Strength and weakness of instruments in IV and GMM estimation of dynamic panel data models

JAN F. KIVIET*
(University of Amsterdam & Tinbergen Institute)

preliminary version: January 2009

JEL-code: C13; C15; C23

Keywords:
concentration parameter, initial conditions, marginal strength of instruments, method of moments, model transformations, unobserved heterogeneity

Abstract

It is shown that efficient GMM (generalized method of moments) estimation of a linear model corresponds to standard IV (instrumental variables) estimation of this model, after transforming it such (as in GLS) that its resulting disturbances have a scalar covariance matrix, while using as instruments the original instruments linearly transformed by the transpose of the inverse of the matrix used to transform the model. This correspondence between efficient GMM and classic IV can be exploited to convert IV measures for the strength of instrumental variables in terms of concentration parameters for use in the more complex GMM context. For inefficient IV estimates in models where the disturbances are dependent, and more generally for GMM employing a suboptimal weighting matrix, such measures can be developed by referring to the asymptotic precision matrix of particular coefficient estimates. These measures for (marginal) instrument strength are then established for various particular implementations of IV and GMM for dynamic panel data models with time-invariant unobserved heterogeneity. Calculations for particular parametrizations allow to better understand aspects of the actual performance of the popular Anderson-Hsiao, Arellano-Bond and Blundell-Bond (system) estimators in samples of empirically relevant sizes.

* Department of Quantitative Economics, Amsterdam School of Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands (j.f.kiviet@uva.nl).

1 Introduction

It is well known that even in very large samples the use of weak instruments leads to (squared) estimation errors with a mean or median substantially away from zero, see Bound et al. (1995). Especially the use of a great number of weak instruments seems counterproductive, see Donald and Newey (2001). Instrumental variable based estimation techniques are very popular to estimate dynamic panel data relationships. Employing panel data is very attractive, because they allow to tackle particular forms of unobserved heterogeneity. However, to neutralize bias due to unobserved individual specific effects that may be correlated with the observed heterogeneity panel data relationships have to be transformed, and in dynamic models this leads to contemporaneous correlation of the transformed lagged-dependent variable regressors and the transformed disturbances and possibly to serial correlation of the disturbances. In microeconometric panels, in which the time-series sample size is typically finite and small while the cross-section sample size may be large, this leads to huge bias of the inconsistent least-squares estimator, but instrumental variables estimators can be designed which are consistent.

The popular IV (instrumental variables, or two-stage least-squares), see Anderson and Hsiao (1982), and GMM (generalized method of moments) estimators, see Arellano and Bond (1991) and Blundell and Bond (1998), for transformed dynamic panel data models do not necessarily exploit external instrumental variables. Internal ones suffice, since higher-order lags of (possibly transformed) regressors constitute an abundance of instruments. However, many of these instruments contain basically the same though lagged information, and therefore the marginal utility of extra higher-order lagged instruments may diminish quickly. Yet another worry is that the actual strength or relevance as such of these internal instruments is to a large extend determined by the true parameter values of the DGP (data generating process) which they should facilitate to estimate accurately. When the lagged dependent variable coefficient is close to either zero or one the instrument strength of particular lagged variables may either be reasonable or just piffling, depending on what particular transformation (first-differencing or not) has been employed.

The adverse effects of using too many instruments when estimating dynamic panel data models have been established in various simulation studies and explained analytically in terms of the order of bias, see Bun and Kiviet (2006). Practitioners seem generally well aware that, although the standard asymptotic efficiency improves with each extra valid instrument, even when it is weak, for good performance in finite sample one better limits the degree of overidentification by simply skipping instruments that involve high-order lags. However, hard analytic evidence on the actual strength of individual or groups of instruments, even in simple AR(1) panel data models, is still very scarce. Some results have been obtained in Blundell and Bond (1998) on the strength of particular instruments when the autoregressive parameter is close to unity, which supports using a system panel data estimator when the autoregressive coefficient is high, because of the strength of lagged first-differenced variables for the untransformed panel model under particular initial conditions of the autoregressive process. However, these results just pertain to the simple case where the sample size in the time-series dimension is so small that the number of instruments equals the number of regressors and GMM simplifies to just identified IV and choosing a weighting matrix becomes irrelevant. Bun and Windmeijer (2007) examine the system GMM estimator and its constituent levels and first-differences components for larger time-series sample sizes, but they do not really focus on panel data estimators, but on cross-section estimators which exploit just one time-shift of the panel. Moreover, they just consider the 2SLS estimator, i.e. focus on a very naive form of weighting matrix, while neglecting the serial dependence of the transformed disturbances.

In this study we shall establish characterizations of the strength of individual instruments, subsets of the instrumental variables and the full set of instruments as they are generally used in the Anderson and Hsiao (1982) and related IV estimators, the Arellano and Bond (1991) efficient GMM dynamic panel data model estimator and in the GMM system estimator, which was put forward first by Arellano and Bover (1995). We focus on IV and 1-step GMM assuming cross-section and time-series homoskedasticity and independence of the underlying idiosyncratic error component. In measuring instrument strength we do take the time-series dependence of the disturbances due to the required model transformation into account, and we consider general forms of GMM covering both optimal and suboptimal weighting of the exploited orthogonality conditions. The major results concern the simple panel AR(1) model with individual specific effects, but our approach allows as well to cover cases which involve higher-order lagged dependent regressor variables and further weakly exogenous or endogenous regressors, employing any choice regarding the set of internal and possibly external valid instruments.

The strength of instruments can be expressed in terms of the so-called concentration parameter, see (references in) Staiger and Stock (1997). However, this approach has only been developed for standard linear IV models with i.i.d. disturbances and not really yet for GMM and not for panels. For linear models with contemporaneous correlation between some of the regressors and the disturbances and where the disturbances have a nonscalar covariance matrix, as is the case in the dynamic panel data models that we are interested in here, GMM improves on the efficiency of IV estimators in a comparable way as GLS (generalized least-squares) does for OLS estimators in linear models with predetermined regressors and disturbances having a nonscalar covariance matrix. We will show that efficient GMM estimation of a linear model corresponds to standard IV estimation of a GLS-type transformed model exploiting the original instruments after transforming these by the transpose of the inverse of the GLS-type transformation. Along these lines standard IV approaches regarding measuring the strength of instrumental variables can be translated into the more complex situation of GMM estimation. However, in the context of estimating dynamic panel data models, it also happens that IV is used whereas the disturbances have a nonscalar covariance matrix, or that GMM is employed using a suboptimal weighting matrix. For these situations the concentration parameter approach regarding measuring instrument strength requires some adaptations, which we provide by referring to correspondences between concentration parameters and the usual asymptotic measure for estimator precision.

After developing these generalizations of concentration parameters for one-dimensional data we will consider them for various forms of IV and GMM implementations for dynamic panel models and analyse the strength of the instrumental variables that correspond to subsets (either determined by lag length or by type of moment condition) of the exploited instrumental variables. We focus on models with cross-section homoskedasticity which in 1-step GMM leads to a very simple operational form of the optimal GMM weighting matrix for the Arellano-Bond estimator, but just to suboptimal operational (and more sophisticated nonoperational) weighting matrices for the system estimator, see Kiviet (2007b). Various calculations and simulations [there are no simulations yet in this version of the paper] allow to better understand the actual performance of the popular Anderson-Hsiao, Arellano-Bond and Blundell-Bond estimators in samples of various sizes. In particular we find that the classic IV based measures are often misleading when applied naively to models with dependent disturbances or which are not estimated by 2SLS but by genuine GMM.

In the next Section 2 we recapitulate standard linear IV and GMM results and establish that efficient GMM corresponds to IV after appropriate transformation of the model and the instruments. This allows to generalize standard notions of instrument weakness, expressed by first-stage F statistics or by so-called concentration parameters, for the linear GMM context. We also suggest simple alternative measures of instrument strength for situations where IV is employed to models where the disturbances have a nonscalar covariance matrix, or where GMM uses a suboptimal weighting matrix. In Section 3 we present the popular IV and GMM dynamic panel data model estimators, paying special attention to the various moment conditions that can be exploited and how they enter variants of the Anderson-Hsiao, Arellano-Bond and Blundell-Bond estimators respectively. In Section 4 we establish for the various particular cases the measures of instrument weakness for all (and for subsets of) the instruments, and present graphs containing numerical results for a few cases of substantial practical interest [this section is yet incomplete]. Finally, Section 5 concludes.

2 IV results and their GMM counterparts

In the first subsection of this section we introduce notation to express some standard textbook results (without proof) on OLS and GLS and on IV and GMM for linear models for one-dimensional data sets. For a fuller treatment see, for instance, Davidson and MacKinnon (2004). In subsection 2.2 we present results developed for the classic IV context to express instrument weakness, see Staiger and Stock (1997). We focus on single equation models with just one endogenous regressor and an arbitrary number of predetermined regressors, because this is the prevalent case in dynamic panel data models. We highlight the correspondence between concentration parameter based measures for instrument strength and the precision of the endogenous regressor coefficient estimate. This correspondence inspires a modification of the classic concentration parameter when IV is applied in models with disturbances that have a nonscalar covariance matrix, as occurs in Anderson-Hsiao dynamic panel data model estimators. Next, in subsection 2.3, we demonstrate what transformations link efficient GMM with standard IV in linear models and how this link can be exploited to express instrument strength in a GMM context by transforming measures developed already for the IV context. These results directly apply to the efficient 1-step Arellano-Bond estimator, but not to 1-step Blundell-Bond estimation, because this involves a suboptimal weighting matrix. However, here again an appropriate instrument strength measure can be developed by referring to the precision of the estimator. Finally, in subsection 2.4, we present a generic form of linear model for two-dimensional panel data sets, focussing on the case where the cross-section sample size N may grow large and the time-series sample size T is relatively small. We introduce a notation such that all measures for instrument strength introduced for one-dimensional data sets are directly applicable for two-dimensional panel data sets and we also address instrument strength measures for 2-step GMM implementations [???].

2.1 Standard results for one-dimensional data sets

Consider the linear regression model for a scalar dependent variable observed for N individual units $i = 1, ..., N$ given by

$$y_i = X_i\beta + u_i, \tag{1}$$

where X_i' and β are $K \times 1$ vectors of explanatory variables and corresponding regression coefficients respectively. Defining the $N \times K$ regressor matrix $X = [X_1' \ X_2' \ ... \ X_N']'$, we can collect all data in the model

$$y = X\beta + u, \tag{2}$$

where the vectors of dependent variables y and of random disturbances u are both $N \times 1$. We assume that matrix X has rank K, whereas $E(u) = 0$ and $Var(u) = \sigma_u^2\Omega$ with Ω positive definite. To avoid underidentification we shall make the innocuous additional assumption $tr(\Omega) = N$. The fixed parameters β and σ_u^2 are unknown, and initially we will assume that Ω is fully known, whereas y and the regressor variables X have been observed.

The OLS (ordinary least-squares) estimator of β is defined by

$$\hat{\beta}_{OLS} \equiv (X'X)^{-1}X'y, \tag{3}$$

and the GLS (generalized least-squares) estimator by

$$\hat{\beta}_{GLS} \equiv (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y. \tag{4}$$

These are obtained as $\arg\min(y - X\beta)'(y - X\beta)$ and $\arg\min(y - X\beta)'\Omega^{-1}(y - X\beta)$ respectively.

As is well known, the $N \times N$ matrix Ω^{-1} can be decomposed such that $\Omega^{-1} = \Psi\Psi'$, where Ψ is in fact non-unique and may be chosen upper or lower triangular. The GLS estimator can also be represented as the OLS estimator of a transformed model, viz. the model that is obtained after premultiplying (2) by the transformation matrix Ψ', i.e.

$$y^* = X^*\beta + u^*, \tag{5}$$

where $y^* = \Psi'y$, $X^* = \Psi'X$, $u^* = \Psi'u$ with $E(u^*) = 0$ and $Var(u^*) = \sigma_u^2\Psi'\Omega\Psi = \sigma_u^2 I_N$, because

$$\begin{aligned} \hat{\beta}^*_{OLS} &\equiv (X^{*\prime}X^*)^{-1}X^{*\prime}y^* \\ &= (X'\Psi\Psi'X)^{-1}X'\Psi\Psi'y = \hat{\beta}_{GLS}. \end{aligned} \tag{6}$$

Note that above we just discussed the algebraic properties and not the statistical properties of $\hat{\beta}_{OLS}$ and $\hat{\beta}_{GLS} \equiv \beta^*_{OLS}$. The statistical properties depend on the joint distribution of (X, u). If X and u are independent (X is exogenous) then both estimators are unbiased, but GLS is generally more efficient. When the regressors are not necessarily exogenous but just contemporaneously uncorrelated with the disturbances, i.e.

$$E(u_i \mid X_i) = 0, \qquad i = 1, ..., N, \tag{7}$$

asymptotic properties can be established under some additional auxiliary assumptions. Then both estimators are consistent and GLS is asymptotically efficient. And, when

Ω is unknown but can be estimated consistently by $\hat{\Omega}$ then the FGLS (feasible GLS) estimator, in which Ω is replaced by $\hat{\Omega}$, can be shown to be asymptotically equivalent to GLS and is therefore efficient too.

For models where some regressors and the disturbance are contemporaneously correlated (some regressors and the regressand are jointly dependent) and where alternative moment conditions in terms of so-called instrumental variables can be made, the following method of moment estimators are available. Let $Z = [Z_1' \; Z_2' \; ... \; Z_N']'$ be an $N \times L$ matrix of rank L, where $L \geq K$. Then, provided $Z'X$ has rank K, the GMM (generalized method of moments) estimator using instruments Z and the $L \times L$ weighting matrix W is

$$\hat{\beta}_{GMM,Z}(W) \equiv (X'ZWZ'X)^{-1}X'ZWZ'y. \tag{8}$$

This is obtained as $\arg\min(y - X\beta)'ZWZ'(y - X\beta)$. The major condition for this estimator to be consistent is instrument validity, i.e.

$$E(u_i \mid Z_i) = 0, \qquad i = 1, ..., N. \tag{9}$$

This estimator is asymptotically efficient when $\text{plim}_{N\to\infty} W$ is proportional to the inverse of the covariance matrix of the limiting distribution of $N^{-1/2}Z'(y - X\beta)$. Assuming that

$$E(u_i u_j \mid Z_i, Z_j) = \sigma_u^2 \Omega_{ij}, \qquad i, j = 1, ..., N \tag{10}$$

and

$$\text{plim}\, N^{-1/2}Z'(y - X\beta) \sim \; \text{N}\left(0, \text{plim}\, N^{-1}Z'\Omega Z\right), \tag{11}$$

we find that

$$\hat{\beta}_{EGMM,Z} \equiv \hat{\beta}_{GMM,Z}((Z'\Omega Z)^{-1}) = [X'Z(Z'\Omega Z)^{-1}Z'X]^{-1}X'Z(Z'\Omega Z)^{-1}Z'y \tag{12}$$

is the efficient GMM estimator exploiting instruments Z. The IV (instrumental variable, or two-stage least-squares) estimator[1] of β in model (2) exploiting instruments Z, defined as

$$\hat{\beta}_{IV,Z} \equiv \hat{\beta}_{GMM,Z}((Z'Z)^{-1}) = [X'Z(Z'Z)^{-1}Z'X]^{-1}X'Z(Z'Z)^{-1}Z'y, \tag{13}$$

is efficient only when $\Omega = I_N$. Note that the IV and EGMM estimators are obtained when using weighting matrix W equal to $(Z'Z)^{-1}$ and $(Z'\Omega Z)^{-1}$ respectively. Probably even more efficient estimators than EGMM can be obtained by extending or changing the set of L valid instruments Z and adapting the weighting matrix accordingly.

Note that when $L = K$, thus $Z'X$ is invertible, $\hat{\beta}_{GMM,Z}(W)$ specializes for any W to $\hat{\beta}_{sIV,Z} = (Z'X)^{-1}Z'y$ and the choice of W is irrelevant, provided it has rank K. The reason is that $\hat{\beta}_{sIV,Z}$ already fully satisfies the K sample moment conditions $Z'(y - X\hat{\beta}_{sIV,Z}) = 0$. Hence, knowledge of Ω cannot be exploited to improve the efficiency of method of moment estimators using instruments Z when $L = K$, although Ω does affect its actual efficiency. In particular, when $Z = X$ we obtain $\hat{\beta}_{sIV,X} = \hat{\beta}_{OLS}$. Whereas, when the instruments $Z = \Omega^{-1}X$ are employed to the untransformed model (2), or when the instruments $\Psi'X$

[1] In the literature this estimator is often addressed as the GIV (generalized instrumental variable) estimator and IV is then used exclusively for the special case where $L = K$ and $\hat{\beta}_{IV,Z} = (Z'X)^{-1}Z'y$. Here we shall use "generalized" just to indicate estimators which aim to take Ω into account, and special results for the $L = K$ case will be addressed as sIV (simple IV).

are employed to the transformed model (5), we obtain $\hat{\beta}_{sIV,\Omega^{-1}X} = \hat{\beta}^*_{sIV,\Psi'X} = \hat{\beta}_{GLS}$. This illustrates that (even when $L = K$) the efficiency may change when one transforms the model and/or the instruments (assuming that the instruments remain valid).

A feasible GMM (FGMM) estimator is asymptotically equivalent with EGMM if it uses $\hat{\Omega}$ instead of Ω in the formula for EGMM provided

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} Z'(\hat{\Omega} - \Omega)Z = O. \tag{14}$$

2.2 Measuring instrument weakness

In a classic linear IV context, where we have $\Omega = I_N$, instrument strength is often measured in terms of concentration parameters or of (population equivalents of noncentrality parameters of) F statistics in reduced form equations. If $K = 1$ the single regressor which is jointly dependent with the disturbances u is simply regressed on all the instruments and their joint significance is tested by an F-test. Generalizations have been discussed in the literature for cases where there are more endogenous and also some exogenous regressors, see for instance the review by Stock et al. (2002) and Stock and Yogo (2005). When we later specialize to the standard first-order dynamic panel data model we will have estimating equations in which we have just one regressor which is jointly dependent with the disturbances, with or without further weakly-exogenous regressors. Therefore we will focus below on the case with just one endogenous regressor. We will also have to pay attention to two cases that as it seems have not yet been addressed in the literature, viz. (i) where $\Omega \neq I_N$ or (ii) where the reduced form equation is not explicitly specified and could also have a nonscalar covariance matrix of its disturbances.

We assume that the first column x_1 of $X = (x_1\ X_2)$ contains the single endogenous regressor, whereas the X_2 can and will be used as instruments. Hence, these X_2 are incorporated in the $N \times L$ matrix $Z = (Z_1\ X_2)$, where Z_1 is an $N \times L_1$ matrix of $L_1 = L - K + 1$ instruments additional to X_2. The focus is now on the F-test which tests the joint significance of Z_1 in the reduced form regression of x_1 on Z, supposing this has i.i.d. disturbances. This F-test statistic is given by

$$F(Z_1; x_1; Z) = \frac{N-L}{L_1} \frac{x_1' M_{X_2} x_1 - x_1' M_Z x_1}{x_1' M_Z x_1} = \frac{N-L}{L_1} \frac{x_1' P_{M_{X_2} Z_1} x_1}{x_1' M_Z x_1}, \tag{15}$$

where for any full column rank $N \times M$ matrix A we define the projections $P_A \equiv A(A'A)^{-1}A'$ and $M_A \equiv I_N - P_A$, and we made use of the general result that if $A = (A_1\ A_2)$ we have $P_A = P_{A_1} + P_{M_{A_1} A_2}$. For the very special case where $K = 1$ and X_2 is void so that $Z_1 = Z$ (15) specializes to

$$F(Z; x_1; Z) = \frac{N-L}{L} \frac{x_1' P_Z x_1}{x_1' M_Z x_1}. \tag{16}$$

Decomposing Z_1 into two partitions $Z_1 = (Z_{11}\ Z_{12})$, where Z_{11} has L_{11} columns and Z_{12} has $L_{12} = L_1 - L_{11}$, one can also express the marginal strength of exploiting the instruments Z_{11} in addition to the instruments $Z_2 = (Z_{12}\ X_2)$ by the following F statistic

$$F(Z_{11}; x_1; Z) = \frac{N-L}{L_{11}} \frac{x_1' M_{Z_2} x_1 - x_1' M_Z x_1}{x_1' M_Z x_1} = \frac{N-L}{L_{11}} \frac{x_1' P_{M_{Z_2} Z_{11}} x_1}{x_1' M_Z x_1}. \tag{17}$$

Population equivalents of the above F-test statistics and corresponding concentration parameter based measures can be expressed as follows. Assuming stationarity of all variables involved, let for $i = 1, 2$ the general $N \times M_i$ data matrices A_i have full column rank, whereas

$$\Sigma_{A_i' A_i} \equiv \underset{N\to\infty}{\text{plim}} \frac{1}{N} A_i' A_i \qquad \text{and} \qquad \Sigma_{A_1' A_2} \equiv \underset{N\to\infty}{\text{plim}} \frac{1}{N} A_1' A_2,$$

with $\Sigma_{A_i' A_i}$ invertible. Then we can consider the concentration parameter based expressions

$$C(Z_1; x_1; Z) \equiv$$
$$\frac{(\Sigma_{x_1' Z_1} - \Sigma_{x_1' X_2} \Sigma_{X_2' X_2}^{-1} \Sigma_{X_2' Z_1})[\Sigma_{Z_1' Z_1} - \Sigma_{Z_1' X_2} \Sigma_{X_2' X_2}^{-1} \Sigma_{X_2' Z_1}]^{-1} (\Sigma_{Z_1' x_1} - \Sigma_{Z_1' X_2} \Sigma_{X_2' X_2}^{-1} \Sigma_{X_2' x_1})}{\Sigma_{x_1' x_1} - \Sigma_{x_1' Z} \Sigma_{Z' Z}^{-1} \Sigma_{Z' x_1}}, \tag{18}$$

$$C(Z; x_1; Z) \equiv \frac{\Sigma_{x_1' Z} \Sigma_{Z' Z}^{-1} \Sigma_{Z' x_1}}{\Sigma_{x_1' x_1} - \Sigma_{x_1' Z} \Sigma_{Z' Z}^{-1} \Sigma_{Z' x_1}} \tag{19}$$

and

$$C(Z_{11}; x_1; Z) \equiv$$
$$\frac{(\Sigma_{x_1' Z_{11}} - \Sigma_{x_1' Z_2} \Sigma_{Z_2' Z_2}^{-1} \Sigma_{Z_2' Z_{11}})[\Sigma_{Z_{11}' Z_{11}} - \Sigma_{Z_{11}' Z_2} \Sigma_{Z_2' Z_2}^{-1} \Sigma_{Z_2' Z_{11}}]^{-1} (\Sigma_{Z_{11}' x_1} - \Sigma_{Z_{11}' Z_2} \Sigma_{Z_2' Z_2}^{-1} \Sigma_{Z_2' x_1})}{\Sigma_{x_1' x_1} - \Sigma_{x_1' Z} \Sigma_{Z' Z}^{-1} \Sigma_{Z' x_1}}, \tag{20}$$

and the population (or noncentrality parameter) equivalents of the F-test statistics

$$\bar{F}(Z_1; x_1; Z) \equiv \frac{N - L}{L_1} C(Z_1; x_1; Z) \text{ and similarly for } \bar{F}(Z; x_1; Z) \text{ and } \bar{F}(Z_{11}; x_1; Z). \tag{21}$$

Stock and Yogo (2005) develop critical values of the above measures regarding their potential to signal serious finite sample bias of IV estimators (as a fraction of the OLS bias) and particular size distortions of standard asymptotic Wald tests. A simple rough and ready rule of thumb resulting from this is that instruments are seriously weak when actual (or population) F statistics are below 10. As it seems the various results on instrument weakness measures obtained for classic linear static IV models have not yet been generalized for GMM. Hall (2005, Section 6.2.1) only discusses results for classic static IV. We will make an attempt here to generalize some of these results for GMM and for IV with nonscalar covariance of the disturbances and implement and verify these for the estimation of transformed dynamic panel data models.

There is a direct correspondence between the above measures and the actual asymptotic precision of the IV estimator for the coefficient of x_1. When $Var(u) = \sigma_u^2 I_n$ then, under regularity sufficient for invoking a central limit theorem, one has

$$\underset{N\to\infty}{\text{plim}} N^{1/2} (\hat{\beta}_{IV,Z} - \beta) \sim \text{ N} \left(0, \sigma_u^2 \underset{N\to\infty}{\text{plim}} \left(\frac{1}{N} X' P_Z X \right)^{-1} \right), \tag{22}$$

and for the coefficient β_1 of regressor x_1 this implies

$$\underset{N\to\infty}{\text{plim}} N^{1/2} (\hat{\beta}_{1,IV,Z} - \beta_1) \sim \text{ N} \left(0, \sigma_u^2 \underset{N\to\infty}{\text{plim}} \left(\frac{1}{N} x_1' P_Z M_{X_2} P_Z x_1 \right)^{-1} \right). \tag{23}$$

Since $x_1' P_Z M_{X_2} P_Z x_1 = x_1'(P_Z - P_{X_2})x_1 = x_1' P_{M_{X_2} Z_1} x_1$ we find

$$AsyVar(\hat{\beta}_{1,IV,Z}) = \sigma_u^2 \operatorname*{plim}_{N\to\infty} \left(\frac{1}{N} x_1' P_{M_{X_2} Z_1} x_1 \right)^{-1},$$

and hence, we can express (18) equivalently in terms of a precision based measure defined as

$$C_I(Z_1; x_1; Z) \equiv \frac{[AsyVar(\hat{\beta}_{1,IV,Z})]^{-1}}{(\Sigma_{x_1'x_1} - \Sigma_{x_1'Z}\Sigma_{Z'Z}^{-1}\Sigma_{Z'x_1})/\sigma_u^2} = C(Z_1; x_1; Z), \tag{24}$$

where the subscript I refers to $\Omega = I$. The denominator of (24) corresponds to the (consistently estimated) ratio of the variances of the reduced form and of the structural form disturbances. Note that $C_I(Z; x_1; Z)$, which concerns the special case $K = 1$, fits into this definition, with $AsyVar(\hat{\beta}_{1,IV,Z})$ simplifying to $\sigma_u^2 \operatorname{plim}_{N\to\infty}(\frac{1}{N} x_1' P_Z x_1)^{-1}$.

The marginal strength measure $C(Z_{11}; x_1; Z)$ of (20) is related to estimator precision as follows. Note that just employing the instruments Z_2 one has

$$\begin{aligned} AsyVar(\hat{\beta}_{1,IV,Z_2}) &= \sigma_u^2 \operatorname*{plim}_{N\to\infty} \left(\frac{1}{N} x_1' P_{Z_2} M_{X_2} P_{Z_2} x_1 \right)^{-1} \\ &= \sigma_u^2 \operatorname*{plim}_{N\to\infty} \left(\frac{1}{N} x_1'(P_{Z_2} - P_{X_2}) x_1 \right)^{-1}. \end{aligned}$$

Hence,

$$\begin{aligned} \sigma_u^2[AsyVar(\hat{\beta}_{1,IV,Z})]^{-1} - \sigma_u^2[AsyVar(\hat{\beta}_{1,IV,Z2})]^{-1} &= \operatorname*{plim}_{N\to\infty} \left(\frac{1}{N} x_1'(P_Z - P_{Z_2}) x_1 \right) \\ &= \operatorname*{plim}_{N\to\infty} \frac{1}{N} x_1' P_{M_{Z_2} Z_{11}} x_1, \end{aligned}$$

and

$$C_I(Z_{11}; x_1; Z) \equiv \frac{[AsyVar(\hat{\beta}_{1,IV,Z})]^{-1} - [AsyVar(\hat{\beta}_{1,IV,Z2})]^{-1}}{(\Sigma_{x_1'x_1} - \Sigma_{x_1'Z}\Sigma_{Z'Z}^{-1}\Sigma_{Z'x_1})/\sigma_u^2} = C(Z_{11}; x_1; Z). \tag{25}$$

These alternative precision based expressions are helpful in developing sensible extensions of the concentration parameter based measures in nonstandard applications of IV.

Suppose that IV has been applied, although $\Omega \neq I_N$. Hence, one should have employed GMM but did actually apply IV because of using a weighting matrix proportional to $(Z'Z)^{-1}$. Because now

$$AsyVar(\hat{\beta}_{IV,Z}) = \sigma_u^2 \operatorname*{plim}_{N\to\infty} N(X'P_Z X)^{-1} X' P_Z \Omega P_Z X (X' P_Z X)^{-1},$$

a precision based modified concentration parameter, defined now more generally as

$$C_\Omega(Z_1; x_1; Z) \equiv \frac{[AsyVar(\hat{\beta}_{1,IV,Z}]^{-1}}{(\Sigma_{x_1'x_1} - \Sigma_{x_1'Z}\Sigma_{Z'Z}^{-1}\Sigma_{Z'x_1})/\sigma_u^2}, \tag{26}$$

seems more appropriate. It differs from (18), and so does the alternative to (25)

$$C_\Omega(Z_{11}; x_1; Z) \equiv \frac{[AsyVar(\hat{\beta}_{1,IV,Z}]^{-1} - [AsyVar(\hat{\beta}_{1,IV,Z_2}]^{-1}}{(\Sigma_{x_1'x_1} - \Sigma_{x_1'Z}\Sigma_{Z'Z}^{-1}\Sigma_{Z'x_1})/\sigma_u^2}. \tag{27}$$

In the above definitions we implicitly assumed that the denominator still properly expresses the ratio of the variances of the disturbances of reduced and structural form. Whether this F-test inspired measure, which naturally leads to measures like $\bar{F}_{\Omega}(Z_1; x_1; Z)$ etc., is suitable for expressing (marginal) instrument strength seems worth examining.

2.3 Algebraic connections between GMM and IV

If we decide to apply IV estimation not to the untransformed model (2), but to the transformed model (5) exploiting some set of instruments collected in an $N \times L$ matrix $\mathring{Z}$, we obtain

$$\begin{aligned}\hat{\beta}^*_{IV,\mathring{Z}} &= [X^{*\prime}\mathring{Z}(\mathring{Z}'\mathring{Z})^{-1}\mathring{Z}'X^*]^{-1}X^{*\prime}\mathring{Z}(\mathring{Z}'\mathring{Z})^{-1}\mathring{Z}'y^* \\ &= [X'\Psi\mathring{Z}(\mathring{Z}'\mathring{Z})^{-1}\mathring{Z}'\Psi'X]^{-1}X'\Psi\mathring{Z}(\mathring{Z}'\mathring{Z})^{-1}\mathring{Z}'\Psi'y.\end{aligned}$$

Note that for the special case

$$\mathring{Z} = Z^* = \Psi^{-1}Z \tag{28}$$

we find

$$\hat{\beta}^*_{IV,Z^*} = \{X'Z[Z'(\Psi\Psi')^{-1}Z]^{-1}Z'X\}^{-1}X'Z[Z'(\Psi\Psi')^{-1}Z]^{-1}Z'y = \hat{\beta}_{EGMM,Z}. \tag{29}$$

Hence, EGMM is equivalent to applying IV to an appropriately GLS-type transformed model upon employing the inverse of the transposed transformation to the instrumental variables. Note that (in)validity of the instruments Z for u implies (in)validity of the instruments Z^* for u^*, and vice versa, since

$$E(Z^{*\prime}u^*) = E[Z'(\Psi^{-1})'\Psi'u] = E(Z'u) \text{ and } E(Z_i'u_i) = E(Z_i^{*\prime}u_i^*).$$

From the above it is obvious how we can proceed to examine and express instrument strength and weakness in a EGMM context when Ω may differ from I_N. Exploiting the transformations $x_1^* = \Psi'x_1$, $X_2^* = \Psi'X_2$ and $Z^* = \Psi^{-1}Z = \Psi^{-1}(Z_1\ X_2)$, we find the EGMM generalization of the standard IV measure (15)

$$\begin{aligned}F^*(Z_1; x_1; Z) &= \frac{N-L}{L_1}\frac{x_1^{*\prime}P_{M_{X_2^*}Z_1^*}x_1^*}{x_1^{*\prime}M_{Z^*}x_1^*} \\ &= \frac{N-L}{L_1}\frac{x_1'\Omega^{-1}QZ_1(Z_1'QZ_1)^{-1}Z_1'Q\Omega^{-1}x_1}{x_1'[\Omega^{-1} - Z(Z'\Omega Z)^{-1}Z']x_1},\end{aligned} \tag{30}$$

where $Q = \Omega - X_2(X_2'\Omega X_2)^{-1}X_2'$. Note that when $K = 1$, thus X_2 is void and $L_1 = L$, this simplifies to

$$F^*(Z; x_1; Z) = \frac{N-L}{L}\frac{x_1'Z(Z'\Omega Z)^{-1}Z'x_1}{x_1'[\Omega^{-1} - Z(Z'\Omega Z)^{-1}Z']x_1}. \tag{31}$$

In the same way we can translate $F(Z_{11}; x_1; Z)$, giving

$$\begin{aligned}F^*(Z_{11}; x_1; Z) &= \frac{N-L}{L_{11}}\frac{x_1^{*\prime}P_{M_{Z_2^*}Z_{11}^*}x_1^*}{x_1^{*\prime}M_{Z^*}x_1^*} \\ &= \frac{N-L}{L_{11}}\frac{x_1'S\Omega Z_{11}[Z_{11}'\Omega S\Omega Z_{11}]^{-1}Z_{11}'\Omega Sx_1}{x_1'[\Omega^{-1} - Z(Z'\Omega Z)^{-1}Z']x_1},\end{aligned} \tag{32}$$

where $S = \Omega^{-1} - Z_2(Z_2'\Omega Z_2)^{-1}Z_2'$, and in the special case where $K = 1$ one has in fact $S = \Omega^{-1} - Z_{12}(Z_{12}'\Omega Z_{12})^{-1}Z_{12}'$.

In the above F-tests it is in fact implicitly assumed that the disturbances in the reduced form equation for x_1^* are i.i.d., which is something which is not self-evident, and which we will concern about in more detail when we consider specific implementations of the above formulas in Section 4.

Note also that the above F^* measures are relevant only for the case where GMM exploits the optimal weighting matrix. In some of the panel GMM implementations that we will introduce in Section 3 an optimal weighting matrix is not always available, and neither EGMM nor IV are used, but GMM (at least in a first step) with some particular suboptimal weighting matrix W, giving $\hat{\beta}_{GMM,Z}(W)$ with

$$AsyVar(\hat{\beta}_{GMM,Z}(W)) = \sigma_u^2 \operatorname*{plim}_{N\to\infty} N(X'ZWZ'X)^{-1}X'ZWZ'\Omega ZWZ'X(X'ZWZ'X)^{-1}.$$

This gives rise to the GMM generalizations of the IV (when $\Omega \neq I_N$) forms of the precision based modified concentration parameter type measures (26) and (27), viz.

$$C_\Omega(Z_1; x_1; Z; W) \equiv \frac{[AsyVar(\hat{\beta}_{1,GMM,Z}(W))]^{-1}}{(\Sigma_{x_1'x_1} - \Sigma_{x_1'Z}\Sigma_{Z'Z}^{-1}\Sigma_{Z'x_1})/\sigma_u^2}, \tag{33}$$

and

$$C_\Omega(Z_{11}; x_1; Z; W; W_2) \equiv \frac{[AsyVar(\hat{\beta}_{1,GMM,Z}(W))]^{-1} - [AsyVar(\hat{\beta}_{1,GMM,Z_2}(W_2))]^{-1}}{(\Sigma_{x_1'x_1} - \Sigma_{x_1'Z}\Sigma_{Z'Z}^{-1}\Sigma_{Z'x_1})/\sigma_u^2}, \tag{34}$$

which can be put into population modified F-form as

$$\bar{F}_\Omega(Z_1; x_1; Z; W) \equiv \frac{N-L}{L_1} C_\Omega(Z_1; x_1; Z; W)$$

and

$$\bar{F}_\Omega(Z_{11}; x_1; Z; W; W_2) \equiv \frac{N-L}{L_{11}} C_\Omega(Z_{11}; x_1; Z; W; W_2).$$

2.4 Generic results for two-dimensional data sets

The above should be generalized now such that we can deal with double-indexed or two-dimensional data variables, where $i = 1, ..., N$ still refers to cross-section units in the sample, whereas $t = 0, ..., T$ refers to the available time-series observations (where $t = 0$ entails the initial values of the variables in the first-order dynamic process). In the dynamic panel data models that we are interested in we want to allow for unobserved individual effects. To deal with these either the model equation has to be transformed or otherwise the variables that are used as instrumental variables, which have to be lagged as well. Such transformations (often, but not exclusively, first differencing) and lag operations will lead to an effectively available time-series sample size in estimation that may deviate form T. Often the time-series sample size in estimation is smaller than T, but the number of observations per individual in estimation can be larger too, when in the system estimator all time-series observations on two differently transformed equations are stacked. Due to first-differencing also the number of regressors and coefficients may change, when the original relationship involved an intercept. All these transformations complicate combining the familiar notation for panel data models with the standard one-dimensional cross-section results of the earlier subsections. We have chosen to deal with this in the following way.

In the generic notation below we simply address the time-series sample size as $\tilde{T}$ to be able to cover all possible cases; often $\tilde{T} = T - 1$ or $\tilde{T} = 2(T-1)$. The total number of observations can now be denoted as $\tilde{N} = N\tilde{T}$, and the linear panel data model that will be estimated by IV or GMM is expressed as

$$\tilde{y} = \tilde{X}\tilde{\beta} + \tilde{u} = (\tilde{x}_1 \ \tilde{X}_2)\tilde{\beta} + \tilde{u},$$

for which we have the $\tilde{N} \times L$ instrument matrix Z. Here $\tilde{y}$ and $\tilde{u}$ are $\tilde{N} \times 1$, $\tilde{X}$ is $\tilde{N} \times \tilde{K}$ and $\tilde{\beta}$ is $\tilde{K} \times 1$, whereas $\tilde{x}_1$ is $\tilde{N} \times 1$ and $E(\tilde{u}) = 0$ and $Var(\tilde{u}) = \sigma^2_{\tilde{u}}\tilde{\Omega}$. All results of the earlier subsections are directly applicable, but just require to reinterpret (and supplement with a "tilde") the scalars, vectors and matrices N, y, u, σ^2_u, β, X, x_1, X_2 and Ω. These vectors and matrices contain in some way or another (transformed and/or lagged, as is indicated in the next section) elements corresponding to the underlying dynamic panel data relationship, which we denote as

$$y_{it} = \gamma y_{i,t-1} + x'_{it}\delta + \eta_i + \varepsilon_{it}, \tag{35}$$

where the vectors x_{it} and δ consist of $K-1$ elements.

After transformation (if any) and stacking of all the time-series observations (possibly of two differently transformed equations) the generic model relation can be written as

$$\tilde{y}_i = \tilde{X}_i\tilde{\beta} + \tilde{u}_i. \tag{36}$$

Each row of the $\tilde{T} \times 1$ vectors $\tilde{y}_i = (\tilde{y}_{i1}, ..., \tilde{y}_{i\tilde{T}})'$ and $\tilde{u}_i = (\tilde{u}_{i1}, ..., \tilde{u}_{i\tilde{T}})'$ and of the $\tilde{T} \times \tilde{K}$ matrix $\tilde{X}_i = (\tilde{X}'_{i1} \ \tilde{X}'_{i2} \ ... \ \tilde{X}'_{i\tilde{T}})'$ results from a particular transformation applied to (35) for some index t. The $\tilde{K} \times 1$ vector $\tilde{\beta}$ is related to $(\gamma, \beta')'$ conformable to the transformations and stacking operations. We assume $E(\tilde{u}_i) = 0$, $Var(\tilde{u}_i) = \sigma^2_{\tilde{u}}\tilde{\Omega}_1$ with $tr(\tilde{\Omega}_1) = \tilde{T}$ and $\tilde{\Omega}_1$ a positive definite $\tilde{T} \times \tilde{T}$ matrix, whereas $E(\tilde{u}_i\tilde{u}'_j) = O$ for $i \neq j$.

We will exploit N observed $\tilde{T} \times L$ matrices Z_i as instruments, which are valid when they satisfy $\forall i$ the L orthogonality conditions

$$E(Z'_i\tilde{u}_i) = E[Z'_i(\tilde{y}_i - \tilde{X}_i\tilde{\beta})] = 0. \tag{37}$$

Defining $\tilde{y} = (\tilde{y}'_1, ..., \tilde{y}'_N)'$, $\tilde{u} = (\tilde{u}'_1, ..., \tilde{u}'_N)'$, $\tilde{X} = (\tilde{X}'_1 \ ... \ \tilde{X}'_N)'$, $Z = (Z'_1 \ ... \ Z'_N)'$ and $\sigma^2_{\tilde{u}}\tilde{\Omega} = \sigma^2_{\tilde{u}}(I_N \otimes \tilde{\Omega}_1)$ all having $\tilde{N}$ rows now, we have in the new "tilde" notation again the original model $y = X\beta + u$ with $E(u) = 0$ and $Var(u) = \sigma^2\Omega$ and L instruments Z which can be used to obtain IV or GMM estimators, provided X, Z and $Z'X$ all have full column rank, hence $\tilde{N} = N\tilde{T} > L \geq \tilde{K}$. The only serious differences with the situation in subsection 2.1 are that we have $N\tilde{T}$ instead of N observations, and that Ω has a block-diagonal structure, but the formulas for $\hat{\beta}_{IV,Z}$, $\hat{\beta}_{GMM,Z}(W)$ and $\hat{\beta}_{EGMM,Z,\Omega} = \hat{\beta}^*_{IV}$ remain the same, and so for the F-test statistics $F(Z_1; x_1; Z)$ and $F(Z_{11}; x_1; Z)$, upon changing N in $N\tilde{T}$. Also all concentration parameter and precision inspired measures C and C_Ω are directly applicable.

3 IV and GMM estimators for dynamic panel data models

In this section we further consider the just introduced linear first-order dynamic panel data model with two unobserved error components, viz. random individual effects and idiosyncratic disturbances, and also an arbitrary number of further regressors which are assumed to be contemporaneously uncorrelated with the idiosyncratic disturbances. From the various model assumptions stated explicitly in subsection 3.1, the linear orthogonality conditions are derived in subsection 3.2, which are exploited in subsection 3.3 to present some variants of the Anderson and Hsiao (1982) estimators, the efficient GMM estimator, see Arellano and Bond (1991), and the system GMM estimator (GMMs), see Blundell and Bond (1998). The latter estimator exploits an additional model assumption, viz. effect stationarity, which gives rise to additional linear orthogonality conditions, but also to complexities regarding the weighting matrix, which are discussed in the final subsection 3.4.

3.1 Model assumptions

Below we shall use the notation

$$\left.\begin{array}{l} y_i^{t-1} \equiv (y_{i,0}, ..., y_{i,t-1}) \\ X_i^t \equiv (x'_{i1}, ..., x'_{it}) \end{array}\right\} t = 1, ..., T,$$

where y_i^{t-1} is $1 \times t$ and X_i^t is $1 \times t(K-1)$. We also define

$$Y^{t-1} \equiv \left[\begin{array}{c} y_1^{t-1} \\ \vdots \\ y_N^{t-1} \end{array}\right], \; X^t \equiv \left[\begin{array}{c} X_1^t \\ \vdots \\ X_N^t \end{array}\right], \; \eta \equiv \left(\begin{array}{c} \eta_1 \\ \vdots \\ \eta_N \end{array}\right),$$

where η is $N \times 1$, Y^{t-1} is $N \times t$ and X^t is $N \times t(K-1)$.

Regarding the two random error components η_i and ε_{it} in model (35) we make the assumptions ($i, j = 1, ..., N$; $t, s = 1, ..., T$):

$$\left.\begin{array}{l} E(\varepsilon_{it} \mid Y^{t-1}, X^t, \eta) = 0, \;\forall i, t \\ E(\varepsilon_{it}^2 \mid Y^{t-1}, X^t, \eta) = \sigma_\varepsilon^2 > 0, \;\forall i, t \\ E(\varepsilon_{it}\varepsilon_{js} \mid Y^{t-1}, X^t, \eta) = 0, \;\forall t < s, \;\forall i, j \text{ and } \forall i \neq j, \;\forall t \leq s \\ E(\eta_i) = 0, \; E(\eta_i^2) = \sigma_\eta^2 \geq 0, \; E(\eta_i\eta_j) = 0 \;\forall i \neq j \end{array}\right\} \quad (38)$$

These assumptions entail that all regressors are predetermined with respect to the idiosyncratic disturbances ε_{it}, which are homoskedastic and serially and contemporaneously uncorrelated.

Additional assumptions that may be made — and which are crucial when the GMMs estimator is employed — involve

$$E(y_{it}\eta_i) = \sigma_{y\eta} \text{ and } E(x_{it}\eta_i) = \sigma_{x\eta}, \;\forall i, t. \quad (39)$$

Here $\sigma_{y\eta}$ is a scalar and $\sigma_{x\eta}$ is a $(K-1) \times 1$ vector. Note that both $\sigma_{y\eta}$ and $\sigma_{x\eta}$ are assumed to be time-invariant. By multiplying the model equation (35) by η_i and taking expectations we find that (39) implies

$$\sigma_{y\eta} = \gamma\sigma_{y\eta} + \sigma'_{x\eta}\delta + \sigma_\eta^2$$

or

$$\sigma_{y\eta} = \frac{\sigma_{x\eta}'\delta + \sigma_\eta^2}{1-\gamma}. \tag{40}$$

This condition we will call for obvious reasons *effect-stationarity*.

3.2 Moment conditions

By taking first-differences the model simplifies in the sense that only one unobservable error component remains. Estimating

$$\Delta y_{it} = \gamma \Delta y_{i,t-1} + (\Delta x_{it})'\delta + \Delta\varepsilon_{it} \tag{41}$$

by OLS would yield inconsistent estimators[2] because

$$E(\Delta y_{i,t-1}\Delta\varepsilon_{it}) = -E(y_{i,t-1}\varepsilon_{i,t-1}) = -\sigma_\varepsilon^2 \neq 0.$$

Unless σ_ε^2 would be known this moment condition cannot directly be exploited in a method of moments estimator. Note, however, that it easily follows from the model assumptions (38) that for $i = 1, ..., N$ we have

$$\left.\begin{array}{r} E(Y^{t-2}\Delta\varepsilon_{it}) = O \\ E(X^{t-1}\Delta\varepsilon_{it}) = O \end{array}\right\} \qquad t = 2, ..., T,$$

which together provide $KNT(T-1)/2$ moment conditions for estimating the K coefficients of (41). Especially when the cross-section units are independent many of these conditions will produce weak or completely ineffective instruments. Then it will be more appropriate to exploit just the $KT(T-1)/2$ moment conditions

$$\left.\begin{array}{r} E(y_i^{t-2}\Delta\varepsilon_{it}) = 0' \\ E(X_i^{t-1}\Delta\varepsilon_{it}) = 0' \end{array}\right\} \qquad t = 2, ..., T,$$

as is done in the Arellano-Bond (1991) estimator. Upon substituting (41) for $\Delta\varepsilon_{it}$ it is obvious that these moment conditions are linear in the unknown coefficients γ and δ, i.e.

$$\left.\begin{array}{r} E\{y_i^{t-2}[\Delta y_{it} - \gamma\Delta y_{i,t-1} - (\Delta x_{it})'\delta]\} = 0' \\ E\{X_i^{t-1}[\Delta y_{it} - \gamma\Delta y_{i,t-1} - (\Delta x_{it})'\delta]\} = 0' \end{array}\right\} \qquad t = 2, ..., T. \tag{42}$$

Blundell and Bond (1998) argue that assumption (39), when valid, may yield relatively strong additional useful instruments for estimating the undifferenced equation (35). These additional instruments are the first-differenced variables. Defining

$$\left.\begin{array}{l} \Delta y_i^{t-1} \equiv (\Delta y_{i1}, ..., \Delta y_{i,t-1}) \\ \Delta X_i^t \equiv (\Delta x_{i2}', ..., \Delta' x_{i,t}) \end{array}\right\} \; t = 2, ..., T,$$

which are $1 \times (t-1)$ and $1 \times (t-1)(K-1)$ respectively, it follows from (39) that (for $i = 1, ..., N$)

$$E(\Delta y_i^{t-1}\eta_i) = 0' \text{ and } E(\Delta X_i^t \eta_i) = 0'.$$

[2] Note that in case x_{it} contains a unit element and β an intercept this model has one unknown coefficient less than suggested in what follows.

From (38) we find

$$E(\Delta y_i^{t-1}\varepsilon_{it}) = 0' \text{ and } E(\Delta X_i^t \varepsilon_{it}) = 0'.$$

Combining these and substituting $\eta_i + \varepsilon_{it} = y_{it} - \gamma y_{i,t-1} - x_{it}'\delta$ yields the $KT(T-1)/2$ linear moment conditions

$$\left.\begin{array}{r} E[\Delta y_i^{t-1}(y_{it} - \gamma y_{i,t-1} - x_{it}'\delta)] = 0' \\ E[\Delta X_i^t(y_{it} - \gamma y_{i,t-1} - x_{it}'\delta)] = 0' \end{array}\right\} \qquad t = 2, ..., T. \tag{43}$$

These can be transformed linearly into two subsets of $K(T-1)$ and $K(T-1)(T-2)/2$ conditions respectively, viz.

$$\left.\begin{array}{r} E[\Delta y_{i,t-1}(y_{it} - \gamma y_{i,t-1} - x_{it}'\delta)] = 0' \\ E[\Delta x_{it}'(y_{it} - \gamma y_{i,t-1} - x_{it}'\delta)] = 0' \end{array}\right\} \qquad t = 2, ..., T, \tag{44}$$

and

$$\left.\begin{array}{r} E\{\Delta y_i^{t-1}[\Delta y_{it} - \gamma \Delta y_{i,t-1} - (\Delta x_{it})'\delta]\} = 0' \\ E\{\Delta X_i^t[\Delta y_{it} - \gamma \Delta y_{i,t-1} - (\Delta x_{it})'\delta]\} = 0' \end{array}\right\} \qquad t = 3, ..., T. \tag{45}$$

The second subset (45), though, can also be obtained by a simple linear transformation of (42). Hence, effect stationarity only leads to the $K(T-1)$ additional linear moment conditions (44). These involve estimation of the undifferenced model (35) by employing all regressor variables of the first-differenced model (41) as instruments.

Due to the i.i.d. assumption regarding ε_{it} further (non-linear) moment conditions do hold in the present dynamic panel data model, but below we will stick to the linear conditions mentioned above.

3.3 Implementation of various IV and GMM estimators

Anderson and Hsiao (1982) introduced IV estimators for dynamic panel data models. They focussed on the case with first-order dynamics, took first-differences of the relationship, yielding $\tilde{y}_i = \tilde{X}_i\beta + \tilde{u}_i$, with

$$\tilde{y}_i = \begin{pmatrix} \Delta y_{i2} \\ \vdots \\ \Delta y_{iT} \end{pmatrix}, \ \tilde{u}_i = \begin{pmatrix} \Delta \varepsilon_{i2} \\ \vdots \\ \Delta \varepsilon_{iT} \end{pmatrix} \text{ and } \tilde{X}_i = \begin{bmatrix} \Delta y_{i1} & \Delta x_{i2}' \\ \vdots & \vdots \\ \Delta y_{i,T-1} & \Delta x_{iT}' \end{bmatrix}. \tag{46}$$

When there is an intercept in δ and corresponding unit value in x_{it} (which vanishes after first-differencing) we have $\tilde{K} = K - 1$, otherwise $\tilde{K} = K$ and $\beta = (\gamma, \delta')'$. Using as many valid instruments as coefficients to be estimated (i.e. accepting that the resulting IV estimators will not have finite moments), they distinguished two cases regarding the instruments to be used for regressor $\Delta y_{i,t-1}$, viz. using the lagged level instrument $y_{i,t-2}$ or the lagged first differenced instrument $\Delta y_{i,t-2}$. So, assuming $E(\Delta x_{it}\Delta\varepsilon_{it}) = 0$, the corresponding instrument matrices Z have N blocks

$$Z_i^{AHl} = \begin{bmatrix} y_{i,0} & \Delta x_{i,2}' \\ y_{i,1} & \Delta x_{i,3}' \\ \vdots & \vdots \\ y_{i,T-2} & \Delta x_{i,T}'' \end{bmatrix} \text{ and } Z_i^{AHd} = \begin{bmatrix} 0 & 0' \\ \Delta y_{i,1} & \Delta x_{i,2}' \\ \vdots & \vdots \\ y_{i,T-2} & \Delta x_{i,T}'' \end{bmatrix} \tag{47}$$

respectively, where the row of zeros in Z_i^{AHd} basically means that for each individual only $\tilde{T} = T - 2$ time-series observations are being used in estimation. We have $L^{AHl} = L^{AHd} = K$ for the number of columns.

We shall also examine below the Anderson-Hsiao type implementations which use one extra instrument, and where $\tilde{T} = T - 2$ and $\tilde{T} = T - 3$ respectively, viz.:

$$Z_i^{AHl1} = \begin{bmatrix} y_{i,1} & y_{i,0} & \Delta x'_{i,3} \\ \vdots & \vdots & \vdots \\ y_{i,T-2} & y_{i,T-3} & \Delta x''_{i,T} \end{bmatrix} \text{ and } Z_i^{AHd1} = \begin{bmatrix} \Delta y_{i,2} & \Delta y_{i,1} & \Delta x'_{i,4} \\ \vdots & \vdots & \vdots \\ \Delta y_{i,T-2} & \Delta y_{i,T-3} & \Delta x''_{i,T} \end{bmatrix}, \quad (48)$$

where $L^{AHl1} = L^{AHd1} = K + 1$. Since $E(\Delta\varepsilon_{it}\Delta\varepsilon_{i,t-1}) \neq 0$ the elements of $\tilde{u}_i$ are not i.i.d., thus the IV estimators AHl1 and AHd1 are not efficient. Not just because not all valid instruments are being used, but also because Ω is not exploited as GMM would.

The Arellano-Bond GMM estimator uses the model transformation given in (46) too, hence $\tilde{T} = T - 1$ and employs all orthogonality conditions (42). Hence, $L^{AB} = KT(T-1)/2$ and the $\tilde{T} \times L^{AB}$ matrix Z_i is taken to be [what to do with X???]

$$Z_i^{AB} = \begin{bmatrix} y_i^0 & 0' & 0' & \Delta X_i^1 & & 0' \\ 0 & \ddots & O & O & \ddots & O \\ 0 & 0' & y_i^{T-2} & 0' & 0' & \Delta X_i^{T-1} \end{bmatrix}. \quad (49)$$

It can be derived that for the EGMM estimator one should use the weighting matrix

$$W_{AB}^{opt} \propto \left(\sum_{i=1}^{N} (Z_i^{AB})' H Z_i^{AB} \right)^{-1} \quad (50)$$

with $(T-1) \times (T-1)$ matrix

$$H \equiv \begin{bmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \ddots & \vdots \\ 0 & -1 & 2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -1 \\ 0 & \cdots & 0 & -1 & 2 \end{bmatrix}. \quad (51)$$

So, under our strong assumptions regarding homoskedasticity and uncorrelatedness of the idiosyncratic disturbances ε_{it} 1-step GMM employing (50) is asymptotically efficient within the class of estimators exploiting the instruments Z_i^{AB}.

In case of GMMs we have $\tilde{K} = K$, $\beta = (\gamma, \delta')$ and $\tilde{T} = 2(T-1)$ with

$$\tilde{y}_i = \begin{pmatrix} \Delta y_{i2} \\ \vdots \\ \Delta y_{iT} \\ y_{i2} \\ \vdots \\ y_{iT} \end{pmatrix}, \ \tilde{u}_i = \begin{pmatrix} \Delta\varepsilon_{i2} \\ \vdots \\ \Delta\varepsilon_{iT} \\ \eta_i + \varepsilon_{i2} \\ \vdots \\ \eta_i + \varepsilon_{iT} \end{pmatrix} \text{ and } \tilde{X}_i = \begin{bmatrix} \Delta y_{i1} & \Delta x'_{i2} \\ \vdots & \vdots \\ \Delta y_{i,T-1} & \Delta x'_{iT} \\ y_{i1} & x'_{i2} \\ \vdots & \vdots \\ y_{i,T-1} & x'_{iT} \end{bmatrix}, \quad (52)$$

and from (42) and (44) it follows that $L^{BB} = K(T-1)(T+2)/2$ whereas the $\tilde{T} \times L^{BB}$ matrix of instruments for block i is

$$Z_i^{BB} = \begin{bmatrix} Z_i^{AB} & 0 & \cdots & 0 & O & \cdots & O \\ 0' & \Delta y_{i1} & 0' & 0 & \Delta x'_{i2} & 0' & 0' \\ \vdots & 0 & \ddots & 0 & 0' & \ddots & O \\ 0' & 0 & 0' & \Delta y_{i,T-1} & 0' & 0' & \Delta x'_{iT} \end{bmatrix}. \tag{53}$$

Note that $E[(\Delta\varepsilon_{it})^2] = 2\sigma_\varepsilon^2$ differs from $E[(\varepsilon_{it} + \eta_i)^2] = \sigma_\varepsilon^2 + \sigma_\eta^2$ when $\sigma_\eta^2 \neq \sigma_\varepsilon^2$, and for $t \neq s$ we find $E[(\varepsilon_{it} + \eta_i)(\varepsilon_{is} + \eta_i)] = \sigma_\eta^2 \geq 0$. Thus, again it does not hold here that $\tilde{u}_{it} \mid Z_i^t$ is i.i.d., so IV is not efficient. However, here an appropriate weighting matrix is not readily available due to the complexity of $\text{Var}(Z_i'\tilde{u}_i)$.

3.4 Weighting matrices in use for 1-step GMMs

The GMMs optimal weighting matrix has been obtained for the no individual effects case $\sigma_\eta^2 = 0$ by Windmeijer (2000), who presents

$$W_{BB}^{FW} \propto \left(\sum_{i=1}^{N} (Z_i^{BB})' D^{FW} Z_i^{BB} \right)^{-1} \tag{54}$$

with

$$D^{FW} = \begin{pmatrix} H & C_1 \\ C_1' & I_{T-1} \end{pmatrix}, \tag{55}$$

where C_1 is the $(T-1) \times (T-1)$ matrix

$$C_1 = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ -1 & 1 & 0 & \ddots & \vdots \\ 0 & -1 & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & -1 & 1 \end{bmatrix}. \tag{56}$$

Blundell, Bond and Windmeijer (2000, footnote 11), and Doornik et al. (2002, p.9) in the computer program DPD, use in 1-step GMMs the operational weighting matrix

$$W_{BB}^{DPD} \propto \left(\sum_{i=1}^{N} (Z_i^{BB})' D^{DPD} Z_i^{BB} \right)^{-1}, \tag{57}$$

with

$$D^{DPD} = \begin{pmatrix} H & O \\ O & I_{T-1} \end{pmatrix}. \tag{58}$$

There is no special situation for which these weights are optimal.

Blundell and Bond (1998) did use (see page 130, 7 lines from bottom) in their first step of 2-step GMMs

$$D^{GIV} = \begin{pmatrix} I_{T-1} & O \\ O & I_{T-1} \end{pmatrix} = I_{2T-2}, \tag{59}$$

which yields the IV or 2SLS estimator, which neglects that $\tilde{u}_i$ has a nonscalar covariance matrix. In Kiviet (2007a) it is demonstrated that different initial weighting matrices can lead to substantial differences in the performance of 1-step and 2-step GMMs. There it is argued that a better (though yet suboptimal and unfeasible) weighting matrix would be

$$W_{BB}^{\sigma_\eta^2/\sigma_\varepsilon^2} \propto \left(\sum_{i=1}^{N}(Z_i^{BB})' D^{\sigma_\eta^2/\sigma_\varepsilon^2} Z_i^{BB}\right)^{-1}, \tag{60}$$

with

$$D^{\sigma_\eta^2/\sigma_\varepsilon^2} = \begin{pmatrix} H & C \\ C' & I_{T-1} + \frac{\sigma_\eta^2}{\sigma_\varepsilon^2}\iota_{T-1}\iota_{T-1}' \end{pmatrix}. \tag{61}$$

This can be made operational by choosing some value for $\sigma_\eta^2/\sigma_\varepsilon^2$. From simulations it was found that this value should not be chosen too low, and reasonably satisfying results were obtained by choosing $\sigma_\eta^2/\sigma_\varepsilon^2 = 10$.

4 Measurements of instrument weakness

In the subsections of this section we establish numerical values for the alternative measures stated in Section 2 for the various estimators of Section 3. First we consider results for the panel AR(1) model in great detail, and next we produce some results for the first-order dynamic panel data model with one extra explanatory variable.

4.1 Results for stationary panel AR(1) models

In the zero-mean panel AR(1) model we have $\tilde{K} = 1$ and $\tilde{\beta} = \gamma$. The data generating process is

$$y_{it} = \gamma y_{i,t-1} + \eta_i + \varepsilon_{it}, \tag{62}$$

and, assuming full stationarity, i.e. restricting $|\gamma| < 1$ and generating the start-up as

$$y_{i0} = \frac{1}{1-\gamma}\eta_i + \frac{1}{\sqrt{1-\gamma^2}}\varepsilon_{i0}, \tag{63}$$

where

$$\eta_i \sim i.i.d.(0, \sigma_\eta^2) \text{ and } \varepsilon_{it} \sim i.i.d.(0, \sigma_\varepsilon^2), \tag{64}$$

we obtain

$$y_{it} = \frac{1}{1-\gamma}\eta_i + \sum_{s=0}^{t-1}\gamma^s\varepsilon_{i,t-s} + \frac{\gamma^t}{\sqrt{1-\gamma^2}}\varepsilon_{i0}$$

and

$$\Delta y_{it} = \varepsilon_{it} - (1-\gamma)\left(\sum_{s=1}^{t-1}\gamma^{s-1}\varepsilon_{i,t-s} + \frac{\gamma^{t-1}}{\sqrt{1-\gamma^2}}\varepsilon_{i0}\right),$$

from which one can derive (for integer $s \geq 1$):

$$\left.\begin{array}{ll} Var(y_{it}) = \frac{\sigma_\eta^2}{(1-\gamma)^2} + \frac{\sigma_\varepsilon^2}{1-\gamma^2} & Var(\Delta y_{it}) = \frac{2\sigma_\varepsilon^2}{1+\gamma} \\ Cov(y_{it}, y_{i,t-s}) = \frac{\sigma_\eta^2}{(1-\gamma)^2} + \frac{\gamma^s\sigma_\varepsilon^2}{1-\gamma^2} & Cov(\Delta y_{it}, \Delta y_{i,t-s}) = -\sigma_\varepsilon^2\gamma^{s-1}\frac{1-\gamma}{1+\gamma} \\ Cov(y_{it}, \Delta y_{i,t-s}) = \sigma_\varepsilon^2\gamma^s\frac{1}{1+\gamma} & Cov(\Delta y_{it}, y_{i,t-s}) = -\sigma_\varepsilon^2\gamma^{s-1}\frac{1}{1+\gamma}. \end{array}\right\} \tag{65}$$

Instead of expressing various of the results below in the error-component variance ratio $\sigma_\eta^2/\sigma_\varepsilon^2$, we find it more instructive (also see Kiviet, 2007a) to employ the nonlinear transformation of the parameters

$$\mu = \frac{1}{1-\gamma}\frac{\sigma_\eta}{\sigma_\varepsilon}. \tag{66}$$

Note that μ^2 is the ratio of the variance component in y_{it} due to the accumulated individual effect η_i and the variance component in y_{it} due to the current shock ε_{it}, and

$$Var(y_{it}) = \sigma_\varepsilon^2 \left[\mu^2 + \frac{1}{1-\gamma^2}\right]. \tag{67}$$

Parameter μ is an autonomous characteristic of the data series y_{it} that is invariant with respect to the speed of the adjustment process, which is completely determined by γ. Thus, in some sense μ and γ are variation free or orthogonal now.

Parameter γ of model (62) is estimated by applying IV or GMM to

$$\Delta y_{i,t} = \gamma \Delta y_{i,t-1} + \Delta\varepsilon_{i,t} \tag{68}$$

In the notation of Section 2 (68) is a model in which X_2 is void. Hence, for the strength of the full set of instruments used, the relevant standard IV measures are $F(Z;x_1;Z)$ of (16) and

$$C(Z;x_1;Z) = \frac{\Sigma_{x_1'Z}\Sigma_{Z'Z}^{-1}\Sigma_{Z'x_1}}{\Sigma_{x_1'x_1} - \Sigma_{x_1'Z}\Sigma_{Z'Z}^{-1}\Sigma_{Z'x_1}} \text{ and } \bar{F}(Z;x_1;Z) = \frac{N-L}{L}C(Z;x_1;Z). \tag{69}$$

However, taking into account that the model has nonscalar covariance matrix of the disturbances, we should better consider

$$C_\Omega(Z;x_1;Z;(Z'Z)^{-1}) = \frac{(\Sigma_{x_1'Z}\Sigma_{Z'Z}^{-1}\Sigma_{Z'x_1})^2(\Sigma_{x_1'Z}\Sigma_{Z'Z}^{-1}\Sigma_{Z'\Omega Z}\Sigma_{Z'Z}^{-1}\Sigma_{Z'x_1})^{-1}}{\Sigma_{x_1'x_1} - \Sigma_{x_1'Z}\Sigma_{Z'Z}^{-1}\Sigma_{Z'x_1}} \tag{70}$$

and

$$\bar{F}_\Omega(Z;x_1;Z;(Z'Z)^{-1}) = \frac{N-L}{L}C_\Omega(Z;x_1;Z;(Z'Z)^{-1}). \tag{71}$$

We should [and in a next version we will] simulate the statistic $F(Z;x_1;Z)$ and the sample equivalent of (71) to examine their information content regarding the finite sample properties of IV panel estimators. Below we examine $\bar{F}(Z;x_1;Z)$ and $\bar{F}_\Omega(Z;x_1;Z;(Z'Z)^{-1})$, and their GMM counterparts, over the relevant parameter space.

4.1.1 Anderson-Hsiao type estimators

In the first-stage regression of the classic Anderson-Hsiao approach the single regressor $\Delta y_{i,t-1}$ is fitted for each t to the same set of instruments (the Z_i matrices do not have a block structure). The rows of Z_i have either lagged first-differences, viz. $(\Delta y_{i,t-2},...,\Delta y_{i,t-L+1})$, or lagged levels of y_{it}, i.e. $(y_{i,t-2},...,y_{i,t-L+1})$. Since

$$\Delta y_{i,t-1} = \gamma \Delta y_{i,t-2} + \Delta\varepsilon_{i,t-1}, \tag{72}$$

at first sight it seems obvious that when $L=1$ using just the level instrument $y_{i,t-2}$ will yield a more moderate fit than using it jointly with $y_{i,t-3}$ when $L=2$, and using just

$\Delta y_{i,t-2}$ seems already appropriate. However, note that $y_{i,t-2}$ and $\Delta y_{i,t-2}$ are correlated with $\Delta\varepsilon_{i,t-1}$. Apparently, unlike the premise of section 2.2, regressing $\Delta y_{i,t-1}$, the regressor of (72), on the instruments Z_i does not represent the reduced form. Because an explicit reduced form is not readily available, the situation is not so clear-cut.

We first examine the case where $L = K = 1$. This simplifies the above formulas regarding instrument strength considerably. Denoting the one and only instrument as z_1, and focussing first on the naive instrument strength measures (69), we have to establish

$$C(z_1; x_1; z_1) = \frac{[Cor(x_1, z_1)]^2}{1 - [Cor(x_1, z_1)]^2}, \tag{73}$$

where the vector x_1 contains $\Delta y_{i,t-1}$ and z_1 contains $y_{i,t-2}$ for AHl and $\Delta y_{i,t-2}$ for AHd. From

$$[\mathrm{Cor}(\Delta y_{i,t-1}, \Delta y_{i,t-2})]^2 = \frac{1}{4}(1-\gamma)^2 \tag{74}$$

we find for the estimator with the first-difference instrument AHd, where the sample size is $\tilde{N} = N(T-2)$, for the population value of the F statistic, $\bar{F}$ (with between brackets indications of: examined instrument(s); jointly dependent regressor; and full set of instruments) the expression

$$\bar{F}(\Delta y_{i,t-2}; \Delta y_{i,t-1}; \Delta y_{i,t-2}) = \frac{(\tilde{N}-1)(1-\gamma)^2}{4-(1-\gamma)^2}. \tag{75}$$

From

$$[\mathrm{Cor}(\Delta y_{i,t-1}, y_{i,t-2})]^2 = \frac{1}{2}\frac{(1-\gamma)^2}{1-\gamma+\mu^2(1+\gamma)} \tag{76}$$

we obtain for the AHl estimator, where $\tilde{N} = N(T-1)$,

$$\bar{F}(y_{i,t-2}; \Delta y_{i,t-1}; y_{i,t-2}) = \frac{(\tilde{N}-1)(1-\gamma)^2}{1-\gamma^2+2\mu^2(1+\gamma)}. \tag{77}$$

Thus, for AHl instrument strength seems to improve for smaller values of μ, whereas AHd is invariant with respect to μ. Note that these AHd and AHl measures are equal for $\mu = 1$, whereas for $\mu < 1$ the AHl measure is larger than for AHd, and vice versa for $\mu > 1$.

However, these measures do not take into account that the equation is estimated by IV whilst it has disturbances with a covariance matrix where $\tilde{\Omega}_1 = H$, as given in (51). Hence, we should examine the modified measure $\bar{F}_\Omega(z_1; x_1, z_1, (z_1'z)^{-1})$. Note that because $K = L$ the weighting matrix has no effect here on the estimator, but $\Omega \neq I_N$ makes a change for its variance. In evaluating

$$\bar{F}_\Omega = (\tilde{N}-1)C_\Omega = (\tilde{N}-1)C\frac{Var(z_1)}{\Sigma_{z_1'\Omega z_1}} = \bar{F}\frac{Var(z_1)}{\Sigma_{z_1'\Omega z_1}}$$

we have to derive $\Sigma_{z_1'\Omega z_1}$ for AHd and AHl and find

$$2[Var(\Delta y_{it}) - \frac{\tilde{T}-1}{\tilde{T}}Cov(\Delta y_{it}, \Delta y_{i,t-1})] = \frac{2\sigma_\varepsilon^2}{1+\gamma}\left[3-\gamma+\frac{1-\gamma}{T-2}\right] \tag{78}$$

and

$$2[Var(y_{it}) - \frac{\tilde{T}-1}{\tilde{T}}Cov(y_{it}, y_{i,t-1})] = \frac{2\sigma_\varepsilon^2}{1-\gamma^2}\left[1-\gamma+\frac{\gamma+\mu^2(1+\gamma)/(1-\gamma)}{T-1}\right] \tag{79}$$

respectively. Hence, $\bar{F}_\Omega$ is determined by the actual value of T, whereas $\bar{F}$ is not. For $Var(z_1)/\Sigma_{z_1'\Omega z_1}$ we find the expressions

$$\left[3-\gamma+\frac{1-\gamma}{T-2}\right]^{-1}$$

and

$$\frac{1}{2}\frac{1+\mu^2(1+\gamma)/(1-\gamma)}{1-\gamma+\frac{\gamma+\mu^2(1+\gamma)/(1-\gamma)}{T-1}}$$

respectively. Thus, for AHd the modification involves multiplying the naive measure with a factor smaller than 0.5, so the naive measure seems too optimistic about instrument strength, but for AHl it can go either way; for $\gamma=0$ the factor is 0.5, but for $\mu=0$ it is greater than 1 for $\gamma>0.5$.

The situation is best examined by comparing graphs, see Figure 1, which are based on calculations of the naive (neglecting the serial correlation) formula $\bar{F}(z_1;x_1;z_1)$ (in the top row) and the modified $\bar{F}_\Omega(z_1;x_1;z_1)$ for $T=6$ (in the bottom row) respectively. The AHl measures have been calculated for $\mu=0.2$, 1, 5. From the graphs one can grasp for which values of the sample size and of γ and μ the measures qualify the instrument as weak, because the F-value is smaller than 10. In all cases this occurs for large positive γ and more severely so for smaller sample size. That the naive AHd measure is systematically too optimistic can clearly be seen. For the cases examined we find that the AHl naive measure is generally too pessimistic, except for $\mu<1$ and $\tilde{N}$ small. The bottom row suggests that AHl has a stronger instrument than AHd only for moderate μ. For $\mu=5$ the AHl instrument is weak for $\gamma>0.5$ but less so for AHd. However, we should note that we have not examined yet whether the value of 10 is an appropriate watershed for the $\bar{F}_\Omega$ criterion in this model.

When we slightly generalize the Anderson-Hsiao approach by using the instrument matrices given in (48) then we can increase the value of L. In Figure 2 $L=2$ and in Figure 3 $L=4$. For increasing L we note that all naive pictures are too optimistic, and that all modified measures seem to converge to a more similar pattern. At $L=4$ the modified F for AHd and AHl seem very close for $\mu=1$, and for $\mu>1$ AHl performs worse though better for $\mu<1$.

These first three figures focussed on (modification of) the population F statistic. If this is small the estimator might suffer from weak instruments with detremental effects on the accuracy of first order asymptotics. A prominent factor of the (modified) concentration parameter is (the inverse) of the asymptotic variance. It might be informative too to scan the asymptotic precision measures of the γ estimators and their relative differences. For that a naive measure, which wrongly takes H as I seems of very limited interest. The actual differences in asymptotic precision of AHd and AHl and its dependence on γ, μ, T and L can be examined directly by comparing $AsyVar(\hat{\gamma}_{AHd})$ and $AsyVar(\hat{\gamma}_{AHl})$ which are given in this model by

$$\frac{\Sigma_{x_1'Z}\Sigma_{Z'Z}^{-1}\Sigma_{Z'\Omega Z}\Sigma_{Z'Z}^{-1}\Sigma_{Z'x_1}}{(\Sigma_{x_1'Z}\Sigma_{Z'Z}^{-1}\Sigma_{Z'x_1})^2},$$

for the respective Z matrices. Indicating these by Z_d and Z_l respectively, and using

$$\bar{F}_\Omega=\frac{\tilde{N}-L}{L}C_\Omega=\frac{\tilde{N}-L}{L}\frac{\sigma_\varepsilon^2[AsyVar(\hat{\gamma})]^{-1}}{\Sigma_{x_1'x_1}-\Sigma_{x_1'Z}\Sigma_{Z'Z}^{-1}\Sigma_{Z'x_1}},$$

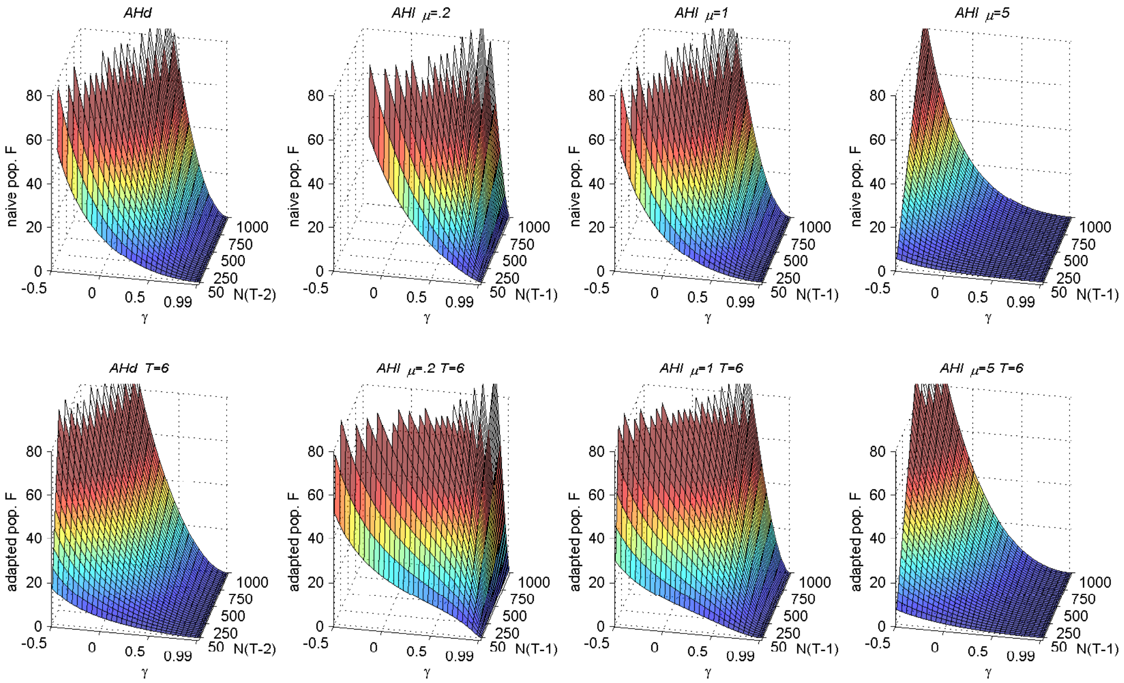

Figure 1: Anderson-Hsiao in fully stationary panel AR(1); $K = 1$, $L = 1$.

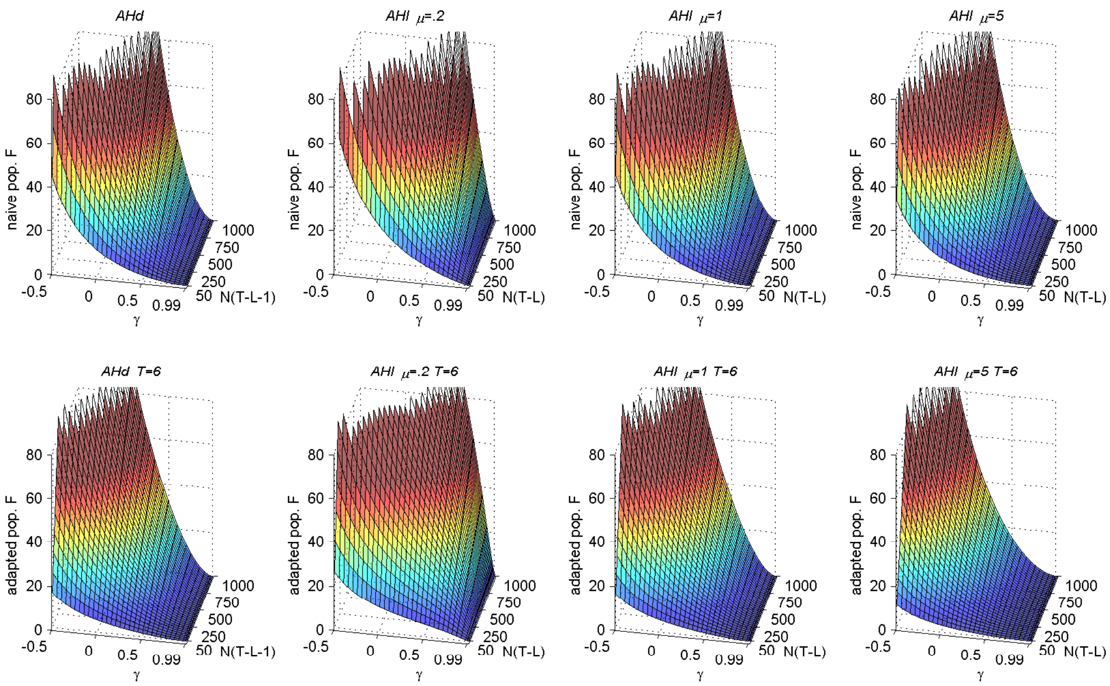

Figure 2: Anderson-Hsiao in fully stationary panel AR(1); $K = 1$, $L = 2$.

we have

$$\frac{AsyVar(\hat{\gamma}_{AHd})}{AsyVar(\hat{\gamma}_{AHl})} = \frac{\bar{F}_{\Omega}(\hat{\gamma}_{AHl})}{\bar{F}_{\Omega}(\hat{\gamma}_{AHd})} \frac{\Sigma_{x_1'x_1} - \Sigma_{x_1'Z_l}\Sigma^{-1}_{Z_l'Z_l}\Sigma_{Z_l'x_1}}{\Sigma_{x_1'x_1} - \Sigma_{x_1'Z_d}\Sigma^{-1}_{Z_d'Z_d}\Sigma_{Z_d'x_1}}$$

In the next graphs we consider the natural logarithm of the square root of this expression (asymptotic standard deviation) and calculate the log of the ratio for two alternative estimators, hence a value of zero means equivalence, ± 0.05 indicates roughly a $\pm 5\%$ difference, ± 0.1 indicates $\pm 10.5\%$ and ± 1 means $\pm 272\%$.

In Figure 4 we see that the difference in asymptotic precision of AHd and AHl can be enormous. For $L = 1$ AHd is better only when T is sufficiently small and μ is sufficiently large. AHl is better, especially when γ is positive, μ not very large and T not very small. Especially for γ very large and μ moderate AHl is substantially more efficient. Figure 5 shows that the picture is less pronounced for $L = 4$. The differences are moderate, except for γ very large and μ small where AHl is again substantially better.

In Figure 6 we examine how AsyStd changes for AHd when L moves from 1 to 4. Quite remarkably there is a difference only for $T = 6$ (where for $L = 4$ only one observation per individual can be used so that the disturbances are not serially correlated) with $L = 4$ yielding a smaller AsyVar, but for $T > 6$ the efficiency does not depend on L.

In Figure 7 we see that increasing L may have a mtigating effect on the asymptotic efficiency of AHl, viz. when γ is postive and μ is small. For negative γ the effect is moderate but very substantial when μ is large and γ positive.

However, we should be more concerned of course about the actual precision of the estimators in finite sample, and in the ability of the instrument strength measures to indicate whether or not asymptotics is accurate regarding the actual quality of estimators. To examine this we ran Monte Carlo simulations. All results are based on 1000 replications of samples of $N = 100$ and various values of $T \geq 6$. In these panel data samples both error components $\varepsilon_{i,t}$ and η_i were drawn randomly from the normal distribution. For each replication these series of random drawings were held constant over all different values of γ and μ examined, and (as much as possible) over the various values of T.

Figures 8 and 9 show how bad asymptotics works in samples of this size, both for $L = 1$ and $L = 4$. Note that it is self-evident that in fully stationary samples these phenomena are invariant with respect to μ. The earlier pictures suggested that asymptotics would work accurately for negative γ and bad for increasing values of γ. These figures suggest that for $L = 1$ there is a large positive γ value where ....

Regarding AHd we find

4.1.2 GMM estimators

[calculations to follow]

4.2 Results for panel ARX(1) models

In a next version of the paper we will also make calculations and perform simulations on panel data models with further predetermined regressors, cf. Bun and Kiviet (2006).

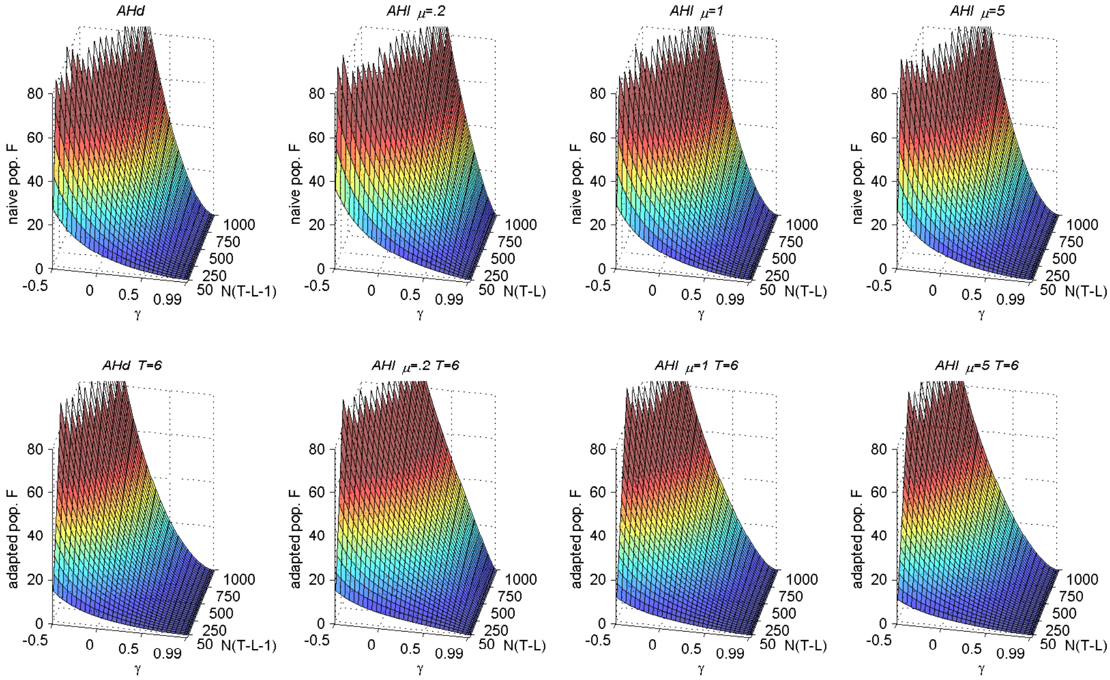

Figure 3: Anderson-Hsiao in fully stationary panel AR(1); $K = 1$, $L = 4$.

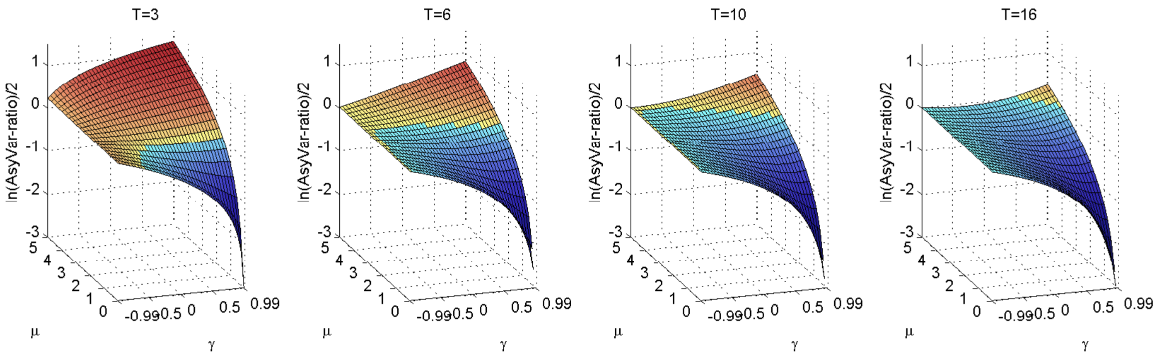

Figure 4: AsyStd of AHl in terms of AHd for $L = 1$ in panel AR(1)

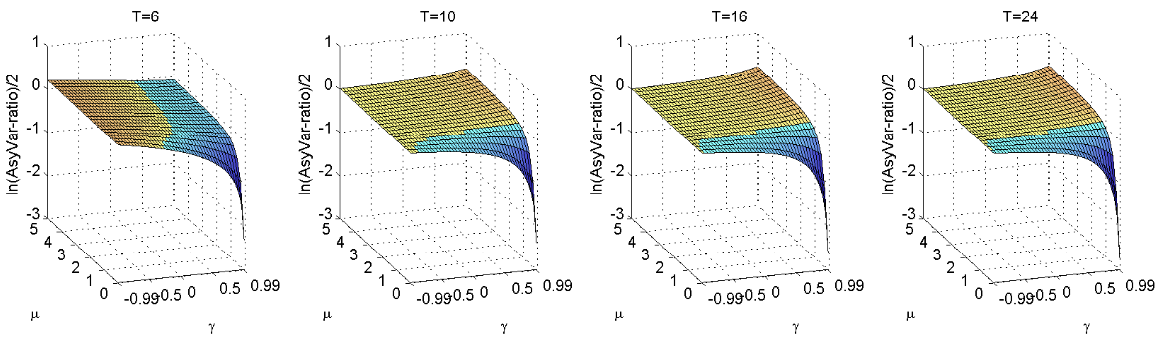

Figure 5: AsyStd of AHl in terms of AHd for $L = 4$ in panel AR(1)

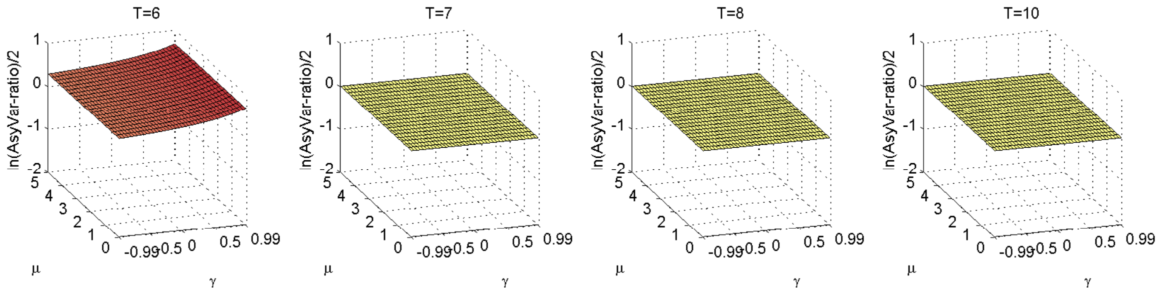

Figure 6: AsyStd of AHd for $L = 1$ in terms of AHd for $L = 4$ in panel AR(1)

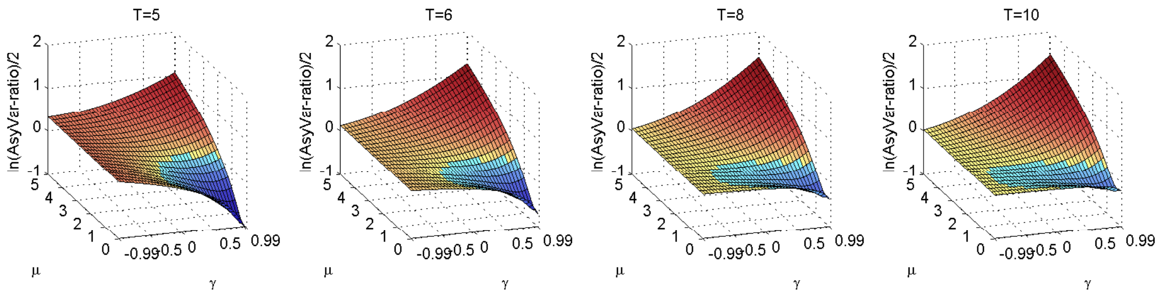

Figure 7: AsyStd of AHd for $L = 1$ in terms of AHl for $L = 4$ in panel AR(1)

5 Conclusions

In this paper we show how measures for instrument weakness based on concentration parameters as developed for single static linear structural equations, with a fully specified reduced form and i.i.d. disturbances and estimated by IV (or 2SLS), can be generalized to cover the much more complex situation met in dynamic panel data models. The relationships estimated in that context contain regressors that are contemporaneously correlated with the disturbances due to transformations that, while achieving their primary purpose, viz. removing unobserved time-constant individual heterogeneity, at the same time lead to disturbance vectors that will not be serially uncorrelated. Hence, ideally such equations should be estimated by efficient GMM, instead of IV, or when still estimated by IV the measures for instrument weakness should be adapted to allow for the effects of dependent disturbances. Not only do we develop measures for instrument weakness in the context of efficient GMM and for IV when errors are serially correlated, but also for GMM employing an arbitrary weighting matrix. We also circumvent the problem that in the context of the transformed dynamic panel data model no explicit reduced form is available. For all the situations considered, we also design measures for the strength of subsets of the full set of instruments used.

For particular popular 1-step GMM panel model estimators, which use an extra set of instruments that are valid only when particular initial conditions hold, no operational optimal weighting matrix is available. It would be of great practical interest if in addition to the techniques available to test the validity of subsets of the instruments, such as incremental Sargan or J-tests, there would be techniques to assess the relevance or

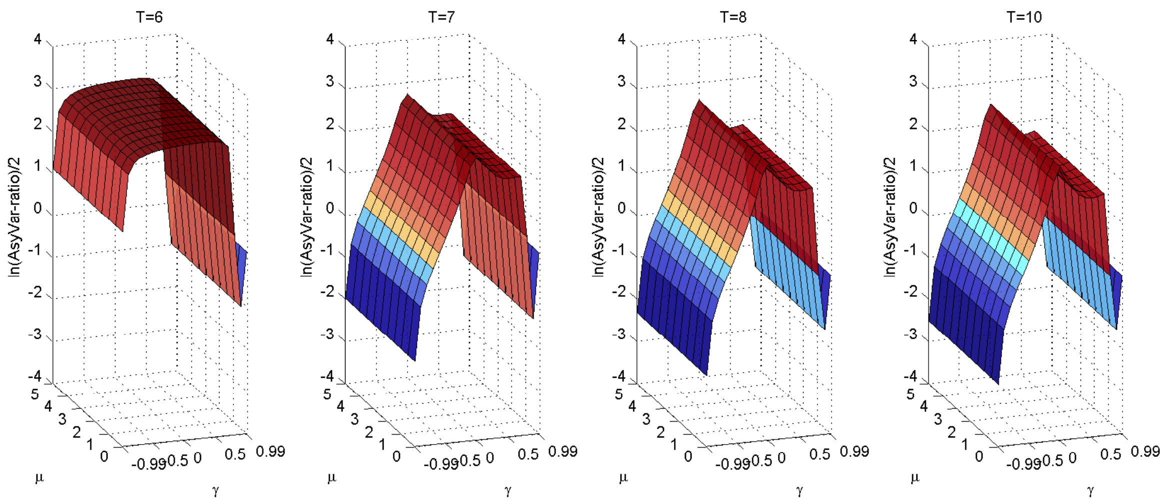

Figure 8: AsyStd of AHd in terms of its actual RMSE at $N = 100$ for $L = 1$ in panel AR(1)

strength (contribution to the precision of estimators in finite samples) of subsets of instruments. The techniques developed in this paper aim to provide these. How useful they are should be analyzed further by evaluating and comparing their results for empirically relevant DGP's, and checking their relevance for strategies regarding the selection of the instruments that one should use in actual practice. This requires further calculations and simulations.

At this stage only some calculations, presented in the form of diagrams, have been produced for measuring instrument strength in Anderson-Hsiao type IV estimators (which neglect disturbance dependence) in panel AR(1) models. These already illustrate that the naive standard IV measures are ineffective in the context of the dynamic panel data model, and that the performance of the here developed alternative measures seems promising.

References

Anderson, T.W., Hsiao, C. (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics* 18, 47-82.

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

Arellano, M., 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., Bover, O., 1995. Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics* 68, 29-51.

Blundell, R., Bond, S., 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87, 115-143.

Bound, J., Jaeger, D.A., Baker, R.M., 1995. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variables is weak. *Journal of the American Statistical Association* 90, 443-450.

Bun, M.J.G, Kiviet, J.F., 2006. The effects of dynamic feedbacks on LS and MM

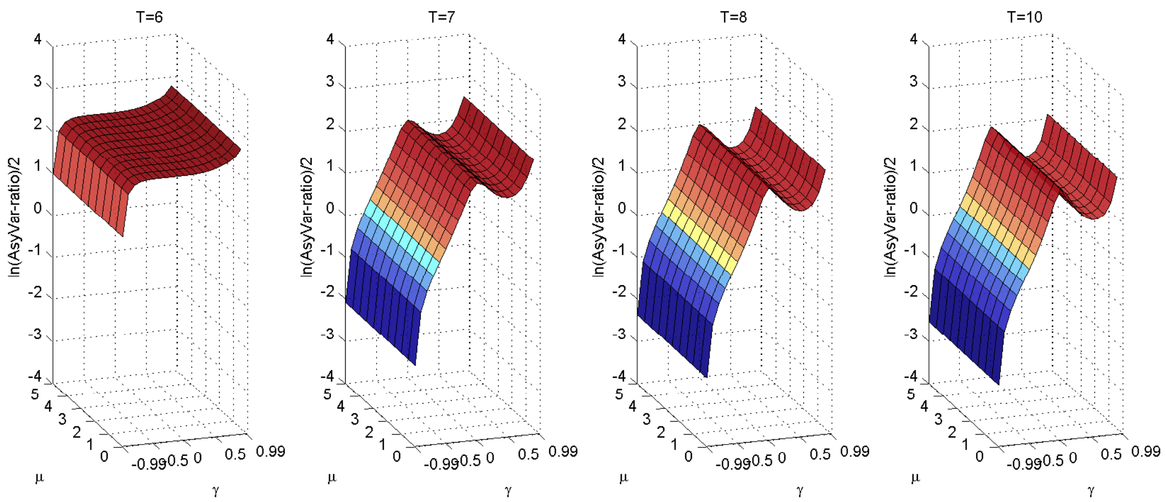

Figure 9: AsyStd of AHd in terms of its actual RMSE at $N = 100$ for $L = 4$ in panel AR(1)

estimator accuracy in panel data models. *Journal of Econometrics* 132, 409-444.

Bun, M.J.G., Windmeijer, F., 2007. The weak instrument problem of the system GMM estimator in dynamic panel data models. *UvA-Econometrics Discussion Paper 2007/01*, University of Amsterdam.

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

Donald, S.G., Newey, W.K., 2001. Choosing the number of instruments. *Econometrica* 69, 1161-1192.

Doornik, J.A., Arellano, M., Bond, S., 2002. Panel data estimation using DPD for Ox. mimeo, University of Oxford.

Hall, A.R., 2005. *Generalized Method of Moments*. Oxford University Press.

Hansen, L.P., 1982. Large sample properties of generalized method of moment estimators. *Econometrica* 50, 1029-1054.

Kiviet, J.F., 2007a. Judging contending estimators by simulation: Tournaments in dynamic panel data models. Chapter 11 (pp.282-318) in Phillips, G.D.A., Tzavalis, E. (Eds.). *The Refinement of Econometric Estimation and Test Procedures; Finite Sample and Asymptotic Analysis*. Cambridge University Press.

Kiviet, J.F., 2007b. On the optimal weighting matrix for the GMM system estimator in dynamic panel data models. *UvA-Econometrics Discussion Paper 2007/08*, University of Amsterdam.

Staiger, D., Stock, J.H., 1997. Instrumental variables regression with weak instruments. *Econometrica* 65, 557-586.

Stock, J.H., Wright, J.H., 2000. GMM with weak identification. Econometrica 68, 1055-1069.

Stock, J.H., Wright, J.H., Yogo, M., 2002. A survey of weak instruments and weak identification in generalized method of moments. Journal of Business and Economic Statistics , 518-529.

Stock, J.H., Yogo, M., 2005. Testing for weak instruments in linear IV regression. Chapter 5 (pp. 80-108) in: *Identification and Inference for Econometric Models, Essays*

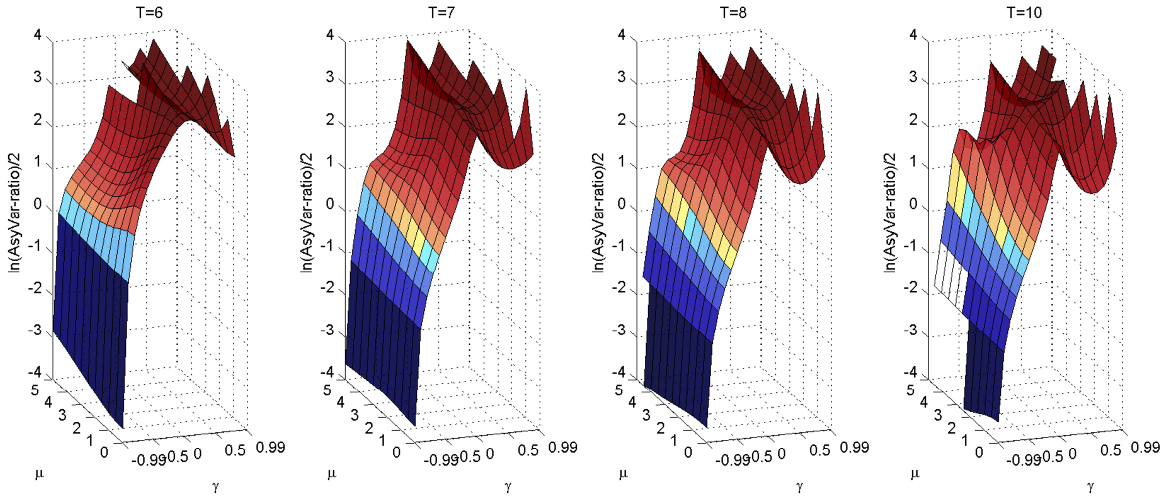

Figure 10: AsyStd of AHl in terms of its actual RMSE at $N = 100$ for $L = 1$ in panel AR(1)

in Honor of Thomas Rothenberg (eds. Andrews, D.W.K., Stock, J.H.). Cambridge University Press.

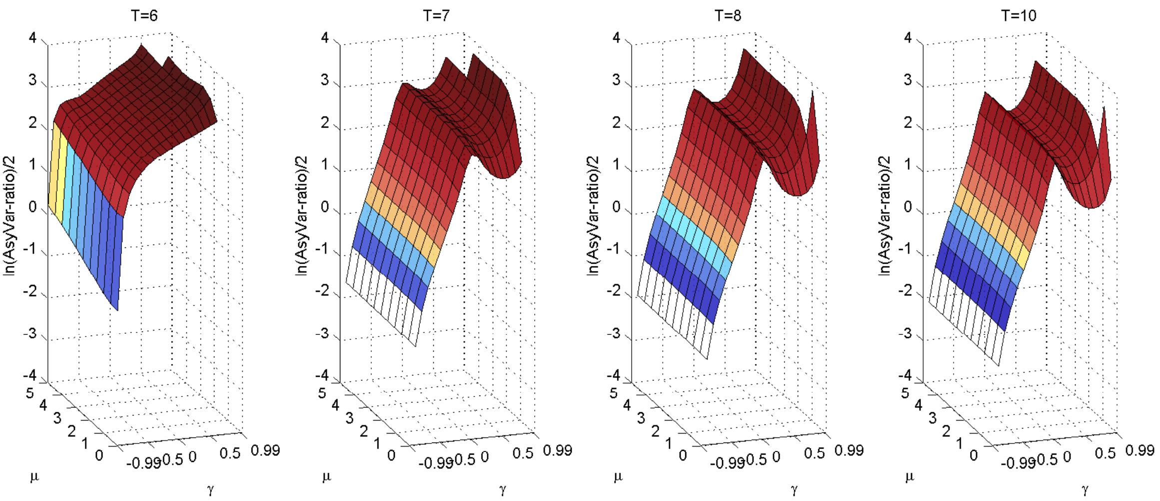

Figure 11: AsyStd of AHl in terms of its actual RMSE at $N = 100$ for $L = 4$ in panel AR(1)