Efficiency profiles of MM estimators in dynamic panel data models

Bun, M.J.G.; Kiviet, J.F.

Publication date
2002

Citation for published version (APA):
http://www1.feb.uva.nl/pp/bin/435fulltext.pdf

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Efficiency Profiles of MM Estimators
in Dynamic Panel Data Models

Maurice J.G. Bun and Jan F. Kiviet
Tinbergen Institute
&
Department of Quantitative Economics
University of Amsterdam
Roetersstraat 11
1018 WB Amsterdam
The Netherlands

preliminary version: 19 September 2002

Abstract

In dynamic panel data models with unobserved individual effects least-squares estimators are inconsistent when the number of cross-section units $N$ gets large while the number of time-series observations $T$ remains finite. For that situation an abundance of method of moments (MM) estimators is available, which differ in the way unobserved heterogeneity is dealt with and regarding the number and nature of instruments that is being exploited. For some stylized models we derive and compare characteristics concerning instrument weakness (or fitness) and the resulting effectiveness with respect to estimator efficiency for $T$ small and $N$ infinite. We make extensive use of graphical methods to show the characteristic qualities of and differences between estimation methods over relevant areas of the parameter space.

1. Introduction

Our aim is to analyze estimator accuracy and efficiency in dynamic panel data models, where the unobserved disturbance term consists of two random components, viz. a white noise error and a time invariant individual specific effect, whereas the explanatory part of the model contains weakly exogenous regressors. More in particular the basic model, its variables and the two parameters of primary interest $\beta$ and $\gamma$ on which we want to focus, is given by

\[ y_{it} = \alpha + \beta x_{it} + \gamma y_{i,t-1} + \eta_i + \varepsilon_{it}, \]  

(1.1)

where

\[ x_{it} = \delta + \rho x_{i,t-1} + \lambda y_{i,t-1} + \phi \eta_i + \xi_{it}, \]  

(1.2)
with error components with first and second moments
\[ \begin{align*}
    &E(\varepsilon_{it}) = 0, \quad E(\xi_{it}) = 0, \quad E(\eta_{it}) = 0 \\
    &V(\varepsilon_{it}) = \sigma^2_{\varepsilon}, \quad V(\xi_{it}) = \sigma^2_{\xi}, \quad V(\eta_{it}) = \sigma^2_{\eta}
\end{align*} \]
\[i = 1, ..., N; \quad t = 1, ..., T \quad (1.3)\]
and \(\varepsilon_{it}, \xi_{it}\) and \(\eta_{it}\) independent of each other and among themselves. In addition, we shall make the simplifying assumptions that the parameters \((\beta, \gamma, \rho, \lambda)\) and the initial conditions are such that for every \(i = 1, ..., N\) the series \(\{y_{it}; \ t = 0, ..., T\}\) and \(\{x_{it}; \ t = 0, ..., T\}\) are mean and covariance stationary through time, and also that the set of \(2(T + 1)\) observations \(\{y_{it}, x_{it}; t = 0, ..., T\}\) is iid (independent and identically distributed) over all cross-section subjects \(i = 1, ..., N\).

The derivation of (higher-order) asymptotic approximations to the bias of least-squares and particular instrumental variables estimators for this type of model has been undertaken in Kiviet (1995, 1999), and in Bun & Kiviet (2001) the accuracy of bias corrected least-squares estimators is compared with (large \(N\) consistent) method of moments type estimators, focussing on the situation where both \(N\) and \(T\) are small. Such comparisons require simulations, because analytical derivation of actual mean squared errors in finite samples is not feasible. Analytical results on instrument fitness and estimator efficiency can however be obtained for \(N \to \infty\). Here we shall make comparisons between the asymptotic variances for \(N \to \infty\) of various method of moments (MM) estimators for particular members of the class of model given above. From these results guidelines follow on how to cope with unobserved heterogeneity in dynamic models, notably on the effective selection of instrumental variables in order to enhance estimator efficiency in panels where \(T\) is small.

In section 2 we present some general properties of the MM estimators for the panel data models to be analyzed. In section 3 we produce results for the simple panel AR(1) model; in section 4 we add a strongly exogenous explanatory variable to this model. In section 5 we consider the model in its full generality, and finally draw some conclusions in section 6.

2. Estimators for dynamic panel models

Least-squares estimators for panel data models with unobserved heterogeneity and weakly-exogenous regressors are inconsistent for finite \(T\) and \(N\) large. Therefore MM techniques are used. A great number of implementations are available, which differ with respect to:

1. whether the model, or how the model, is transformed in order to get rid of the individual effects \(\eta_{it}\);

2. the number, form and qualities of the employed instrumental variables that are implied by the moment conditions exploited;

3. the way in which attention is paid to any correlation structure of the disturbances of the (transformed) estimation equation.

Before we examine particular MM implementations for our model of interest, we first present some general results on MM estimators for panel regression models where \(T\)
is kept fixed\(^1\) and then specialize these for the case where all variables are stationary through time and are iid for the \(N\) individual units. We start off from the following situation. Let the estimation equation be

\[
y_{it}^* = w_{it} \theta + u_{it}, \quad i = 1, \ldots, N; \quad t = t^*, \ldots, T
\]

where the dependent variable \(y_{it}^*\) can either be \(y_{it}\) or \(\Delta y_{it}\) or another transformation, \(w'_{it}\) and \(\theta\) are \(K \times 1\) vectors and \(u_{it}\) is the disturbance, where

\[
E(u_{it}) = 0, \quad E(u_{i1} u_{j,s}) = 0 \quad i \neq j = 1, \ldots, N, \quad E(u_{it} u_{is}) = \omega_{t-s}, \quad t, s = t^*, \ldots, T. \tag{2.2}
\]

The time-index of the initial dependent variable used in estimation \(t^*\) may exceed 1 due to the "loss" of data when we use first differences, or when no (lagged) instrumental variables are available for the initial observation(s). Let the \(L \times 1\) vector \(z'_{it}\) contain the \(L \geq K\) instruments we employ to estimate equation (2.1). Collecting the data per individual, we can construct for each \(i\) the \(T^* \times 1\) vector \(y_{it}^* = (y_{it}, \ldots, y_{iT})'\), where \(T^* = T - t^* + 1\), and the \(T^* \times K\) matrices \(W_i = (w'_{it}, \ldots, w'_{iT})'\) and the \(T^* \times L\) matrices \(Z_i = (z'_{it}, \ldots, z'_{iT})'\), so that

\[
y_{it}^* = W_i \theta + u_i, \quad i = 1, \ldots, N, \tag{2.3}
\]

where \(E(u_i u_j') = \Omega\), with \(\Omega_{ii} = \omega_{t-i}\), and \(E(u_i u_j') = O\), for \(i \neq j\). For the instruments to be valid, it is required that

\[
E(Z_i' u_i) = 0 \quad (\forall i) \quad \text{and} \quad \text{rank}(\sum_{i=1}^N Z_i' W_i) = K \text{ with probability 1}. \tag{2.4}
\]

In line with the foregoing, we assume that the matrices \(Z_i\) are iid for \(i = 1, \ldots, N\). For reasons that become fully clear later, we shall often have instrumental variables, where \(z_{it}\) is not covariance stationary through time (because a vector time-series \(z_{it}, \ldots. z_{iT}\) will often contain many zero elements).

In case \(L = K\) the ordinary instrumental variables (IV) estimator of \(\beta\) using the full panel is

\[
\hat{\beta}_{\text{IV}} \equiv (\sum_{i=1}^N Z_i' W_i)^{-1} \sum_{i=1}^N Z_i' y_i^* = \theta + (\sum_{i=1}^N Z_i' W_i)^{-1} \sum_{i=1}^N Z_i' u_i, \tag{2.5}
\]

with

\[
\text{plim}_{N \to \infty} \hat{\beta}_{\text{IV}} = \theta + \left(\text{plim}_{N \to \infty} \frac{1}{N} \sum_{i=1}^N Z_i' W_i\right)^{-1} \text{plim}_{N \to \infty} \sum_{i=1}^N \frac{1}{N} Z_i' u_i
\]

\[
= \theta + \Sigma_{ZW}^{-1} \theta = \theta,
\]

where we used (2.4) and defined

\[
\Sigma_{ZW} \equiv \text{plim}_{N \to \infty} \frac{1}{N} \sum_{i=1}^N Z_i' W_i. \tag{2.6}
\]

\(^1\)For a study on estimators for this type of model where both \(T\) and \(N\) may get large, but at different rates, see Alvarez & Arellano (1998).
From
\[ N^{1/2} (\hat{\theta}_V - \theta) = (N^{-1} \sum_{i=1}^{N} Z_i' W_i)^{-1} N^{-1/2} \sum_{i=1}^{N} Z_i' u_i \]
we find, upon defining \( \Sigma_{WZ} \equiv \Sigma_{ZW} \) and \( \Sigma_{Z_0 Z} \), for the asymptotic variance
\[
\text{AV} (\hat{\theta}_V) = \Sigma_{ZW}^{-1} \left[ \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} Z_i' u_i' u_j Z_j \right] \Sigma_{WZ}^{-1}
= \Sigma_{ZW}^{-1} \Sigma_{Z_0 Z} \Sigma_{WZ}^{-1}.
\] (2.7)

Making use of the fact that we assumed that the matrices \( W_i \) and \( Z_i \) are identically distributed over all individuals, we have
\[
\Sigma_{ZW} = E(Z_i' W_i).
\] (2.8)

Employing the predeterminedness of the instruments and the iid property of the individual units, we obtain
\[
\Sigma_{Z_0 Z} \equiv \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} Z_i' u_i' u_j Z_j
= \sum_{t=1}^{T} \sum_{s=1}^{T} \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} Z_i' u_i' u_j Z_j
= \sum_{t=1}^{T} \sum_{s=1}^{T} \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} u_i u_i' u_j u_j Z_i Z_j
= \sum_{t=1}^{T} \sum_{s=1}^{T} \omega \sum_{i=1}^{N} \omega \sum_{t=1}^{T} \omega \sum_{s=1}^{T} \omega \sum_{i=1}^{N} \omega \sum_{i=1}^{N} Z_i Z_i' Z_i Z_i' + Z_i Z_i' Z_i Z_i' Z_i Z_i'.
\] (2.9)

Substitution in (2.7) yields for the \( L = K \) case
\[
\text{AV} (\hat{\theta}_V) = \left[ E(Z_i' W_i) \right]^{-1} \left[ \omega \sum_{i=1}^{N} \omega \sum_{t=1}^{T} \omega \sum_{s=1}^{T} \omega \sum_{i=1}^{N} \omega \sum_{i=1}^{N} Z_i Z_i' Z_i Z_i' + Z_i Z_i' Z_i Z_i' Z_i Z_i' \right] \left[ E(W_i' Z_i) \right]^{-1}.
\] (2.10)

In case \( L \geq K \) the generalized instrumental variables estimator GIV of \( \beta \) is
\[
\hat{\beta}_{\text{GIV}} \equiv [\sum_{i=1}^{N} W_i Z_i (\sum_{i=1}^{N} Z_i' Z_i)^{-1} \sum_{i=1}^{N} Z_i' W_i]^{-1} \sum_{i=1}^{N} W_i Z_i (\sum_{i=1}^{N} Z_i' Z_i)^{-1} \sum_{i=1}^{N} Z_i' y_i
= [W' Z (Z' Z)^{-1} Z' W]^{-1} W' Z (Z' Z)^{-1} Z' y_i.
\] (2.11)

In the latter expression we stacked all data by defining the \( NT^* \times K \) matrix \( W = (W_1', ..., W_K') \), and in the same way the \( NT^* \times L \) matrix \( Z \) and the \( NT^* \times 1 \) vector \( y^* \). Of course, \( \lim_{N \to \infty} \hat{\beta}_{\text{GIV}} = \theta \), and defining
\[
\Sigma_{ZZ} \equiv \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} Z_i' Z_i
\] (2.12)

it follows that for the general case
\[
\text{AV} (\hat{\beta}_{\text{GIV}}) = (\Sigma_{WZ} \Sigma_{ZZ}^{-1} \Sigma_{ZW}^{-1})^{-1} \Sigma_{WZ} \Sigma_{ZZ}^{-1} \Sigma_{Z_0 Z} \Sigma_{ZZ}^{-1} \Sigma_{ZW} (\Sigma_{WZ} \Sigma_{ZZ}^{-1} \Sigma_{ZW})^{-1}.
\] (2.13)

This may be simplified, like (2.10), under the stationarity and iid assumptions by substituting (2.8), (2.9) and \( \Sigma_{ZZ} = E(Z_i' Z_i) \).
In case the variance matrix of $u$ is non-scalar and known the efficiency can be improved by employing the generalized method of moments estimator GMM. With $E(u'\varepsilon) = I_N \otimes \Omega$, where $\Omega_{i\cdot} = \omega_{i\cdot i}$, this equals

$$\hat{\theta}_{GMM} = \{W'Z[Z'(I_N \otimes \Omega)Z]^{-1}Z'W\}^{-1}W'Z[Z'(I_N \otimes \Omega)Z]^{-1}Z'y^*, \quad (2.14)$$

with

$$AV(\hat{\theta}_{GMM}) = (\Sigma_{WZ}^{-1}\Sigma_{Z\Omega}^{-1}\Sigma_{ZW})^{-1}, \quad (2.15)$$

which, in the present context, again can be simplified using (2.8) and (2.9). Usually $\Omega$ is unknown and only a two-stage GMM estimator is feasible, which employs an estimator of $\Omega$ obtained from GIV estimation results. This has an asymptotic variance equivalent with (2.15). If $Z$ contains all instruments that follow from the adopted moment conditions then the resulting estimator is most efficient in that class and is called the efficient GMM estimator.

### 3. The simple panel AR(1) model

We start by examining a very special simple case of model (1.1), viz. the stationary zero-mean panel AR(1) model

$$y_{it} = \gamma y_{i,t-1} + \eta_t + \varepsilon_{it} \quad |\gamma| < 1. \quad (3.1)$$

Special aspects of least-squares (with dummy variables) estimation of this model have been investigated in Nickell (1981); a recent study on various aspects of MM estimators in this model can be found in Blundell & Bond (1998). We assume, like Nickell (but unlike Blundell & Bond, who take $y_{i1}$ as the first actual observation on $y_{it}$), that data for $y_{it}$ are available for $i = 1, ..., N$ and $t = 0, ..., T$. Hence, $y_{i0}$ are start-up values, and for estimation of the parameters of the untransformed equation (3.1) in fact $NT$ observations on the dependent variable are available. To achieve variance stationarity we could (generalizing an approach not uncommon in time-series AR(1) models) assume

$$y_{i0} = \frac{1}{1 - \gamma} \eta_0 + \frac{1}{\sqrt{1 - \gamma^2}} \varepsilon_{i0}, \quad (3.2)$$

since this yields $V(y_{it}) = V(y_{i0})$ for $t > 0$. However, note that the covariance stationarity implies inter alia $E(y_{it}\varepsilon_{it-1}) = \gamma \sigma^2$, but (3.2) would imply $E(y_{i1}\varepsilon_{i0}) = \gamma (1 - \gamma^2)^{-1/2} \sigma^2$. Hence, to have correct values for all the cross-covariances relevant for the various MM estimators, we shall always assume that data and disturbances have been generated since the infinite past.

Over a dozen different MM implementations are available for estimating the single parameter $\gamma$, where the instruments are sets of lagged or differenced $y_{it}$ series and the estimation equation is either in levels or in first differences, or forms a system of both. Since all these estimators are determined by the parameters $\gamma$, $\sigma^2$ and $\sigma^2$ only, whereas $\sigma$ is in fact a simple scaling factor, the asymptotic variances of all these estimators are

\footnote{For system estimation see Blundell, Bond & Windmeijer (2000).}
functions of just three characteristics, viz. $\gamma$, $\sigma_\eta^2/\sigma_\varepsilon^2$ and $T$. We embark on assessing these functions to make further comparisons.

We examine the (auto-)covariances of the series involved. First we find

$$
E(y_{t\eta}\eta_t) = \gamma E(y_{t-1}\eta_t) + \sigma_\eta^2 = \frac{1}{1-\gamma}\sigma_\eta^2,
$$

(3.3)
due to the assumed stationarity, and then the same type of reasoning yields

$$
V(y_{t\varepsilon}) = \gamma^2 V(y_{t-1}) + 2\gamma E(y_{t-1}\eta_t) + \sigma_\eta^2 + \sigma_\varepsilon^2
$$

(3.4)

$$
= \gamma^2 V(y_{t\varepsilon}) + \frac{1+\gamma}{1-\gamma}\sigma_\eta^2 + \sigma_\varepsilon^2
$$

$$
= \sigma_\eta^2 \left(1 + \frac{1}{1-\gamma}\frac{\sigma_\eta^2}{1-\gamma}\sigma_\varepsilon^2\right).
$$

Hence, like the disturbance $\eta_t + \varepsilon_t$, the series $y_{t\varepsilon}$ has two variance components. The ratio between these two components can be seen as an important characterization of the data generating process. Hence, it seems sensible to parametrize it. We define

$$
\mu^2 = \frac{\sigma_\eta^2 (1-\gamma)^2}{\sigma_\varepsilon^2 (1-\gamma)} = \frac{1+\gamma}{1-\gamma}\sigma_\eta^2
$$

(3.5)

which enables to substitute

$$
\sigma_\eta^2 = \frac{1-\gamma}{1+\gamma}\mu^2 \sigma_\varepsilon^2.
$$

(3.6)

This will help to simplify many formulas to be derived below. The results (3.4) and (3.3) can now be rewritten as

$$
V(y_{t\varepsilon}) = \frac{1+\mu^2}{1-\gamma^2}\sigma_\varepsilon^2
$$

(3.7)

and

$$
E(y_{t\eta}\eta_t) = \frac{\mu^2}{1+\gamma}\sigma_\varepsilon^2.
$$

(3.8)

For integer $l \neq 0$ we obtain

$$
E(y_{t\eta} y_{t-l\varepsilon}) = \gamma E(y_{t-l\eta} y_{t-l\varepsilon}) + E(\eta_t y_{t-l\varepsilon})
$$

$$
= \gamma E(y_{t-l\eta} y_{t-l\varepsilon}) + \frac{\mu^2}{1+\gamma}\sigma_\varepsilon^2.
$$

Hence,

$$
E(y_{t\eta} y_{t-l\varepsilon}) = \frac{\gamma + \mu^2}{1-\gamma^2}\sigma_\varepsilon^2 + \frac{\mu^2}{1-\gamma^2}\sigma_\varepsilon^2 = \frac{\gamma + \mu^2}{1-\gamma^2}\sigma_\varepsilon^2
$$

and from this it is easily established that for any integer $l$

$$
E(y_{t\eta} y_{t-l\varepsilon}) = \frac{\gamma^{|l|} + \mu^2}{1-\gamma^2}\sigma_\varepsilon^2.
$$

(3.9)

\footnote{Note that Kiviet (1995) and Blundell and Bond (1998) choose a different relationship, viz. $\sigma_\eta^2 = (1-\gamma)^2 \mu^2 \sigma_\varepsilon^2$, where, as we understand now, $\mu^2$ has a less clearcut interpretation.}
For the estimators to be investigated we need some further results, which can be obtained now in a straightforward fashion. We mention
\[
E(y_{it} \Delta y_{i,t-1}) = \sigma_e^2 \gamma \frac{1}{1+\gamma} \\
V(\Delta y_{it}) = \sigma_e^2 \gamma \frac{1}{1+\gamma} \\
E(y_{it} \Delta y_{i,t-1} \Delta y_{it}) = -\sigma_e^2 \gamma \frac{1}{1+\gamma} \\
E(\Delta y_{it} \Delta y_{i,t-1} \Delta y_{it}) = -\sigma_e^2 \gamma \frac{1}{1+\gamma},
\]
(3.10)

Now we shall obtain expressions for asymptotic moments of particular MM estimators in this model. For curiosity we first picture the inconsistency of the LSDV (least-squares dummy variables) or Within Groups estimator. This has been obtained for this model by Nickell (1981) and can be expressed as
\[
\text{plim}_{N \to \infty} (\hat{\gamma}_{LSDV} - \gamma) = -\frac{1 + \gamma}{T} \frac{1 - \frac{1 - \gamma T}{1 - \gamma}}{1 - \frac{1 + \gamma}{1 - \gamma} + \frac{2 \gamma}{(1 - \gamma)^2}};
\]
(3.11)

which is visualized in Figure 3.1. Note that this inconsistency is invariant with respect

\[\text{Figure 3.1: LSDV asymptotic bias in panel AR(1) model}\]

\[\text{Figure 3.2: LSDV relative asymptotic bias (%) in stationary panel AR(1) model}\]
to $\mu^2$ (the relative magnitude of the individual effects) and that it is very small for $\gamma$ close to $-1$ and/or $T$ not too small, but substantially negative for $T$ a one digit number especially when $\gamma$ is positive, i.e. in the range of most practical interest. We examine that specific region more closely in Figure 3.2 which presents the relative asymptotic bias, i.e. $\lim_{N \to \infty} (\hat{\gamma}_{LSDV} - \gamma) / \gamma$. Self-evidently, the relative asymptotic bias is enormous for $\gamma$ close to zero; it is also shown that biases of -50% or worse occur for any $0 < \gamma \leq 0.99$ at $T \leq 5$. Hence, it seems most appropriate to turn to MM estimation techniques which avoid such inconsistencies because these are (large $N$) consistent.

### 3.1. IV estimation

First we consider three simple IV estimators which use just one instrument for estimating the single parameter $\gamma$. These are denoted by IVld, IVdl and IVdd respectively (where "l" refers to levels and "d" to first differences; the first letter indicates the form of the model estimated (levels or differences) and the next one the form of the instrument. The three estimators are defined as

\[
\hat{\gamma}_{\text{IVld}} = \left( \sum_{t=1}^{N} \sum_{j=1}^{T} \Delta y_{j,t-1} y_{j,t-1} \right)^{-1} \left( \sum_{t=1}^{N} \sum_{j=1}^{T} \Delta y_{j,t-1} \Delta y_{j,t} \right) \tag{3.12}
\]

\[
\hat{\gamma}_{\text{IVdl}} = \left( \sum_{t=1}^{N} \sum_{j=1}^{T} \Delta y_{j,t-1} \Delta y_{j,t-1} \right)^{-1} \left( \sum_{t=1}^{N} \sum_{j=1}^{T} \Delta y_{j,t-1} \Delta y_{j,t} \right) \tag{3.12}
\]

\[
\hat{\gamma}_{\text{IVdd}} = \left( \sum_{t=1}^{N} \sum_{j=1}^{T} \Delta y_{j,t-2} \Delta y_{j,t-2} \right)^{-1} \left( \sum_{t=1}^{N} \sum_{j=1}^{T} \Delta y_{j,t-2} \Delta y_{j,t} \right) \tag{3.12}
\]

It is well known that these are consistent (for $N \to \infty$), i.e. the instruments are valid. IVld has been examined before (in a model which contains exogenous explanatory variables too) in Arellano & Bover (1995) and in Kiviet (1995); IVd and IVdd have been proposed by Anderson & Hsiao (1982). To examine the quality of the instruments we derive the plim of the square of the correlation coefficient between regressor and instrument, which we will call the IQ (instrument quality) coefficient. We obtain

\[
\text{IQ}_{\text{IVld}} = \lim_{N \to \infty} R^2_{y_{j,t-1}, \Delta y_{j,t-1}} = \left( \frac{\sigma^2_{\epsilon}}{\sigma^2_{\epsilon}} \right)^2 = \frac{1 - \gamma}{2\mu^2} \tag{3.13}
\]

\[
\text{IQ}_{\text{IVdl}} = \lim_{N \to \infty} R^2_{\Delta y_{j,t-1}, \Delta y_{j,t-2}} = \left( \frac{-\sigma^2_{\epsilon}}{\sigma^2_{\epsilon}} \right)^2 = \frac{1 - \gamma}{2\mu^2} \tag{3.14}
\]

\[
\text{IQ}_{\text{IVdd}} = \lim_{N \to \infty} R^2_{\Delta y_{j,t-2}, \Delta y_{j,t-2}} = \left( \frac{-\sigma^2_{\epsilon}}{\sigma^2_{\epsilon}} \right)^2 = \frac{1}{4} (1 - \gamma)^2 \tag{3.15}
\]

In Figure 3.3 we examine this IQ for IVld and IVdl, which are similar in this respect\(^4\), and for IVdd, which is invariant with respect to $\mu$. We note that for both IVld and IVdl their instrument gets weaker for larger $\gamma$ and for larger $\mu$. For $\gamma$ close to one and/or $\mu > 2$ the IQ is very small. Note that their instrument weakness is more serious than for IVdd when $\mu$ is moderate or large. In fact,

\[
\text{IQ}_{\text{IVdd}} > \text{IQ}_{\text{IVdl}} = \text{IQ}_{\text{IVld}} \quad \text{if } \mu^2 > \frac{\gamma}{1 - \gamma}, \text{ i.e. } \sigma^2_{\eta} > \sigma^2_{\epsilon}, \tag{3.16}
\]

\(^4\)Note that Blundell & Bond (1998) seem to conclude that for IVdl the instrument is much weaker than for IVld. We do not find support for that.
which is remarkable, because among practitioners IVdd is believed to be inferior to IVdl, generally speaking. We shall see, however, that indeed this squared correlation is not the only determining factor of the effectiveness of an IV estimator.

We proceed and examine the estimator efficiency, i.e. the asymptotic variance (AV) for \( N \to \infty \), of the estimators \( \hat{\gamma}_{IVdl} \), \( \hat{\gamma}_{IVdl} \) and \( \hat{\gamma}_{IVdd} \). Using (2.7), and noting that the estimation equation for \( \hat{\gamma}_{IVdl} \) consists of only \( (T - 1)N \) observations (i.e. \( t^* = 2 \)), whereas in the levels equation \( \omega_0 = \sigma_y^2 + \sigma_\epsilon^2 \) and \( \omega_l = \sigma_y^2 \) for \( l \geq 1 \), we find

\[
\text{AV}(\hat{\gamma}_{IVdl}) = \left[ \sum_{t=2}^{T} \text{E}(\Delta y_{l,t-1} y_{l,t-1}) \right]^{-2} \sum_{t=2}^{T} \sum_{s=2}^{T} \text{E}[\Delta y_{l,t-1} (\eta_k + \varepsilon_d) \Delta y_{l,s-1} (\eta_k + \varepsilon_s)]
\]

\[
= \left( T - 1 \right) \frac{\sigma_\epsilon^2}{1 + \gamma} \left[ (T - 1) \frac{2\omega_0 \sigma_\epsilon^2}{1 + \gamma} - \sum_{s=1}^{T-2} 2(T - s - 1) \omega_s \sigma_y^2 \gamma^{s-1} \frac{1 - \gamma}{1 + \gamma} \right]
\]

\[
= \left( T - 1 \right) \frac{\sigma_\epsilon^2}{1 + \gamma} \left[ \frac{\sigma_\epsilon^4}{1 + \gamma} \frac{2}{1 + \gamma} + \frac{\sigma_y^2 \sigma_\epsilon^2}{1 - \gamma^2} \right]
\]

\[= \frac{2(1 + \gamma)}{T - 1} \left( 1 + \frac{\mu^2}{T - 1} \frac{1 - \gamma^{T-1}}{1 + \gamma} \right). \tag{3.17} \]

Hence, \( \text{AV}(\hat{\gamma}_{IVdl}) \) decreases with \( T \) and increases with \( \mu^2 \) (but at a slower rate in terms of \( T \)) and increases with \( \gamma \) (if \( \mu^2 < 1/\gamma^{T-2} \)).

For \( \hat{\gamma}_{IVdl} \) again \( (T - 1)N \) observations are available for estimation (because \( y_{k,-1} \) is not available) and now \( \omega_0 = 2\sigma_\epsilon^2 \), \( \omega_1 = -\sigma_\epsilon^2 \) and \( \omega_l = 0 \) for \( l \geq 2 \). We obtain

\[
\text{AV}(\hat{\gamma}_{IVdl}) = \left[ \sum_{t=2}^{T} \text{E}(\Delta y_{l,t-2} \Delta y_{l,t-1}) \right]^{-2} \sum_{t=2}^{T} \sum_{s=2}^{T} \text{E}(y_{l,t-2} \Delta \varepsilon_d y_{l,s-2} \Delta \varepsilon_s)
\]

\[
= \left[ - \sum_{t=2}^{T} \frac{\sigma_\epsilon^2}{1 + \gamma} \right]^{-2} \left[ (T - 1) \omega_0 \sigma_\epsilon^2 \frac{1 + \mu^2}{1 - \gamma^2} - 2(T - 2) \omega_1 \sigma_\epsilon^2 \frac{\gamma + \mu^2}{1 - \gamma^2} \right]
\]
\begin{align*}
&= \left[ (T-1) \frac{\sigma^2}{1+\gamma} \right]^{-2} \left[ (T-1) 2\sigma^4 \frac{1-\gamma}{1-\gamma^2} + 2\sigma^2 \frac{\gamma + \mu^2}{1-\gamma^2} \right] \\
&= \frac{2 (1+\gamma)}{T-1} \left( 1 + \frac{1}{T-1} \frac{\gamma + \mu^2}{1-\gamma} \right).
\end{align*}

(3.18)

Hence, for $\text{AV}(\hat{\gamma}_{\text{IVdl}})$ we find a slightly different expression than for $\text{AV}(\hat{\gamma}_{\text{IVdl}})$, although their instruments showed similar weakness, but the same tendencies emerge: it decreases with $T$ and increases with $\mu^2$ and $\gamma$ (except for rather pathological cases).

Self-evidently, $\text{AV}(\hat{\gamma}_{\text{IVdl}})$ will not depend on $\mu^2$. Now only $(T-2)N$ observations are available ($t^* = 3$). The auto-covariances $\omega_t$ of the disturbances of the estimation equation are as for $\hat{\gamma}_{\text{IVdl}}$. We find

\begin{align*}
\text{AV}(\hat{\gamma}_{\text{IVdl}}) &= \left[ \sum_{t=3}^{T} \mathbb{E}(\Delta y_{t-2} \Delta y_{t-1}) \right]^{-2} \sum_{t=3}^{T} \sum_{s=3}^{T} \mathbb{E}(\Delta y_{t-2} \Delta \varepsilon_{it} \Delta y_{s-2} \Delta \varepsilon_{is}) \\
&= \left[ -\sum_{t=3}^{T} \sigma^2 \frac{1-\gamma}{1+\gamma} \right]^{-2} \left[ (T-2) \frac{2\omega_0 \sigma^2}{1+\gamma} - 2 (T-3) \omega_1 \sigma^2 \frac{1-\gamma}{1+\gamma} \right] \\
&= \left[ (T-2) \sigma^2 \frac{1-\gamma}{1+\gamma} \right]^{-2} \left[ (T-2) \frac{4\sigma^4}{1+\gamma} + 2 (T-3) \sigma^2 \frac{1-\gamma}{1+\gamma} \right] \\
&= \frac{2}{T-2} \frac{1+\gamma}{1-\gamma} \left( \frac{3-\gamma}{1+\gamma} - \frac{1}{T-2} \right),
\end{align*}

(3.19)

so like for the other two we find that efficiency generally increases with $T$ and decreases with $\gamma$. Note that for $T$ large IVdd is less efficient than the other two, which use level variables in either the equation or the instrument, whereas the efficiencies of IVld and IVdl are equivalent asymptotically. For $\gamma \uparrow 1$ $\text{AV}(\hat{\gamma}_{\text{IVdl}})$ tends to $4/(T-1)$, whereas the asymptotic variance of either of the IV estimators of the differenced equation is unbounded.

Examining the efficiency ratios, we find

\begin{equation}
\frac{\text{AV}(\hat{\gamma}_{\text{IVdl}})}{\text{AV}(\hat{\gamma}_{\text{IVdl}})} = \frac{1 + \frac{\mu^2}{T-1} \frac{1-\gamma}{1+\gamma} T^{-1}}{1 + \frac{\mu^2}{T-1} \frac{1-\gamma}{1+\gamma} T^{-1}}.
\end{equation}

(3.20)

We make comparisons for two particular small values of $T$. For the minimal number of required observations ($T = 2$) the ratio is

$$
\frac{1 + \frac{\mu^2}{T-1} \frac{1-\gamma}{1+\gamma} T^{-1}}{1 + \frac{\mu^2}{T-1} \frac{1-\gamma}{1+\gamma} T^{-1}} \bigg|_{T=2} = \frac{1 - \gamma^2 + (1-\gamma)^2 \mu^2}{(1+\gamma)(1+\mu^2)},
$$

giving Figure 3.4, which shows that the effect of $\mu$ is only marginal. We find that only for negative $\gamma$ the value of $\text{AV}(\hat{\gamma}_{\text{IVdl}})$ will be larger than $\text{AV}(\hat{\gamma}_{\text{IVdl}})$. The efficiency of IVdl can be very much worse than that of IVld for positive $\gamma$ and is already appalling when $\gamma > 0.5$ irrespective of the value of $\mu$. A similar, but slightly less dramatic picture emerges for a larger value of $T$. For $T = 10$ we obtain Figure 3.5, which shows that the estimator $\hat{\gamma}_{\text{IVdl}}$ is again more efficient for $\gamma > 0$. 10
Next we compare the efficiency of IVdd with that of IVld. We find

$$\frac{\text{AV} (\hat{\gamma}_{\text{IVdd}})}{\text{AV} (\hat{\gamma}_{\text{IVld}})} = \frac{T - 2 \left[ 1 + \frac{\mu^2}{T-1} \frac{1-\gamma^{-1}}{1+\gamma} \right]}{T - 1 \left[ \frac{3-\gamma}{1-\gamma^2} - \frac{1}{T-2} \right]}$$

(3.21)

Hence, for large $T$ the ratio is $(1 - \gamma)^2 / (3 - \gamma) < 1$ and $\text{AV} (\hat{\gamma}_{\text{IVdd}})$ will be more efficient. The IVdd technique requires at least 3 observations (apart from the start-up) and then the ratio is

$$\frac{T - 2 \left[ 1 + \frac{\mu^2}{T-1} \frac{1-\gamma^{-1}}{1+\gamma} \right]}{T - 1 \left[ \frac{3-\gamma}{1-\gamma^2} - \frac{1}{T-2} \right]} \bigg|_{T=3} = \frac{(1 - \gamma)^2}{8} \left[ 2 + \mu^2 (1 - \gamma) \right],$$

and from Figure 3.6 we immediately see that IVld is then less efficient for small $\gamma$ and large $\mu$. For $T = 10$ we see from Figure 3.7 that for $\mu$ really large there remains some scope for IVdd, but for larger $T$ and large $\gamma$ IVdd appears inferior.

From the above analysis we conclude that when a large $N$ consistent IV estimator is used in the stationary panel AR(1) model, i.e. one sticks to employing just one in-
Figure 3.6: $\text{AV}(\hat{\gamma}_{IVld})/\text{AV}(\hat{\gamma}_{IVdd})$ for $T = 3$ in panel AR(1)

Figure 3.7: $\text{AV}(\hat{\gamma}_{IVld})/\text{AV}(\hat{\gamma}_{IVdd})$ for $T = 10$ in panel AR(1)

strument, then IVld seems to be preferable from a large $N$ asymptotic efficiency point of view when $\gamma$ is positive, especially so when $\mu$ has a moderate or small value. This is only partly explained by the IQ coefficient of weakness of the instrument; the signal-to-noise ratio in the estimating equation (which changes considerably after taking first differences) seems to be another.

3.2. GIV and GMM estimation

Next we examine the effect of adding an extra lagged instrument to the three settings just examined (at the expense of loosing another observation per individual in the estimation equation). This leads to three GIV methods with one degree of overidentification. We
label them GIVld1, GIVdl1 and GIVdd1 respectively. For GIVld1 we have

\[ \Sigma_{ZW} = (T - 2) \frac{\sigma^2}{2 + \gamma} \begin{pmatrix} 1 & \gamma \\ \gamma & \frac{1}{\gamma} \end{pmatrix}, \]

\[ \Sigma_{ZZ} = (T - 2) \frac{\sigma^2}{2 + \gamma} \begin{pmatrix} 2 & -(1 - \gamma) \\ -(1 - \gamma) & 2 \end{pmatrix}, \tag{3.22} \]

\[ \Sigma_{WW} = (T - 2) \sigma^2 \frac{1 + \mu^2}{2 + \gamma}. \]

From these we can establish the quality of the two instruments for the single regressor. We define the relevant IQ as the plim of the \( R^2 \) of the regression of the single column of \( W \) on \( Z \), i.e. \( \Sigma_{WZ} \Sigma_{ZZ}^{-1} \Sigma_{ZW} / \Sigma_{WW} \), and find

\[ \text{IQ}_{GIVdl1} = \frac{2(1 - \gamma)}{(1 + \mu^2)(3 - \gamma)} = \frac{4}{3 - \gamma} \text{ IQ}_{IVd1} > \text{ IQ}_{IVdl1}. \tag{3.23} \]

Hence, for \( \gamma \downarrow -1 \) there is no effect of the extra instrument on \( R^2 \) and for \( \gamma \uparrow 1 \) the maximum effect is achieved which tends to double the \( R^2 \) value, but note that as \( \gamma \uparrow 1 \) this converges to zero anyhow. Hence, the effects do not seem marvelous as is seen from Figure 3.8.

![Figure 3.8: IQ_{GIVdl1} in the panel AR(1) model](image)

For the asymptotic variance we have to assess \( \Sigma_{Z \in Z} \). Note that

\[ E(z_{11}' z_{11}) = \frac{\sigma^2}{2 + \gamma} \begin{pmatrix} 2 & -(1 - \gamma) \\ -(1 - \gamma) & 2 \end{pmatrix}, \]

\[ E(z_{11}' z_{12}) = -\sigma^2 \frac{(1 - \gamma)}{2 + \gamma} \begin{pmatrix} 1 & -\gamma \\ \gamma & 1 \end{pmatrix}, \]

\[ E(z_{11}' z_{1,1+s}) = -\sigma^2 \frac{(1 - \gamma)}{2 + \gamma} \begin{pmatrix} \gamma^{s-1} & \gamma^{s-2} \\ \gamma^{s} & \gamma^{s-1} \end{pmatrix}. \]
for $s = 2, ..., T - 3$. From these we find

$$
\Sigma_{Z_{\cdot}Z} = (T - 2) \frac{\sigma^4}{1 + \gamma} \begin{pmatrix}
  -2 & - (1 - \gamma) \\
  2 & 2 - (1 - \gamma) \gamma T^{-2}
\end{pmatrix} + \frac{\sigma^4 \mu^2}{(1 + \gamma)^2} \begin{pmatrix}
  2 - 2 \gamma T^{-2} & 2 + (1 - \gamma^2) \gamma T^{-3} \\
  2 + (1 - \gamma^2) \gamma T^{-3} & 2 - 2 \gamma T^{-2}
\end{pmatrix}
$$

and now we can obtain the asymptotic variance of GIVld1 and GMMld1 by evaluating (2.13) and (2.15). We find

$$
\text{AV}(\hat{\gamma}_{\text{GIVld1}}) = \\
\text{AV}(\hat{\gamma}_{\text{GMMld1}}) =
$$

For GIVld1 we have

$$
\Sigma_{ZW} = (T - 2) \frac{\sigma^2}{1 + \gamma} \begin{pmatrix}
  -1 \\
  -\gamma
\end{pmatrix}
$$

$$
\Sigma_{ZZ} = (T - 2) \frac{\sigma^2}{1 + \gamma^2} \begin{pmatrix}
  1 + \mu^2 & \gamma + \mu^2 \\
  \gamma + \mu^2 & 1 + \mu^2
\end{pmatrix}
$$

(3.24)

$$
\Sigma_{WW} = (T - 2) \frac{2 \sigma^2}{1 + \gamma}.
$$

From these we can establish

$$
\text{IQ}_{\text{GIVld1}} = \frac{1 - \gamma + \gamma + \mu^2 (1 - \gamma)}{2} \frac{1 + \gamma + 2 \mu^2}{1 + \gamma + 2 \mu^2} \text{IQ}_{\text{GIVld1}} > \text{IQ}_{\text{GIVld1}}.
$$

(3.25)

For $\gamma \rightarrow 1$ and for $\mu = 0$ the effect of the extra instrument on $R^2$ is nil. Plotting in Figure 3.9 both $\text{IQ}_{\text{GIVld1}}$ and $\text{IQ}_{\text{GIVld1}}$ we find that the latter is often much larger, whereas earlier we found $\text{IQ}_{\text{IVld1}} \equiv \text{IQ}_{\text{IVld1}}$. We also find

$$
\text{AV}(\hat{\gamma}_{\text{GIVld1}}) = \\
\text{AV}(\hat{\gamma}_{\text{GMMld1}}) =
$$

Moving now to $\text{IQ}_{\text{GIVld1}}$ we obtain

$$
\Sigma_{ZW} = (T - 3) \frac{\sigma^2}{1 + \gamma} \begin{pmatrix}
  1 \\
  \gamma
\end{pmatrix},
$$

$$
\Sigma_{ZZ} = (T - 3) \frac{\sigma^2}{1 + \gamma} \begin{pmatrix}
  2 & - (1 - \gamma) \\
  - (1 - \gamma) & 2
\end{pmatrix},
$$

(3.26)

$$
\Sigma_{WW} = (T - 3) \frac{2 \sigma^2}{1 + \gamma}.
$$
Figure 3.9: IQ_{GIVdd1} and IQ_{GIVd1} (again) in panel AR(1)

yielding

\[ IQ_{GIVdd1} = \frac{1}{4} (1 - \gamma)^2 \frac{1 + \gamma}{1 - \frac{1}{4} (1 - \gamma)^2} > IQ_{GIVd1}, \] (3.27)

which is again invariant with respect to \( \mu^2 \) and multiplies IQ_{GIVd1}, which wasn’t all that bad, by a factor running from 1 (for \( \gamma \to -1 \)) to 2 (for \( \gamma \to 1 \)). We compare it with IQ_{GIVdd1} by examining the ratio IQ_{GIVd1} / IQ_{GIVdd1} in Figure 3.10 and see that IQ_{GIVdd1}

Figure 3.10: IQ_{GIVd1} / IQ_{GIVdd1} in panel AR(1)

is superior almost everywhere. Comparing it in the same way with IQ_{GIVd1} we find in Figure 3.11 an even more striking difference in IQ. In this respect GIVd1 is only inferior for \( \gamma \) very close to one and \( \mu \) close to zero.

That the IQ values show the relative performance just established can be explained as follows. Fitting \( \Delta y_{t-1} \) by \( \Delta y_{t-2} \) and \( \Delta y_{t-3} \) (GIVd1) is of course more successful than fitting \( \Delta y_{t-1} \) by the levels \( y_{t-2} \) and \( y_{t-3} \) (GIVd1), and the fit of \( y_{t-1} \) by the differences \( \Delta y_{t-1} \) and \( \Delta y_{t-2} \) (GIVd1) will naturally be poor.
We also find

\[ \text{AV}(\hat{\gamma}_{\text{GMMd1l}}) = \]
\[ \text{AV}(\hat{\gamma}_{\text{GMMd1l}}) = \]

4. The panel ARX(1) model with a strongly exogenous regressor

To examine and compare the asymptotic variance of particular MM estimators in more general models we need to evaluate the appropriate matrices \( \Sigma_{ZW} \), \( \Sigma_{ZZ} \) and \( \Sigma_{Z\omega} \). For model (1.1) these are the variance, covariance and auto-covariances of the random variables \( y_{it} \) and \( x_{it} \) and of their lags or (lags of) their first-differences. To derive these all manually (i.e. mentally) is very involved. We have tried to use the formula manipulation options of SWP\(^5\) to obtain these. However, SWP cannot simplify the other elements of this vector without mental help. The lengthy expressions provided are pretty useless for construction of the relevant \( \Sigma_{ZW} \), \( \Sigma_{WX} \) and \( \Sigma_{Z\omega} \) matrices and inverting them. Therefore we resume, as far as feasible, mental derivations. In that way we obtain for the first element \( E y_{it}^2 \) the ‘factorized’ expression

\[
V(y_{it}) = \frac{(1 + \gamma \rho) \beta^2}{(1 - \gamma^2)(1 - \rho^2)(1 - \gamma \rho)} \sigma^2 \xi + \frac{1}{(1 - \gamma)^2} \sigma^2 \eta + \frac{1}{1 - \gamma^2} \sigma^2 \varepsilon.
\]

Evidently, \( y_{it} \) consists now of three variance components, and again we may simplify some of the expressions by parametrizing the ratio between the variance component stemming from the exogenous regressor and that from the white noise disturbance. We define

\[
\kappa^2 = \frac{(1 + \gamma \rho) \beta^2}{1 - \gamma^2 \sigma^2 \varepsilon} = \frac{(1 + \gamma \rho) \beta^2 \sigma^2 \xi}{(1 - \rho^2)(1 - \gamma \rho) \sigma^2 \varepsilon},
\]

which yields

\[
V(y_{it}) = \frac{1}{1 - \gamma^2} \left( \kappa^2 + \mu^2 + 1 \right) \sigma^2 \varepsilon.
\]

\(^5\)Scientific Work Place 3.50, www.mackichan.com
From the basic model assumptions one can also obtain the following auto-covariances:

\[
\begin{align*}
E(y_{a,t} y_{i,t-1}) &= \frac{1}{1 + \gamma} \left( \frac{\gamma}{1 + \rho \kappa^2} + \mu^2 + \gamma \right) \sigma^2 \\
E(x_{a,t} x_{i,t-1}) &= \frac{\beta \rho^2 \sigma^2}{1 - \rho^2} \xi \\
E(y_{a,t} \Delta y_{i,t}) &= \frac{1}{1 + \gamma} \left( \frac{1}{1 + \rho \kappa^2} + 1 \right) \sigma^2 \\
E(x_{a,t} \Delta x_{i,t}) &= \frac{2}{1 + \gamma} \left( \frac{1}{1 + \rho \kappa^2} + 1 \right) \sigma^2 \\
V(\Delta y_{a,t}) &= \frac{\beta}{1 - \rho^2} \left( \frac{1}{1 + \rho \kappa^2} + 1 \right) \sigma^2 \\
V(\Delta x_{a,t}) &= \frac{2}{1 + \gamma} \left( \frac{1}{1 + \rho \kappa^2} + 1 \right) \sigma^2 \\
E(y_{i,t-1} \Delta y_{i,t}) &= \frac{1}{1 + \gamma} \left( \frac{1}{1 + \rho \kappa^2} + 1 \right) \sigma^2 \\
E(x_{a,t} \Delta y_{i,t}) &= \frac{\beta \rho^2}{1 - \rho^2} \left( \frac{1}{1 + \rho \kappa^2} + 1 \right) \sigma^2 \\
E(\Delta x_{a,t} \Delta y_{i,t-1}) &= -\frac{\beta}{1 - \rho^2} \left( \frac{1}{1 + \rho \kappa^2} + 1 \right) \sigma^2 \\
E(y_{i,t} x_{i,t}) &= \frac{\beta}{1 - \rho^2} \left( \frac{1}{1 + \rho \kappa^2} + 1 \right) \sigma^2 \\
E(x_{a,t} \Delta x_{i,t-1}) &= \frac{\beta}{1 - \rho^2} \left( \frac{1}{1 + \rho \kappa^2} + 1 \right) \sigma^2 \\
E(\Delta x_{a,t} \Delta x_{i,t-1}) &= \frac{\beta \rho^2}{1 - \rho^2} \left( \frac{1}{1 + \rho \kappa^2} + 1 \right) \sigma^2 .
\end{align*}
\]

These allow to examine the instrument quality and the asymptotic variance of some MM estimators. We will restrict ourselves to simple IV estimators where two instruments are used to estimate the two coefficients \( \beta \) and \( \gamma \). A similar exercise has been undertaken in Arellano (1989). He considers the two estimators which we will (again) indicate by IVdl (where the regressors are \( \Delta x_{a,t} \) and \( \Delta y_{i,t-1} \) and the instruments \( \Delta x_{a,t} \) and \( y_{i,t-2} \)) and IVdd (where the regressors are \( x_{a,t} \) and \( y_{i,t-1} \) and the instruments \( \Delta x_{a,t} \) and \( \Delta y_{i,t-2} \)). Note that \( \Delta x_{a,t} \) is instrumented by itself and not by its (lagged) level. We shall also examine IVld (where the regressors are \( x_{a,t} \) and \( y_{i,t-1} \) and the instruments \( x_{i,t} \) and \( \Delta y_{i,t-1} \)). For obtaining \( \Sigma_{ZW} \) we examine \( E(z_{i1} w_{i1}) \) and find for IVld

\[
E(z_{i1} w_{i1}) = \begin{bmatrix}
E(x_{i1} x_{11}) \\
E(x_{i1} y_{10}) \\
E(y_{10} \Delta y_{10})
\end{bmatrix}
= \begin{bmatrix}
\frac{1}{1 - \rho^2} \sigma^2 \\
\frac{1}{1 - \rho^2} \sigma^2 \\
\frac{\beta \rho^2}{1 - \rho^2} \left( \frac{1}{1 + \rho \kappa^2} + 1 \right) \sigma^2
\end{bmatrix}
= \frac{\sigma^2}{1 - \rho^2} \left[ \begin{array}{ccc}
1 - \gamma \rho & \rho \beta \\
\rho \beta & 1 - \rho + \frac{1 + \gamma \rho}{\kappa^2}
\end{array}\right].
\]

This estimator would lead to estimation problems and an infinite asymptotic variance if this matrix is singular. We examine its determinant and find it to be

\[
\det[E(z_{i1} w_{i1})] = \frac{\beta^2 \sigma^4}{(1 - \rho^2)^2 (1 - \gamma \rho)} \left[ 1 - \rho + \frac{1 + \gamma \rho}{\kappa^2} \right] - \frac{\rho^2}{1 - \gamma \rho}.
\]

Note that singularity problems occur for \( \beta = 0 \) and when the factor is square brackets is zero. We examine it for \( \kappa^2 = 1 \) and find solutions for

\[
\frac{2 - \rho + \gamma \rho}{1 + \gamma} - \frac{\rho^2}{1 - \gamma \rho} = 0 \quad (4.3)
\]

graphically. From Figure 4.1 we observe that IVld may face problems similar to those established by Arellano (1989) for IVdl (where he chose the specific parametrization \( \beta = \sigma^2 = \sigma^2 = 1 \) and \( \sigma^2 = 0 \), because for empirically relevant values of \( \gamma \) and \( \rho \) the determinant is zero.

Without making such restrictions on \( \beta, \sigma^2 \) and \( \sigma^2 \) (and hence generalizing the Arellano (1989) results, we shall also examine singularity of \( \Sigma_{ZW} \) for IVdl and IVdd. For the latter
we find
\[
E(z'_{1w11}) = \begin{bmatrix}
E(\Delta x_{11}^2) & E(\Delta x_{11} \Delta y_{10}) \\
E(\Delta x_{12} \Delta y_{10}) & E(\Delta y_{11} \Delta y_{10})
\end{bmatrix} = \frac{\sigma^2}{(1 + \rho)(1 - \gamma \rho)} \begin{bmatrix}
2(1 - \gamma \rho) & -\rho (1 - \rho) \beta - \frac{\rho \beta}{1 - \gamma} \left(1 - \rho \frac{1 + \gamma}{\kappa(1 - \rho)} + \frac{(1 + \gamma) \beta}{\kappa^2(1 - \rho)} \right)
\end{bmatrix}.
\]

Hence for the determinant we here obtain
\[
det[E(z'_{1w11})] = -\frac{\beta^2 \sigma^4}{(1 + \rho)^2 (1 - \gamma \rho)} \left[ \frac{2}{1 + \gamma} - 2 \rho + \frac{2}{\kappa^2 (1 + \gamma)} - \frac{\rho^2}{1 - \gamma \rho} \right].
\]

For \( \kappa^2 = 1 \) singularity problems arise when
\[
\frac{2}{1 + \gamma} - 2 \rho + \frac{2}{\kappa^2 (1 + \gamma)} - \frac{\rho^2}{1 - \gamma \rho} = 0. \tag{4.4}
\]

Figure 4.2 shows when this occurs. For \( \kappa^2 = 10 \) the situation is depicted in Figure 4.3 and singularities are found to arise in cases of even more practical relevance.

Finally we examine IVd1. We obtain
\[
E(z'_{1w11}) = \begin{bmatrix}
E(\Delta x_{11}^2) & E(\Delta x_{11} \Delta y_{10}) \\
E(\Delta x_{12} \Delta y_{10}) & E(\Delta y_{11} \Delta y_{10})
\end{bmatrix} = \begin{bmatrix}
\frac{2}{1 + \gamma} - \frac{(1 - \rho) \beta}{1 + \gamma} \left(1 - \rho \frac{1 + \gamma}{\kappa(1 - \rho)} + \frac{(1 + \gamma) \beta}{\kappa^2(1 - \rho)} \right) \\
\frac{2}{1 + \gamma} - \frac{(1 - \rho) \beta}{1 + \gamma} \left(1 - \rho \frac{1 + \gamma}{\kappa(1 - \rho)} + \frac{(1 + \gamma) \beta}{\kappa^2(1 - \rho)} \right)
\end{bmatrix}.
\]
The determinant is here
\[
\det[E(z_{11}' w_{11})] = -\frac{\beta^2 \sigma^4_i}{(1 - \gamma \rho)(1 + \rho)^2} \left[ \frac{2}{1 + \gamma} \left( 1 + \frac{(1 + \gamma \rho)}{\kappa^2 (1 - \rho)} \right) + \frac{\rho (1 - \rho)^2}{(1 - \gamma \rho)} \right]
\]
and from Figure 4.4 we see that the term in square brackets leads for $\kappa^2 = 1$ to singularities only when $\rho$ is strongly negative. Examining the effect of a larger $\kappa^2$ (which may frequently occur in practice) in Diagram 4.5 we see that this has little effect.

To find parametric expressions in this model for the asymptotic variance of the various estimators seems quite involved, and at present is seems more practicable to write programming code that enables direct numerical calculation instead of attempting to find closed explicit parametric expressions first.

5. The panel ARX(1) model with a weakly exogenous regressor

This is the case where $\lambda \neq 0$ ($x_{it}$ is weakly exogenous) and $\phi \neq 0$ ($x_{it}$ affected by the individual effects, even if $\lambda = 0$).
Figure 4.4: Singularities of IVdl for $\kappa^2 = 1$ in panel AR(1)

Figure 4.5: Singularities of IVdl for $\kappa^2 = 10$ in panel AR(1)

### yet to be explored ###

6. Conclusions

Main findings at this stage of the study:

- Instrument Quality (IQ) is just one determinant of estimator efficiency (asymptotic variance); the signal-to-noise ratio seems more important;

- In panel AR(1) models, employing one instrument ($L = 1$), we find $AV(\hat{\gamma}_{IVdl}) < AV(\hat{\gamma}_{IVdl})$ for $\gamma > 0$, but $AV(\hat{\gamma}_{IVdl}) < AV(\hat{\gamma}_{IVdl})$ for large $\mu$ (moderate $\gamma$ and $T$);

- In a stationary time-series AR(1) process we have for $T \to \infty$ that $AV(\hat{\gamma}_{OLS}) = 1 - \gamma^2$, hence $\lim_{T \to \infty} AV(\hat{\gamma}_{OLS}) = 0$, whereas in stationary panel AR(1) models we find for $T$ finite and $N \to \infty$ that $\lim_{T \to \infty} AV(\hat{\gamma}_{IVdl}) = \frac{4}{T^2}$, $\lim_{T \to \infty} AV(\hat{\gamma}_{IVdl}) = \infty$ and $\lim_{T \to \infty} AV(\hat{\gamma}_{IVdl}) = \infty$;
• In panel AR(1) and \( L = 2 \) we find \( IQ_{GIV\dot{d}} \) to be superior; \( AV( \) is ???? Also to be examined for \( L > 2 \)?;

• The panel ARX(1) model is much more complex; computer algebra yet of little help;

• In panel ARX(1) model rank condition fails not only for IVdd (as was already known), but also for IVld, whereas IVdl is safe for \( \gamma > 0 \).

References


