.9808	0.9520	0.9946	0.9772	1.0061	0.9528
	(0.0068)	(0.0065)	(0.0197)	(0.0231)	(0.0063)	(0.0067)	(0.0170)	(0.0243)
IV quality statistics:								
R^2	0.0719	0.2186	0.1400	0.2023	0.0468	0.1889	0.0863	0.1482
PS R^2	0.0719	0.2186	0.1398	0.2016	0.0468	0.1889	0.0862	0.1476
J-test,p=	0.4361	0.6832	0.5871	0.7153	0.0861	0.3815	0.1939	0.3676

Table 4: Static equation in differences. GMM with IVs in levels
$N = 61, T = 10$. Standard errors in parenthesis.

R^2 = Standard R^2. PS R^2= Pesaran-Smith-corrected R^2. J-test=Hansen-Sargan orthogonality test

A. Memory in latent regressor (level) only

	Memory: $(N_\nu, N_u, N_\eta, N_\xi) = (0,0,0,3)$		Memory: $(N_\nu, N_u, N_\eta, N_\xi) = (0,0,0,4)$	
IVs in levels	xs	xs & ys	xs	xs & ys
β= coef. of Δx	0.0153	0.0216	0.0095	0.0138
	(0.0148)	(0.0150)	(0.0135)	(0.0127)
IV quality statistics:				
R^2	0.2044	0.4283	0.2360	0.4704
PS R^2	0.0026	0.0060	0.0012	0.0032
J-test,p=	0.4510	0.6177	0.2297	0.9223

B. Memory in both latent regressor and errors (levels)

	Memory: $(N_\nu, N_u, N_\eta, N_\xi) = (1,1,1,4)$		Memory: $(N_\nu, N_u, N_\eta, N_\xi) = (2,2,2,5)$	
IVs in levels	xs	xs & ys	xs	xs & ys
β= coef. of Δx	0.0102	0.0128	0.0030	0.0157
	(0.0107)	(0.0098)	(0.0108)	(0.0108)
IV quality statistics:				
R^2	0.1547	0.6277	0.1173	0.6813
PS R^2	0.0014	0.0025	0.0001	0.0030
J-test,p=	0.3849	0.9337	0.1127	0.9525

Table 5: AR(1) equation in differences. GMM with IVs in levels
$N = 61, T = 10$. Standard errors in parenthesis.

R^2 = Standard R^2. PS R^2= Pesaran-Smith-corrected R^2. J-test=Hansen-Sargan orthogonality test

A. Memory in latent regressor only

	Memory: $(N_\nu, N_u, N_\eta, N_\xi) = (0,0,0,3)$		Memory: $(N_\nu, N_u, N_\eta, N_\xi) = (0,0,0,4)$	
IVs in levels	xs	xs & ys	xs	xs & ys
β= coef. of Δx	0.0078	-0.0010	0.0053	-0.0004
	(0.0143)	(0.0128)	(0.0137)	(0.0133)
λ= coef. of Δk_{-1}	0.3619	0.3061	0.3982	0.3058
	(0.1842)	(0.0747)	(0.1762)	(0.0751)
IV quality statistics:				
R^2	0.2370	0.4394	0.2709	0.4736
PS R^2	0.0225	0.1038	0.0273	0.1038
J-test,p=	0.3581	0.1965	0.3961	0.3982

B. Memory in both latent regressor and errors

	Memory: $(N_\nu, N_u, N_\eta, N_\xi) = (1,1,1,4)$		Memory: $(N_\nu, N_u, N_\eta, N_\xi) = (2,2,2,5)$	
IVs in levels	xs	xs & ys	xs	xs & ys
β= coef. of Δx	0.0010	0.0080	0.0171	0.0117
	(0.0142)	(0.0142)	(0.0246)	(0.0145)
λ= coef. of Δk_{-1}	0.4982	0.4406	0.5113	0.5095
	(0.1060)	(0.1254)	(0.1130)	(0.1325)
IV quality statistics:				
R^2	0.1692	0.3857	0.1270	0.5626
PS R^2	0.0501	0.0930	0.0404	0.1375
J-test,p=	0.3390	0.2719	0.2165	0.3088

Appendix:

Proof of Equations (2.22)–(2.24)

This appendix contains, for the time series Example 5 in Section 2, the proof of Equations (2.22)–(2.24).

First, since

$$\mathsf{E}\left[\xi_{t-s}\ \frac{\xi_t}{1-\gamma\mathsf{L}}\ \right] = \mathsf{E}[\textstyle\sum_{i=0}^{\infty}\gamma^i\xi_{t-s}\xi_{t-i}],$$

$$\mathsf{E}\left[\xi_t\frac{\xi_{t-s}}{1-\gamma\mathsf{L}}\right] = \mathsf{E}[\textstyle\sum_{i=0}^{\infty}\gamma^i\xi_t\xi_{t-i-s}], \qquad s=0,1,2\dots\ ,$$

it follows by using (2.14) that

$$\mathsf{E}\left[\xi_t\frac{\xi_{t-s}}{1-\gamma\mathsf{L}}\right] = \begin{cases} \sigma_{\xi\xi(0)}+\gamma\sigma_{\xi\xi(1)}+\gamma^2\sigma_{\xi\xi(2)}, & s=0,\\ \sigma_{\xi\xi(1)}+\gamma\sigma_{\xi\xi(2)}, & s=1,\\ \gamma\sigma_{\xi\xi(2)}, & s=2,\\ 0, & s=3,4,\dots\ , \end{cases} \tag{A.1}$$

$$\mathsf{E}\left[\xi_{t-s}\frac{\xi_t}{1-\gamma\mathsf{L}}\right] = \begin{matrix} \gamma\sigma_{\xi\xi(0)}+(1+\gamma^2)\sigma_{\xi\xi(1)}+\gamma^3\sigma_{\xi\xi(2)}, & s=1,\\ \gamma^s\sigma_{\xi\xi(0)}+\gamma^{s-1}(1+\gamma^2)\sigma_{\xi\xi(1)}+\gamma^{s-2}(1+\gamma^4)\sigma_{\xi\xi(2)}, & s=2,3,\dots\ . \end{matrix} \tag{A.2}$$

Using

$$\Lambda(\gamma) = \sigma_{\xi\xi(0)}+\gamma\sigma_{\xi\xi(1)}+\gamma^2\sigma_{\xi\xi(2)}, \tag{A.3}$$

(A.2) can be rewritten as

$$\mathsf{E}\left[\xi_{t-s}\frac{\xi_t}{1-\gamma\mathsf{L}}\right] = \begin{matrix} \gamma\Lambda(\gamma)+\sigma_{\xi\xi(1)}, & s=1,\\ \gamma^{s-2}[\Lambda(\gamma)+\gamma\sigma_{\xi\xi(1)}+\sigma_{\xi\xi(2)}], & s=2,3,\dots\ . \end{matrix} \tag{A.4}$$

Next, since

$$\mathsf{E}\left[\ \frac{\xi_t}{1-\gamma\mathsf{L}}\ \ \frac{\xi_{t-s}}{1-\gamma\mathsf{L}}\ \right] = \mathsf{E}[\textstyle\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\gamma^{i+j}\xi_{t-i}\xi_{t-j-s})], \quad s=0,1,2\dots\ ,$$

it also follows that

$$\mathsf{E}\left[\ \frac{\xi_t}{1-\gamma\mathsf{L}}\ \ \frac{\xi_{t-s}}{1-\gamma\mathsf{L}}\ \right] = \begin{cases} \frac{1}{1-\gamma^2}[\sigma_{\xi\xi(0)}+2\gamma\sigma_{\xi\xi(1)}+2\gamma^2\sigma_{\xi\xi(2)}], & s=0,\\ \frac{1}{1-\gamma^2}[\gamma\sigma_{\xi\xi(0)}+(1+\gamma^2)\sigma_{\xi\xi(1)}+\gamma(1+\gamma^2)\sigma_{\xi\xi(2)}], & s=1,\\ \frac{1}{1-\gamma^2}[\gamma^s\sigma_{\xi\xi(0)}+\gamma^{s-1}(1+\gamma^2)\sigma_{\xi\xi(1)} & \\ \qquad\qquad +\gamma^{s-2}(1+\gamma^4)\sigma_{\xi\xi(2)}], & s=2,3,\dots\ , \end{cases} \tag{A.5}$$

which, when again using (A.3), can be rewritten as

$$\mathsf{E}\left[\ \frac{\xi_t}{1-\gamma\mathsf{L}}\ \ \frac{\xi_{t-s}}{1-\gamma\mathsf{L}}\ \right] = \begin{cases} \frac{1}{1-\gamma^2}[\Lambda(\gamma)+\gamma\sigma_{\xi\xi(1)}+\gamma^2\sigma_{\xi\xi(2)}], & s=0.\\ \frac{1}{1-\gamma^2}[\gamma\Lambda(\gamma)+\sigma_{\xi\xi(1)}+\gamma\sigma_{\xi\xi(2)}], & s=1,\\ \frac{1}{1-\gamma^2}[\gamma^s\Lambda(\gamma)+\gamma^{s-1}\sigma_{\xi\xi(1)}+\gamma^{s-2}\sigma_{\xi\xi(2)}], & s=2,3,\dots . \end{cases} \tag{A.6}$$

Combining (2.18) with (A.1), (A.3), (A.4) and (A.6) we find that this implies

$$\mathsf{E}(\xi_t\mu_{t-s}) = \begin{cases} \beta\Lambda(\gamma), & s=0,\\ \beta(\sigma_{\xi\xi(1)}+\gamma\sigma_{\xi\xi(2)}), & s=1,\\ \beta\sigma_{\xi\xi(2)}, & s=2, \end{cases} \tag{A.7}$$

$$\mathsf{E}(\xi_{t-s}\mu_t) = \begin{matrix} \beta[\gamma\Lambda(\gamma)+\sigma_{\xi\xi(1)}], & s=1,\\ \beta[\gamma^2\Lambda(\gamma)+\gamma\sigma_{\xi\xi(1)}+\sigma_{\xi\xi(2)}], & s=2, \end{matrix} \tag{A.8}$$

$$\mathsf{E}(\mu_t\mu_{t-s}) = \begin{cases} \frac{\beta^2}{1-\gamma^2}[\Lambda(\gamma)+\gamma\sigma_{\xi\xi(1)}+\gamma^2\sigma_{\xi\xi(2)}]+(1+\chi^2)\sigma_v^2, & s=0,\\ \frac{\beta^2}{1-\gamma^2}[\gamma\Lambda(\gamma)+\sigma_{\xi\xi(1)}+\gamma\sigma_{\xi\xi(2)}]+[\gamma(1+\chi^2)+\lambda]\sigma_v^2, & s=1,\\ \frac{\beta^2}{1-\gamma^2}[\gamma^2\Lambda(\gamma)+\gamma\sigma_{\xi\xi(1)}+\sigma_{\xi\xi(2)}]+\gamma[\gamma(1+\chi^2)+\lambda]\sigma_v^2, & s=2. \end{cases} \tag{A.9}$$

This completes the proof.

References

Ahn, S.C., and Schmidt, P. (1995): Efficient Estimation of Models for Dynamic Panel Data. *Journal of Econometrics* **68**, 5-27.

Altonji, J.G., and Segal, L.M. (1996): Small-Sample Bias in GMM Estimation of Covariance Structures *Journal of Business and Economic Statistics* **14**, 353–366.

Anderson, T.W., and Hsiao, C. (1981): Estimation of Dynamic Models with Error Components. *Journal of the American Statistical Association* **76**, 598–606.

Anderson, T.W., and Hsiao, C. (1982): Formulation and Estimation of Dynamic Models Using Panel Data. *Journal of Econometrics* **18**, 47–82.

Arellano, M. (2003): *Panel Data Econometrics.* Oxford: Oxford University Press.

Arellano, M., and Bond, S. (1991): Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies* **58**, 277–297.

Arellano, M., and Bover, O. (1995): Another Look at the Instrumental Variable Estimation of Error-Components Models. *Journal of Econometrics* **68**, 29–51.

Arellano, M., and Honoré, B. (2001): Panel Data Models: Some Recent Developments. Chapter 53 in: Heckman, J.J. and Leamer, E. (eds.): *Handbook of Econometrics*, vol. 5. Amsterdam: North-Holland.

Balestra, P., and Nerlove, M. (1966): Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas. *Econometrica* **34**, 585–612.

Biørn, E. (1992): The Bias of Some Estimators for Panel Data Models with Measurement Errors. *Empirical Economics* **17**, 51–66.

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

Biørn, E. (2000): Panel Data with Measurement Errors. Instrumental Variables and GMM Estimators Combining Levels and Differences. *Econometric Reviews* **19**, 391–424.

Biørn, E. (2003): Handling the Measurement Error Problem by Means of Panel Data: Moment Methods Applied on Firm Data. Chapter 24 in *Econometrics and the Philosophy of Economics*, ed. by B. Stigum. Princeton: Princeton University Press, 2003.

Biørn, E., and Krishnakumar, J. (2008): Measurement Errors and Simultaneity. Chapter 10 in *The Econometrics of Panel Data. Fundamentals and Recent Developments in Theory and Practice*, ed. by L. Mátyás and P. Sevestre. Springer-Verlag.

Blundell, R., and Bond, S. (1998): Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. *Journal of Econometrics* **87**, 115–143.

Davidson, R., and MacKinnon, J.G. (1993): *Estimation and Inference in Econometrics.* Oxford: Oxford University Press.

Davidson, R., and MacKinnon, J.G. (2004): *Econometric Theory and Methods.* Oxford: Oxford University Press.

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.

Griliches, Z., and Hausman, J.A. (1986): Errors in Variables in Panel Data. *Journal of Econometrics* **31**, 93–118.

Hansen, L.P. (1982): Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica* **50**, 1029–1054.

Harris, D., and Mátyás, L. (1999): Introduction to the Generalized Method of Moments Estimation. Chapter 1 in *Generalized Method of Moments Estimation*, ed. by L. Mátyás. Cambridge: Cambridge University Press.

Holtz-Eakin, D., Newey, W. and Rosen, H.S. (1988): Estimating Vector Autoregressions with Panel Data. *Econometrica* **56**,1371–1395.

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.

Newey, W.K. (1985): Generalized Method of Moments Specification Testing. *Journal of Econometrics* **29**, 229–256.

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.

Pesaran, M.H. (2006): Estimation and Inference in Large Heterogeneous Panels With a Multifactor Structure. *Econometrica* **74**, 967–1012.

Pesaran, M.H. (2007): A Simple Panel Unit Root Test in the Presence of Cross-Section Dependence, *Journal of Applied Economics* **22**, 265–312.

Pesaran, M.H., and Smith, R.J. (1994): A Generalized R^2 Criterion for Regression Models Estimated by the Instrumental Variables Method. *Econometrica* **62**, 705–710.

Roodman, D. (2009): A Note on the Theme of Too Many Instruments. *Oxford Bulletin of Economics and Statistics* **71**, 135–158.

Sevestre, P., and Trognon, A. (1985): A Note on Autoregressive Error Components Models. *Journal of Econometrics* **28**, 231–245.

Sevestre, P., and Trognon, A. (1996): Dynamic Linear Models. Chapter 7 in: Mátyás L. and Sevestre, P. (eds.): *The Econometrics of Panel Data. A Handbook of the Theory with Applications*. Dordrecht: Kluwer.

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

Wansbeek, T.J. (2001): GMM Estimation in Panel Data Models with Measurement Error. *Journal of Econometrics* **104**, 259–268.

Wansbeek, T.J., and Koning, R.H. (1991): Measurement Error and Panel Data. *Statistica Neerlandica* **45**, 85–92.

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

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

White, H. (1986): Instrumental Variables Analogs of Generalized Least Squares Estimators. In *Advances in Statistical Analysis and Statistical Computing. Theory and Applications, vol. 1*, ed. by R.S. Mariano, JAI Press, pp. 173–227.

Ziliak, J.P. (1997): Efficient Estimation with Panel Data When Instruments Are Predetermined: An Empirical Comparison of Moment-Condition Estimators. *Journal of Business and Economic Statistics* **15**, 419–431.