MEMORANDUM

No 02/2012

The Measurement Error Problem in Dynamic Panel Data Analysis: Modeling and GMM Estimation

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 01/12	Erik Hernæs and Zhiyang Jia *Earning Distribution and Labour Supply after a Retirment Earnings Test Reform*
No 27/11	Jørgen Heibø Modalsli *Polarization, Risk and Welfare in General Equilibrium*
No 26/11	Jo Thori Lind and Dominic Rhoner *Knowledge is Power: A Theory of Information, Income, and Welfare Spending*
No 25/11	Kjell Arne Brekke and Tore Nilssen *The Market for Fake Observations*
No 24/11	Berhe Mekonnen Beyene *Determinants of Internal and International Migration in Ethiopia*
No 23/11	Ragnar Nymoen, Anders Rygh Swensen, Eivind Tveter *Interpreting the evidence for New Keynesian models of inflation dynamics*
No 22/11	Finn R. Førsund, Jon Vislie *From Macro Growth to Disaggregated Production Studies*
No 21/11	Jørgen Heibø Modalsli *Solow meets Marx: Economic growth and the emergence of social class*
No 20/11	Nils Chr. Framstad *On free lunches in random walk markets with short-sale constraints and small transaction costs, and weak convergence to Gaussian continuous-time processes*
No 19/11	Atle Seierstad *Pareto Improvements of Nash Equilibria in Differential Games*

Previous issues of the memo-series are available in a PDF® format at:
http://www.sv.uio.no/econ/forskning/publikasjoner/memorandum

THE MEASUREMENT ERROR PROBLEM
IN DYNAMIC PANEL DATA ANALYSIS:
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

Abstract: The Generalized Method of Moments (GMM) is discussed for handling the joint occurrence of fixed effects and random measurement errors in an autoregressive panel data model. Finite memory of disturbances, latent regressors and measurement errors is assumed. Two specializations of GMM are considered: (i) using instruments (IVs) in levels for a differenced version of the equation, (ii) using IVs in differences for an equation in levels. Index sets for lags and lags are convenient in examining how the potential IV set, satisfying orthogonality and rank conditions, changes when the memory pattern changes. The joint occurrence of measurement errors with long memory may sometimes give an IV-set too small to make estimation possible. On the other hand, problems of 'IV proliferation' and 'weak IVs' may arise unless the time-series length is small. An application based on data for (log-transformed) capital stock and output from Norwegian manufacturing firms is discussed. Finite sample biases and IV quality are illustrated by Monte Carlo simulations. Overall, with respect to bias and IV strength, GMM inference using the level version of the equation seems superior to inference based on the equation in differences.

Keywords: Panel data, Measurement error, Dynamic modeling, ARMA model, GMM, Monte Carlo simulation

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

Acknowledgements: Substantially revised version of paper presented at: Conference on Factor Structures for Panel and Multivariate Time Series Data, Maastricht, September 2008; the North American Summer Meeting of the Econometric Society, Boston, June 2009; the 15th International Conference on Panel Data, Bonn, July 2009; the 64th Econometric Society European Meeting, Barcelona, August 2009, and seminars at the University of Oslo and Statistics Norway. I am grateful to Xuehui Han for excellent assistance in the programming and testing of the numerical procedures and to Terje Skjerpen and conference and seminar participants for comments and suggestions.

1 Introduction

It is well known that endogenous right-hand side variables correlated with the disturbance in a static equation biases Ordinary Least Squares (OLS) – the 'simultaneity problem' – and that similar problems arise (A) in a static equation where random measurement errors affect the regressors and (B) in a dynamic equation where *lagged endogenous variables* and *autocorrelated disturbances* jointly occur. In the (A) case – unless extraneous information, say parameter restrictions or valid instrument variables (IVs), exist – coefficients cannot be identified from uni-dimensional data. In the (B) case, consistent estimation is possible by using lagged values of the exogenous, and the endogenous variables, as IVs. Panel data, exhibiting two-dimensional variation, may improve the situation, as data transformations can be performed along one dimension to eliminate heterogeneity, leaving one dimension available for constructing IVs. Such ideas have been explored for 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), Biørn and Krishnakumar (2008, Section 10.2), and Xiao *et al.* (2007, 2010). For the (B) case consistent estimation with finite time series length is discussed in Balestra and Nerlove (1966), Anderson and Hsiao (1981, 1982), Harris, Mátyás, and Sevestre (2008), Holtz-Eakin *et al.* (1988), Arellano and Bond (1991), Arellano and Bover (1995), Ahn and Schmidt (1995), Kiviet (1995), and Blundell and Bond (1998).[1]

In this paper we discuss the Generalized Method of Moments (GMM) to handle the joint occurrence of random measurement errors and autoregressive mechanisms in panel data. Motivating examples are: (1) an equilibrium-correcting mechanism for a firm's capital-labour ratio motivated by an error-inflicted labour/capital service price ratio, (2) an autoregressive household consumption function with inadequately measured income and net wealth, (3) an autoregressive equation explaining individual wage rates by noisy measures of ability, education achievement or work experience. The errors may well have memory. We will consider autoregressive equations with additive, random measurement errors (errors in variables, EIV) for confrontation with balanced panel data. It is possible to estimate such equations consistently while eliminating fixed and unstructured individual heterogeneity in two different ways: letting y and q denote the error-infected endogenous and exogenous variables, respectively, we can: (a) Keep the equation in level form and use *q- or y-values in differences as IVs* for the level q-values. (b) Transform the equation to differences and use *q-or y-values in levels as IVs* for the differenced q-values.

GMM estimators which are valid for situations with white noise errors can be modified to account for *finite memory of errors or disturbances* by reducing the IV set. The essence is, loosely speaking, to remove IVs valid for the zero memory case such that all remaining IVs 'get clear of' the memory of the error process, in order to ensure that the IVs and the error/disturbances are uncorrelated (*the orthogonality condition*), while ensuring that

[1]For *time-series 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.

they are also correlated with the variables for which they serve as IVs (*the rank condition*). Certain extensions of *static* EIV models to cases with finite error memory 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 we present the basic model and describe its relation to the literature. The orthogonality and rank conditions and their implied IV sets are considered and GMM estimation equations in levels and in differences is discussed in Section 3. An application based on panel data for measures of capital stock and output volume from manufacturing firms with 10 observations of each unit, supplemented with Monte Carlo (MC) simulations to illustrate finite sample biases, are presented in Section 4. We show on the one hand how the transformation from levels to differences and the following changes in the IV set – *inter alia* inclusion of IVs based on lagged endogenous variables to supplement IVs based on exogenous variables – affect the estimation results. On the other hand the results illustrate how the IV set and its quality and the coefficient estimates they lead to change when the memory pattern of the latent regressor and error elements change. Section 5 concludes.

2 An ARMAX-EIV model for panel data

The model under consideration is first-order autoregressive and allows for individual heterogeneity and measurement errors in both endogenous and exogenous variables, in a panel of N individuals, indexed by i, which are observed in T periods, indexed by t:

$$(2.1)\quad \begin{array}{ll} \mu_{it} = \alpha_i + \boldsymbol{\xi}_{it}\boldsymbol{\beta} + \mu_{i,t-1}\lambda + u_{it}, \qquad |\lambda|<1, & \\ \boldsymbol{q}_{it} = \boldsymbol{\xi}_{it} + \boldsymbol{\eta}_{it}, & \\ y_{it} = \mu_{it} + \nu_{it}, & \\ \boldsymbol{\xi}_{it} \perp \boldsymbol{\eta}_{it} \perp u_{it} \perp \nu_{it}, & \\ \mathsf{E}(u_{it}) = 0, \quad \mathsf{E}(u_{it}u_{i,t+s}) \neq 0 \quad \text{for } |s| \leq N_u, \quad = 0 \text{ for } |s| > N_u, & \\ \mathsf{E}(\nu_{it}) = 0, \quad \mathsf{E}(\nu_{it}\nu_{i,t+s}) \neq 0, \quad \text{for } |s| \leq N_\nu, \quad = 0 \text{ for } |s| > N_\nu, & i = 1,\dots,N, \\ \mathsf{E}(\boldsymbol{\eta}_{it}) = \mathbf{0}, \quad \mathsf{E}(\boldsymbol{\eta}'_{it}\boldsymbol{\eta}_{i,t+s}) \neq \mathbf{0}, \quad \text{for } |s| \leq N_\eta, \quad = \mathbf{0} \text{ for } |s| > N_\eta, & t = 1,\dots,T, \end{array}$$

where α_i is a *fixed* effect, $(\mu_{it}, \boldsymbol{\xi}_{it})$ are latent variables where $\boldsymbol{\xi}_{it}$ has finite memory N_ξ, $(y_{it}, \boldsymbol{q}_{it})$ are their observable counterparts, $(\nu_{it}, \boldsymbol{\eta}_{it})$ are errors with memories (N_ν, N_η), u_{it} is a disturbance with memory N_u, and $(\lambda, \boldsymbol{\beta}', k)$ are constants. Boldface letters denote vectors with K elements, *rows for variables, columns for coefficients.* In general, all structural and error variables are then allowed to have memory.

Allowing for memory of errors can be motivated by some examples. First, the equation may include a stock variable, *e.g.*, of finished goods or of production capital constructed by cumulated flows, in which case (measurement) errors tend to vary cyclically. Second, for a flow variable like sales, cash-flow and income, the periodization may be incorrect, creating serial correlation, often negative, between errors which are close in time. Third, a latent non-stationary variable in levels, integrated of order P with white noise measurement errors, will have an observed counterpart which after differencing P times is stationary with MA(P) errors. An IID property for the latent regressors would have been detrimental to identification of slope coefficients, even in the case with zero memory if all errors, as demonstrated for a related static model in Biørn (2000, Section 2.b).[2]

Eliminating μ_{it} and $\boldsymbol{\xi}_{it}$ from (2.1) we obtain

$$y_{it} = \alpha_i + \boldsymbol{q}_{it}\boldsymbol{\beta} + y_{i,t-1}\lambda + w_{it}, \tag{2.2}$$

$$w_{it} = u_{it} + \nu_{it} - \nu_{i,t-1}\lambda - \boldsymbol{\eta}_{it}\boldsymbol{\beta}, \tag{2.3}$$

which after a one-period differencing yields

$$\Delta y_{it} = \Delta\boldsymbol{q}_{it}\boldsymbol{\beta} + \Delta y_{i,t-1}\lambda + \Delta w_{it}, \tag{2.4}$$

Since (2.2) implies

$$y_{it} = (1-\lambda)^{-1}\alpha_i + \textstyle\sum_{s=0}^{\infty} \lambda^s[\boldsymbol{q}_{i,t-s}\boldsymbol{\beta} + w_{i,t-s}], \tag{2.5}$$

$(y_{it}, \boldsymbol{q}_{i,t+\tau})$ and $w_{i,t+\theta}$, and hence $(\Delta y_{it}, \Delta\boldsymbol{q}_{i,t+\tau})$ and $\Delta w_{i,t+\theta}$, will be correlated for some (τ, θ), uncorrelated for others. A closer examination is needed to delimit valid IV sets. This follows in Section 3. We assume that the panel has a finite number of time periods, T; all asymptotics refer to the $N \to \infty$ case.

Notable special cases of this setup are:[3]

(i) EIV static, with memory in errors and disturbances: $\lambda = 0$.
(ii) EIV static, without memory in errors and disturbances: $\lambda = N_u = N_\nu = N_\eta = 0$.
(iii) ARMAX without errors: $\boldsymbol{\eta}_{it} = \boldsymbol{0}$.
(iv) ARX(1) without errors: $\boldsymbol{\eta}_{it} = \boldsymbol{0}$; $N_u = N_\nu = 0$.
(v) EIV-ARX(1) without memory in errors: $N_u = N_\nu = N_\eta = 0$.

Cases (i) and (v) include Case (ii). Case (iii) includes Case (iv). These special cases relate to the literature in several ways. Case (ii), and to some extent Case (i), are discussed in Griliches and Hausman (1986), Wansbeek and Koning (1991), Biørn (1992, 1996, 2000), Wansbeek (2002), Biørn and Krishnakumar (2008, Section 10.2), and Xiao *et al.* (2007, 2010). Case (iv) is related to Anderson and Hsiao (1981, 1982), Harris, Mátyás, and Sevestre (2008), Arellano and Bond (1991), and Arellano and Bover (1995). Holtz-Eakin *et al.* (1988) consider Case (iii) and a generalization which allows for higher-order autoregression, and also include errors without memory, although in a bivariate context. Notice also that by specifying the latent regressor vector such that cross-sectional dependence is allowed for, elements from the literature on *factor models* for panel data, see, *e.g.*, Pesaran (2006), can be accounted for by this setup.

3 GMM estimation

To delimit potential IVs we need to establish the set of ys and $\boldsymbol{q}$s that satisfy jointly rank conditions – ensuring correlation between the IVs and the variables in (2.2) or (2.4)

[2] "... 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)].

[3] ARX and ARMAX are acronyms for AR and ARMA models with exogenous variables.

for which they serve – and orthogonality conditions for the IVs and their composite errors/disturbances. How these sets change when the model's memory pattern changes, is of particular interest. Another issue is how to address problems related to an excessive number of 'weak IVs'.

Index sets for leads and lags

Let the integer τ index the lags($-$)/leads($+$) and let for arbitrary pairs of variables, $(a,b),(c,d)$, $\boldsymbol{Z}_{a\bullet b}$ denote the set of τ indexes which ensures $a_{i,t+\tau}$ and b_{it} to be *orthogonal*, and let $\boldsymbol{S}_{c\bullet d}$ denote the set of τ indexes which ensures $c_{i,t+\tau}$ and d_{it} to be *correlated*. For our problem, writing (2.2) and (2.4) as

$$\begin{aligned} y_{it} &= \alpha_i + \boldsymbol{x}_{it}\boldsymbol{\gamma} + w_{it}, & \boldsymbol{x}_{it} &= (\boldsymbol{q}_{it}, y_{i,t-1}), \\ \Delta y_{it} &= \Delta\boldsymbol{x}_{it}\boldsymbol{\gamma} + \Delta w_{it}, & \Delta\boldsymbol{x}_{it} &= (\Delta\boldsymbol{q}_{it}, \Delta y_{i,t-1}), \end{aligned} \qquad \boldsymbol{\gamma} = \begin{bmatrix} \boldsymbol{\beta} \\ \lambda \end{bmatrix},$$

the relevant index sets are:

$$\begin{array}{lll} [1] & \boldsymbol{Z}_{\Delta q\bullet w} \equiv \{\tau : \mathrm{cov}(\Delta\boldsymbol{q}_{i,t+\tau}, w_{it}) = \boldsymbol{0}\}, & \boldsymbol{Z}_{\Delta y\bullet w} \equiv \{\tau : \mathrm{cov}(\Delta y_{i,t+\tau}, w_{it}) = 0\}, \\ [2] & \boldsymbol{Z}_{q\bullet\Delta w} \equiv \{\tau : \mathrm{cov}(\boldsymbol{q}_{i,t+\tau}, \Delta w_{it}) = \boldsymbol{0}\}, & \boldsymbol{Z}_{y\bullet\Delta w} \equiv \{\tau : \mathrm{cov}(y_{i,t+\tau}, \Delta w_{it}) = 0\}, \\ [3] & \boldsymbol{S}_{\Delta q\bullet x} \equiv \{\tau : \mathrm{cov}(\Delta\boldsymbol{q}_{i,t+\tau}, \boldsymbol{x}_{it}) \neq \boldsymbol{0}\}, & \boldsymbol{S}_{\Delta y\bullet x} \equiv \{\tau : \mathrm{cov}(\Delta y_{i,t+\tau}, \boldsymbol{x}_{it}) \neq \boldsymbol{0}\}, \\ [4] & \boldsymbol{S}_{q\bullet\Delta x} \equiv \{\tau : \mathrm{cov}(\boldsymbol{q}_{i,t+\tau}, \Delta\boldsymbol{x}_{it}) \neq \boldsymbol{0}\}, & \boldsymbol{S}_{y\bullet\Delta x} \equiv \{\tau : \mathrm{cov}(y_{i,t+\tau}, \Delta\boldsymbol{x}_{it}) \neq \boldsymbol{0}\}. \end{array}$$

The sets [1] and [3] relate to the orthogonality and rank conditions for (2.2), while [2] and [4] relate to the orthogonality and rank conditions for (2.4),[4] and can be derived from

$$(3.1) \qquad \boldsymbol{Z}_{q\bullet w} \equiv \{\tau : \mathrm{cov}(\boldsymbol{q}_{i,t+\tau}, w_{it}) = \boldsymbol{0}\}, \qquad \boldsymbol{Z}_{y\bullet w} \equiv \{\tau : \mathrm{cov}(y_{i,t+\tau}, w_{it}) = 0\},$$

$$(3.2) \qquad \boldsymbol{S}_{q\bullet x} \equiv \{\tau : \mathrm{cov}(\boldsymbol{q}_{i,t+\tau}, \boldsymbol{x}_{it}) \neq \boldsymbol{0}\}, \qquad \boldsymbol{S}_{y\bullet x} \equiv \{\tau : \mathrm{cov}(y_{i,t+\tau}, \boldsymbol{x}_{it}) \neq \boldsymbol{0}\},$$

since in general

$$(3.3) \qquad \begin{aligned} \boldsymbol{Z}_{\Delta a\bullet b} &= \boldsymbol{Z}_{a\bullet b} \cap \boldsymbol{Z}_{a(-1)\bullet b}, \\ \boldsymbol{Z}_{a\bullet\Delta b} &= \boldsymbol{Z}_{a\bullet b} \cap \boldsymbol{Z}_{a\bullet b(-1)}, \end{aligned}$$

$$(3.4) \qquad \begin{aligned} \boldsymbol{S}_{\Delta c\bullet d} &= \boldsymbol{Z}_{c\bullet d} \cup \boldsymbol{Z}_{c(-1)\bullet d}, \\ \boldsymbol{S}_{c\bullet\Delta d} &= \boldsymbol{Z}_{c\bullet d} \cup \boldsymbol{Z}_{c\bullet d(-1)}. \end{aligned}$$

Therefore:

The sets $\boldsymbol{Z}_{\Delta q\bullet w} \cap \boldsymbol{S}_{\Delta q\bullet x}$ and $\boldsymbol{Z}_{\Delta y\bullet w} \cap \boldsymbol{S}_{\Delta y\bullet x}$ define the τ-indexes for which $(\Delta\boldsymbol{q}_{i,t+\tau}, \Delta y_{i,t+\tau})$ give valid IVs for $\boldsymbol{x}_{it} = (\boldsymbol{q}_{it}, y_{i,t-1})$ in (2.2).

The sets $\boldsymbol{Z}_{q\bullet\Delta w} \cap \boldsymbol{S}_{q\bullet\Delta x}$ and $\boldsymbol{Z}_{y\bullet\Delta w} \cap \boldsymbol{S}_{y\bullet\Delta x}$ define the τ-indexes for which $(\boldsymbol{q}_{i,t+\tau}, y_{i,t+\tau})$ give valid IVs for $\Delta\boldsymbol{x}_{it} = (\Delta\boldsymbol{q}_{it}, \Delta y_{i,t-1})$ in (2.4).

Memory pattern and potential IVs

Let (N_q, N_w) and $(N_{\Delta q}, N_{\Delta w})$ be the memory of $(\boldsymbol{q}_{it}, w_{it})$ and $(\Delta\boldsymbol{q}_{it}, \Delta w_{it})$, respectively. We have

$$(3.5) \qquad \begin{aligned} N_q &= \max[N_\xi, N_\eta] \\ N_w &= \max[N_u, N_\nu{+}1, N_\eta] \end{aligned} \implies \begin{aligned} N_{\Delta q} &= N_q{+}1, \\ N_{\Delta w} &= N_w{+}1. \end{aligned}$$

Also, let $N_\omega = \max[N_u, N_\nu{+}1]$ be the memory of the 'gross disturbance' in (2.2), $\omega_{it} = u_{it}{+}\nu_{it}{-}\nu_{i,t-1}\lambda$. Since (3.1)–(3.2) imply

[4]Subscript convention: IVs are placed before $\bullet$, the composite error/disturbance and the instrumented variable are placed after $\bullet$ in the $\boldsymbol{Z}$-sets and the $\boldsymbol{S}$-sets, respectively.

$$\begin{aligned}
\boldsymbol{Z}_{q\bullet w} &= \{\tau : |\tau| \geq N_\eta+1\},\\
\boldsymbol{Z}_{y\bullet w} &= \{\tau : \tau \leq -(N_\omega+1)\},\\
\boldsymbol{S}_{q\bullet q} &= \{\tau : |\tau| \leq N_q\},\\
\boldsymbol{S}_{y\bullet q} &= \{\tau : \tau \geq -N_\xi\},\\
\boldsymbol{S}_{q\bullet y} &= \{\tau : \tau \leq N_\xi\},
\end{aligned}$$

we find, using (3.3)–(3.4),

$$\begin{aligned}
\boldsymbol{Z}_{\Delta q\bullet w} &= \{\tau : \tau \notin [-N_\eta, N_\eta+1]\},\\
\boldsymbol{Z}_{\Delta y\bullet w} &= \{\tau : \tau \leq -(N_\omega+1)\},\\
\boldsymbol{Z}_{q\bullet \Delta w} &= \{\tau : \tau \notin [-(N_\eta+1), N_\eta]\},\\
\boldsymbol{Z}_{y\bullet \Delta w} &= \{\tau : \tau \leq -(N_\omega+2)\},
\end{aligned} \tag{3.6}$$

$$\begin{aligned}
\boldsymbol{S}_{\Delta q\bullet q} &= \{\tau : -N_q \leq \tau \leq N_q+1\},\\
\boldsymbol{S}_{\Delta y\bullet q} &= \{\tau : \tau \geq -N_\xi\},\\
\boldsymbol{S}_{\Delta q\bullet y(-1)} &= \{\tau : \tau \leq N_\xi\},\\
\boldsymbol{S}_{q\bullet \Delta q} &= \{\tau : -(N_q+1) \leq \tau \leq N_q\},\\
\boldsymbol{S}_{y\bullet \Delta q} &= \{\tau : \tau \geq -(N_\xi+1)\},\\
\boldsymbol{S}_{q\bullet \Delta y(-1)} &= \{\tau : \tau \leq N_\xi-1\}.
\end{aligned} \tag{3.7}$$

The potentially valid IVs thus obtained are:[5]

Equation in levels, IVs in differences. General case

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

[A]: $\Delta\boldsymbol{q}_{i,t+\tau}$ for $\tau \in (\boldsymbol{Z}_{\Delta q\bullet w} \cap \boldsymbol{S}_{\Delta q\bullet q}) \Longrightarrow \tau \notin [-N_\eta, N_\eta+1]$ & $\tau \in [-N_q \leq \tau \leq N_q+1]$.

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

Potential IVs for $y_{i,t-1}$*:*

[C]: $\Delta\boldsymbol{q}_{i,t+\tau}$ for $\tau \in (\boldsymbol{Z}_{\Delta q\bullet w} \cap \boldsymbol{S}_{\Delta q\bullet y(-1)}) \Longrightarrow \tau \notin [-N_\eta, N_\eta+1]$ & $\tau \leq N_\xi$.

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

Equation in differences, IVs in levels. General case

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

[A]: $\boldsymbol{q}_{i,t+\tau}$ for $\tau \in (\boldsymbol{Z}_{q\bullet \Delta w} \cap \boldsymbol{S}_{q\bullet \Delta q}) \Longrightarrow \tau \notin [-(N_\eta+1), N_\eta]$ & $\tau \in [-(N_q+1), N_q]$.

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

Potential IVs for $\Delta y_{i,t-1}$*:*

[C]: $\boldsymbol{q}_{i,t+\tau}$ for $\tau \in (\boldsymbol{Z}_{q\bullet \Delta w} \cap \boldsymbol{S}_{q\bullet \Delta y(-1)}) \Longrightarrow \tau \notin [-(N_\eta+1), N_\eta]$ & $\tau \leq N_\xi-1$.

[D]: $y_{i,t+\tau}$ for $\tau \in (\boldsymbol{Z}_{y\bullet \Delta w} \cap \boldsymbol{S}_{y\bullet \Delta y(-1)}) \Longrightarrow \tau \leq -(N_\omega+2)$.

From (3.5) it follows that in general

Increasing $N_\xi \Longrightarrow$ *Increasing* $N_q \Longrightarrow$ *Potential* IV set extended.
Increasing $N_\omega \Longrightarrow$ *Increasing* $N_w \Longrightarrow$ *Potential* IV set diminished.
Increasing $N_\eta \Longrightarrow$ *Increasing* $N_w \Longrightarrow$ Effect on *Potential* IV set indeterminate.

For the *potential* IV set to be sufficiently large, the latent regressors must have a sufficiently long and the disturbances and errors in the endogenous variables must have a sufficiently short memory. Since an increase in the error memory of the exogenous variables increases the memory of both $\boldsymbol{q}$ and w in general, its effect on the potential IV set is indeterminate.

[5] Note that the sets [C] and [D] have elements in common with the sets [A] and [B].

If $N_\eta \geq N_\xi$, no valid $\Delta\boldsymbol{q}$-IV for $\boldsymbol{q}$ and no valid $\boldsymbol{q}$-IV for $\Delta\boldsymbol{q}$ exist. If $N_\omega \geq N_\xi$, no Δy can be an IV for $\boldsymbol{q}$ and no y can be an IV for $\Delta\boldsymbol{q}$.

If the disturbance and the errors have no memory ($N_\eta = N_u = N_\nu = 0 \implies N_q = N_\xi > 0$, $N_w = 1$), confer Case (v) above, we get

$$\begin{aligned} \boldsymbol{Z}_{\Delta q \bullet w} &= \{\tau : \tau \neq 0, 1\}, \\ \boldsymbol{Z}_{\Delta y \bullet w} &= \{\tau : \tau \leq -2\}, \\ \boldsymbol{Z}_{q \bullet \Delta w} &= \{\tau : \tau \neq -1, 0\}, \\ \boldsymbol{Z}_{y \bullet \Delta w} &= \{\tau : \tau \leq -3\}, \end{aligned} \tag{3.8}$$

$$\begin{aligned} \boldsymbol{S}_{\Delta q \bullet q} &= \{\tau : -N_\xi \leq \tau \leq N_\xi + 1\}, \\ \boldsymbol{S}_{\Delta y \bullet q} &= \{\tau : \tau \geq -N_\xi\}, \\ \boldsymbol{S}_{\Delta q \bullet y(-1)} &= \{\tau : \tau \leq N_\xi\}, \\ \boldsymbol{S}_{q \bullet \Delta q} &= \{\tau : -(N_\xi + 1) \leq \tau \leq N_\xi\}, \\ \boldsymbol{S}_{y \bullet \Delta q} &= \{\tau : \tau \geq -(N_\xi + 1)\}, \\ \boldsymbol{S}_{q \bullet \Delta y(-1)} &= \{\tau : \tau \leq N_\xi - 1\}, \end{aligned} \tag{3.9}$$

so that the potentially valid IVs become:

Equation in levels, IVs in differences. No error memory case

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

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

Potential IVs for $y_{i,t-1}$:

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

Equation in differences, IVs in levels. No error memory case

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

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

Potential IVs for $\Delta y_{i,t-1}$:

[C]: $\boldsymbol{q}_{i,t+\tau}$ for $\tau \in (\boldsymbol{Z}_{q \bullet \Delta w} \cap \boldsymbol{S}_{q \bullet \Delta y(-1)}) \implies \tau = \dots, -4, -3, -2, 1, 2, \dots, N_\xi - 1$,
[D]: $y_{i,t+\tau}$ for $\tau \in (\boldsymbol{Z}_{y \bullet \Delta w} \cap \boldsymbol{S}_{y \bullet \Delta y(-1)}) \implies \tau \leq -3$.

For several (N, T) combinations, the number of orthogonality conditions may be very large, which suggests that some potential IVs should be omitted. The 'weak instrument' problem for AR(1) panel data models using variables in differences, combined with what Roodman (2009) characterizes as 'instrument proliferation .... an underappreciated problem .... in the application of difference and system GMM', give reasons for 'curtailing' the IV sets. Aspects of the problem, related, *inter alia*, to small-sample bias and estimation efficiency, are also discussed by Altonji and Segal (1996) and Ziliak (1997). The way the IV set is 'curtailed' in the empirical application will be described in Section 4.

Let $\boldsymbol{x}_{it} = (\boldsymbol{q}_{it}, y_{i,t-1})$, $\boldsymbol{\gamma} = (\boldsymbol{\beta}', \lambda)'$ and write (2.2) and (2.4) as

$$y_{it} = \alpha_i + \boldsymbol{x}_{it}\boldsymbol{\gamma} + w_{it}, \tag{3.10}$$

$$\Delta y_{it} = \Delta\boldsymbol{x}_{it}\boldsymbol{\gamma} + \Delta w_{it}. \tag{3.11}$$

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 equations, in system format, subscripts L and D denoting *Level* and *Difference*, respectively, read

$$\boldsymbol{y}_{Li} = \boldsymbol{\alpha}_i + \boldsymbol{X}_{Li}\boldsymbol{\gamma} + \boldsymbol{w}_{Li}, \tag{3.12}$$

$$\boldsymbol{y}_{Di} = \boldsymbol{X}_{Di}\boldsymbol{\gamma} + \boldsymbol{w}_{Di}. \tag{3.13}$$

For $\boldsymbol{x}_{it}$ in (3.10), we use an IV vector *in differences*, and for $\Delta\boldsymbol{x}_{it}$ in (3.11), we use an IV vector *in levels*, defined as respectively,

$$\boldsymbol{z}_{Di(t)} = \boldsymbol{Q}_t \Delta\boldsymbol{x}_{it}, \tag{3.14}$$

$$\boldsymbol{z}_{Li(t,t-1)} = \boldsymbol{P}_{t,t-1}\boldsymbol{x}_{it}, \tag{3.15}$$

where $\boldsymbol{Q}_t$ and $\boldsymbol{P}_{t,t-1}$ are selection matrices which comply with the above conditions. The implied IV matrices for $\boldsymbol{X}_{Li}$ and $\boldsymbol{X}_{Di}$ are, respectively,

$$\boldsymbol{Z}_{Di} = \text{diag}[\boldsymbol{x}'_{Di(1)}, \boldsymbol{x}'_{Di(2)}, \dots, \boldsymbol{x}'_{Di(T)}], \tag{3.16}$$

$$\boldsymbol{Z}_{Li} = \text{diag}[\boldsymbol{x}'_{Li(2,1)}, \boldsymbol{x}'_{Li(3,2)}, \dots, \boldsymbol{x}'_{Li(T,T-1)}], \tag{3.17}$$

where 'diag' denotes block-diagonalization.

The estimators for $\boldsymbol{\gamma}$ in (3.12) and (3.13) considered are 'step-two' GMM, of the form

$$\begin{aligned} \widetilde{\boldsymbol{\gamma}}_L = \{[\textstyle\sum_i \boldsymbol{X}'_{Li}\boldsymbol{Z}_{Di}][\sum_i \boldsymbol{Z}'_{Di}\widehat{\boldsymbol{w}}_{Li}\widehat{\boldsymbol{w}}'_{Li}\boldsymbol{Z}_{Di}]^{-1}[\sum_i \boldsymbol{Z}'_{Di}\boldsymbol{X}_{Li}]\}^{-1} \\ \times \{[\textstyle\sum_i \boldsymbol{X}'_{Li}\boldsymbol{Z}_{Di}][\sum_i \boldsymbol{Z}'_{Di}\widehat{\boldsymbol{w}}_{Li}\widehat{\boldsymbol{w}}'_{Li}\boldsymbol{Z}_{Di}]^{-1}[\sum_i \boldsymbol{Z}'_{Di}\boldsymbol{y}_{Li}]\}, \end{aligned} \tag{3.18}$$

$$\begin{aligned} \widetilde{\boldsymbol{\gamma}}_D = \{[\textstyle\sum_i \boldsymbol{X}'_{Di}\boldsymbol{Z}_{Li}][\sum_i \boldsymbol{Z}'_{Li}\widehat{\boldsymbol{w}}_{Di}\widehat{\boldsymbol{w}}'_{Di}\boldsymbol{Z}_{Li}]^{-1}[\sum_i \boldsymbol{Z}'_{Li}\boldsymbol{X}_{Di}]\}^{-1} \\ \times \{[\textstyle\sum_i \boldsymbol{X}'_{Di}\boldsymbol{Z}_{Li}][\sum_i \boldsymbol{Z}'_{Li}\widehat{\boldsymbol{w}}_{Di}\widehat{\boldsymbol{w}}'_{Di}\boldsymbol{Z}_{Li}]^{-1}[\sum_i \boldsymbol{Z}'_{Li}\boldsymbol{y}_{Di}]\}. \end{aligned} \tag{3.19}$$

where $\widehat{\boldsymbol{w}}_{Li}$ and $\widehat{\boldsymbol{w}}_{Di}$ are residuals from 'step-one' GMM estimators; see Davidson and MacKinnon (2004, Sections 9.2-9.3).

The present application of GMM can be specialized to a standard ARMAX model and to a static EIV model. The similarities in and the differences between the way GMM is applied in these boundary cases can be summarized as follows:

Common features: (a) GMM exploits *orthogonality and rank conditions* delimiting potential IVs, and can be implemented on the equation in *differences* with IVs in *levels*, or on the equation in *levels* with IVs in *differences*. At least one should be in differences to account for the heterogeneity represented by the fixed effects. (b) In choosing level IVs for an equation in differences, values for periods other than those defining the differences should be selected, while in choosing difference IVs for an equation in levels, the periods represented by the levels should be excluded. (c) When memory of the errors or disturbances are allowed for, the IV set should be diminished to ensure that all IVs 'get clear of' the memory of the composite error processes, while the IVs are correlated with the variables for which they serve.

Discrepancies: (a) For the ARMAX model, *only lagged* exogenous and endogenous variables are potential IVs. Their validity *follows from* the model's memory pattern. For the static EIV model, *lagged and leaded* values of endogenous and exogenous variables can serve as IVs, provided that certain mild *additional conditions* are satisfied. (b) The ARMAX model has a memory pattern which rationalizes the orthogonality and rank conditions, while the static EIV model has no such 'intrinsic' memory pattern. (c) While for the static EIV model, (leaded or lagged) y-values may serve as IVs for other ys, for the ARMAX model, no y-values can be valid IVs for other ys.

4 An illustration

Empirical example

For our illustrative application we utilize data from $N = 136$ Norwegian firms in the Pulp & Paper/Chemicals/Basic Metals industries, observed in $T = 10$ years (1984–1993). We consider a relationship between an assumed endogenous measure of capital in machinery, k, and an assumed exogenous measure of the sector's output, q, both log-transformed ($K = 1$) and believed to be error-ridden. The *capital-output relationship* is particularly relevant for this purpose. First, it exemplifies well dynamic modelling with panel data, in a case where it is disputable whether to express the relationship in levels or in changes – contrast the acceleration principle in some business-cycle theories with the constant capital coefficient (elasticity) assumption in certain economic growth theories. Second, both capital stocks and output flows are notoriously difficult to measure. To prevent an excessive number of potentially weak lagged y-IVs, only the q- and y-IVs listed under [A] and [B] in Section 3 are included.[6] The results are presented in Tables 1 through 4.

Four constellations of memory length for the errors are considered: $(N_\nu, N_\eta, N_\xi) = (0,0,4), (0,1,4), (1,1,4), (0,2,4)$; for the disturbance a zero memory ($N_u = 0$) is assumed throughout. The latter is no essential limitation, as $N_\omega = \max[N_u, N_\nu + 1]$, For each constellation, parallel results when including only q-IVs and including both q- and y-IVs as described above, are provided – giving a total of eight variants. In columns 1–2, only the latent regressor ξ is assumed to have a 4-period memory, in columns 3–8 error memory, shorter than the regressor memory, is also allowed for. For each set of results are reported p-values for *Sargan-Hansen $\mathcal{J}$-tests for orthogonality*, confer Hansen (1982), Newey (1985), and Arellano and Bond (1991); and p-values for *$\mathcal{F}$-tests for IV validity/strength based on a 'concentration measure'*, confer Staiger and Stock (1997) and Bun and Windmeijer (2010). Standard R^2 indexes of goodness of fit, are supplemented by R^2 measures based on pre-

[6] A computer program in the Gauss software code, version 7.0, cf. Gauss (2006), constructed by Xuehui Han in cooperation with the author, is applied. The reported standard errors are calculated from the GMM formulae as described in Biørn and Krishnakumar (2008, Section 10.2.5).

diction errors for equations estimated by IVs, as proposed by Pesaran and Smith (1994), the latter denoted as PS R^2 in the tables. The orthogonality $\mathcal{J}$-test statistics are, for the level version and the difference version, respectively,

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

Under the null hypotheses they are asymptotically distributed as χ^2 with 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.

Consider first estimation from a *static equation in levels*. All estimates of the capital elasticity β exceed one (Table 1). Introducing a one- or two-period memory in the measurement errors, while retaining the four-period memory of the latent regressor, has a small impact on the coefficient estimates, and the p-values of the $\mathcal{J}$-test indicate non-rejection of the orthogonality conditions when only qs are included in the IV set (p-values above 0.1). Including also y-IVs (columns 2, 4, 6 and 8), both R^2 measures are increased, while the p-values of the $\mathcal{J}$-tests are below 5 %, when zero or one-period memory of the measurement error in output is assumed ($N_\eta = 0$ or 1), which indicates rejection. However, if the measurement error memory is increased to 2, non-rejection of the (diminished) set of orthogonality conditions follows for both selections of IVs (column 8, p-values 28 % and 43 %, respectively).

For a corresponding *AR(1) equation in levels* (Table 2) we get β estimates which depart substantially from those for the static equation, while the λ estimates are close to one, which – not surprisingly – signalizes substantial persistence in the capital adjustment. Overall, the β estimates are negative although not significantly different from zero. The conclusions from the $\mathcal{J}$-tests are similar to those for the static equation when only q-IVs are included. A notable difference between the AR(1) equation and the static equation, is that for the memory constellation $(N_\nu, N_\eta, N_\xi) = (1, 1, 4)$ when both q- and y-IVs are included, the orthogonality conditions are not rejected in the former case and rejected in the latter (p-value 17.0 % against 2.7 %).

For the *static equation in differences* (Table 3)[7] all β estimates are positive, but none exceed 0.04, and only a few are significantly positive (5% one-sided test). The estimates are higher when error memory is allowed for than when it is neglected. The goodness of fit, according the PS R^2, is poor, and markedly worse than indicated by standard R^2. As all $\mathcal{J}$-tests have p-values less than 5 %, rejection of the orthogonality conditions is indicated. For memory constellations $(N_\nu, N_\eta, N_\xi) = (0, 1, 4), (1, 1, 4)$, however, the p-values are only slightly below 5%.

The corresponding results for an *AR(1)-equation in differences* (Table 4) depart substantially from the estimates based on the level equation: the λ estimates are in the range 0.6–0.8 and depart significantly from both 0 and 1. Both the β and the λ estimates are higher when error memory is allowed for than when it is neglected. Overall, the p-values

[7] Because of (3.5), $N_\xi - N_\nu \geq 3$ must hold to ensure an IV set sufficiently large to make GMM estimation feasible.

for the $\mathcal{J}$-tests when only q-IVs are used, suggest non-rejection of the orthogonality conditions, but when they are supplemented with y-IVs, rejection is indicated. Unlike the corresponding results for the level version of the equation (Table 2), the p-values of the $\mathcal{F}$-tests for IV strength in Table 4 clearly signalize a 'weak IV' problem. The latter comes as no surprise, however, since capital stock time series usually have strong persistence. Blundell and Bond (1998), discussing GMM estimation of a more restrictive AR(1) equation, transformed to first-differences, in a non-EIV context, conclude that this procedure can have very poor finite sample properties in terms of bias and estimator precision when the series considered are persistent, because then the IVs are weak predictors for the differenced endogenous variable.

The two overall conclusions with respect to IV-validity we can draw from the capital-output example in Tables 1–4 then become: (i) Inclusion of y-IVs tends to violate the orthogonality conditions. The tendency is clearest for the static equation in differences with IVs in levels. (ii) Among the four model versions, the IV-set seems to be weakest for the AR(1)-equation in differences. This holds both when only q-IVs are used and when they are combined with y-IVs.

Simulation experiments

To examine *finite sample properties* of the GMM estimators considered, we supplement the above results with *Monte Carlo (MC) simulations.* Basic statistics are collected in Tables 5 and 6. The experiments relate to a panel design with dimension parameters $K=1$, $N=100$, $T=10$, rather similar to the empirical example in Tables 1–4. The number of replications is set to $R=200$. Two coefficient sets resembling the estimates from the empirical example are considered, where part A mimics a static equation and part B an ARX(1) equation with the same unitary long-run coefficient:

A: $\beta = 1$, $\lambda = 0$,
B: $\beta = 0.3$, $\lambda = 0.7$.

The latent regressor ξ_{it} is generated as the sum of two independent components, a time invariant, normally distributed part with expectation 5, and an MA(4)-process with zero mean. For the measurement error in the exogenous variable, η_{it}, both a white noise, an MA(1) process, and an MA(2) process are considered. For the measurement error in the endogenous variable ν_{it} and the disturbance u_{it}, white noise is assumed in all experiments. The individual heterogeneity (assumed fixed in the model in Section 2) is generated by a similar process. The memory parameters, the other distributional assumptions and the parameter values are given below (IIN denoting i.i.d. normal):[8]

$$\begin{aligned}
&\chi_i \sim \mathsf{IIN}(5,0.1),\ \alpha_i \sim \mathsf{IIN}(0,0.1),\ u_{it} \sim \mathsf{IIN}(0,0.1),\ \nu_{it} \sim \mathsf{IIN}(0,0.1),\\
&N_\xi = 4:\ \xi_{it} = \chi_i + \psi_{it} + 0.8\psi_{i,t-1} + 0.6\psi_{i,t-2} + 0.4\psi_{i,t-3} + 0.2\psi_{i,t-4}, \qquad \psi_{it} \sim \mathsf{IIN}(0,0.1),\\
&N_\eta = 0:\ \eta_{it} = \epsilon_{it},\\
&N_\eta = 1:\ \eta_{it} = \epsilon_{it} + 0.5\epsilon_{i,t-1},\\
&N_\eta = 2:\ \eta_{it} = \epsilon_{it} + 0.5\epsilon_{i,t-1} + 0.25\epsilon_{i,t-2}, \qquad \epsilon_{it} \sim \mathsf{IIN}(0,0.1).
\end{aligned}$$

[8] This implies: $\mathrm{var}(\xi_{it}) = 0.32$ and $\mathrm{var}(\eta_{it}) = 0.1$, $= 0.125$, $= 0.13125$ for $N_\eta = 0,1,2$, respectively.

Simulations: Equation in levels. The means of the simulated β-estimates from the static equation in levels (Table 5, part A) are lower than the values assumed, *i.e.*, negative finite sample bias in $\mathsf{E}(\widehat{\beta})$. The bias is smallest in the no error memory case (mean of estimates 0.9573 and 0.9479 in the q-IV and the qy-IV cases, respectively, roughly 5% below the assumed value of 1.) When data are generated from the static equation ($\lambda = 0$), the means of the simulated λ-estimates are negative, although with so large standard errors that 'non-rejection' of $\lambda = 0$ at conventional significance levels is indicated. In the ARX equation in levels (Table 5, part B), the mean of the simulated β-estimates, in the range 0.32–0.34, exceed, for all six combinations of error memory and IV selection, the value assumed, 0.3, indicating positive finite sample bias in $\mathsf{E}(\widehat{\beta})$ of about 10%. The counterpart is that the means of the simulated λ estimates, in the range 0.65–0.68, indicate an underestimation of 3–6%, as the assumed value is 0.7. The bias is smaller when both q-IVs and y-IVs are included than when only q-IVs are used.

The means of the standard error estimates of both the β and the λ estimates are for both the static and the ARX-equation very close to the empirical standard deviations of the estimates. Which measure is used to assess estimator precision hence seems to be of minor importance. Supplementing q-IVs with y-IVs reduces the standard errors both when the data are generated from the static equation (panel A) and from the ARX equation (panel B). For the former, both the orthogonality $\mathcal{J}$-test statistics and the $\mathcal{F}$-test statistics for IV-strength are satisfactory, as judged from the mean values of their respective p-values. While for the ARX equation the orthogonality $\mathcal{J}$-test statistics are satisfactory, the overall strength of the IV-set, as indicated by the $\mathcal{F}$-test statistics, seems weak. We can interpret this as indicating that adding y-IVs to q-IVs 'weakens' the IV set when the AR parameter is a large as 0.7, even for the equation in levels.

From Table 5 we also see that the range of the 200 simulated estimates ('MAX' minus 'MIN') increases when the IV set is diminished following from an increased memory of the 'noise' in the exogenous variable (N_η). This result may also reflect the increased variance of the 'noise' relative to the variance of the 'signal' (N_ξ).

Simulations: Equation in differences. Considering the equation in differences (Table 6), we find for both the static equation (part A) and the ARX-equation a sizable negative finite sample bias for both β and λ. For the static equation, the means of standard error estimates come out with values which are very close to the empirical standard deviations of the estimates for both β and λ. Adding y-IVs to q-IVs for the equation in differences, reduces the standard errors of the estimates both when the data are generated from the static equation (panel A) and from the ARX equation (panel B). In this case, however, the means of $se(\widehat{\beta})$ lie somewhat below the empirical standard deviations of the $\widehat{\beta}$ estimates (0.1425 against 0.1701 for $N_\eta = 0$, 0.1574 against 0.1605 for $N_\eta = 1$, and 0.1775 against 0.1903 for $N_\eta = 2$). On the other hand, for the λ estimates, we find no similar discrepancy between the two ways of computing standard errors.

The bias, indicated by the difference between the mean of the estimates and the value assumed, is overall larger when using the equation in differences than when keeping it in levels. When neither disturbances nor measurement errors are assumed to have a memory,

$N_u = N_\nu = N_\eta = 0$, we obtain, for example, a mean equal to $\widehat{\beta} = 0.65$ when the data are generated by $(\beta, \lambda) = (1, 0)$ (panel A) and a mean equal to $\widehat{\beta} \approx 0.2$ when $(\beta, \lambda) = (0.3, 0.7)$ is assumed in the data generation (panel B), *i.e.*, a negative bias exceeding 30%. When the data are generated from $(\beta, \lambda) = (1, 0)$, the negative bias when using q-IVs is larger when the measurement error in the exogenous variable is an MA process than when it is white noise (mean $\widehat{\beta} = 0.65$, 0.54, 0.53 for $N_\eta = 0, 1, 2$, respectively). Also for λ there is a negative bias, as the mean of the λ estimates is negative (the two last columns of part A).

Also with respect to IV-quality, the test statistics show a marked difference between the static equation and the ARX-equation. For the static equation (Table 6, panel A), both the orthogonality $\mathcal{J}$-test statistics and the IV-strength $\mathcal{F}$-test statistics for IV-strength come out with satisfactory values when only q-IVs are included, again as judged from the means of their respective p-values. Including also y-IVs a 'weak IV problem' is indicated. For the ARX equation (Table 6, panel B), the p-values of the orthogonality tests seem acceptable only when q-IVs are used alone. The means of the p-values for the $\mathcal{F}$-tests based on the concentration measure indicate a 'weak IV problem' even when only q-IVs are considered. Overall, these simulations suggest 'poor IV quality' of the chosen procedures for GMM estimation of an ARX equation in differences.

Again, the range of the 200 simulated estimates ('MAX' minus 'MIN') increases when the IV set is diminished, following from the increased memory of the noise.

Overall, these Monte Carlo simulations indicate that when performing GMM estimation of a panel data equation with individual heterogeneity, it is preferable to keep the equation in levels and using IVs in differences rather than transforming the equation to differences and using IVs in levels. This is the recommendation both with regard to finite sample bias – in both cases there is a non-negligible underestimation of the coefficients in the mean – and with regard to instrument validity and reliability. The results also suggest that one should be careful not to rely too strongly on y-IVs as supplements to q-IVs and to select y-IV candidates with care. It therefore seems that our decision of curtailing, in a somewhat informal way, the potential set of y-IVs when constructing the actual set of q- and y-IVs, and hence restricting 'IV proliferation', is well founded.

5 Concluding remarks

In this paper we have considered an application of the GMM that can handle jointly the heterogeneity problem and the measurement error problem in static and dynamic panel data models under different assumptions for the error memory. These problems are often intractable when only pure (single or repeated) cross section data or pure time series data are available, in the sense that consistent estimators are unavailable because of lack of identification. Estimators using either equations in differences with level values as IVs, or equations in levels with differenced values as IVs seem *a priori* useful candidates when panel data are at hand. GMM exploits certain orthogonality conditions and certain rank conditions, which jointly delimit a class of *potential* IVs. Transformation to differences is performed to eliminate fixed heterogeneity (or random heterogeneity correlated with regressors). GMM estimation can be implemented: (a) on the equation transformed to differences with IVs in levels, or (b) on the equation in levels using IVs in differences. It is important that at least one of the two are in differences in order to get clear of the nuisance created by the individual effects.

For the pure ARMAX model, only a selection of lagged endogenous and exogenous variables can be IVs. For the static EIV model a selection of lagged and leaded endogenous and exogenous variables are potential candidates. For the mixed ARMAX-EIV model here considered it is convenient to let index sets represent the leads and the lags in the observed endogenous and exogenous variables which jointly satisfy the orthogonality and rank conditions. 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 special case with a static EIV model, a memory pattern which satisfies the rank condition must be postulated as a supplement. For the present mixed model a trade-off exists between the memory lengths 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 strategy for GMM estimation, is known to raise 'weak instruments' problems in pure AR or ARX models with no measurement errors. The same is true for the more complex model considered here, but not to the same extent when using the equation in levels and when using the version in differences. Operationalizing procedures to identify such weak IVs within the potential IV set, the number of which may be a large even for moderate time series length, is a challenge for future research. The simulation experiments performed, considering their obvious limitations, suggest that a GMM estimation strategy in which the equation in kept in levels and the IVs used in differences performs far better than the frequently used difference transformation, not least with respect to finite sample bias and IV strength. Also for the capital-output relationship example considered, the former seems to outperform the latter by a considerable margin. The results again indicate that care should be taken not to rely too strongly on lagged values of endogenous variables as IVs even if they satisfy IV-requirements on *a priori* grounds. A strategy curtailing the set of potential IVs is well founded.

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., 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.

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.

Bun, M.J.G., and Windmeijer, F. (2010): The Weak Instrument Problem of the System GMM Estimator in Dynamic Panel Data Models. *Econometrics Journal* **13**, 95–126.

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.

Gauss (2006): Gauss^{TM}. Version 7.0. User's Guide. Maple Valley: Aptech Systems.

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, M.N., Mátyás. L. and Sevestre, P. (2008): Dynamic Models for Short Panels. Chapter 8 in *The Econometrics of Panel Data. Fundamentals and Recent Developments in Theory and Practice*, ed. by L. Mátyás and P. Sevestre. Berlin: Springer-Verlag.

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

Kiviet, J.F. (1995): On Bias, Inconsistency, and Efficiency of Various Estimators in Dynamic Panel Data Models. *Journal of Econometrics* **68**, 53–78.

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., 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.

Staiger, D., and Stock, J.H. (1997): Instrumental Variables Regression with Weak Instruments. *Econometrica* **65**, 557–586.

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.

Xiao, Z., Shao, J., and Xu, R. (2007): Efficiency of GMM Estimation in Panel Data Models with Measurement Error. *Sankhya: The Indian Journal of Statistics* **69**, 101–118.

Xiao, Z., Shao, J., and Palta, M. (2010): Instrumental Variables and GMM Estimation for Panel Data with Measurement Errors. *Statistica Sinica* **20**, 1725–1747.

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.

Table 1: Static capital adjustment equation in levels. IVs in differences
$N = 136, T = 10$. Different memory pattern and IV selection. Standard errors in parenthesis.

	Memory constellation: $N_u = 0$, $(N_\nu, N_\eta, N_\xi) =$							
	(0, 0, 4)		(0, 1, 4)		(1, 1, 4)		(0, 2, 4)	
	q-IV	qy-IV	q-IV	qy-IV	q-IV	qy-IV	q-IV	qy-IV
β= coef. of q	1.1378	1.1716	1.1408	1.1862	1.1408	1.1750	1.1405	1.1996
	(0.0368)	(0.0284)	(0.0387)	(0.0272)	(0.0387)	(0.0295)	(0.0405)	(0.0262)
R^2	0.0944	0.1412	0.0619	0.1077	0.0619	0.0931	0.0388	0.0868
PS R^2	0.0926	0.1390	0.0606	0.1061	0.0606	0.0916	0.0379	0.0855
J-test,p=	0.1182	0.0433	0.1499	0.0434	0.1499	0.0270	0.2791	0.4276
F-test,p=	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0005	0.0001
# IVs	52	70	36	54	36	47	22	40

J-test = Hansen-Sargan orthogonality test.
F-test = Staiger-Stock IV Validity test based on concentration parameter

Table 2: AR(1) capital adjustment equation in levels. IVs in differences
$N = 136, T = 10$. Different memory pattern and IV selection. Standard errors in parenthesis.

	Memory constellation: $N_u = 0$, $(N_\nu, N_\eta, N_\xi) =$							
	(0, 0, 4)		(0, 1, 4)		(1, 1, 4)		(0, 2, 4)	
	q-IV	qy-IV	q-IV	qy-IV	q-IV	qy-IV	q-IV	qy-IV
β= coef. of q	-0.0020	-0.0059	-0.0031	-0.0079	-0.0031	-0.0059	-0.0057	-0.0095
	(0.0049)	(0.0053)	(0.0064)	(0.0065)	(0.0064)	(0.0061)	(0.0056)	(0.0056)
λ= coef. of y_{-1}	1.0014	1.0063	1.0015	1.0081	1.0015	1.0060	1.0039	1.0100
	(0.0040)	(0.0045)	(0.0052)	(0.0055)	(0.0052)	(0.0051)	(0.0050)	(0.0049)
R^2	0.0709	0.1231	0.0452	0.0966	0.0452	0.0790	0.0282	0.0806
PS R^2	0.0709	0.1231	0.0452	0.0966	0.0452	0.0790	0.0282	0.0806
J-test,p=	0.1840	0.0282	0.5283	0.0407	0.5283	0.1699	0.4306	0.0080
F-test,p=	0.0039	0.0012	0.0058	0.0009	0.0058	0.0016	0.0409	0.0039
# IVs	44	59	30	45	30	39	18	33

J-test = Hansen-Sargan orthogonality test.
F-test = Staiger-Stock IV Validity test based on concentration parameter

Table 3: Static capital adjustment equation in differences. IVs in levels
$N = 136, T = 10$. Different memory pattern and IV selection. Standard errors in parenthesis.

	Memory constellation: $N_u = 0$, (N_ν, N_η, N_ξ) =							
	(0, 0, 4)		(0, 1, 4)		(1, 1, 4)		(0, 2, 4)	
	q-IV	qy-IV	q-IV	qy-IV	q-IV	qy-IV	q-IV	qy-IV
β= coef. of Δq	0.0152	0.0167	0.0260	0.0312	0.0260	0.0286	0.0248	0.0346
	(0.0106)	(0.0101)	(0.0153)	(0.0151)	(0.0153)	(0.0151)	(0.0194)	(0.0199)
R^2	0.1260	0.1661	0.0754	0.1192	0.0754	0.1032	0.0542	0.1012
PS R^2	0.0042	0.0053	0.0046	0.0077	0.0046	0.0061	0.0030	0.0073
J-test,p=	0.0226	0.0080	0.0462	0.0009	0.0462	0.0004	0.0044	0.0000
F-test,p=	0.2311	0.0476	0.0001	0.0000	0.0001	0.0000	0.0045	0.0000
# IVs	52	70	36	54	36	47	22	40

J-test = Hansen-Sargan orthogonality test.
F-test = Staiger-Stock IV Validity test based on concentration parameter

Table 4: AR(1) capital adjustment equation in differences. IVs in levels
$N = 136, T = 10$. Different memory pattern and IV selection. Standard errors in parenthesis.

	Memory constellation: $N_u = 0$, (N_ν, N_η, N_ξ) =							
	(0, 0, 4)		(0, 1, 4)		(1, 1, 4)		(0, 2, 4)	
	q-IV	qy-IV	q-IV	qy-IV	q-IV	qy-IV	q-IV	qy-IV
β= coef. of Δq	0.0126	0.0098	0.0214	0.0169	0.0214	0.0178	0.0168	0.0149
	(0.0121)	(0.0111)	(0.0131)	(0.0106)	(0.0131)	(0.0104)	(0.0163)	(0.0129)
λ= coef. of Δy_{-1}	0.6117	0.6007	0.6943	0.6330	0.6943	0.6453	0.7749	0.6384
	(0.0795)	(0.0841)	(0.1088)	(0.0938)	(0.1088)	(0.0834)	(0.0826)	(0.1044)
R^2	0.1367	0.2007	0.0804	0.1572	0.0804	0.1158	0.0563	0.1396
PS R^2	0.0535	0.0908	0.0455	0.0916	0.0455	0.0662	0.0323	0.0820
J-test,p=	0.2034	0.0409	0.0878	0.0099	0.0878	0.0181	0.0420	0.0058
F-test,p=	0.7564	0.6255	0.0825	0.0934	0.0825	0.0556	0.1596	0.1530
# IVs	44	62	30	48	30	41	18	36

J-test = Hansen-Sargan orthogonality test.
F-test = Staiger-Stock IV Validity test based on concentration parameter

Table 5: MC simulations of GMM. Equation in levels, IVs in differences

A. $\beta = 1, \lambda = 0, N = 100, T = 10, R = 200$

$(N_u, N_\nu, N_\eta, N_\xi) = (0,0,0,4)$

	q-IV		qy-IV		q-IV		qy-IV	
	$\widehat{\beta}$	se($\widehat{\beta}$)	$\widehat{\beta}$	se($\widehat{\beta}$)	$\widehat{\lambda}$	se($\widehat{\lambda}$)	$\widehat{\lambda}$	se($\widehat{\lambda}$)
MEAN	0.9573	0.0568	0.9479	0.0525	0.0397	0.0540	0.0484	0.0511
STDEV	0.0593		0.0550		0.0566		0.0540	
MAX	1.1313		1.1002		0.1783		0.1943	
MIN	0.7952		0.8129		-0.1330		-0.1138	
J-test, mean $p =$	0.4313		0.4136					
F-test, mean $p =$	0.0174		0.0017					

$(N_u, N_\nu, N_\eta, N_\xi) = (0,0,1,4)$

	q-IV		qy-IV		q-IV		qy-IV	
	$\widehat{\beta}$	se($\widehat{\beta}$)	$\widehat{\beta}$	se($\widehat{\beta}$)	$\widehat{\lambda}$	se($\widehat{\lambda}$)	$\widehat{\lambda}$	se($\widehat{\lambda}$)
MEAN	0.8309	0.1276	0.8238	0.0990	0.1685	0.1265	0.1736	0.0983
STDEV	0.1422		0.0976		0.1398		0.0968	
MAX	1.3060		1.1155		0.5305		0.3937	
MIN	0.4495		0.5983		-0.2530		-0.1147	
J-test, mean $p =$	0.4850		0.4572					
F-test, mean $p =$	0.0002		0.0000					

$(N_u, N_\nu, N_\eta, N_\xi) = (0,0,2,4)$

	q-IV		qy-IV		q-IV		qy-IV	
	$\widehat{\beta}$	se($\widehat{\beta}$)	$\widehat{\beta}$	se($\widehat{\beta}$)	$\widehat{\lambda}$	se($\widehat{\lambda}$)	$\widehat{\lambda}$	se($\widehat{\lambda}$)
MEAN	0.8417	0.1762	0.8304	0.1180	0.1511	0.1758	0.1631	0.1178
STDEV	0.1873		0.1230		0.1865		0.1240	
MAX	1.3328		1.1165		0.8391		0.4874	
MIN	0.1607		0.5139		-0.3170		-0.1149	
J-test, mean $p =$	0.5393		0.4822					
F-test, mean $p =$	0.0023		0.0001					

B. $\beta = 0.3, \lambda = 0.7, N = 100, T = 10, R = 200$

$(N_u, N_\nu, N_\eta, N_\xi) = (0,0,0,4)$

	q-IV		qy-IV		q-IV		qy-IV	
	$\widehat{\beta}$	se($\widehat{\beta}$)	$\widehat{\beta}$	se($\widehat{\beta}$)	$\widehat{\lambda}$	se($\widehat{\lambda}$)	$\widehat{\lambda}$	se($\widehat{\lambda}$)
MEAN	0.3296	0.0515	0.3189	0.0354	0.6591	0.0651	0.6786	0.0399
STDEV	0.0531		0.0367		0.0664		0.0410	
MAX	0.4797		0.4175		0.8141		0.7724	
MIN	0.2011		0.2378		0.4896		0.5645	
J-test, mean $p =$	0.4838		0.4150					
F-test, mean $p =$	0.0003		1.0000					

$(N_u, N_\nu, N_\eta, N_\xi) = (0,0,1,4)$

	q-IV		qy-IV		q-IV		qy-IV	
	$\widehat{\beta}$	se($\widehat{\beta}$)	$\widehat{\beta}$	se($\widehat{\beta}$)	$\widehat{\lambda}$	se($\widehat{\lambda}$)	$\widehat{\lambda}$	se($\widehat{\lambda}$)
MEAN	0.3306	0.0670	0.3186	0.0395	0.6577	0.0849	0.6784	0.0443
STDEV	0.0712		0.0371		0.0926		0.0418	
MAX	0.6307		0.4513		0.8492		0.7695	
MIN	0.1273		0.2356		0.3219		0.5310	
J-test, mean $p =$	0.4840		0.4136					
F-test, mean $p =$	0.0027		1.0000					

$(N_u, N_\nu, N_\eta, N_\xi) = (0,0,2,4)$

	q-IV		qy-IV		q-IV		qy-IV	
	$\widehat{\beta}$	se($\widehat{\beta}$)	$\widehat{\beta}$	se($\widehat{\beta}$)	$\widehat{\lambda}$	se($\widehat{\lambda}$)	$\widehat{\lambda}$	se($\widehat{\lambda}$)
MEAN	0.3358	0.0936	0.3174	0.0433	0.6508	0.1190	0.6808	0.0484
STDEV	0.0874		0.0458		0.1097		0.0507	
MAX	0.7242		0.5148		0.8851		0.8144	
MIN	0.0968		0.1908		0.2289		0.4640	
J-test, mean $p =$	0.5392		0.4501					
F-test, mean $p =$	0.0139		1.0000					

Table 6: MC simulations of GMM. Equation in differences, IVs in levels

A. $\beta = 1, \lambda = 0, N = 100, T = 10, R = 200$

$(N_u, N_\nu, N_\eta, N_\xi) = (0,0,0,4)$

	q-IV		qy-IV		q-IV		qy-IV	
	$\widehat{\beta}$	se($\widehat{\beta}$)	$\widehat{\beta}$	se($\widehat{\beta}$)	$\widehat{\lambda}$	se($\widehat{\lambda}$)	$\widehat{\lambda}$	se($\widehat{\lambda}$)
MEAN	0.6504	0.1641	0.6708	0.1348	-0.0954	0.0707	-0.0554	0.0441
STDEV	0.1502		0.1267		0.0707		0.0464	
MAX	1.1245		1.0097		0.0665		0.0746	
MIN	0.2265		0.3537		-0.3039		-0.1766	
J-test, mean $p =$	0.4069		0.3811					
F-test, mean $p =$	0.1488		0.9995					

$(N_u, N_\nu, N_\eta, N_\xi) = (0,0,1,4)$

	q-IV		qy-IV		q-IV		qy-IV	
	$\widehat{\beta}$	se($\widehat{\beta}$)	$\widehat{\beta}$	se($\widehat{\beta}$)	$\widehat{\lambda}$	se($\widehat{\lambda}$)	$\widehat{\lambda}$	se($\widehat{\lambda}$)
MEAN	0.5375	0.1863	0.6116	0.1497	-0.3408	0.1628	-0.0498	0.0472
STDEV	0.2007		0.1609		0.1397		0.0500	
MAX	1.1840		1.0376		0.0031		0.0741	
MIN	0.0050		0.1675		-0.7168		-0.2013	
J-test, mean $p =$	0.5028		0.4134					
F-test, mean $p =$	0.0005		0.9999					

$(N_u, N_\nu, N_\eta, N_\xi) = (0,0,2,4)$

	q-IV		qy-IV		q-IV		qy-IV	
	$\widehat{\beta}$	se($\widehat{\beta}$)	$\widehat{\beta}$	se($\widehat{\beta}$)	$\widehat{\lambda}$	se($\widehat{\lambda}$)	$\widehat{\lambda}$	se($\widehat{\lambda}$)
MEAN	0.5349	0.2448	0.6217	0.1729	-0.3390	0.2150	-0.0336	0.0496
STDEV	0.2404		0.1795		0.1890		0.0470	
MAX	1.4225		1.0134		0.1765		0.1003	
MIN	-0.1379		0.0779		-0.8865		-0.1616	
J-test, mean $p =$	0.5161		0.4397					
F-test, mean $p =$	0.0053		1.0000					

B. $\beta = 0.3, \lambda = 0.7, N = 100, T = 10, R = 200$

$(N_u, N_\nu, N_\eta, N_\xi) = (0,0,0,4)$

	q-IV		qy-IV		q-IV		qy-IV	
	$\widehat{\beta}$	se($\widehat{\beta}$)	$\widehat{\beta}$	se($\widehat{\beta}$)	$\widehat{\lambda}$	se($\widehat{\lambda}$)	$\widehat{\lambda}$	se($\widehat{\lambda}$)
MEAN	0.2054	0.1820	0.1853	0.1425	0.6392	0.0298	0.5379	0.0302
STDEV	0.1863		0.1701		0.0270		0.0305	
MAX	0.6243		0.6747		0.7037		0.6245	
MIN	-0.3033		-0.3058		0.5616		0.4132	
J-test, mean $p =$	0.2215		0.0310					
F-test, mean $p =$	1.0000		1.0000					

$(N_u, N_\nu, N_\eta, N_\xi) = (0,0,1,4)$

	q-IV		qy-IV		q-IV		qy-IV	
	$\widehat{\beta}$	se($\widehat{\beta}$)	$\widehat{\beta}$	se($\widehat{\beta}$)	$\widehat{\lambda}$	se($\widehat{\lambda}$)	$\widehat{\lambda}$	se($\widehat{\lambda}$)
MEAN	0.1953	0.2155	0.1605	0.1574	0.6549	0.0300	0.5549	0.0305
STDEV	0.2250		0.1893		0.0282		0.0284	
MAX	0.9139		0.7928		0.7260		0.6126	
MIN	-0.6706		-0.5070		0.5856		0.4759	
J-test, mean $p =$	0.2857		0.0128					
F-test, mean $p =$	1.0000		1.0000					

$(N_u, N_\nu, N_\eta, N_\xi) = (0,0,2,4)$

	q-IV		qy-IV		q-IV		qy-IV	
	$\widehat{\beta}$	se($\widehat{\beta}$)	$\widehat{\beta}$	se($\widehat{\beta}$)	$\widehat{\lambda}$	se($\widehat{\lambda}$)	$\widehat{\lambda}$	se($\widehat{\lambda}$)
MEAN	0.2006	0.2849	0.1777	0.1775	0.6729	0.0302	0.5678	0.0309
STDEV	0.2737		0.1903		0.0315		0.0330	
MAX	0.9096		0.7465		0.7621		0.6345	
MIN	-0.4063		-0.4825		0.5729		0.4644	
J-test, mean $p =$	0.3530		0.0048					
F-test, mean $p =$	1.0000		1.0000					