Bias Corrected Instrumental Variables Estimation for Dynamic Panel Models with Fixed Effects

Jinyong Hahn	Jerry Hausman	Guido Kuersteiner
UCLA	MIT	Boston University

July, 2005

Acknowledgment: We are grateful to Yu Hu for exceptional research assistance. Helpful comments by Guido Imbens, Whitney Newey, James Powell, and participants of workshops at Arizona State University, Berkeley, Harvard/MIT, Maryland, Triangle, UCL, UCLA, UCSD are appreciated. Hahn gratefully acknowledges finacial support from NSF grant SES-0313651. Kuersteiner gratefully acknowledges financial support from NSF grant SES-0095132 and SES-0523186.

Abstract

This paper proposes a new instrumental variables estimator for a dynamic panel model with fixed effects with good bias and mean squared error properties even when identification of the model becomes weak near the unit circle. We adopt a weak instrument asymptotic approximation to study the behavior of various estimators near the unit circle. We show that an estimator based on long differencing the model is much less biased than conventional implementations of the GMM estimator for the dynamic panel model. We also show that under the weak instrument approximation such conventional estimators are dominated in terms of mean squared error by an estimator with far less moment conditions. The long difference estimator mimics the infeasible optimal procedure through its reliance on a small set of moment conditions.

Keywords: dynamic panel, bias correction, second order, unit root, weak instrument

JEL: C13, C23, C51

1 Introduction

We are concerned with estimation of the dynamic panel model with fixed effects. Under large n, fixed T asymptotics it is well known from Nickell (1981) that the standard maximum likelihood estimator suffers from an incidental parameter problem leading to inconsistency. In order to avoid this problem the literature has focused on instrumental variables estimation (GMM) applied to first differences. Examples include Anderson and Hsiao (1982), Holtz-Eakin, Newey, and Rosen (1988), and Arellano and Bond (1991). Ahn and Schmidt (1995), Hahn (1997), and Blundell and Bond (1998) considered further moment restrictions.

Unfortunately, the standard GMM estimator obtained after first differencing has been found to suffer from substantial finite sample biases, especially when the autoregressive parameter is close to the unit circle. See Alonso-Borrego and Arellano (1996). Motivated by this problem, modifications of likelihood based estimators emerged in the literature. See Kiviet (1995), Lancaster (1997), Hahn and Kuersteiner (2002). The likelihood based estimators do reduce finite sample bias compared to the standard maximum likelihood estimator, but the remaining bias is still substantial for T relatively small.

We attempt to solve these problems by using the "long difference technique" of Griliches and Hausman (1986). Griliches and Hausman noted that bias is reduced when long differences are used in the errors in variable problem, and a similar result works here with the second order bias. Long differences also increase the explanatory power of the instruments which further reduces the finite sample bias and also decreases the MSE of the estimator. Monte Carlo results demonstrate that the long difference estimator performs quite well, even for high positive values of the lagged variable coefficient where previous estimators are badly biased.

In order to analyze the properties of the long difference estimator, we consider a local to non-identification asymptotic approximation. We use the weak instrument approximation of Staiger and Stock (1997) and Stock and Wright (2000). Here we let the autoregressive parameter tend to unity as the number of cross-sectional observations n tends to infinity. Our limiting distribution for the GMM estimator shows that only moment conditions involving initial conditions are asymptotically relevant. We define a class of estimators based on linear combinations of asymptotically relevant moment conditions and show that a bias minimal estimator within this class can approximately be based on taking long differences of the dynamic panel model. In general, it turns out that under near non-identification asymptotics the optimal procedures of Alvarez and Arellano (1998), Arellano and Bond (1991) , Ahn and Schmidt (1995, 1997) are suboptimal from a bias point of view, and optimal inference should be based on a smaller than the full set of moment conditions. We show that a bias minimal estimator can be obtained by using a particular linear combination of the original moment conditions. As far as bias minimization under the alternative asymptotics is concerned this optimality result shows that only a single moment condition should optimally be used. We also analyze the form of the optimal estimator in terms of minimal mean squared error (MSE) under weak identificaiton asymptotics. First order asymptotically efficient procedures are again found to be suboptimal. The optimal number of moment conditions minimizing the MSE under alternative asymptotics can be up to a factor of 10 smaller than the set of first order optimal instruments.

Optimal procedures under weak instrument asymptotics are difficult to implement because they depend on unobservable nuisance parameters that can not be readily estimated. We show that the long difference estimator we propose in this paper is a good heuristic approximation to the optimal procedures. It captures the essence of our theoretical analysis which shows that the selection of a few highly significant instruments is called for to minimize bias and MSE.

2 Long Difference: Intuitive Motivation

Consider the standard dynamic panel model with fixed effects:

$$y_{it} = \alpha_i + \beta y_{i,t-1} + \varepsilon_{it}, \qquad i = 1, \ldots, n; \quad t = 1, \ldots, T \tag{1}$$

Throughout the paper we assume that the moment conditions put forward by Ahn and Schmidt (1997) and summarized in 4 hold. In addition we impose distributional assumptions for specific theoretical results. In this section we impose the following somewhat strong conditions, which we partly relax in Section 3.

Condition 1 $\varepsilon_{it} \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \sigma^2\right)$ *over* i *and* t.

Condition 2 $y_{i0} | \alpha_i \sim \mathcal{N}\left(\frac{\alpha_i}{1-\beta}, \frac{\sigma^2}{1-\beta^2}\right)$ *and* $\alpha_i \sim N\left(0, \sigma_\alpha^2\right)$.

The usual GMM estimator[1] is based on the first difference form of the model

$$y_{it} - y_{i,t-1} = \beta\left(y_{i,t-1} - y_{i,t-2}\right) + \left(\varepsilon_{it} - \varepsilon_{i,t-1}\right),$$

where the instruments are based on the orthogonality $E\left[y_{i,s}\left(\varepsilon_{it} - \varepsilon_{i,t-1}\right)\right] = 0$, $s = 0, \ldots, t-2$. Unfortunately, the standard GMM estimator obtained after first differencing has been found to

[1] Condition 2 is a stationarity condition that is imposed for analytical convenience. Anderson and Hsiao (1982) show that estimators exploiting this initial condition can be very sensitive to misspecification. For this reason we do not exploit the initial condition to construct our estimator.

suffer from substantial finite sample biases, especially when β is close to the unit circle. Such behavior can be understood by using the higher order bias formula of 2SLS (GMM)[2], which depends on 4 factors: "Explained" variance of the first stage reduced form equation, covariance between the stochastic disturbance of the structural equation and the reduced form equation, the number of instruments, and sample size:

$$E\left[\hat{\beta}_{2SLS}-\beta\right] \approx \frac{1}{n} \frac{(\text{number of instruments}) \times (\text{“covariance”})}{\text{“Explained” variance of the first stage reduced form equation}}$$

We now use this formula can explain the bias of GMM in the first difference set up.[3] Assume that $T = 4$. The first difference set up considers the following equation among others:

$$y_4 - y_3 = \beta\left(y_3 - y_2\right) + \varepsilon_4 - \varepsilon_3 \tag{2}$$

For the RHS variables it uses the instrument equation:

$$y_3 - y_2 = \left(\beta - 1\right) y_2 + \alpha + \varepsilon_3$$

Now calculate the R^2 for the first stage equation using the Ahn-Schmidt (AS) moments under "ideal conditions" where the researcher knows β in the sense that the nonlinear restrictions become linear restrictions: We would then use (y_2, y_1, y_0) as instruments. Assuming stationarity for symbols, but not using it as additional moment information, we can write $y_0 = \alpha/\left(1-\beta\right) + \eta_0$, where $\eta_0 \sim \left(0, \sigma_\varepsilon^2/\left(1-\beta^2\right)\right)$. It can be shown that the covariance between the structure error and the first stage error[4] is $-\sigma_\varepsilon^2$, and the "explained variance" in the first stage is equal to $\sigma_\varepsilon^2\left(1-\beta\right)/\left(\beta+1\right)$. Therefore, the ratio that determines the bias of 2SLS is equal to $-\left(1+\beta\right)/\left(1-\beta\right)$, which is equal to -19 for $\beta = .9$. For $n = 100$, this implies a percentage bias of -105.56.

We now turn to the long difference (LD) technique of Griliches and Hausman (1986). The LD technique is based on a single equation

$$y_{iT} - y_{i1} = \beta\left(y_{iT-1} - y_{i0}\right) + \left(\varepsilon_{iT} - \varepsilon_{i1}\right) \tag{3}$$

It is easy to see that the initial observation y_{i0} would serve as a valid instrument. To increase further the explanatory power of the instruments, we can use similar intuition as in Hausman

[2] As explained in Donald and Newey (2001) the formula is valid strictly speaking only when the number of instruments tends to infinity. It provides however a good approximation to the finite sample behavior of GMM estimators as far as the impact of additional instruments on bias is concerned. Kuersteiner (2000) shows that the same expression holds in a time series context if appropriate adjustments are made for the covariance term.

[3] It is possible to estimate the higher order bias, and subtract it off from the GMM estimator. Monte Carlo simulations reveal that this procedure does not work well when β is close to one. See Appendix A.

[4] This covariance is conditional on the set of instruments.

and Taylor (1983) or Ahn and Schmidt (1995), and see that the residuals $y_{iT-1} - \beta y_{iT-2}$, ..., $y_{i2} - \beta y_{i1}$ are valid instruments as well. We call this estimator the long difference estimator.

In order to understand why the LD technique may improve on the first difference, we use the higher order bias formula again. Assume that $T = 4$ as before. The LD set up is based on the equation

$$y_4 - y_1 = \beta (y_3 - y_0) + \varepsilon_4 - \varepsilon_1$$

It can be shown that the covariance between the first stage and second stage errors is $-\beta^2 \sigma_\varepsilon^2$, and the "explained variance" in the first stage is given by

$$-\sigma_\varepsilon^2 \frac{\left(2\beta^6 - 4\beta^4 - 2\beta^5 + 4\beta^2 + 4\beta - 2\beta^3 + 6\right)\sigma^2 + \beta^6 - \beta^4 + 2 - 2\beta^3}{\left(-2\beta - 3 + \beta^2\right)\sigma^2 - 1 + \beta^2},$$

where $\sigma^2 = \sigma_\alpha^2 / \sigma_\varepsilon^2$. Therefore, the ratio that determines the bias is equal to

$$-.37408 + \frac{2.5703 \times 10^{-4}}{\sigma^2 + 4.8306 \times 10^{-2}}$$

for $\beta = .9$. Note that the maximum value that this ratio can take in absolute terms is $-.37$, which is much smaller than -19 for the Ahn and Schmidt estimator. We therefore conclude that the long difference increases R^2 but decreases the covariance thus alleviating the "weak instruments" problem. Further, the number of instruments is smaller in the long difference specification so we should expect even smaller bias.

The LD estimator as discussed above uses additional instruments, which requires knowledge of β. A feasible version of the LD estimator therefore requires a preliminary consistent estimator of β. An obvious choice is Arellano and Bover's (1995) estimator $\widehat{b}_{GMM}$. We call such an estimator $\widehat{b}_{LD-AB}$: It is 2SLS on (3) using $\left(y_{i0}, y_{iT-1} - \widehat{b}_{GMM} \cdot y_{iT-2}, \ldots, y_{i2} - \widehat{b}_{GMM} \cdot y_{i1}\right)$ as instruments. Such instruments were previously used in Hausman, Newey and Taylor (1987). Using the preliminary consistent estimator $\widehat{b}_{LD-AB}$ to obtain residuals as instruments, we can also come up with another long difference estimator. Call it $\widehat{b}_{LD1}$. We can iterate it further, and come up with $\widehat{b}_{LD2}$, $\widehat{b}_{LD3}$, ... Another possibility is to interpret the LD as a GMM estimator based on

$$\begin{aligned}
E\left[y_{i0} \cdot \left(y_{iT} - y_{i1} - \beta\left(y_{iT-1} - y_{i0}\right)\right)\right] &= 0 \\
E\left[\left(y_{iT-1} - \beta y_{iT-2}\right) \cdot \left(y_{iT} - y_{i1} - \beta\left(y_{iT-1} - y_{i0}\right)\right)\right] &= 0 \\
&\vdots \\
E\left[\left(y_{i2} - \beta y_{i1}\right) \cdot \left(y_{iT} - y_{i1} - \beta\left(y_{iT-1} - y_{i0}\right)\right)\right] &= 0
\end{aligned}$$

We will call such estimator $\widehat{b}_{LDGMM}$.

In Table 1, we compare Monte Carlo[5] properties of various estimators with the LD. We examine $\widehat{b}_{AS}$, the estimator proposed by Ahn and Schmidt (1995), $\widehat{b}_{BB}$, the estimator proposed by Blundell and Bond (1998), and $\widehat{b}_{GMM}$, the estimator proposed by Arellano and Bover (1995), among others. Precise definitions of the estimators are given in Section 3.[6] We find that LD does better than the other estimators for large β and not significantly worse for moderate β.

3 Near Unit Root Approximation

In order to analyze the properties of the long difference estimator, we consider a local to non-identification asymptotic approximation.[7] We use the weak instrument approximation as in Staiger and Stock (1997) and Stock and Wright (2000). It is based on letting the correlation between instruments and regressors decrease at a prescribed rate as the sample size increases. In the case of the dynamic panel model it is the autoregressive parameter that determines the quality of lagged observations used as instruments. Here we consider model (1) for T fixed and $n \to \infty$ when also β_n tends to unity. We want to keep certain aspects of the model, in our case the quality of the instruments, constant as the sample size increases. The purpose of alternative asymptotic sequences is to obtain asymptotic approximations that better reflect finite sample properties, in our case poor identification, than standard asymptotics with fixed parameters would deliver.

The fact that $\beta_n \to 1$ with $n \to \infty$ is not meant to be a reflection of some realistic data generating process but is an analytical device. The asymptotics that we propose are also quite different from the near unit root literature in a pure time series context. Here, T is fixed and the degree of temporal dependence modelled by β_n is not directly relevant for our analysis. What matters are the implications for identification as discussed before.

We analyze the bias of the associated weak instrument limit distribution. We analyze the class of GMM estimators that exploit Ahn and Schmidt's (1997) complete set of moment conditions and show that a strict subset of the full set of moment restrictions should be used in

[5] We set $\sigma^2 = \sigma_\alpha^2 = 1$ for our Monte Carlo simulation. The results are based on 5000 runs.

[6] The version of Ahn and Schmidt's estimator (1995) adopted in this Monte Carlo simulation does not utlize homoscedasticity.

[7] The long difference technique was motivated by the higher order theory, and the related higher order theory is presented in Appendix A. Unfortunately, the second order bias calculations do not predict well the performance of the estimator near the unit circle where the model suffers from a near non-identification problem. See Blundell and Bond (1998), who related the problem to the analysis by Staiger and Stock (1997).

estimation in order to minimize bias. We show that the long difference estimator is a good approximation to the bias minimal procedure.

Following Ahn and Schmidt (1995,1997) we exploit the moment conditions

$$\begin{aligned} E\left[u_i u_i'\right] &= \sigma_\varepsilon^2 I + \sigma_\alpha^2 \mathbf{1}\mathbf{1}' \\ E\left[u_i y_{i0}\right] &= \sigma_{\alpha y_0}\mathbf{1} \end{aligned}$$

with $\mathbf{1} = [1, ..., 1]'$ a vector of dimension T and $u_i = [u_{i1}, ..., u_{iT}]'$. The moment conditions can be written more compactly as

$$d = \begin{bmatrix} \operatorname{vech} E\left[u_i u_i'\right] \\ E\left[u_i y_{i0}\right] \end{bmatrix} = \sigma_\varepsilon^2 \begin{bmatrix} \operatorname{vech} I \\ 0 \end{bmatrix} + \sigma_\alpha^2 \begin{bmatrix} \operatorname{vech} \mathbf{1}\mathbf{1}' \\ 0 \end{bmatrix} + \sigma_{\alpha y_0} \begin{bmatrix} 0 \\ \mathbf{1} \end{bmatrix} \tag{4}$$

where the redundant moment conditions have been eliminated by use of the vech operator which extracts the upper diagonal elements from a symmetric matrix. Representation (4) makes it clear that the vector $d.\in \mathbb{R}^{T(T+1)/2+T}$ is contained in a 3 dimensional subspace which is another way of stating that there are $G = T(T+1)/2 + T - 3$ restrictions imposed on d. This statement is equivalent to Ahn and Schmidt's (1997) analysis of the number of moment conditions.

GMM estimators are obtained from the moment conditions by eliminating the unknown parameters $\sigma_\varepsilon^2, \sigma_\alpha^2$ and $\sigma_{\alpha y_0}$. The set of all GMM estimators leading to consistent estimates of β can therefore be described by a $(T(T+1)/2+T) \times G$ matrix A which contains all the vectors spanning the orthogonal complement of $d.$. This matrix A satisfies

$$d'A = 0.$$

For our purposes it will be convenient to choose A such that

$$d'A = \left[E\left[u_{it}\Delta u_{is}\right], E\left[u_{iT}\Delta u_{ij}\right], E\left[\overline{u}_i \Delta u_{ik}\right], E\left[\Delta u_i' y_{i0}\right]\right],$$
$$s = 2, ..., T; t = 1, ...s-2; j = 2, ..., T-1; k = 2, ..., T$$

where $\Delta u_i = [u_{i2} - u_{i1}, ..., u_{iT} - u_{iT-1}]'$ and $\overline{u}_i = T^{-1}\sum_{t=1}^T u_{it}$. It becomes transparent that any other representation of the moment conditions can be obtained by applying a corresponding nonsingular linear operator $\tilde{C}$ to the matrix A. It can be checked that there exists a nonsingular matrix $\tilde{C}$ such that $d'AC = 0$ is identical to the moment conditions (4a)-(4c) in Ahn and Schmidt (1997).

We investigate the properties of (infeasible) GMM estimators based on

$$E\left[u_{it}\Delta u_{is}(\beta)\right] = 0, \quad E\left[u_{iT}\Delta u_{ij}(\beta)\right] = 0, \quad E\left[\overline{u}_i \Delta u_{ik}(\beta)\right] = 0, \quad E\left[y_{i0}\Delta u_{it}(\beta)\right] = 0$$

obtained by setting $\Delta u_{it}(\beta) \equiv \Delta y_{it} - \beta \Delta y_{it-1}$. Here, we assume that the instruments u_{it} are observable. We partition the moment conditions into two groups where the second group only contains instruments involving y_{i0}. In particular, let $g_{i1}(\beta)$ denote a column vector consisting of $u_{it}\Delta u_{is}(\beta), u_{iT}\Delta u_{ij}(\beta), \overline{u}_i \Delta u_{ik}(\beta)$. Also let $g_{i2}(\beta) \equiv [y_{i0}\Delta u_i(\beta)]$. It will turn out that for most estimators used in practice only these instruments are asymptotically relevant under near unit root asymptotics. Finally, let $g_n(\beta) \equiv n^{-3/2}\sum_{i=1}^n \left[g_{i1}(\beta)', g_{i2}(\beta)'\right]'$ with the optimal weight matrix $\Omega_n \equiv E\left[g_i(\beta_n) g_i(\beta_n)'\right]$. The infeasible GMM estimator of a possibly transformed set of moment conditions $C' g_n(\beta)$ then solves

$$b_{2SLS}\left(\tilde{C}\right) = \underset{\beta}{\operatorname{argmin}}\, g_n(\beta)' \tilde{C}\left(\tilde{C}'\Omega_n\tilde{C}\right)^+ \tilde{C}' g_n(\beta) \tag{5}$$

where $\tilde{C}$ is a $G \times r$ matrix for $1 \leq r \leq G$ such that $\tilde{C}$ is of full column rank r and rank $\left(\tilde{C}\left(\tilde{C}'\Omega\tilde{C}\right)^+ \tilde{C}'\right) \geq 1$. We use $\left(\tilde{C}'\Omega\tilde{C}\right)^+$ to denote the Moore-Penrose inverse. We thus allow the use of a singular weight matrix. Choosing r less than G allows to exclude certain moment conditions. Let $f_{i,1} \equiv -\partial g_{i1}(\beta)/\partial\beta$, $f_{i,2} \equiv -\partial g_{i2}(\beta)/\partial\beta$, and $f_n \equiv n^{-3/2}\sum_{i=1}^n \left[f'_{i1}, f'_{i,2}\right]'$. The infeasible 2SLS estimator can be written as

$$b_{2SLS}\left(\tilde{C}\right) - \beta_n = \left(f'_n\tilde{C}\left(\tilde{C}'\Omega_n\tilde{C}\right)^+ \tilde{C}' f_n\right)^{-1} f'_n\tilde{C}\left(\tilde{C}'\Omega_n\tilde{C}\right)^+ \tilde{C}' g_n(\beta_n). \tag{6}$$

We are now analyzing the behavior of $b_{2SLS} - \beta_n$ under local to unity asymptotics. We make the following assumptions.[8]

Condition 3 *Let*[9] $y_{it} = \alpha_i + \beta_n y_{it-1} + \varepsilon_{it}$ *where* $\beta_n = \exp(-c/n)$ *for some* $\infty > c > 0$ *where* $y_{i0} = \eta_{i0} + \alpha_i/(\beta_n - 1)$ *and* $\eta_{i0} = \varepsilon_{i0}/\left(1 - \beta_n^2\right)^{1/2}$.

[8] Kruiniger (2000) considers similar local-to-unity asymptotics.

[9] An alternative specification is the model $y_{it} = (1 - \beta_n)\alpha_i + \beta_n y_{it-1} + \varepsilon_{it}$. We analyze this model in an auxiliary appendix available on http://econ-www.mit.edu/faculty/jhausman/papers.htm.

We find that for this model, if the estimator b_{2SLS} is based solely on the condition

$$E\left[\overline{u}_i \Delta u_{ik}(\beta_0)\right] = 0 \tag{7}$$

then b_{2SLS} is consistent and thus weak instrument problems can be avoided in this case. If we consider estimators that do use all the moment conditions except (7) then we are in a situation where again the moment conditions in $g_{i1}(\beta)$ become asymptotically redundant. Now, however the limiting distribution of the non-redundant conditions has a noncentrality parameter.

More explicitly, if b_{2SLS} is the GMM estimator based on $g_{i2}(\beta)$ then it has the following nonstandard limiting distribution under the weak identification asymptotics adopted here

$$b_{2SLS} - \beta_0 \xrightarrow{d} \frac{(\mu + \xi_x)' C(C'\Omega C)^{-1} C'\xi_y}{(\mu + \xi_x)' C(C'\Omega C)^{-1} C'(\mu + \xi_x)} = X$$

where $\mu = -\mathbf{1}_{T-1}\sigma_\varepsilon^2/2$. It turns out that this limiting distribution is the same as the one ob-

Condition 4 *Assume that $\varepsilon_{it}, \eta_{i0}$ and α_i are independent and identically distributed across i for all t. Assume that ε_{it} satisfies $E[\varepsilon_{it}] = 0$ and $E\left[|\varepsilon_{it}|^4\right] < \infty$ for all i and $t = 1, ..., T$. Similarly, $E[\eta_{i0}] = 0$, $E[\alpha_i] = 0$, $E\left[|\eta_{i0}|^4\right] < \infty$ and $E[|\alpha_i|] < \infty$ for all i. Let $\tau = (\tau_1, ..., \tau_k) \in \mathbb{R}^k$ and $\varepsilon = (\varepsilon_{it_1}, ..., \varepsilon_{it_k})$, then $\phi_{t_1,...,t_k}(\tau) \equiv E\left[e^{i\tau'\varepsilon}\right]$ is the joint characteristic function with corresponding cumulant generating function $\ln \phi_{t_1,...,t_k}(\tau)$. The joint k-th order cross-cumulant function is*

$$\operatorname{cum}(\varepsilon_{it_1}, ..., \varepsilon_{it_k}) \equiv \left.\frac{\partial^k}{\partial\tau_1 \cdots \partial\tau_k} \ln \phi_{t_1,...,t_k}(\tau)\right|_{\tau=0}.$$

In the same way define the cumulant function $\operatorname{cum}(\varepsilon_{it_1}, ..., \varepsilon_{it_k}, \eta_{i0})$, $\operatorname{cum}(\varepsilon_{it_1}, ..., \varepsilon_{it_k}, \alpha_{i0})$ and $\operatorname{cum}(\varepsilon_{it_1}, ..., \varepsilon_{it_k}, \eta_{i0}, \alpha_i)$. Assume that $E\left[\varepsilon_{it}^2\right] = \sigma_\varepsilon^2$, $E\left[\eta_{i0}^2\right] = \sigma_\varepsilon^2 / \left(1 - \beta_n^2\right)$, $E\left[\alpha_i^2\right] = \sigma_\alpha^2$, $E[\varepsilon_{it}\varepsilon_{is}] = 0$ for $t \neq s$, $E[\varepsilon_{it}\eta_{i0}] = 0$, $E[\varepsilon_{it}\alpha_i] = 0$ for all t and i, $E[\alpha_i\eta_{i0}] = 0$ for all i.

Condition 5 *In addition to Condition 5 assume that ε_{it} and α_i are all mutually independent for all $t = 1, ..., T$.*

Remark 1 *The moment conditions summarized in Condition 4 comprise the moment conditions (A.1) to (A.4) of Ahn and Schmidt (1997). In addition we impose stationarity of y_{i0} in our approximating sequence indexed by β_n but do not exploit it in estimation.*

The critical assumption in Condition (3) is that $\beta_n = \exp(-c/n) = 1 - c/n + o(n^{-1})$. Note that linearity of the model implies that one could specify $\beta_n = -c/n$ whithout affecting the results. The parameter β_n controls the degree of covariation between regressors and instruments. For the case of GMM estimators based on first differences this covariance amounts to $E[\Delta y_t y_{t-2}] = O(\beta_n - 1) = -c/n$. Our specification thus implies an asymptotic lack of correlation more stringent than the one used by Staiger and Stock (1997). In Hahn and Kuersteiner (2002) it is argued that the faster n^{-1}-rate corresponds to a situation of near-nonidentification. Numerical and analytical calculations[10] in the context of our dynamic panel model show that $n^{-1/2}$-rate asymptotics do not capture the type of finite sample biases that we find important for the behavior of GMM estimators and that the specification $\beta_n = \exp(-c/n)$ delivers more accurate predictions.

Under the generating mechanism described in the previous definition the following Lemma can be established.

tained when $\beta = \exp(-c/\sqrt{n})$ as is shown in a separate Appendix also available at http://econ-www.mit.edu/faculty/jhausman/papers.htm.

[10] Available at http://econ-www.mit.edu/faculty/jhausman/papers.htm

Lemma 1 *Assume that Conditions 3 and 4 hold. Use the notation* $c_{t,s,\alpha,\alpha} = \operatorname{cum}(\varepsilon_{it}, \varepsilon_{is}, \alpha_i, \alpha_i)$. *For* T *fixed and as* $n \to \infty$

$$n^{-3/2} \sum_{i=1}^{n} f_{i,1} \xrightarrow{p} 0, n^{-3/2} \sum_{i=1}^{n} g_{i,1}(\beta_0) \xrightarrow{p} 0$$

and

$$n^{-3/2} \sum_{i=1}^{n} \left[f_{i,2}', g_{i,2}'(\beta_n)\right]' \xrightarrow{d} \left[\xi_x', \xi_y'\right]'.$$

The joint distribution of $\left[\xi_x', \xi_y'\right]' \sim N(0, \Sigma)$ *with* $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$ *where* $\Sigma_{11} = \delta I + \mathbb{K}_1$, $\Sigma_{12} = \delta M_1' + \mathbb{K}_3$, $\Sigma_{22} = \delta M_2 + \mathbb{K}_2$, $\Sigma_{12} = \Sigma_{21}'$, *where* $\mathbb{K}_1, \mathbb{K}_2$ *and* $\mathbb{K}_3$ *are defined in Appendix C.*

$$\delta = \frac{\sigma_\alpha^2 \sigma_\varepsilon^2}{c^2},$$

$$M_1 = \begin{bmatrix} -1 & 1 & & 0 \\ & \ddots & \ddots & \\ & 0 & \ddots & 1 \\ & & & -1 \end{bmatrix}, \qquad M_2 = \begin{bmatrix} 2 & -1 & & 0 \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2 \end{bmatrix}.$$

We also have

$$\frac{1}{n^2}\Omega_n = \begin{bmatrix} 0 & 0 \\ 0 & \Sigma_{22} \end{bmatrix} + o(1).$$

If in addition, Condition 5 holds then $\Sigma_{11} = \delta I$, $\Sigma_{12} = \delta M_1'$, $\Sigma_{22} = \delta M_2$.

Proof. See Appendix B. ■

It can be shown that the limiting normal distribution of Lemma 1 is singular. An alternative representation of the limiting distribution can be given by eliminating redundant dimensions from $\left[\xi_x', \xi_y'\right]'$. Let $\xi_y = \left[\xi_{y0}', \xi_{y1}\right]'$ where ξ_{y1} is the last element of ξ_y and $\xi_u = \left[\xi_x', \xi_{y1}\right]'$. Then $\xi_y = H\xi_u$, where H is a $(T-1) \times T$ matrix defined in Appendix C. This leads to an alternative representation of the limiting distribution.

Lemma 2 *Assume that Conditions 3 and 4 hold. Then*

$$n^{-3/2} \sum_{i=1}^{n} \left[f_{i,2}', g_{i,2}'(\beta_0)\right]' \xrightarrow{d} \left[\xi_x', \xi_u' H'\right]'.$$

Lemmas 1 and 2 imply that the moment conditions inovolving the initial condition dominate the limiting behavior of GMM estimators. As our analysis below shows, moment conditions of the form $E\left[y_{i,s}\left(\varepsilon_{i,t}-\varepsilon_{it-1}\right)\right]=0$ are asymptotically equivalent to conditions involving $y_{i,0}$ because the contribution of the innovations contained in $y_{i,s}$ asymptotically disappears. The reason for this result lies in the fact that asymptotically all instruments are weak but also in the stationarity condition which means that asymptotically $y_{i,0}$ is the dominating source of variation in $y_{i,t}$.

Using Lemma 1 the limiting distribution of $b_{2SLS}\left(\tilde{C}\right)-\beta_n$ is stated in the next corollary. For this purpose we define the augmented vectors $\xi_x^{\#}=\left[0,...,0,\xi_x'\right]'$ and $\xi_y^{\#}=\left[0,...,0,\xi_y\right]$ and partition $\tilde{C}=\left[\tilde{C}_0',\tilde{C}_1'\right]'$ such that $\tilde{C}'\xi_x^{\#}=\tilde{C}_1'\xi_x$. Let r_1 denote the rank of $\tilde{C}_1$.

Corollary 1 *Let $b_{2SLS}\left(\tilde{C}\right)-\beta_n$ be given by (6). If Conditions 3 and 4 are satisfied and if $r_1\geq 1$ then*

$$b_{2SLS}\left(\tilde{C}\right)-1\xrightarrow{d}\frac{\xi_x'\tilde{C}_1\left(\tilde{C}_1'\Sigma_{22}\tilde{C}_1\right)^{+}\tilde{C}_1'H\xi_u}{\xi_x'\tilde{C}_1\left(\tilde{C}_1'\Sigma_{22}\tilde{C}_1\right)^{+}\tilde{C}_1'\xi_x}=\xi\left(\tilde{C}_1,\Sigma_{22}\right)\tag{8}$$

Remark 2 *The restriction $r_1\geq 1$ insures that $b_{2SLS}\left(\tilde{C}\right)$ is based on moment conditions that use y_{i0} as an instrument. All estimators used in practice that we are aware of satisfy this condition.*[11]

Unlike the limiting distribution for the standard weak instrument problem, $\xi\left(\tilde{C},\Sigma_{22}\right)$, as defined in (8), is based on normal vectors that have zero mean. This degeneracy is generated by the presence of the fixed effect in the initial condition, scaled up appropriately to satisfy the stationarity requirement for the process y_{it}. Inspection of the proof shows that the usual concentration parameter appearing in the limit distribution is dominated by a stochastic component related to the fixed effect.

Lemma 1 and Corollary 1 are indicative of the fact, that under the weak instrument asymptotic approximation it makes sense to consider estimators that use only moment restrictions based on initial conditions. Conventional estimators for the dynamic panel model such as the estimators proposed by Arellano and Bover (1995) or Ahn and Schmidt (1997) on the other hand are based on both types of moment conditions. To have a benchmark of procedures under weak instrument asymptotics we define the following class of 2SLS estimators for the dynamic panel model.

[11] In an auxiliary appendix available from the authors we investigate the properties of $b_{2SLS}(C)$ when $r_1=0$.

Definition 1 *Let b^1_{2SLS} be defined as $b^1_{2SLS}\left(C_1,\tilde{\Omega}\right) \equiv \operatorname{argmin}_\beta g_{2,n}(\beta)' C_1 \left(C_1'\tilde{\Omega}C_1\right)^{-1} C_1' g_{2,n}(\beta)$, where $g_{2,n}(\beta) \equiv n^{-3/2}\sum_{i=1}^n g_{i2}(\beta)$, $\tilde{\Omega}$ is a symmetric positive definite $(T-1)\times(T-1)$ matrix of constants and C_1 is a $(T-1)\times r_1$ matrix of full column rank $r_1 \leq T-1$ such that $C_1'C_1 = I_{r_1}$.*

The limiting distribution of the class b^1_{2SLS} of estimators now follows immediately from Lemma 1 and Corollary 1. This result is summarized in the next Corollary.

Corollary 2 *Let b^1_{2SLS} be defined as in Definition 1. Then $b^1_{2SLS} - 1 \xrightarrow{d} \xi\left(C_1,\tilde{\Omega}\right)$.*

In Definition 1 the matrix C_1 is restricted to be of full column rank. As the next result shows, this restriction is without loss of generality.

Theorem 1 *Let $\tilde{C}$ be any matrix of dimension $G\times r$ and rank r such that $1 \leq r \leq G$. Partition $\tilde{C} = \left[\tilde{C}_0', \tilde{C}_1'\right]'$ where $\tilde{C}_0$ is a $q_3\times r$ matrix of rank r_0 and $\tilde{C}_1$ is a $(T-1)\times r$ matrix of rank $r_1 \geq 1$. Then there exists a $(T-1)\times r_1$ matrix C_1 of full column rank such $\xi\left(\tilde{C}_1,\Sigma_{22}\right) \overset{d}{=} \xi\left(C_1,\Sigma_{22}\right)$ and $C_1'C_1 = I_{r_1}$.*

This result is important for the analysis of the bias minimal estimator in the class of all estimators based on the Ahn and Schmidt moment conditions as well as for the analysis of the estimator that minimizes quadratic loss under near unit root asymptotic approximations. It shows that for these optimality considerations, the class b^1_{2SLS} fully describes the set of all possible estimators that implicitly or explicitly use y_{i0} as an instrument. As we show later, under the near non-identification asymptotics commonly used estimators are not optimal in this class.

Next we turn to the analysis of the asymptotic bias for the estimators b^1_{2SLS} of the dynamic panel model. Since the limit only depends on zero mean normal random vectors we can apply the results of Smith (1993).

Theorem 2 *Let $\xi\left(C_1,\tilde{\Omega}\right)$ be the limiting distribution of $b^1_{2SLS}-1$ in Definition 1 under Conditions 3 and 4. Let $\bar{D} \equiv (D+D')/2$, where $D = W\Sigma_{21}\Sigma_{11}^{-1}$ and $W = C_1\left(C_1'\tilde{\Omega}C_1\right)^{-1}C_1'$. Define L such that $\Sigma_{11} = LL'$. Let Γ be the matrix of eigenvectors with corresponding eigenvalues λ_i of $L'WL$. Assume $\lambda_1 \geq ... \geq \lambda_{T-1}$. Let $\Lambda = \operatorname{diag}(\lambda_1,..,\lambda_{T-1})$ and Λ_1 be the upper right block of Λ containing all the non-zero eigenvalues. Partition $\Gamma = [\Gamma_1,\Gamma_2]$ where Γ_1 are the eigenvectors corresponding to the nonzero eigenvalues of $L'WL$ such that $\Gamma_1\Lambda_1\Gamma_1' = L'WL$ and $\Gamma_1\Gamma_1' + \Gamma_2\Gamma_2' = I$. Then*

$$E\left[\xi\left(C_1,\tilde{\Omega}\right)\right] = M1\left(\bar{\lambda}^{-1},\Gamma_1'L'\bar{D}L\Gamma_1,\Lambda_1,r\right) \equiv \bar{\lambda}^{-1}\sum_{k=0}^{\infty}\frac{(1)_k\left(\frac{1}{2}\right)_{1+k}}{\left(\frac{r_1}{2}\right)_{1+k}k!}C^{1,k}_{1+k}\left(\Gamma_1'\bar{D}\Gamma_1, I_r - \bar{\lambda}^{-1}\Lambda_1\right)$$

where $r = \operatorname{rank}(W)$, $(a)_b$ *is the Pochhammer symbol* $\Gamma(a+b)/\Gamma(b)$, $C^{1,k}_{p+k}(.,.)$ *is a top order invariant polynomial defined by Davis (1980) and* $\bar{\lambda}$ *is the largest eigenvalue of* $L'WL$. *The mean* $E\left[\xi\left(C_1, \tilde{\Omega}\right)\right]$ *exists for* $r \geq 1$.
If in addition Condition 5 holds then

$$E\left[\xi\left(C_1, \tilde{\Omega}\right)\right] = M1\left(\bar{\lambda}^{-1}, \Gamma_1' \bar{D} \Gamma_1, \Lambda_1, r\right)$$

where $D = WM_1$ *and* $L = I$.

Proof. See Appendix B. ■

The Theorem shows that the bias of b^1_{2SLS} both depends on the choice of C_1 and the weight matrix $\tilde{\Omega}$. Note for example that $E\left[\xi(C_1, I_{T-1})\right] = \operatorname{tr} \bar{D}/r_1$.

The problem of minimizing the bias of b^1_{2SLS} by choosing optimal matrices C_1 and $\tilde{\Omega}$ does not seem to lead to an analytical solution but could in principle be carried out numerically for a given number of time periods T. For our purpose we are not interested in such an exact minimum. Instead we analyze optimal choices of C_1 minimizing bias for a given weight matrix. Two choices of $\tilde{\Omega}$ are most relevant in practice, namely $\tilde{\Omega} = I_{T-1}$ or $\tilde{\Omega} = \Sigma_{22}$. Under certain additional distributional restrictions we are able to obtain an analytical solution for the bias minimal estimator in the class of estimators defined in Definition 1. We use this analytical optimum as a bench mark against which we judge existing estimators that have been proposed in the literature.

Theorem 3 *Let* $\xi\left(C_1, \tilde{\Omega}\right)$ *be as defined in Definition 1. Assume that Conditions 3, 4 and 5 are satisfied. Let* $\bar{D} = (D + D')/2$ *where* $D = C_1\left(C_1' \tilde{\Omega} C_1\right)^{-1} C_1' M_1$. *Let* Γ_1 *be as defined in Theorem 2. Then*

$$\min_{\substack{C_1 \ s.t. \ C_1'C_1 = I_{r_1} \\ r_1 = 1,..,T-1}} \left|E\left[\xi(C_1, I_{T-1})\right]\right| = \min_{\substack{C_1 \ s.t. \ C_1'C_1 = I_{r_1} \\ r_1 = 1,..,T-1}} \left|E\left[\xi(C_1, \Sigma_{22})\right]\right|.$$

Moreover,

$$E\left[\xi\left(C_1, I_{T-1}\right)\right] = \operatorname{tr} \bar{D}/r_1.$$

Let $C_1^* = \operatorname{argmin}_{C_1} \left|E\left[\xi(C_1, I_{T-1})\right]\right|$ *subject to* $C_1'C_1 = I_{r_1}, r_1 = 1, .., T-1$. *Then* $C_1^* = \rho_i$ *where* ρ_i *is the eigenvector corresponding to the smallest eigenvalue* l_i *of* M_2. *Thus,* $\min_{C_1} \operatorname{tr} \bar{D}/r_1 = -\min l_i/2$. *As* $T \to \infty$ *the smallest eigenvalue of* M_2, $\min l_i \to 0$.

Theorem 3 shows that the estimator that minimizes the bias is based only on a single moment condition which is a linear combination of the moment conditions involving y_{i0} as instrument where the weights are the elements of the eigenvector ρ_i corresponding to the smallest eigenvalue of M_2.

Optimal estimators based on C_1^* are unintuitive to implement. Moreover, optimality only holds near the unit circle. For these reasons we do not recommend direct implementation of these optimal procedures. We instead use them as benchmarks against which we measure the performance of more easily implementable estimators.

We turn now to an analysis of bias and mean squared error of commonly used estimators under near unit root asymptotics. All the estimators we consider can be described in terms of linear transformations of the basic moment conditions discussed by Ahn and Schmidt (1995). In this section we carry out the analysis assuming that the innovations u_{it} are observable instruments. The case of feasible estimators where u_{it} is replaced with an estimate is analyzed in Section 4.

To represent the Arellano and Bond estimator based on the moment conditions $E\left[y_{i,s}\left(\varepsilon_{it}-\varepsilon_{i,t-1}\right)\right]=0$, $s=0,\ldots,t-2$ we define the matrices $C_0^{AB,n}$ and $C_1^{AB,n}$ in Appendix (C) such that the moment conditions for the Arellano and Bond estimator can be expressed as

$$\begin{aligned} g_{i,AB} &= C_0^{AB,n\prime}B\left(\Delta u_i\otimes u_i\right)+C_1^{AB,n\prime}y_{i0}\Delta u_i \\ &= C^{AB,n\prime}\begin{bmatrix} g_{i,1} \\ g_{i,2}\end{bmatrix}\end{aligned}$$

where $g_{i,AB}=\left[y_{i,0}\Delta u_{i2},y_{i,0}\Delta u_{i3},y_{i1}\Delta u_{i3},y_{i,0}\Delta u_{i4}...,y_{i2}\Delta u_{i4},....,y_{i0}\Delta u_{iT},....,y_{iT-2}\Delta u_{iT}\right]'$ and B is defined in Equation 20 in Appendix C. The Arellano and Bond estimator can therefore be written as

$$\hat{b}_{AB}-\beta_n=\left(f_n'C^{AB,n}\left(C^{AB,n\prime}\Omega_nC^{AB,n}\right)^{+}C^{AB,n\prime}f_n\right)^{-1}f_n'C^{AB,n}\left(C^{AB,n\prime}\Omega_nC^{AB,n}\right)^{+}C^{AB,n\prime}g_n\left(\beta_n\right).$$

One of the first instrumental variables estimators proposed in the literature is the estimator of Anderson and Hsiao (1981) based on the moment condition $E\left[\sum_{t=2}^{T-2}y_{i,t-2}\left(\varepsilon_{i,t}-\varepsilon_{i,t-1}\right)\right]=0$. It can be expressed in terms of a matrix $C^{AH,n}$ defined in the appendix as

$$\hat{b}_{AH}-\beta_n=\frac{C^{AH,n\prime}g_n\left(\beta_n\right)}{C^{AB,n\prime}f_n}.$$

To represent the estimator of Arellano and Bover (1995) define the $T-1\times T-1$ non-singular matrix B^* that transforms Δu into the Helmert transform of u_i. An explicit definition of B^*

is given in Equation (21) in Appendix C. The moment conditions of the Arellano and Bover (1995) estimator $\hat{b}_{GMM}$ defined in Appendix A, Equation (11) can be represented as

$$\begin{aligned} g_{i,GMM} &= C_0^{AB,n\prime} B \left(-B^* \otimes I_T\right) \left(\Delta u_i \otimes u_i\right) - C_1^{AB,n\prime} B^* y_{i0} \Delta u_i \\ &= C^{AB,n\prime} \begin{bmatrix} g_{i,1}^* \\ g_{i,2}^* \end{bmatrix} \end{aligned}$$

where

$$g_{i,GMM} = \left[y_{i,0} u_{i1}^*, y_{i,0} u_{i2}^*, y_{i1} u_{i2}^*, y_{i,0} u_{i3}^* ..., y_{i2} u_{i3}^*, ..., y_{i0} u_{iT-1}^*, ..., y_{iT-2} u_{iT-1}^*\right]',$$

and we have defined

$$g_{i,1}^* = B\left(u_i^* \otimes u_i\right) \text{ and } g_{i,2}^* = y_{i0} u_i^*$$

with $u_i^* = \left[u_{i1}^*, ..., u_{iT-1}^*\right]'$. Also define $g_n^* = n^{-3/2} \sum_{i=1}^n \left[g_{i,1}^{*\prime}, g_{i,2}^{*\prime}\right]'$ and $f_n^* = n^{-3/2} \sum_{i=1}^n \left[f_{i,1}^{*\prime}, f_{i,2}^{*\prime}\right]'$ where

$$f_{i,1}^* = B\left(I_T \otimes -B^*\right)\left(u_i \otimes \Delta y_{i,-1}\right) \text{ and } f_{i,2}^* = B^* y_{i0} \Delta y_{i,-1}$$

The Arellano and Bover estimator can therefore be written as

$$\hat{b}_{GMM} - \beta_n = \left(f_n^{*\prime} C^{AB,n} \left(C^{AB,n\prime} \Omega_n^* C^{AB,n}\right)^+ C^{AB,n\prime} f_n^*\right)^{-1} f_n^{*\prime} C^{AB,n} \left(C^{AB,n\prime} \Omega_n^* C^{AB,n}\right)^+ C^{AB,n\prime} g_n^*$$

where $\Omega_n^* = E g_n^* g_n^{*\prime}$.

We analyze the infeasible version of the GMM estimator based on the moment conditions (4a) and (4b) of Ahn and Schmidt (1997). We do not use the moment conditions (4c) to insure comparability with the Long Difference estimator where we do not exploit (4c). We call this estimator the Ahn and Schmidt estimator[12]. The moment conditions can be represented by defining the selector matrix $C^{AS,n}$ such that

$$\begin{aligned} g_{i,AS} &= C_0^{AS,n\prime} B \left(\Delta u_i \otimes u_i\right) + C_1^{AS,n\prime} y_{i0} \Delta u_i \\ &= C^{AS,n\prime} \begin{bmatrix} g_{i,1} \\ g_{i,2} \end{bmatrix} \end{aligned}$$

where a more explicit representation of $g_{i,AS}$ is given as

$$g_{i,AS} = \left[y_{i,0} \Delta u_{i2}, y_{i,0} \Delta u_{i3}, y_{i1} \Delta u_{i3}, ..., y_{i0} \Delta u_{iT}, ..., y_{iT-2} \Delta u_{iT}, u_{iT} \Delta u_{i2}, ..., u_{iT} \Delta u_{iT-1}\right]'.$$

[12] In this section we use observed innovations as instruments. In other words, for the near unit root asymptotics we do not consider nonlinear estimation of the AS estimator in order to simplify the limit theory while for the Monte Carlo simulations we do implement a nonlinear version of this estimator. The moment conditions used for estimation are the same in both cases.

The long difference estimator is based on the moments $Ey_{i0}(u_{iT} - u_{i1}) = 0$ as well as $Eu_{it}(u_{iT} - u_{i1}) = 0$ for $t = 2, ..., T-1$. We define

$$B^{LD} = \mathbf{1}'_{T-1} \otimes \begin{bmatrix} 0_{T-2,1} & I_{T-2} & 0_{T-2,1} \end{bmatrix} \tag{9}$$

such that

$$g_{i,1}^{LD} = B^{LD}(\Delta u_i \otimes u_i), \; g_{i,2}^{LD} = g_{i,2} = y_{i0}\Delta u_i.$$

Then define

$$C^{LD\prime} = \begin{bmatrix} I_{T-2} & 0_{T-2,T-1} \\ 0_{1,T-2} & \mathbf{1}'_{T-1} \end{bmatrix} \equiv \left[C_0^{LD\prime}, C_1^{LD\prime}\right]$$

such that

$$g_{i,LD} = C^{LD\prime} \begin{bmatrix} g_{i,1}^{LD} \\ g_{i,2}^{LD} \end{bmatrix}.$$

Let $f_{i,1}^{LD} = B^{LD}(\Delta y_i \otimes u_i)$, $f_{i,2}^{LD} = y_{i0}\Delta y_i$, $f_i^{LD\prime} = \left[f_{i,1}^{LD\prime}, f_{i,2}^{LD\prime}\right]$, $f_n^{LD} = n^{-3/2}\sum_{i=1}^n f_i^{LD}$ and $g_n^{LD} = n^{-3/2}\sum_{i=1}^n \left[g_{i,1}^{LD\prime}, g_{i,2}^{LD\prime}\right]$.The Long Difference estimator can then be written as

$$\hat{b}_{LD} - \beta_n = \left(f_n^{LD\prime} C^{LD,n}\left(C^{LD,n\prime}\Omega_n^{LD} C^{LD,n}\right)^+ C^{LD,n\prime} f_n^{LD}\right)^{-1} f_n^{LD\prime} C^{LD,n}\left(C^{LD,n\prime}\Omega_n^{LD} C^{LD,n}\right)^+ C^{LD,n\prime} g_n^{LD}.$$

where $\Omega_n^{LD} = E\left[g_n^{LD} g_n^{LD\prime}\right]$. The limiting distributions of the estimators $\hat{b}_{AB}$, $\hat{b}_{AH}$, $\hat{b}_{GMM}$, $\hat{b}_{AS}$ and $\hat{b}_{LD}$ are summarized in the following Lemma.

Lemma 3 *Assume Conditions 3, 4 and 5 are satisfied. Then*

$$\hat{b}_{AB} - 1 \xrightarrow{d} \frac{\xi_x' C_1^{AB}\left(C_1^{AB\prime}\Sigma_{22} C_1^{AB}\right)^+ C_1^{AB\prime} H\xi_u}{\xi_x' C_1^{AB}\left(C_1^{AB\prime}\Sigma_{22} C_1^{AB}\right)^+ C_1^{AB\prime}\xi_x} \equiv \xi_{AB},$$

$$\hat{b}_{AH} - 1 \xrightarrow{d} \frac{C_1^{AH\prime} H\xi_u}{C_1^{AH\prime}\xi_x} \equiv \xi_{AH},$$

$$\hat{b}_{GMM} - 1 \xrightarrow{d} \frac{\xi_x' B^{*\prime} C_1^{AB}\left(C_1^{AB\prime} B^*\Sigma_{22} B^{*\prime} C_1^{AB}\right)^+ C_1^{AB\prime} B^* H\xi_u}{\xi_x' B^{*\prime} C_1^{AB}\left(C_1^{AB\prime} B^*\Sigma_{22} B^{*\prime} C_1^{AB}\right)^+ C_1^{AB\prime} B^*\xi_x} \equiv \xi_{GMM},$$

$$\hat{b}_{AS} - 1 \xrightarrow{d} \frac{\xi_x' C_1^{AS}\left(C_1^{AS\prime}\Sigma_{22} C_1^{AS}\right)^+ C_1^{AS\prime} H\xi_u}{\xi_x' C_1^{AS}\left(C_1^{AS\prime}\Sigma_{22} C_1^{AS}\right)^+ C_1^{AS\prime}\xi_x} \equiv \xi_{AS},$$

$$\hat{b}_{LD} - 1 \xrightarrow{d} \frac{\xi_x' C_1^{LD}\left(C_1^{LD\prime}\Sigma_{22} C_1^{LD}\right)^+ C_1^{LD\prime} H\xi_u}{\xi_x' C_1^{LD}\left(C_1^{LD\prime}\Sigma_{22} C_1^{LD}\right)^+ C_1^{LD\prime}\xi_x} \equiv \xi_{LD}.$$

It follows that $\xi_{AS} \overset{d}{=} \xi_{AB} \overset{d}{=} \xi_{GMM}$ *and* $\xi_{LD} = \xi_{AH}$ *where* $\overset{d}{=}$ *means equal in distribution.*

One implication of Lemma 3 is that estimators which are asymptotically equivalent under standard asymptotics where identification is strong continue to be equivalent under the weak instrument asymptotics imployed here as long as they implicitly or explicitly use the same moment conditions involving y_{i0}. To be more concrete, consider the Arellano and Bond estimator based on the moment conditions $E\left[y_{is}\Delta u_{it}\left(\beta_n\right)\right]=0$ and an alternative estimator based on the moment conditions $E\left[y_{i0}\Delta u_{it}\left(\beta_n\right)\right]=0$ as well as $E\left[u_{is}\Delta u_{it}\left(\beta_n\right)\right]=0$ for $s=0,...,t-2$ and $t=2,...,T$. Both estimators are equivalent under standard asymptotic approximations. Under the alternative near unit root asymptotics of this Section the first version of the estimator converges in distribution to ξ_{AB} while the second estimator converges in distribution to the random variable $\xi_x'\Sigma_{22}^{-1}H\xi_u/\xi_x'\Sigma_{22}^{-1}\xi_x$. Note that the column rank of C_1^{AS} is $T-1$. By Theorem 1 it then follows that $\xi_x'\Sigma_{22}^{-1}H\xi_u/\xi_x'\Sigma_{22}^{-1}\xi_x=\xi_{AB}$. In other words the limiting distribution is unchanged in this example because both sets of moment conditions 'span' the same moment conditions involving y_{i0}. In Section 4 we see that this equivalence may break down once u_{it} is replaced with an estimated instrument.

The limiting distributions of the estimators $\hat{b}_{AB}$, $\hat{b}_{GMM}$, $\hat{b}_{AS}$ and $\hat{b}_{LD}$ can be expressed in terms of the genreal result of Corollary 1. We have $\xi_{AB}\overset{d}{=}\xi\left(C_1^{AB},\Sigma_{22}\right)$, $\xi_{AS}\overset{d}{=}\xi\left(C_1^{AS},\Sigma_{22}\right)$, $\xi_{GMM}\overset{d}{=}\xi\left(B^{*\prime}C_1^{AB},\Sigma_{22}\right)$ and $\xi_{LD}\overset{d}{=}\xi\left(C_1^{LD},\Sigma_{22}\right)$ where now $W^{LD}=C_1^{LD}\left(C_1^{LD\prime}\Sigma_{22}C_1^{LD}\right)^{+}C_1^{LD\prime}$ is singular of rank 1. The next result gives explicit formulas for the bias of the asymptotic limiting distribution of the Ahn and Schmidt, Arellano and Bond, Arellano and Bover and long difference estimators.

Corollary 3 *Let* $\xi\left(C_1^{AS},\Sigma_{22}\right)$, $\xi\left(C_1^{AB},\Sigma_{22}\right)$, $\xi\left(C_1^{GMM},\Sigma_{22}\right)$ *and* $\xi\left(C_1^{LD},\Sigma_{22}\right)$ *be the limiting distribution of the Ahn-Schmidt, Arellano-Bond, Arellano-Bover and Long Difference estimators under Condition 3. Let* $\bar{D}^a\equiv\left(D^a+D^{a\prime}\right)/2$*, where* $D^a=W^aM_1'$ *and* $W^a=C_1^a\left(C_1^{a\prime}\Sigma_{22}C_1^a\right)^{+}C_1^{a\prime}$ *for* $a=\{"AS","AB","GMM","LD"\}$*. Let* $\bar{\lambda}_a$ *be the largest eigenvalue of* W^a*. Then for* $a=\{"AS","AB","GMM"\}$*,*

$$E\left[\xi\left(C_1^a,\Sigma_{22}\right)\right]=M1\left(\bar{\lambda}_a^{-1},\bar{D}^a,W^a,T-1\right).$$

Let Γ *the matrix of eigenvectors of* W^{LD} *with* $\Gamma=[\Gamma_1,\Gamma_2]$ *where* Γ_1 *is the eigenvector corresponding to the nonzero eigenvalue of* W^{LD} *such that* $\Gamma_1\Lambda_1\Gamma_1'=W^{LD}$ *and* $\Gamma_1\Gamma_1'+\Gamma_2\Gamma_2'=I$*. Define* $z_1=L^{-1}\Gamma_1'\xi_x$ *and* $z_2=L^{-1}\Gamma_2'\xi_x$*. Then, for* $a=\{\text{“LD”}\}$*,*

$$E\left[\xi\left(C_1^a,\Sigma_{22}\right)\right]=M1\left(\bar{\lambda}_a^{-1},\Gamma_1'\bar{D}^a\Gamma_1,\Lambda_1,1\right).$$

Numerical procedures to evaluate the top order invariant polynomials appearing in the expressions for the bias are discussed in Smith (1993). Implementation of these methods is quite complicated and in our numerical work we rely on Monte Carlo integration to evaluate the moments of various procedures instead.

4 Near Unit Root Approximations for Feasible GMM

The long difference and Ahn and Schmidt (1997) estimators depend on first stage parameter estimates for the estimation of instruments and weight matrices. When identification of the model is weak these objects can no longer be estimated consistently and their presence potentially affects the limiting distribution. In this section we derive results for a general class of feasible estimators and then specialize these results to a few procedures of interest.

For the Arellano and Bover (1995) estimator we note that the optimal weight matrix Ω_n^{AB} is the block diagonal matrix with blocks $Ez_{it}z_{it}'$ where z_{it} has the representation

$$z_{it} = \begin{bmatrix} 0 & \cdots & & 0 \\ \beta_n^0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \beta_n^{t-2} & \cdots & \beta_n^0 & 0 \end{bmatrix} u_i + \begin{bmatrix} 1 \\ \beta_n \\ \\ \beta_n^{t-1} \end{bmatrix} y_{i0}.$$

Note that $E\left[u_i u_i'\right] = \sigma_\alpha^2 \mathbf{1}_T \mathbf{1}_T' + \sigma_\varepsilon^2 I_T$, $E\left[y_{i0}^2\right] = \sigma_\varepsilon^2 / \left(1-\beta_n^2\right) + \sigma_\alpha^2 / \left(\beta_n - 1\right)^2$ and $E\left[u_i y_{i0}\right] = \sigma_\alpha^2 / \left(\beta_n - 1\right)$. It follows that $n^{-2} E\left[z_{it} z_{it}'\right] \to \sigma_\alpha^2 / c^2 \mathbf{1}_t \mathbf{1}_t'$. Then by a strong law of large numbers it follows that $n^{-3} \sum_{i=1}^n z_{it} z_{it} \to_p \sigma_\alpha^2 / c^2 \mathbf{1}_t \mathbf{1}_t'$ and $\Omega_n^{AB} \to_p \Omega^{AB} = \sigma_\alpha^2 / c^2 \operatorname{diag}(1, \mathbf{1}_2 \mathbf{1}_2', ..., \mathbf{1}_{T-1} \mathbf{1}_{T-1}')$. The feasible version of the Arellano and Bover (1995) estimator can be written as

$$\begin{aligned} \hat{b}_{FGMM} - \beta_n &= \left(f_n^{*\prime} C^{AB,n} \left(\hat{\Omega}_n^{AB}\right)^+ C^{AB,n\prime} f_n^* \right)^{-1} f_n^{*\prime} C^{AB,n} \left(\hat{\Omega}_n^{AB}\right)^+ C^{AB,n\prime} g_n^* \\ & \overset{d}{\to} \frac{\xi_x' B^{*\prime} C_1^{AB} \left(\Omega^{AB}\right)^+ C_1^{AB\prime} B^* H \xi_u}{\xi_x' B^{*\prime} C_1^{AB} \left(\Omega^{AB}\right)^+ C_1^{AB\prime} B^* \xi_x} \equiv \xi_{FGMM}. \end{aligned}$$

Note that $\sigma_\varepsilon^2 \Omega^{AB} = C_1^{AB\prime} B^* \Sigma_{22} B^{*\prime} C_1^{AB}$ such that $\xi_{FGMM} = \xi_{GMM}$. Using the limiting representation for the Arellano and Bover estimator we consider feasible two stage estimators where the vector u_i is replaced by $\hat{u}_i = \left[y_{i1} - \hat{b}_{GMM} y_{i0}, ..., y_{iT} - \hat{b}_{GMM} y_{iT-1}\right]'$. It follows that $\hat{u}_{it} = u_{it} + \left(\hat{b}_{GMM} - \beta_n\right) y_{it-1}$. We can then write

$$\hat{u}_i = u_i + \left(\hat{b}_{GMM} - \beta_n\right) \begin{bmatrix} 0 & \cdots & & 0 \\ \beta_n^0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \beta_n^{T-2} & \cdots & \beta_n^0 & 0 \end{bmatrix} u_i + \left(\hat{b}_{GMM} - \beta_n\right) \begin{bmatrix} 1 \\ \beta_n \\ \\ \beta_n^{T-1} \end{bmatrix} y_{i0}. \quad (10)$$

Using more compact notation this can be written as $\hat{u}_i = u_i + \left(\hat{b}_{GMM} - \beta_n\right)\left(\bar{B}_n u_i + \bar{b}_n y_{i0}\right)$ where $\bar{B}_n$ and $\bar{b}_n$ are defined corresponding to (10). The feasible moment conditions then are

$$\hat{g}_{i1} = B\left(I_{T(T-1)} + \left(\hat{b}_{GMM} - \beta_n\right)\left(I_{T-1} \otimes \bar{B}_n\right)\right)(\Delta u_i \otimes u_i) + \left(\hat{b}_{GMM} - \beta_n\right) B\left(I_{T-1} \otimes \bar{b}_n\right) g_{i2}.$$

In the same way the derivative $\hat{f}_{i,1} = \partial \hat{g}_{i1}/\partial\beta$ is given by

$$\hat{f}_{i,1} = B\left(I_{T(T-1)} + \left(\hat{b}_{GMM} - \beta_n\right)\left(I_{T-1} \otimes \bar{B}_n\right)\right)(\Delta y_{i,-1} \otimes u_i) + \left(\hat{b}_{GMM} - \beta_n\right) B\left(I_{T-1} \otimes \bar{b}_n\right) f_{i2}.$$

These representations are suggestive of the fact that under the near unit root asymptotics, two stage estimators are not feasible in the usual sense because the first stage estimator $\hat{b}_{GMM}$ is not consistent and thus affects the limiting distribution of the two stage estimator. This problem is in fact the same as the inconsistency found in 2SLS under weak instrument asymptotics as shown by Staiger and Stock (1997). The next lemma establishes the distributional properties of sample moments that are used to define the two stage estimator.

Lemma 4 *Assume Conditions 3 and 4 are satisfied. Then*

$$n^{-3/2}\sum_{i=1}^{n}\left[\hat{f}'_{i,1}, f'_{i,2}, \hat{g}'_{i,1} g'_{i,2}\right]' \xrightarrow{d} \begin{bmatrix} \xi_{GMM} B\left(I_{T-1} \otimes \mathbf{1}_T\right) & 0 \\ I_{T-1} & 0 \\ 0 & \xi_{GMM} B\left(I_{T-1} \otimes \mathbf{1}_T\right) \\ 0 & I_{T-1} \end{bmatrix} \begin{bmatrix} \xi_x \\ H\xi_u \end{bmatrix}.$$

The feasible 2SLS estimator not only relies on estimated residuals but also is based on an estimate of the optimal weight matrix. For the estimate of the weight matrix we need to keep in mind that g_i is a function of β_n because it depends on $\Delta u_i(\beta_n)$. To construct estimators of the second moments of g_i β_n needs to be replaced by an estimate. Thus the weight matrix is defined as

$$\hat{\Omega}_n = n^{-3}\sum_{i=1}^{n}\left[\hat{g}'_{i,1}\left(\hat{b}_{GMM}\right), g'_{i,2}\left(\hat{b}_{GMM}\right)\right]'\left[\hat{g}'_{i,1}\left(\hat{b}_{GMM}\right), g'_{i,2}\left(\hat{b}_{GMM}\right)\right]$$

where it becomes transparant that under near unit root asymptotics the estimation error does not vanish even asymptotically.

Lemma 5 *Assume Conditions 3 and 4 are satisfied. Let*

$$R_\xi\left(\xi_{GMM}\right) = \begin{bmatrix} \xi_{GMM} B\left(I_{T-1} \otimes \mathbf{1}_T\right) & -\xi_{GMM}^2 B\left(I_{T-1} \otimes \mathbf{1}_T\right) \\ -\xi_{GMM} I_{T-1} & I_{T-1} \end{bmatrix}.$$

It then follows that

$$\hat{\Omega}_n \xrightarrow{d} R_\xi\left(\xi_{GMM}\right)\Sigma R_\xi\left(\xi_{GMM}\right)'.$$

The general form of the feasible two stage estimator is given by

$$\hat{b}_{F2SLS} - \beta_n = \left(\hat{f}_n' C \left(C'\hat{\Omega}_n C\right) C'\hat{f}_n\right)^{-1} \hat{f}_n' C \left(C'\hat{\Omega}_n C\right)^{+} C'\hat{g}_n\left(\beta_n\right)$$

where $\hat{f}_n = n^{-3/2}\sum_{i=1}^n \left[\hat{f}_{i,1}', f_{i,2}'\right]'$ and $\hat{g}_n\left(\beta_n\right) = n^{-3/2}\sum_{i=1}^n \left[\hat{g}_{i,1}'\left(\beta_n\right), g_{i,2}'\left(\beta_n\right)\right]'$. The next Corollary summarizes the limiting distribution of the feasible two stage estimator which is based on the Arellano and Bover estimator in the first stage.

Theorem 4 *Assume Conditions 3 and 4 are satisfied. Let B be as defined in Equation 20 in Appendix C. Let*

$$\hat{\xi}_x\left(\xi_{GMM}\right) = \begin{bmatrix} \xi_{GMM} B\left(\xi_x \otimes \mathbf{1}_T\right) \\ \xi_x \end{bmatrix}, \ \hat{\xi}_y\left(\xi_{GMM}\right) = \begin{bmatrix} \xi_{GMM} B\left(H\xi_u \otimes \mathbf{1}_T\right) \\ H\xi_u \end{bmatrix}.$$

Then it follows that

$$\hat{b}_{F2SLS} - \beta_n \xrightarrow{d} \frac{\hat{\xi}_x\left(\xi_{GMM}\right)' C\left(C' R_\xi\left(\xi_{GMM}\right)\Sigma R_\xi\left(\xi_{GMM}\right)' C\right)^{+} C'\hat{\xi}_y\left(\xi_{GMM}\right)}{\hat{\xi}_x\left(\xi_{GMM}\right)' C\left(C' R_\xi\left(\xi_{GMM}\right)\Sigma R_\xi\left(\xi_{GMM}\right)' C\right)^{+} C'\hat{\xi}_x\left(\xi_{GMM}\right)} = \xi_{F2SLS}.$$

As before the long difference estimator requires a slightly modified treatment because the moment conditions are more easily expressed directly as functions of the underlying innovations. The long difference estimator is based on the moments $E\left[y_{i0}\left(u_{iT} - u_{i1}\right)\right] = 0$ as well as $E\left[u_{it}\left(u_{iT} - u_{i1}\right)\right] = 0$ for $t = 2, ..., T-1$. As before, we replace unobserved innovations u_{it} by estimates $\hat{u}_{it}$ based on $\hat{b}_{GMM}$ such that $g_{i,1}^{LD} = B^{LD}\left(\Delta u_i \otimes \hat{u}_i\right)$ and $\hat{f}_{i,1} = B^{LD}\left(\Delta y_i \otimes \hat{u}_i\right)$ where B^{LD} is defined in (9). Define $\hat{f}_n^{LD} = n^{-3/2}\sum_{i=1}^n \left[\hat{f}_{i,1}^{LD\prime}, f_{i,2}'\right]'$ and $\hat{g}_n\left(\beta_n\right) = n^{-3/2}\sum_{i=1}^n \left[\hat{g}_{i,1}^{LD\prime}\left(\beta_n\right), g_{i,2}'\left(\beta_n\right)\right]'$. The feasible long difference estimator then is defined as

$$\hat{b}_{FLD} - \beta_n = \frac{\hat{f}_n^{LD\prime} C^{LD,n}\left(C^{LD,n\prime}\hat{\Omega}_n^{LD} C^{LD,n}\right)^{+} C^{LD,n\prime}\hat{g}_n^{LD}}{\hat{f}_n^{LD\prime} C^{LD,n}\left(C^{LD,n\prime}\hat{\Omega}_n^{LD} C^{LD,n}\right)^{+} C^{LD,n\prime}\hat{f}_n^{LD}}$$

where the weight matrix is estimated as

$$\hat{\Omega}_n^{LD} = n^{-3}\sum_{i=1}^n \left[\hat{g}_{i,1}^{LD\prime}\left(\hat{b}_{GMM}\right), g_{i,2}'\left(\hat{b}_{GMM}\right)\right]' \left[\hat{g}_{i,1}^{LD\prime}\left(\hat{b}_{GMM}\right), g_{i,2}'\left(\hat{b}_{GMM}\right)\right].$$

The next result establishes the limiting distribution of the feasible long difference estimator.

Theorem 5 *Assume Conditions 3 and 4 are satisfied. Let B^{LD} be as defined in (9). As before we use the notation*

$$\hat{\xi}_x^{LD}\left(\xi_{GMM}\right) = \begin{bmatrix} \xi_{GMM} B^{LD}\left(\xi_x \otimes \mathbf{1}_T\right) \\ \xi_x \end{bmatrix}, \ \hat{\xi}_y^{LD}\left(\xi_{GMM}\right) = \begin{bmatrix} \xi_{GMM} B^{LD}\left(H\xi_u \otimes \mathbf{1}_T\right) \\ H\xi_u \end{bmatrix}.$$

Then, it follows that

$$
\begin{aligned}
& \hat{b}_{FLD} - \beta_n \\
& \xrightarrow{d} \frac{\hat{\xi}_x^{LD}(\xi_{GMM})' C^{LD} \left(C^{LD\prime} R_\xi^{LD}(\xi_{GMM}) \Sigma R_\xi^{LD}(\xi_{GMM})' C^{LD} \right)^+ C^{LD\prime} \hat{\xi}_y^{LD}(\xi_{GMM})}{\hat{\xi}_x^{LD}(\xi_{GMM})' C^{LD} \left(C^{LD\prime} R_\xi^{LD}(\xi_{GMM}) \Sigma R_\xi^{LD}(\xi_{GMM})' C^{LD} \right)^+ C^{LD\prime} \hat{\xi}_x^{LD}(\xi_{GMM})} \\
\equiv \; & \xi_{FLD}.
\end{aligned}
$$

Furthermore, it follows that $\xi_{FLD} = \xi_{LD}$.

Iterated versions of the long difference estimator have limiting distributions that can be defined recursively for the j-th step. If $\hat{b}_{FLD,j}$ is the feasible long difference estimator based on the j-th iteration then its limiting distribution is given by

$$
\begin{aligned}
& \hat{b}_{FLD,j} - \beta_n \\
& \xrightarrow{d} \frac{\hat{\xi}_x^{LD}(\xi_{FLD,j-1})' C^{LD} \left(C^{LD\prime} R_\xi^{LD}(\xi_{FLD,j-1}) \Sigma R_\xi^{LD}(\xi_{FLD,j-1})' C^{LD} \right)^+ C^{LD\prime} \hat{\xi}_y^{LD}(\xi_{FLD,j-1})}{\hat{\xi}_x^{LD}(\xi_{FLD,j-1})' C^{LD} \left(C^{LD\prime} R_\xi^{LD}(\xi_{FLD,j-1}) \Sigma R_\xi^{LD}(\xi_{FLD,j-1})' C^{LD} \right)^+ C^{LD\prime} \hat{\xi}_x^{LD}(\xi_{FLD,j-1})} \\
\equiv \; & \xi_{FLD,j}
\end{aligned}
$$

From Theorem 5 it follows immediately, that $\xi_{FLD,j}$ does not depend on $\xi_{FLD,j-1}$ and in fact $\xi_{FLD,j} = \xi_{LD}$ for all j. We use the results in Theorem 4 and 5 to evaluate bias and MSE of feasible estimators. Due to the complex nature of the limiting distribution for the general feasible 2SLS estimator these evaluations can not be done analytically. Numerical evaluation on the other hand is possible.

5 Numerical Evaluation of Near Unit Root Approximations

We calibrate our numerical evaluations to $\beta = .95$ and $n = 100$ which leads to a value of $c = 100(-\log .95)$. We set $\sigma_\alpha^2 = \sigma_\varepsilon^2 = 1$ and assume that ε_{it} and α_i are independent as in Condition 5 in order to eliminate nuisance parameters related to higher order moments of the innovations. We then generate 10,000 random draws from the distribution of ξ and ζ. Using these values we compute the asymptotic distribution of various estimators as established in Corollary 1, Lemma 3 and Theorems 4 and 5.

Table 2 reports numerical values for the biases of the Arellano and Bond, Arellano and Bover and Ahn and Schmidt estimators under near unit root asymptotics and compares them with biases for the long difference estimator as well as the bias minimal estimator. We report median biases for numerically evaluated biases in columns AB/AS/GMM and LD/AH while the last column reports theoretical lowerbounds for the mean bias based on Theorem 3. For AB/AS/GMM median and mean are almost identical while for LD/AH the mean does not seem to be reliably evaluated by simulation techniques although it exists theoretically. Despite the fact that long difference does not achieve the same bias reduction as the fully optimal estimator, it has significantly less bias than the more commonly used implementations of the GMM estimator.

Lemma 3 also shows that under the near unit root asymptotics adopted here, there is no difference between the Ahn and Schmidt (AS) estimator, the Arellano and Bover (GMM) and the Arellano and Bond (AB) procedure. We now compare predictions of the near unit root approximation and the higher order approximation reported in Section A.1 to the actual finite sample performance of different estimators. Since there is no difference between AS, AB and GMM under local to unity asymptotics we focus on the GMM procedure for which we have reported Monte Carlo results in Table 9. Considering the results in Table 9 we note that for GMM the bias predicted by higher order asymptotic theory for $\beta = .95$, $n = 100$ and $T = 5$ is -3.84 while the prediction of the near unit root approximation in Table 2 is -.66. The actual small sample bias found in Monte Carlo experiments and reported in Table 9 is -.598. For the LD estimator the higher order theory predicts a bias of -1.8, the near unit root theory predicts a bias of -.25 and the actual small sample bias is -.185. These results suggest that the near unit root theory does a much better job of approximating the bias near the unit circle than the higher order theory which tends to overestimate bias by a large margin. In Table 3 we report the median bias of the feasible estimators under near unit root approximations derived in Section 4. For the feasible version of the Arellano and Bover estimator (FGMM) the results are the same as for the infeasible version, while the feasible Ahn and Schmidt[13] estimator shows an increase in the median bias relative to the infeasible version. This is an indication that the additional overidentifying restrictions, when combined with estimated instruments, further increase the bias problems of this particular specification. The feasible long difference estimator on the other hand is invariant to first stage estimation of the instruments and has the same distribution as the infeasible version.

Turning now to the same comparison for the MSE we note that the higher order prediction for the MSE of GMM in Table 9 is 10.24, for the near unit root approximation in Table 4 it is .7411 and the actual MSE from Table 9 is .4761. As before for the median bias feasible GMM has the same mean squared error as the infeasible version, a fact that is predicted by our theory.

[13] The poor properties of the limiting distribution for the feasible version of the Ahn and Schmidt estimator may be due to the fact that for the near unit root asymptotics we use estimated residuals rather than a nonlinear estimator.

We see again that even though the near unit root approximation is not entirely accurate it does much better in terms of predicting the right magnitude of the MSE than is the case for the higher order theory. It also should be noted that in the case of the near unit root approximation a number of moment conditions become irrelevant asymptotically. The associated drop in the degree of overidentification leads to an increase in the size of the variance. This is also the reason why for the LD estimator the MSE does not exist under the near unit root approximation (this follows by checking the conditions for the existence of moments in Smith 1993, p.272). Under these asymptotics, the LD estimator asymptotically does only depend on one instrument, the initial observation y_{i0}, and thus does not posses enough overidentifying restrictions for the MSE to exist. This remains true for the feasible version of the estimator which, under near unit root asymptotics, has the same distributional properties as the infeasible version.

In order to better understand the trade off between bias reduction and quadratic error loss under the near unit root asymptotics we now turn to a more detailed analysis of the mean squared error implied by the asymptotic distribution under Condition 3. Since $b_{2SLS}^1 - 1 \xrightarrow{d} \xi(C_1, \Sigma_{22})$ we take $E\left[\xi(C_1, \Sigma_{22})\right]^2$ as a measure for the MSE of b_{2SLS}^1. Unlike in the case of bias it is not feasible to find analytical optima for the problem of minimizing $E\left[\xi\left(C_1, \Sigma_{22}\right)\right]^2$ with respect to C_1.

Nevertheless, for a given value of T and a particular point in the parameter space, such optimization can be carried out numerically by Monte Carlo integration techniques and subsequent numerical optimization. We carry out such a numerical exercise to establish that conventional estimators which are optimal under standard first order asymptotic theory fail to be admissible under weak identification circumstances. In our numerical examples we calibrate the asymptotic distribution to the case where $\beta = .95$ and $n = 100$. This implies a value for $c = 100(-\log .95)$. In Table 4 we compare $E\left[\xi\left(C_1^*, \Sigma_{22}\right)\right]^2$ to $E\left[\xi\left(C_1^{AS}, \Sigma_{22}\right)\right]^2$ for different values of T. Here

$$C_1^* = \underset{\substack{C_1 \text{ s.t. } C_1' C_1 = I_{r_1} \\ r_1 = 3, .., T-1}}{\arg\min} E\left[\xi\left(C_1, \Sigma_{22}\right)\right]^2 .$$

Note that the minimal dimension of C_1 is 3 to guarantee the existence of the second moment. The results of Table 4 show that the MSE of the AS, AB and GMM estimators is clearly dominated by the optimal procedure. It also shows that the optimal number of moment conditions increases slowly with T and is much lower than the number of moments used by Ahn and Schmidt. In fact, the optimal procedure uses only about 1/10 of the available moments.

These results show that the optimality arguments under standard asymptotic approximations break down when the model is weakly identified. While it is optimal from an asymptotic variance and thus MSE point of view under standard asymptotics to use as many instruments as possible, this argument is invalid under the near non-identification asymptotics we use here. The GMM estimator that minimizes the asymptotic MSE under these circumstances uses a subset of moment conditions that is much smaller than the set of all available moment conditions. Some intuition can be gained from our analytical results regarding bias minimization. In Theorem 3 it is shown that the bias minimal procedure in fact only uses a single moment condition. When focusing on the MSE this bias reduction needs to be balanced off against efficiency considerations. Our numerical evaluation of the MSE minimization problem shows that the optimum is an interior one with respect to the number of moment conditions used. This implies that neither the Bias minimal estimator that uses only one instrument, nor the first order efficient estimator that uses all available instruments is optimal in the MSE sense under the local to unity asymptotics adopted here.

However, we can not compare the MSE of the Ahn and Schmidt and Arellano and Bond estimators to the long difference estimator analytically because under the near unit root asymptotic limiting distribution the long difference estimator has unbounded second moments. In order to gain some insight into the dispersion of the long difference estimator we report inter quartile ranges of various procedures in Table 5. The results are further evidence of the fact that feasible GMM is the same as infeasible GMM while on the other hand the feasible Ahn and Schmidt estimator is more dispersed than the infeasible version. The lack of moments problem for the long difference estimator manifests itself in higher inter quartile ranges. The findings with respect to the relationship between bias and dispersion of estimators are somewhat similar to our findings in Hahn, Hausman and Kuersteiner (2004) where we found that 2SLS estimators do best in terms of IQR but have significantly larger biases while estimators that do better with bias such as LIML or Fuller's estimator, tend to have larger IQR than 2SLS.

At the same time the Monte Carlo results are consistent with the analytical result that the optimal estimator does not use all the moment conditions. Indeed, the optimal estimator uses only a small fraction of the moment conditions. The long difference estimator does well in terms of MSE because of its small amount of bias and its use of a limited set of moment conditions.

Implementation of a procedure that is optimal in terms of its MSE under non-standard asymptotics is not recommended because such a procedure depends on the unknown parameter β through δ. Instead, our Monte Carlo results show that the long difference estimator is a reasonable, very easily implementable approximation to such an optimal estimator.

6 Conclusion

We have considered the dynamic panel data model where it has been recognized that the usual IV (GMM) estimators have substantial bias for a large positive β, which commonly occurs in applied research. Our approach is an IV estimator that uses a reduced set of instruments, in particular "long differences". This estimator leads to a significant reduction in bias as predicted by a "weak instrument" analysis. Further, the long difference estimator does well in MSE comparisons with existing estimators.

We then conduct a "local to unity" analysis to analyze why the long difference estimator does significantly better than traditional second order theory predicts. We find that near the unit circle the optimal estimator in terms of bias uses only one moment restriction, similar to the long difference estimator. We also find that the optimal estimator in terms of MSE uses a much restricted set of moment restrictions, typically about 1/10 of the available restrictions. Thus, our analysis demonstrates that the usual first order asymptotic advice of using all of the moment restrictions does not provide proper guidance in the dynamic panel data model for large β. Instead a restricted set of moment conditions lead to a better estimator. The long difference estimator, which uses a restricted set of moment conditions, provides such an improved estimator as our Monte-Carlo results demonstrate.

Appendix

A Higher Order Theory

We consider a version of the GMM estimator developed by Arellano and Bover (1995), which simplifies the characterization of the "weight matrix" in GMM estimation. We define the innovation $u_{it} \equiv \alpha_i + \varepsilon_{it}$. Arellano and Bover (1995) eliminate the fixed effect α_i in (1) by applying Helmert's transformation

$$u_{it}^* \equiv \sqrt{\frac{T-t}{T-t+1}} \left[u_{it} - \frac{1}{T-t} \left(u_{i,t+1} + \cdots + u_{iT} \right) \right], \qquad t = 1, \ldots, T-1$$

instead of first differencing.[14] The transformation produces

$$y_{it}^* = \beta x_{it}^* + \varepsilon_{it}^*, \qquad t = 1, \ldots, T-1,$$

where $x_t^* \equiv y_{i,t-1}^*$. Let $z_{it} \equiv (y_{i0}, \ldots, y_{it-1})'$. Our moment restriction is summarized by $E\left[z_{it}\varepsilon_{it}^*\right] = 0$, $t = 1, \ldots, T-1$. It can be shown that, with the homoskedasticity assumption on ε_{it}, the optimal "weight matrix" is proportional to a block-diagonal matrix, with typical diagonal block equal to $E\left[z_{it}z_{it}'\right]$. Therefore, the optimal GMM estimator is equal to

$$\widehat{b}_{GMM} \equiv \frac{\sum_{t=1}^{T-1} x_t^{*\prime} P_t y_t^*}{\sum_{t=1}^{T-1} x_t^{*\prime} P_t x_t^*} = \frac{\sum_{t=1}^{T-1} x_t^{*\prime} P_t x_t^* \cdot \widehat{b}_{2SLS,t}}{\sum_{t=1}^{T-1} x_t^{*\prime} P_t x_t^*} \tag{11}$$

where $\widehat{b}_{2SLS,t}$ denotes the 2SLS of y_t^* on x_t^*:

$$\widehat{b}_{2SLS,t} \equiv \frac{x_t^{*\prime} P_t y_t^*}{x_t^{*\prime} P_t x_t^*}, \qquad t = 1, \ldots, T-1$$

and $x_t^* \equiv (x_{1t}^*, \cdots, x_{nt}^*)'$, $y_t^* \equiv (y_{1t}^*, \cdots, y_{nt}^*)'$, $Z_t \equiv (z_{1t}, \cdots, z_{nt})'$, and $P_t \equiv Z_t \left(Z_t' Z_t\right)^{-1} Z_t'$. Therefore, the GMM estimator $\widehat{b}_{GMM}$ may be understood as a linear combination of the 2SLS estimators $\widehat{b}_{2SLS,1}, \ldots, \widehat{b}_{2SLS,T-1}$. It has long been known that the 2SLS may be subject to substantial finite sample bias. See Nagar (1959), Rothenberg (1983), Bekker (1994), and Donald and Newey (2001), Hahn and Hausman (2002) and Kuersteiner (2000) for related discussion. It is therefore natural to conjecture that a linear combination of the 2SLS may be subject to quite substantial finite sample bias. This is indeed the case. In the first column of Table 6, we summarized finite sample bias of $\widehat{b}_{GMM}$ as a fraction of β approximated by 5,000 Monte Carlo runs.

[14] Arellano and Bover (1995) note that the efficiency of the resultant GMM estimator is not affected whether or not Helmert's transformation is used instead of first differencing.

A.1 Bias Correction

Many econometric estimators, say $\widehat{\theta}$ for θ based on a sample of size n allow for an expansion of the form:

$$\sqrt{n}\left(\widehat{\theta}-\theta\right)=\theta^{(1)}+\frac{1}{\sqrt{n}}\theta^{(2)}+\frac{1}{n}\theta^{(3)}+o_p\left(\frac{1}{n}\right) \tag{12}$$

where $\theta^{(1)}$, $\theta^{(2)}$ and $\theta^{(3)}$ are of order $O_p(1)$. Typically $\theta^{(1)}$ has mean zero and converges in distribution to a normal distribution. Ignoring the $o_p\left(n^{-1}\right)$ term, and taking an expectation, we obtain the "approximate mean" of $\sqrt{n}\left(\widehat{\theta}-\theta\right)$ equals $\frac{1}{\sqrt{n}}E\left[\theta^{(2)}\right]$. We can therefore understand $\frac{1}{n}E\left[\theta^{(2)}\right]$ as the "second order bias" of $\widehat{\theta}$. Based on the mapping between $E\left[\theta^{(2)}\right]$ and primitive parameters, we can $\sqrt{n}$–consistently estimate $E\left[\theta^{(2)}\right]$ by $\widehat{E}\left[\theta^{(2)}\right]$, say. The bias corrected estimator $\widehat{\theta}-\frac{1}{n}\widehat{E}\left[\theta^{(2)}\right]$ has zero second order bias. It is not always a priori clear if the second order bias $\frac{1}{n}E\left[\theta^{(2)}\right]$ approximately equals the actual bias $E\left[\widehat{\theta}-\theta\right]$. It is also not a priori clear if the bias corrected estimator $\widehat{\theta}-\frac{1}{n}\widehat{E}\left[\theta^{(2)}\right]$ indeed reduces much of the bias.

For dynamic panel model of interest, we addressed these questions by Monte Carlo approximation. For such purpose, we first derive the second order bias of $\widehat{b}_{GMM}$:

Theorem 6 *Under Conditions 1-2, the second order bias of $\widehat{b}_{GMM}$ is equal to*

$$\frac{B_1+B_2+B_3}{n}+o\left(\frac{1}{n}\right), \tag{13}$$

where

$$B_1 \equiv \Upsilon_1^{-1}\sum_{t=1}^{T-1}\operatorname{trace}\left(\left(\Gamma_t^{zz}\right)^{-1}\Gamma_{t,t}^{\varepsilon xzz}\right)$$
$$B_2 \equiv -2\Upsilon_1^{-2}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}\Gamma_t^{zx\prime}\left(\Gamma_t^{zz}\right)^{-1}\Gamma_{t,s}^{\varepsilon xzz}\left(\Gamma_s^{zz}\right)^{-1}\Gamma_s^{zx}$$
$$B_3 \equiv \Upsilon_1^{-2}\sum_{t=1}^{T-1}\sum_{s=t}^{T-1}\Gamma_t^{zx\prime}\left(\Gamma_t^{zz}\right)^{-1}B_{3,1}(t,s)\left(\Gamma_s^{zz}\right)^{-1}\Gamma_s^{zx}.$$

where $\Gamma_t^{zz}\equiv E\left[z_{it}z_{it}'\right], \Gamma_t^{zx}\equiv E\left[z_{it}x_{it}^*\right], \Gamma_{t,s}^{\varepsilon xzz}\equiv E\left[\varepsilon_{it}^*x_{is}^*z_{it}z_{is}'\right], B_{3,1}(t,s)\equiv E\left[\varepsilon_{it}^*z_{it}\Gamma_s^{zx\prime}\left(\Gamma_s^{zz}\right)^{-1}z_{is}z_{is}'\right]$ *and* $\Upsilon_1\equiv\sum_{t=1}^{T-1}\Gamma_t^{zx\prime}\left(\Gamma_t^{zz}\right)^{-1}\Gamma_t^{zx}$.

Proof. Available at http://econ-www.mit.edu/faculty/jhausman/papers.htm. ■

In Table 6, we compare the actual performance of $\widehat{b}_{GMM}$ and the prediction of its bias based on Theorem 6. Table 6 tabulates the actual bias of the estimator approximated by 5,000 Monte Carlo runs, and compares it with the second order bias based on the formula (13). It is clear that the second order theory does a reasonably good job except when β is close to the unit circle and n is small.

Theorem 6 suggests a natural way of eliminating the bias. Suppose that $\widehat{B}_1, \widehat{B}_2, \widehat{B}_3$ are $\sqrt{n}$-consistent estimators of B_1, B_2, B_3. Then it is easy to see that

$$\widehat{b}_{BC} \equiv \widehat{b}_{GMM} - \frac{1}{n}\left(\widehat{B}_1 + \widehat{B}_2 + \widehat{B}_3\right) \tag{14}$$

is first order equivalent to $\widehat{b}_{GMM}$, and has second order bias equal to zero. Define $\hat{\Gamma}_t^{zz} = n^{-1}\sum_{i=1}^n z_{it}z_{it}'$, $\hat{\Gamma}_t^{zx} = n^{-1}\sum_{i=1}^n z_{it}x_{it}^*$, $\hat{\Gamma}_{t,s}^{\varepsilon xzz} = n^{-1}\sum_{i=1}^n e_{it}^* x_{is}^* z_{it}z_{is}'$ and $\hat{B}_{3,1}(t,s) = n^{-1}\sum_{i=1}^n e_{it}^* z_{it}'\hat{\Gamma}_t^{zx}\left(\hat{\Gamma}_t^{zz}\right)^{-1} z_{is}z_{is}'$, where $e_{it}^* \equiv y_{it}^* - x_{it}^*\widehat{b}_{GMM}$. Let $\hat{B}_1, \hat{B}_2$ and $\hat{B}_3$ be defined by replacing $\Gamma_t^{zz}, \Gamma_t^{zx}$,$\Gamma_{t,s}^{\varepsilon xzz}$ and $B_{3,1}(t,s)$ by $\hat{\Gamma}_t^{zz}, \hat{\Gamma}_t^{zx}$,$\hat{\Gamma}_{t,s}^{\varepsilon xzz}$ and $\hat{B}_{3,1}(t,s)$ in B_1, B_2 and B_3.

Second order asymptotic theory predicts approximately that $\widehat{b}_{BC}$ would be relatively free of bias. We examined whether this prediction is reasonably accurate in finite sample by 5,000 Monte Carlo runs. Table 6 summarizes the properties of $\widehat{b}_{BC}$. We have seen in Table 6 that the second order theory is reasonably accurate unless β is close to one. It is therefore sensible to conjecture that $\widehat{b}_{BC}$ would have a reasonable finite sample bias property as long as β is not too close to one. We verify this conjecture in Table 6.

We do not expect other methods of bias correction to be much superior to $\widehat{b}_{BC}$. Hahn, Kuersteiner, and Newey (2002) recently showed that higher order mean squared errors of bias corrected maximum likelihood estimators are invariant to methods of bias correction. In particular, they considered a third order expansion of various bias corrected estimators, and showed that bootstrap bias corrected MLE, jackknife bias corrected MLE, and analytically bias corrected MLE all have the same higher order MSE. Their result crucially depends on the fact that the MLE is first order efficient. Because $\widehat{b}_{GMM}$ is first order efficient, we expect their result to carry over, and MSE of bootstrap bias corrected estimator, say, is not expected to be substantially different from that of analytically bias corrected estimator $\widehat{b}_{BC}$.

However the second order asymptotics "fails" to be a good approximation around $\beta \approx 1$. This phenomenon can be explained by the "weak instrument" problem. See Staiger and Stock (1997). Blundell and Bond (1998) argued that the weak instrument problem can be alleviated by assuming stationarity on the initial observation y_{i0}. The stationarity assumption turns out to be a predominant source of information around $\beta \approx 1$ as noted by Hahn (1999). The stationarity condition may or may not be appropriate for particular applications, and substantial finite sample biases due to inconsistency will result under violation of stationarity.[15]

[15] Under stationarity, we would have $y_{i0} \sim N\left(\frac{\alpha_i}{1-\beta}, \frac{\sigma_\varepsilon^2}{1-\beta^2}\right)$. We analyze departures from the stationary initial condition by assuming that instead $y_{i0} \sim N\left(\frac{\alpha_i}{1-\beta_F}, \frac{\sigma_\varepsilon^2}{1-\beta^2}\right)$. Because the asymptotic bias depends on the weight

A.2 Higher Order Bias and MSE of Long Difference Estimator

In this section, we examine the third order expansion of the long difference estimator and compare it with two other estimators discussed in the literature. Note that unlike the approximate bias calculations of the previous section the expansions considered here are valid for any fixed number of instruments. Exact formulas tend to be very complicated and are not reported here.[16] Many econometric estimators, say $\widehat{\theta}$ for θ based on a sample of size n allow for an expansion of the form (12). Ignoring the $o_p\left(n^{-1}\right)$ term and taking expectations of $\left(\sqrt{n}\left(\widehat{\theta}-\theta\right)\right)^2$, we obtain the approximate MSE of $\sqrt{n}\left(\widehat{\theta}-\theta\right)$ equal to

$$E\left[\left(\theta^{(1)}\right)^2\right]+\frac{1}{n}E\left[\left(\theta^{(2)}\right)^2\right]+\frac{2}{\sqrt{n}}E\left[\theta^{(1)}\theta^{(2)}\right]+\frac{2}{n}E\left[\theta^{(1)}\theta^{(3)}\right]+o\left(\frac{1}{n}\right)$$

Ignoring the $o\left(n^{-1}\right)$ term, we may understand

$$\frac{1}{n}E\left[\left(\theta^{(1)}\right)^2\right]+\frac{1}{n^2}E\left[\left(\theta^{(2)}\right)^2\right]+\frac{2}{n\sqrt{n}}E\left[\theta^{(1)}\theta^{(2)}\right]+\frac{2}{n^2}E\left[\theta^{(1)}\theta^{(3)}\right] \tag{15}$$

as the approximate MSE of $\widehat{\theta}$.

We first examine the higher order bias and MSE for $\widehat{b}_{AS}$,[17] the estimator proposed by Ahn and Schmidt (1995) and $\widehat{b}_{BB}$, the estimator proposed by Blundell and Bond (1998). Both are GMM estimators, and the third order expansion of the usual two step estimator is critically dependent on the initial weight matrices. It can be shown that the third order expansion of the "four step" GMM estimator is invariant to the initial weight matrix[18]. Rilstone, Srivastava, and Ullah (1996) obtained higher order bias and MSE formulas for method of moments estimators. Their expressions are difficult to adapt to our situation and we have developed our own expansions. We computed (15) analytically for $T=5$, and summarized it along with the second order bias in Table 8. Table 8 also contains the higher order properties of LD estimators. Given the two step IV nature of the LD estimators, this required a separate expansion for IV estimator along with a

matrix, we consider two versions of GMM estimators. The first one uses an identity matrix for initial weight matrix, whose asymptotic bias is summarized in the upper half of Table 7. The second one uses Blundell and Bond's recommendation, whose asymptotic bias is summarized in the lower half of Table 7. The results show that the bias of Blundell and Bond's estimator can be substantial under misspecification.

[16] Details are available at http://econ-www.mit.edu/faculty/jhausman/papers.htm.

[17] The version of Ahn and Schmidt's estimator (1995) adopted in this comparison does not utlize homoscedasticity.

[18] The details are available at http://econ-www.mit.edu/faculty/jhausman/papers.htm. On a related note, Newey and Smith (2001) point out the second order expansion of "three" step GMM estimator is invariant to the initial weight matrix.

third order expansion[19] of $\widehat{b}_{GMM}$. Unfortunately, analytic calculation of (15) for LD estimators was infeasible. We used a stochastic approximation to each term in (15) instead.[20]

We can see that theoretical higher order MSEs of $\widehat{b}_{LD2}$ and $\widehat{b}_{LD3}$ are comparable to that of $\widehat{b}_{AS}$. We can also see that theoretical bias properties of $\widehat{b}_{LD2}$ and $\widehat{b}_{LD3}$ are moderately superior to that of $\widehat{b}_{AS}$. In Table 9, we compare the theoretical higher order properties of selected estimators with their actual properties. Interestingly, we find that actual properties of long difference estimators are much better than those predicted by higher order theory, especially around the unit circle. This suggests that Monte Carlo comparison should be preferred to any higher order expansion for ranking various estimators. In Table 1, we examine Monte Carlo properties of those estimators. We find that LD does better than the other estimators for large β and not significantly worse for moderate β.

B Proofs

Proof of Lemma 1. Note that $\eta_{i0} \equiv y_{i0} - \alpha_i / (\beta_n - 1)$ such that

$$\Delta y_{it} = \beta_n^{t-1} (\beta_n - 1) \eta_{i0} + \varepsilon_{it} + (\beta_n - 1) \sum_{s=1}^{t-1} \beta_n^{s-1} \varepsilon_{it-s}$$

and

$$\begin{aligned} E[|u_{it} \Delta y_{is-1}|] &\leq \sqrt{E[u_{it}^2]} \sqrt{E\left[(\Delta y_{is-1})^2\right]} \\ &= \sqrt{\sigma_\varepsilon^2 + \sigma_\alpha^2} \sqrt{E\left[\left(\beta_n^{s-2} (\beta_n - 1) \eta_{i0} + \varepsilon_{is-1} + (\beta_n - 1) \sum_{r=1}^{s-2} \beta_n^{r-1} \varepsilon_{is-1-r}\right)^2\right]} \\ &= \sqrt{\sigma_\varepsilon^2 + \sigma_\alpha^2} \sqrt{\beta_n^{2(s-2)} \frac{\sigma_\varepsilon^2 (\beta_n - 1)^2}{1 - \beta_n^2} + \sigma_\varepsilon^2 + (\beta_n - 1)^2 \sigma_\varepsilon^2 \sum_{r=1}^{s-2} \beta_n^{2(r-1)}} = O(1) \end{aligned}$$

By independence of $u_{it} \Delta y_{is-1}$ across i, it therefore follows that $n^{-3/2} \sum_{i=1}^n u_{it} \Delta y_{is-1} = o_p(1)$. By the same reasoning, we obtain $n^{-3/2} \sum_{i=1}^n u_{iT} \Delta y_{ij-1} = o_p(1)$, and $n^{-3/2} \sum_{i=1}^n \overline{u}_i \Delta y_{ik-1} = o_p(1)$. We therefore obtain $n^{-3/2} \sum_{i=1}^n f_{i,1} = o_p(1)$. We can similarly obtain $n^{-3/2} \sum_{i=1}^n g_{i,1} = o_p(1)$.

[19] All derivations referred to in this paragraph are available at http://econ-www.mit.edu/faculty/jhausman/papers.htm.

[20] We perform Monte Carlo integration by generating i.i.d. sequences of $\left(\theta_j^{(1)}, \theta_j^{(2)}, \theta_j^{(3)}\right)$ $j = 1, \ldots, J$, and approximating $E\left[\left(\theta^{(1)}\right)^2\right]$ by $\frac{1}{J} \sum_{j=1}^J \left(\theta_j^{(1)}\right)^2$ for example. Our choice of J was 5,000.

Next we consider $n^{-3/2}\sum_{i=1}^{n} f_{i,2}$ and$n^{-3/2}\sum_{i=1}^{n} g_{i,2}$. Note that

$$\begin{aligned}
E\left[\Delta y_{it} y_{i0}\right] &= E\left[y_{i0}\left(\beta_n^{t-1}\left(\beta_n-1\right)\eta_{i0}+\varepsilon_{it}+\left(\beta_n-1\right)\sum\nolimits_{r=1}^{t-1}\beta_n^{r-1}\varepsilon_{it-1-r}\right)\right] \\
&= \beta_n^{t-1}\sigma_\varepsilon^2\frac{\beta_n-1}{1-\beta_n^2}=\sigma_\varepsilon^2\frac{-c/n}{2c/n}+o(1)=\frac{-\sigma_\varepsilon^2}{2}+o(1),
\end{aligned}$$

and

$$\begin{aligned}
E\left[\left(\Delta y_{it} y_{i0}\right)^2\right] &= E\left[y_{i0}^2\left(\beta_n^{t-1}\left(\beta_n-1\right)\eta_{i0}+\varepsilon_{it}+\left(\beta_n-1\right)\sum\nolimits_{r=1}^{t-1}\beta_n^{r-1}\varepsilon_{it-1-r}\right)^2\right] \\
&= \beta_n^{2(t-1)}\frac{\left(\beta_n-1\right)^2\sigma_\varepsilon^2}{1-\beta_n^2}\frac{\sigma_\alpha^2}{\left(1-\beta_n\right)^2}+3\beta_n^{2(t-1)}\frac{\sigma_\varepsilon^4(\beta_n-1)^2}{\left(1-\beta_n^2\right)^2} \\
&\quad+\left(\sigma_\varepsilon^2+\left(\beta_n-1\right)^2\sigma_\varepsilon^2\sum\nolimits_{r=1}^{s-2}\beta_n^{2(r-1)}\right)\left(\frac{\sigma_\alpha^2}{\left(1-\beta_n\right)^2}+\frac{\sigma_\varepsilon^2}{\left(1-\beta_n^2\right)}\right) \\
&\quad+\frac{\operatorname{cum}\left(\varepsilon_{it},\varepsilon_{it},\alpha_i,\alpha_i\right)}{c^2}n^2+O(n) \\
&= \frac{\sigma_\varepsilon^2\sigma_\alpha^2+\operatorname{cum}\left(\varepsilon_{it},\varepsilon_{it},\alpha_i,\alpha_i\right)}{c^2}n^2+O(n)
\end{aligned}$$

such that $\operatorname{Var}\left(n^{-3/2}\sum_{i=1}^{n}\Delta y_{it}y_{i0}\right)=O(1)$. Use the notation $c_{t,t,\alpha,\alpha}=\operatorname{cum}\left(\varepsilon_{it},\varepsilon_{it},\alpha_i,\alpha_i\right)$, $c_{t,s,\alpha,\xi}=\operatorname{cum}\left(\varepsilon_{it},\varepsilon_{is},\alpha_i,\eta_{i0}\right)$ etc. For $n^{-3/2}\sum_{i=1}^{n} g_{i,2}\left(\beta_0\right)$ we have from the moment conditions that $E\left[g_{i,2}\left(\beta_0\right)\right]=0$ and

$$\begin{aligned}
\operatorname{Var}\left(\Delta u_{it}(\beta_0)y_{i0}\right) &= \frac{2\sigma_\varepsilon^2\sigma_\alpha^2}{\left(1-\beta_n\right)^2}+\frac{c_{t,t,\alpha,\alpha}+2c_{t,t-1,\alpha,\alpha}+c_{t-1,t-1,\alpha,\alpha}}{\left(1-\beta_n\right)^2}+O(n) \\
&= \frac{2\sigma_\varepsilon^2\sigma_\alpha^2+c_{t,t,\alpha,\alpha}+2c_{t,t-1,\alpha,\alpha}+c_{t-1,t-1,\alpha,\alpha}}{c^2}n^2+O(n).
\end{aligned}$$

The joint limiting distribution of $n^{-3/2}\sum_{i=1}^{n}\left[f_{i,2}'-E\left[f_{i,2}'\right],g_{i,2}(\beta_0)'\right]'$ can now be obtained from a triangular array CLT. By previous arguments

$$E\left[f_{i,2}',g_{i,2}(\beta_0)'\right]=\left[\begin{array}{cccc}\mu' & 0 & \cdots & 0\end{array}\right]$$

with $\mu=-\sigma_\varepsilon^2/2\iota+O(n^{-1})$ where ι is the $T-1$ dimensional vector with elements 1. Then

$$E\left[\left(f_{i,2}'-E\left[f_{i,2}'\right],g_{i,2}(\beta_0)'\right)'\left(f_{i,2}'-E\left[f_{i,2}'\right],g_{i,2}(\beta_0)'\right)\right]=\Sigma_n$$

where

$$\Sigma_n=\left[\begin{array}{cc}\Sigma_{11,n} & \Sigma_{12,n}\\ \Sigma_{21,n} & \Sigma_{22,n}\end{array}\right]$$

By previous calculations we have found the diagonal elements of $\Sigma_{11,n}$ and $\Sigma_{22,n}$ to be $\left(\sigma_\varepsilon^2\sigma_\alpha^2+c_{t,t,\alpha,\alpha}\right)c^{-2}n^2$ and $\left(2\sigma_\varepsilon^2\sigma_\alpha^2+c_{t,t,\alpha,\alpha}+2c_{t,t-1,\alpha,\alpha}+c_{t-1,t-1,\alpha,\alpha}\right)c^{-2}n^2$. The off-diagonal elements of $\Sigma_{11,n}$ are found to be

$$\begin{aligned} E\left[\Delta y_{it}\Delta y_{is}y_{i0}^2\right] &= E\Big[y_{i0}^2\left(\beta_n^{s-1}\left(\beta_n-1\right)\eta_{i0}+\varepsilon_{is}+\left(\beta_n-1\right)\sum\nolimits_{r=1}^{s-1}\beta_n^{r-1}\varepsilon_{is-1-r}\right) \\ &\qquad\times\left(\beta_n^{t-1}\left(\beta_n-1\right)\eta_{i0}+\varepsilon_{it}+\left(\beta_n-1\right)\sum\nolimits_{r=1}^{t-1}\beta_n^{r-1}\varepsilon_{it-1-r}\right)\Big] \\ &= \beta_n^{t-1}\beta_n^{s-1}\frac{\left(\beta_n-1\right)^2}{\left(1-\beta_n^2\right)}\left(\frac{\sigma_\varepsilon^2\sigma_\alpha^2+c_{t,s,\alpha,\alpha}}{\left(1-\beta_n\right)^2}+3\frac{\sigma_\varepsilon^4}{\left(1-\beta_n^2\right)}\right)+O(1) \\ &= \frac{\sigma_\varepsilon^2\sigma_\alpha^2}{2c}n+\frac{c_{t,s,\alpha,\alpha}}{c^2}n^2+O(1) \end{aligned}$$

which is of lower order of magnitude while $n^{-1}\left(E\left[\Delta y_{it}y_{i0}\right]\right)^2 = O(1)$. Thus $n^{-2}\Sigma_{11,n} \to \operatorname{diag}(\frac{\sigma_\varepsilon^2\sigma_\alpha^2}{c^2},...,\frac{\sigma_\varepsilon^2\sigma_\alpha^2}{c^2})+\mathbb{K}_1$. The off-diagonal elements of $\Sigma_{22,n}$ are obtained from

$$E\left[\Delta u_{it}\Delta u_{is}y_{i0}^2\right] = \begin{cases} -\frac{\sigma_\varepsilon^2\sigma_\alpha^2}{\left(1-\beta_n\right)^2}+\frac{c_{t,s,\alpha,\alpha}-c_{t,s-1,\alpha,\alpha}-c_{t-1,s,\alpha,\alpha}+c_{t-1,s-1,\alpha,\alpha}}{\left(1-\beta_n\right)^2}+O(n) & t=s+1 \text{ or } t=s-1 \\ \frac{c_{t,t,\alpha,\alpha}+2c_{t,t-1,\alpha,\alpha}+c_{t-1,t-1,\alpha,\alpha}}{\left(1-\beta_n\right)^2} & \text{otherwise} \end{cases}$$

such that $n^{-2}\Sigma_{22,n} = \delta M_2+\mathbb{K}_2$. For $\Sigma_{12,n}$, we consider

$$E\left[\Delta y_{it}\Delta u_{is}y_{i0}^2\right] = \begin{cases} \frac{\sigma_\varepsilon^2\sigma_\alpha^2+c_{t,s,\alpha,\alpha}-c_{t,s-1,\alpha,\alpha}}{c^2}n^2+O(n) & \text{if } t=s \\ \frac{-\sigma_\varepsilon^2\sigma_\alpha^2+c_{t,s,\alpha,\alpha}-c_{t,s-1,\alpha,\alpha}}{c^2}n^2+O(n) & \text{if } t=s-1 \\ \frac{c_{t,s,\alpha,\alpha}-c_{t,s-1,\alpha,\alpha}}{c^2}n^2 & \text{otherwise} \end{cases}$$

where $\Sigma_{12,n} \to \delta M_1+\mathbb{K}_3$. It then follows that for $\ell \in \mathbb{R}^{T(T+1)/+2T-6}$ such that $\ell'\ell = 1$, $n^{-3/2}\sum_{i=1}^n \ell'\Sigma_n^{-1/2}\left[f_{i,2}'-Ef_{i,2},g_{i,2}(\beta_0)'\right]' \xrightarrow{d} N\left(0,1\right)$ by the Lindeberg-Feller CLT for triangular arrays. It then follows from a straightforward application of the Cramer-Wold theorem and the continuous mapping theorem that $n^{-3/2}\sum_{i=1}^n\left[f_{i,2}',g_{i,2}(\beta_0)'\right]' \xrightarrow{d} \left[\xi_x',\xi_y'\right]'$ where $\left[\xi_x',\xi_y'\right]' \sim N\left(0,\Sigma\right)$. Note that $n^{-3/2}\sum_{i=1}^n \ell' E\left[f_{i,2}\right] = O(n^{-1/2})$ and thus does not affect the limit distribution.

Finally note that $E\left[g_{i,1}g_{i,1}'\right] = O\left(1\right)$. It therefore follows that

$$\frac{1}{n^2}E\left[g_ig_i'\right] = \begin{bmatrix} 0 & 0 \\ 0 & \Sigma_{22} \end{bmatrix} + o\left(1\right).$$

■

Proof of Lemma 2. From the previous results it follows that

$$n^{-3/2}\sum_{i=1}^n\left[f_{i,2}-E\left[f_{i,2}\right]\right] = n^{-3/2}\sum_{i=1}^n\left[\varepsilon_{i1},...,\varepsilon_{iT-1}\right]'y_{i0}+o_p(1)$$

as well as

$$n^{-3/2}\sum_{i=1}^n g_{i,2}(\beta_0) = n^{-3/2}\sum_{i=1}^n\left[\Delta\varepsilon_{i2},...,\Delta\varepsilon_{iT}\right]'y_{i0}+o_p\left(1\right)$$

which implies that

$$n^{-3/2}\sum_{i=1}^{n}\left[f_{i,2}'-E\left[f_{i,2}'\right],g_{i,2}(\beta_0)'\right]'=n^{-3/2}\sum_{i=1}^{n}\left[\begin{array}{cc} I_{T-1} & 0_{T-1,1} \\ H & \end{array}\right]\left[\varepsilon_{i1},...,\varepsilon_{iT-1},\Delta\varepsilon_{iT}\right]'y_{i0}+o_p(1)$$

where

$$n^{-3/2}\sum_{i=1}^{n}\left[\begin{array}{cc} I_{T-1} & 0_{T-1,1} \\ H & \end{array}\right]\left[\varepsilon_{i1},...,\varepsilon_{iT-1},\Delta\varepsilon_{iT}\right]'y_{i0}\xrightarrow{d}\left[\xi_x,H\xi_u\right]$$

and H is defined in (19). ■

Proof of Theorem 1. From the properties of the Moore-Penrose inverse, see Magnus and Neudecker (1988, p.33), it follows that

$$\tilde{C}_1\left(\tilde{C}_1'\Sigma_{22}\tilde{C}_1\right)^{+}\tilde{C}_1'=\Sigma_{22}^{-1/2}\Sigma_{22}^{1/2}\tilde{C}_1\left(\Sigma_{22}^{1/2}\tilde{C}_1\right)^{+}\left(\tilde{C}_1\Sigma_{22}^{1/2}\right)^{+\prime}\tilde{C}_1'\Sigma_{22}^{1/2}\Sigma_{22}^{-1/2}$$

where $\Sigma_{22}^{1/2}\tilde{C}_1$ is a $(T-1)\times r$ matrix of rank $r_1\leq T-1$. By Theorem 1.16 of Magnus and Neudecker (1988) there exist matrices V and U of dimensions $(T-1)\times r_1$ and and $r\times r_1$ and a diagonal matrix S of dimension $r_1\times r_1$ such that $V'V=U'U=I_{r_1}$ and $\Sigma_{22}^{1/2}\tilde{C}_1=VS^{1/2}U'$. The unique representation of $\left(\Sigma_{22}^{1/2}\tilde{C}_1\right)^{+}$ then is $\left(\Sigma_{22}^{1/2}\tilde{C}_1\right)^{+}=US^{-1/2}V'$. This implies that

$$\tilde{C}_1\left(\tilde{C}_1'\Sigma_{22}\tilde{C}_1\right)^{+}\tilde{C}_1'=\Sigma_{22}^{-1/2}VV'\Sigma_{22}^{-1/2}.$$

Next set $C_1=\Sigma_{22}^{-1/2}V\left(V'\Sigma_{22}^{-1}V\right)^{1/2}$ where C_1 is a matrix of dimension $(T-1)\times r_1$ of full column rank such that

$$\begin{aligned}
C_1\left(C_1'\Sigma_{22}C_1\right)^{+}C_1' &= \Sigma_{22}^{-1/2}V\left(V'\Sigma_{22}^{-1}V\right)^{1/2}\left(\left(V'\Sigma_{22}^{-1}V\right)^{1/2}V'V\left(V'\Sigma_{22}^{-1}V\right)^{1/2}\right)^{+}\Sigma_{22}^{1/2}V'\Sigma_{22}^{-1/2} \\
&= \Sigma_{22}^{-1/2}V\left(V'\Sigma_{22}^{-1}V\right)^{1/2}\left(V'\Sigma_{22}^{-1}V\right)^{-1}\left(V'\Sigma_{22}^{-1}V\right)^{1/2}V'\Sigma_{22}^{-1/2} \\
&= \Sigma_{22}^{-1/2}VV'\Sigma_{22}^{-1/2}
\end{aligned}$$

and $C_1'C_1=\left(V'\Sigma_{22}^{-1}V\right)^{1/2}V'\Sigma_{22}^{-1}V\left(V'\Sigma_{22}^{-1}V\right)^{1/2}=I_{r_1}$ from which it follows that

$$\xi\left(C_1,\Sigma_{22}\right)=\xi\left(\tilde{C}_1,\Sigma_{22}\right).$$

■

Proof of Theorem 2. Use the transformation $z=L^{-1}\zeta_x$ with $\Sigma_{11}=LL'$. Define $W=C_1\left(C_1'\bar{\Omega}C_1\right)^{-1}C_1'$. Then $\xi\left(C_1,\tilde{\Omega}\right)=z'L'WH\xi_u/\,z'L'WLz$. Let $F=\Sigma_{21}\Sigma_{11}^{-1}$ and use the fact that $E\left[H\xi_u|\,\xi_x\right]=F\xi_x=FLz$. Using a conditioning argument it then follows that

$$E\left[\xi\left(C_1,\tilde{\Omega}\right)\right]=E\left[\frac{z'L'DLz}{z'L'WLz}\right].$$

Note that $z'L'DLz = z'L'D'Lz = z'\bar{D}z$ where $\bar{D} = \frac{1}{2}L'(D + D')L$ is symmetric. Define $z_1 = \Gamma_1'z$ and $z_2 = \Gamma_2'z$ where Γ_1 are the eigenvectors corresponding to the nonzero eigenvalues of $L'WL$ such that $\Gamma_1\Lambda_1\Gamma_1' = L'WL$ and $\Gamma_1\Gamma_1' + \Gamma_2\Gamma_2' = I$. It now follows that

$$E\left[\frac{z'L'DLz}{z'LWLz}\right] = E\left[\frac{z_1'\Gamma_1'L'DL\Gamma_1 z_1 + z_1'L'\Gamma_1'DL\Gamma_2 z_2}{z_1'\Lambda_1 z_1}\right] = E\left[\frac{z_1'\Gamma_1'L'DL\Gamma_1 z_1}{z_1'\Lambda_1 z_1}\right]$$

where the first equality follows from $\Gamma_2'W = 0$ and the second equality follows from independence between z_1 and z_2. The result then follows from Smith (1993, Eq. 2.4, p. 273).

When Condition 5 is note that we can take $L = \sqrt{\delta}I_{r_1}$ and that $F = \Sigma_{21}\Sigma_{11}^{-1} = M_1'$ where the second equality is based on $\Sigma_{11}^{-1} = \delta^{-1}I_{r_1}$. The last statement then follows by the same arguments as before. ■

Proof of Theorem 3. We first analyze $E[\xi(C_1, \Sigma_{22})]$. Note that in this case $W = C_1(C_1'\Sigma_{22}C_1)^{-1}C_1' = \delta^{-1}C_1(C_1'M_2C_1)^{-1}C_1'$ such that

$$E[\xi(C_1, \Sigma_{22})] = E\left[\frac{z_1'\Gamma_1'\bar{D}\Gamma_1 z_1}{z_1'\Lambda_1 z_1}\right]$$

and $\delta \operatorname{tr} \bar{D} = \delta/2 \operatorname{tr} W(M_1' + M_1) = -r_1/2$ since $M_1' + M_1 = -M_2$ where r_1 is the rank of C_1. Then it follows from Smith (1993, p. 273 and Appendix A)

$$E\left[\frac{z_1'\Gamma_1'\bar{D}\Gamma_1 z_1}{z_1'\Lambda_1 z_1}\right] = \bar{\lambda}^{-1}\sum_{k=0}^{\infty}\frac{(1)_k\left(\frac{1}{2}\right)_{1+k}}{\left(\frac{r_1}{2}\right)_{1+k}k!}C_{1+k}^{1,k}\left(\Gamma_1'\bar{D}\Gamma_1, I_{r_1} - \bar{\lambda}^{-1}\Lambda_1\right), \tag{16}$$

where for any two real symmetric matrices Y_1, Y_2

$$C_{1+k}^{1,k}(Y_1, Y_2) = \frac{k!}{2\left(\frac{1}{2}\right)_{k+1}}\sum_{i=0}^{k}\frac{\left(\frac{1}{2}\right)_{k-i}}{(k-i)!}\operatorname{tr}\left(Y_1Y_2^i\right)C_{k-i}(Y_2)$$

and

$$\begin{aligned} C_k(Y_2) &= \frac{k!}{\left(\frac{1}{2}\right)_k}d_k(Y_2) \\ d_k(Y_2) &= k^{-1}\sum_{j=0}^{k-1}\frac{1}{2}\operatorname{tr}\left(Y_2^{k-j}\right)d_j(Y_2) \\ d_0(Y_2) &= 1. \end{aligned}$$

Since all elements $\bar{\lambda}^{-1}\lambda_i$ in $\bar{\lambda}^{-1}\Lambda_1$ satisfy $0 < \bar{\lambda}^{-1}\lambda_i \leq 1$ it follows that $I_{r_1} - \bar{\lambda}^{-1}\Lambda_1$ is positive semidefinite implying that $C_k\left(I_{r_1} - \bar{\lambda}^{-1}\Lambda_1\right) \geq 0$, which holds with equality only if all eigenvalues λ_i are the same. Also note that, if Y_2, Y_3 are diagonal matrices and Y_1 any conformable matrix, then $\operatorname{tr} Y_3 Y_1 Y_2^i = \operatorname{tr} Y_1 Y_3 Y_2^i$. Therefore,

$$\begin{aligned}
\operatorname{tr} \Gamma_1' \bar{D} \Gamma_1 (I_{r_1} - \bar{\lambda}^{-1}\Lambda_1)^i &= \operatorname{tr} \Gamma_1' \left(W M_1' + M_1 W\right) \Gamma_1 \left(I_{r_1} - \bar{\lambda}^{-1}\Lambda_1\right)^i \\
&= \frac{1}{2} \operatorname{tr} \left(\Lambda_1 \Gamma_1' M_1' \Gamma_1 + \Gamma_1' M_1 \Gamma_1 \Lambda_1\right) \left(I_{r_1} - \bar{\lambda}^{-1}\Lambda_1\right)^i \\
&= \frac{1}{2} \operatorname{tr} \Gamma_1' \left(M_1' + M_1\right) \Gamma_1 \Lambda_1 \left(I_{r_1} - \bar{\lambda}^{-1}\Lambda_1\right)^i \\
&= -\frac{1}{2} \operatorname{tr} \Gamma_1' M_2 \Gamma_1 \Lambda_1 \left(I_{r_1} - \bar{\lambda}^{-1}\Lambda_1\right)^i .
\end{aligned}$$

Next, note that for two positive semi-definite matrices Y_1 and Y_2 it follows that $\operatorname{tr} Y_1 Y_2 = \operatorname{tr} Y_2^{1/2} Y_1 Y_2^{1/2} \geq 0$ because $Y_2^{1/2} Y_1 Y_2^{1/2}$ is well defined and positive semi-definite by positive semi-definiteness of Y_1. Next note that $\Lambda_1 \left(I_{r_1} - \bar{\lambda}^{-1}\Lambda_1\right)^i$ is positive semi-definite by previous arguments. Also, $\Gamma_1' M_2 \Gamma_1$ is positive definite. Therefore, every $\operatorname{tr}\left(\Gamma_1' \bar{D} \Gamma_1 \left(I_{r_1} - \bar{\lambda}^{-1}\Lambda_1\right)^i\right) C_{k-i}\left(I_{r_1} - \bar{\lambda}^{-1}\Lambda_1\right)$ has the same sign, and so does every $C_{1+k}^{1,k}\left(\Gamma_1' \bar{D} \Gamma_1, I_{r_1} - \bar{\lambda}^{-1}\Lambda_1\right)$. Therefore, we have

$$|E\left[\xi\left(C_1, \Sigma_{22}\right)\right]| \geq \left| \bar{\lambda}^{-1} \sum_{k=0}^{0} \frac{(1)_k \left(\frac{1}{2}\right)_{1+k}}{\left(\frac{r_1}{2}\right)_{1+k} k!} C_{1+k}^{1,k}\left(\Gamma_1' \bar{D} \Gamma_1, I_{r_1} - \bar{\lambda}^{-1}\Lambda_1\right)\right|$$

For $k = 0$, we have

$$C_1^{1,0}\left(\Gamma_1' \bar{D} \Gamma_1, I_{r_1} - \bar{\lambda}^{-1}\Lambda_1\right) = \operatorname{tr} \Gamma_1' \bar{D} \Gamma_1 \geq -\delta^{-1} r_1/2. \tag{17a}$$

For the last inequality let $\lambda_i^{\bar{D}}$ be the eigenvalues of $\bar{D}$. By Magnus and Neudecker (1988, Theorem 13, p.211), $\operatorname{tr} \Gamma_1' \bar{D} \Gamma_1 \geq \sum_{i=1}^{r_1} \lambda_i^{\bar{D}}$. Note that $\bar{D}$ is of reduced rank r_1 such that $\operatorname{tr} \bar{D} = \sum_{i=1}^{r_1} \lambda_i^{\bar{D}} = -\delta^{-1} r_1/2$. Also, $(1)_0 \left(\frac{1}{2}\right)_1 / \left(\frac{r_1}{2}\right)_1 = \frac{1}{r_1}$. This shows that $|E\left[\xi\left(C_1, \Sigma_{22}\right)\right]| \geq \bar{\lambda}^{-1}/2$ for all C_1 such that $C_1' C_1 = I_{r_1}$ and all $r_1 \in \{1, 2, .., T-1\}$. Then

$$\min_{\substack{C_1 \text{ s.t. } C_1' C_1 = I_{r_1} \\ r_1 \in \{1,2,..,T-1\}}} |E\left[\xi\left(C_1, \Sigma_{22}\right)\right]| \geq \min_{\substack{C_1 \text{ s.t. } C_1' C_1 = I_{r_1} \\ r_1 \in \{1,2,..,T-1\}}} \frac{\bar{\lambda}^{-1}}{2\delta} \geq \frac{\min l_j}{2}, \tag{18}$$

where $\min l_j$ is the smallest eigenvalue of M_2. Since $\bar{\lambda}$ is the largest eigenvalue of W and W has the same nonzero eigenvalues as $\delta^{-1}\left(C_1' M_2 C_1\right)^{-1}$ by Zhang (1999, Theorem 2.8, p.51), it follows that $\bar{\lambda}$ is the largest eigenvalue of $\delta^{-1}\left(C_1' M_2 C_1\right)^{-1}$. The last inequality in (18) follows from Magnus and Neudecker (1988, Theorem 10, p. 209) as well as the fact that $\bar{\lambda}$ is δ^{-1} times the largest eigenvalue of $\left(C_1' M_2 C_1\right)^{-1}$. Now let $r_1 = 1$ and $C_1 = \rho_i$, where ρ_i is the eigenvector corresponding to $\min l_j$. Note that for this case $W = \delta^{-1} \rho_i \rho_i' / \min l_j$ such that $\Gamma_1 = \rho_i$ and $\bar{\lambda} = \delta^{-1}\left(\min l_j\right)^{-1}$. Then $I_{r_1} - \bar{\lambda}^{-1}\Lambda_1 = 1 - 1 = 0$ and $\Gamma_1' \bar{D} \Gamma_1 = 1/\left(2\delta\right)$ such that $|E\left[\xi\left(C_1, \Sigma_{22}\right)\right]| = \bar{\lambda}^{-1}\left|\operatorname{tr} \Gamma_1' \bar{D} \Gamma_1\right| = \left(1/\min l_j\right)^{-1}/2 = \min l_j/2$. Inequality (18) therefore holds with equality.

Next consider $E\left[\xi\left(C_1, I_{T-1}\right)\right]$. Note that now $W = C_1(C_1'C_1)^{-1}C_1' = C_1C_1'$ since $C_1'C_1 = I_{r_1}$ such that $W = \Gamma_1\Lambda_1\Gamma_1'$ with $\Lambda_1 = I_{r_1}$ and $\Gamma_1 = C_1$. Then, $\bar{\lambda} = 1$ and $I_{r_1} - \bar{\lambda}\Lambda_1 = 0$ such that (16) reduces to $E\left[\xi\left(C_1, I_{T-1}\right)\right] = \frac{1}{r_1}C_1^{1,0}\left(\Gamma_1'\bar{D}\Gamma_1, 0\right) = \operatorname{tr}\Gamma_1'\bar{D}\Gamma_1/r_1$. But then $\bar{D} = 2^{-1}\left(C_1C_1'M_1 + M_1'C_1C_1'\right)$ such that $\operatorname{tr}\Gamma_1'\bar{D}\Gamma_1/r_1 = \operatorname{tr}C_1'\bar{D}C_1/r_1 = -2^{-1}\operatorname{tr}C_1'M_2C_1/r_1 = \operatorname{tr}\bar{D}/r_1$. We analyze

$$\min_{\substack{C_1 \text{ s.t. } C_1'C_1 = I_{r_1} \\ r_1 \in \{1,2,..,T-1\}}} \left|-2^{-1}\operatorname{tr}C_1'M_2C_1r_1\right|.$$

It can be checked easily that M_2 is positive definite symmetric. We can therefore minimize $\operatorname{tr}\left(C_1'M_2C_1\right)$. It is now useful to choose an orthogonal matrix R with j-th row ρ_j such that $R'R = RR' = I$ and $M_2 = R\mathbb{L}R'$ where $\mathbb{L}$ is the diagonal matrix of eigenvalues of $M_2 = \sum_{j=1}^{T-1} l_j\rho_j\rho_j'$. Then it follows that $\operatorname{tr}\left(C_1'M_2C_1\right) = \sum_{j=1}^{T-1} l_j\rho_j'C_1C_1'\rho_j$. Next note that all the eigenvalues of C_1C_1' are either zero or one such that $0 \leq \rho_j'C_1C_1'\rho_j \leq 1$. The minimum of $\operatorname{tr}\left(C_1'M_2C_1\right)$ is then found by choosing $r_1 = 1$ and C_1 such that $C_1'\rho_j = 0$ except for the eigenvector ρ_i corresponding to $\min l_j$. To show that $\operatorname{tr}\left(\bar{D}/r_1\right)$ is also minimized for $r_1 = 1$ and $C_1 = \rho_i$, where $\operatorname{tr}\left(\bar{D}/r_1\right) = \min l_j/2$, consider augmenting C_1 by a column vector x such that $x'x = 1$ and $\rho_i'x = 0$. Then $C_1'C_1 = I_2$, $r_2 = 2$ and $\operatorname{tr}C_1'M_2C_1 = l_i + \sum_{j\neq i}^{T-1} l_j\left(\rho_j'x\right)^2$. By Parseval's equality $\sum_{j\neq i}^{T-1}\left(\rho_j'x\right)^2 = 1$. Since $l_j \geq l_i$ we can bound $\operatorname{tr}\left(C_1'M_2C_1\right) \geq 2l_i$ but then $\operatorname{tr}\left(C_1'M_2C_1/2\right) \geq l_i$. This argument can be repeated to more than one orthogonal addition x. It now follows that $E\left[\xi\left(C_1, I_{T-1}\right)\right] = \operatorname{tr}\left(\bar{D}/r_1\right)$ is minimized for $r_1 = 1$ and $C = \rho_i$, where ρ_i is the eigenvector corresponding to the smallest eigenvalue.

Next note that from $x'x = 1$ such that $\min l_i \leq x'M_2x \leq \max l_i$ it follows that

$$\min l_i \leq \mathbf{1}'M_2\mathbf{1}/(\mathbf{1}'\mathbf{1}) = 2\left(T-1\right)^{-1}$$

for $\mathbf{1} = [1, ..., 1]'$ which shows that the smallest eigenvalue is bounded by a monotonically decreasing function of the number of moment conditions. ■

Proof of Lemma 3. The result for $\hat{b}_{AS}$ follows directly from Lemma (1) and the fact that the finite dimensional matrix $C^{AB,n}$ converges to a limit $C^{AB} = \left[C_0^{AB}, C_1^{AB}\right]$ where C_0^{AB} is defined in the obvious way. The result for $\hat{b}_{AB}$ follows directly from Corollary (1). For $\hat{b}_{GMM}$ we note that $g_{i,2}^* = B^*g_{i,2}$ and $f_{i,2}^* = B^*f_{i,2}$ such that the asymptotic properties of $n^{-3/2}\sum_{i=1}^n\left[f_{i,2}^{*\prime}, g_{i,2}^*\right]'$ follow directly from previous results where

$$n^{-3/2}\sum_{i=1}^n\left[f_{i,2}^{*\prime}, g_{i,2}^{*\prime}\right]' \xrightarrow{d} \left[\xi_x'B^{*\prime}, \xi_y'B^{*\prime}\right]'.$$

For $g_{i,1}^* = B(-B^* \otimes I_T)(\Delta u_i \otimes u_i)$ note that

$$
\begin{aligned}
E\left\|g_{i,1}^*\right\| &\leq \|B(I_T \otimes -B^*)\| E\|u_i \otimes \Delta u_i\| \\
&\leq \|B(I_T \otimes -B^*)\| \left(\operatorname{tr} E\left(u_i u_i' \otimes \Delta u_i \Delta u_i'\right)\right)^{1/2} < \infty
\end{aligned}
$$

as well as

$$
\begin{aligned}
E\left\|f_{i,1}^*\right\| &\leq \|B(I_T \otimes -B^*)\| E\|u_i \otimes \Delta y_{i,-1}\| \\
&\leq \|B(I_T \otimes -B^*)\| \left(\operatorname{tr} E\left(u_i u_i' \otimes \Delta y_{i,-1} \Delta y_{i,-1}'\right)\right)^{1/2} < \infty
\end{aligned}
$$

such that $n^{-1}\sum_{i=1}^n \left[f_{i,1}^*, g_{i,1}^*\right] = O_p(1)$ and $n^{-3/2}\sum_{i=1}^n \left[f_{i,1}^*, g_{i,1}^*\right] = o_p(1)$. These results imply as before that

$$
\hat{b}_{GMM} - 1 \xrightarrow{d} \frac{\xi_x' B^{*\prime} C_1^{AB} \left(C_1^{AB\prime} B^* \Sigma_{22} B^{*\prime} C_1^{AB}\right)^+ C_1^{AB\prime} B^* H \xi_u}{\xi_x' B^{*\prime} C_1^{AB} \left(C_1^{AB\prime} B^* \Sigma_{22} B^{*\prime} C_1^{AB}\right)^+ C_1^{AB\prime} B^* \xi_x} \equiv \xi_{GMM}.
$$

For $\hat{b}_{LD}$ we have

$$
Eu_{it}u_{is}\left(u_{iT} - u_{i1}\right)^2 = \begin{cases} 2\sigma_\alpha^2 \sigma_\varepsilon^2 & t \neq s \\ 2\left(\sigma_\alpha^2 + \sigma_\varepsilon^2\right)\sigma_\varepsilon^2 & t = s \end{cases}
$$

such that $n^{-3/2}\sum_{i=1}^n g_{i,1}^{LD} \to_p 0$. For $Eu_{it}\left(y_{iT-1} - y_{i0}\right)$ note that

$$
y_{iT-1} - y_{i0} = \left(\beta_n^{T-1} - 1\right)\eta_{i0} + \sum_{s=1}^{T-1} \beta_n^{s-1} \varepsilon_{iT-s}
$$

leading to

$$
Eu_{it}\left(y_{iT-1} - y_{i0}\right) = \beta_n^{T-t-1}\sigma_\varepsilon^2
$$

and

$$
\begin{aligned}
Eu_{it}u_{is}\left(y_{iT-1} - y_{i0}\right)^2 &= \left(\beta_n^{T-1} - 1\right)^2 E\left(\eta_{i0}^2 u_{it} u_{is}\right) + \sum_{s_1, s_2 = 1}^{T-1} \beta_n^{s_1 - 1}\beta_n^{s_1 - 1} E\left(\varepsilon_{iT-s_1}\varepsilon_{iT-s_2} u_{it} u_{is}\right) \\
&= \left(\beta_n^{T-1} - 1\right)^2 \frac{\sigma_\varepsilon^2\left(\sigma_\varepsilon^2 + \sigma_\alpha^2\right)}{1 - \beta_n^2} + \sum_{s_1, s_2 = 1}^{T-1} \beta_n^{s_1 - 1}\beta_n^{s_1 - 1} E\left(\varepsilon_{iT-s_1}\varepsilon_{iT-s_2} u_{it} u_{is}\right) \\
&= \sum_{s_1, s_2 = 1}^{T-1} \beta_n^{s_1 - 1}\beta_n^{s_1 - 1} E\left(\varepsilon_{iT-s_1}\varepsilon_{iT-s_2} u_{it} u_{is}\right) + O(n^{-1}).
\end{aligned}
$$

This implies that $n^{-3/2}\sum_{i=1}^n u_{it}\left(y_{iT-1} - y_{i0}\right) = o_p(1)$.

Finally note that

$$
C_1^{AS\prime} = \begin{bmatrix} C_1^{AB\prime} \\ 0_{T-2,T-1} \end{bmatrix}
$$

sucht that

$$C_1^{AS\prime}\Sigma_{22}C_1^{AS} = \begin{bmatrix} C_1^{AB\prime}\Sigma_{22}C_1^{AB} & 0 \\ 0 & 0 \end{bmatrix}$$

with

$$\left(C_1^{AS\prime}\Sigma_{22}C_1^{AS}\right)^+ = \begin{bmatrix} \left(C_1^{AB\prime}\Sigma_{22}C_1^{AB}\right)^+ & 0 \\ 0 & 0 \end{bmatrix}.$$

This implies $\xi_{AS} \overset{d}{=} \xi_{AB}$. To show that $\xi_{AB} \overset{d}{=} \xi_{GMM}$ note that C_1^{AB} has rank $T-1$. Then, by Theorem 1 there exists a $(T-1)\times(T-1)$ matrix C_{AB} of full rank such that $C_{AB}'C_{AB} = I$ and $\xi\left(C_1^{AB},\Sigma_{22}\right) \overset{d}{=} \xi\left(C_{AB},\Sigma_{22}\right)$. But then $\xi\left(C_{AB},\Sigma_{22}\right) = \xi\left(C_{AB}K,\Sigma_{22}\right)$ for any nonsingluar matrix K. In the same way, $C_1^{AB\prime}B$ is of rank $T-1$. By the same argument as before there exists a full rank matrix C_{GMM} such that $\xi_{GMM} \overset{d}{=} \xi\left(C_1^{AB},\Sigma_{22}\right) \overset{d}{=} \xi\left(C_{GMM},\Sigma_{22}\right)$. Then choose $K = C_{AB}^{-1}C_{GMM}$ to show that $\xi\left(C_1^{AB},\Sigma_{22}\right) \overset{d}{=} \xi\left(C_{GMM},\Sigma_{22}\right)$. For the second equality note that

$$C_1^{LD\prime} = \begin{bmatrix} 0_{T-2,T-1} \\ \mathbf{1}_{T-1}' \end{bmatrix}$$

such that

$$\left(C_1^{LD\prime}\Sigma_{22}C_1^{LD}\right)^+ = \begin{bmatrix} 0_{T-2,T-2} & 0_{T-2,1} \\ 0_{1,T-2} & \frac{1}{\mathbf{1}_{T-1}'\Sigma_{22}\mathbf{1}_{T-1}} \end{bmatrix}.$$

This implies that

$$\xi_x'C_1^{LD}\left(C_1^{LD\prime}\Sigma_{22}C_1^{LD}\right)^+ = \left[0_{1,T-2},\xi_x'\mathbf{1}_{T-1}\right]\left(C_1^{LD\prime}\Sigma_{22}C_1^{LD}\right)^+ = \left[0_{1,T-2},\frac{\xi_x'\mathbf{1}_{T-1}}{\mathbf{1}_{T-1}'\Sigma_{22}\mathbf{1}_{T-1}}\right]$$

and

$$C_1^{LD\prime}H\xi_u = \begin{bmatrix} 0_{T-2,T-1} \\ \mathbf{1}_{T-1}'H\xi_u \end{bmatrix}$$

such that

$$\xi_{LD} = \frac{\xi_x'C_1^{LD}\left(C_1^{LD\prime}\Sigma_{22}C_1^{LD}\right)^+C_1^{LD\prime}H\xi_u}{\xi_x'C_1^{LD}\left(C_1^{LD\prime}\Sigma_{22}C_1^{LD}\right)^+C_1^{LD\prime}\xi_x} = \frac{\xi_x'\mathbf{1}_{T-1}\mathbf{1}_{T-1}'H\xi_u}{\mathbf{1}_{T-1}'\Sigma_{22}\mathbf{1}_{T-1}}\frac{\mathbf{1}_{T-1}'\Sigma_{22}\mathbf{1}_{T-1}}{\left(\xi_x'\mathbf{1}_{T-1}\right)^2} = \xi_{AH}.$$

■

Proof of Corollary 3:. For $a = \{"AS","AB"\}$ the matrix W^a is of full rank $T-1$. Thus, for $L = \sqrt{\delta}I_{T-1}$, $z = L^{-1}\xi_x \sim \mathcal{N}(0,I_{T-1})$ and

$$E\left[\xi(C_1^a,\Sigma_{22})\right] = E\left[\frac{z'L'W^aHE\left[\xi_u|\xi_x\right]}{z'L'W^aLz}\right] = E\left[\frac{z'L'W^a\delta^{-1}\Sigma_{21}Lz}{z'L'W^aLz}\right]$$

and the result follows from Smith (1993, Eq 2.4, p.273). For $a = \{"LD"\}$ note that

$$\frac{\xi_x' W^a H \xi_u}{\xi_x' W^a \xi_x} = \frac{z_1' \Gamma_1' W^a H \xi_u}{z_1' \Lambda_1 z_1}$$

by the same arguments as in the proof of Theorem (2). The remainder of the proof is the same as in that Theorem. ■

Proof of Lemma 4. We note that

$$\bar{B}_n \rightarrow \begin{bmatrix} 0 & \cdots & & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 1 & \cdots & 1 & 0 \end{bmatrix} \equiv M_5$$

and $\bar{b}_n \rightarrow \mathbf{1}_T$ where $\bar{B}_n$ and $\bar{b}_n$ are defined in (10). Also define $\tilde{y}_{it} = \sum_{j=0}^{t-1} \beta_n^j u_{it-j}$ and $\tilde{g}_{i1} = B\left(\bar{B}_n \otimes I_{T-1}\right)(\Delta u_i \otimes u_i)$ such that from the definition of B in (20) it follows that $\tilde{g}_{i1} = \left[\tilde{y}_{it-1}\Delta u_{is}, T^{-1}\left(\sum_{t=1}^{T-1}\tilde{y}_{it} + y_{i0}\right)\Delta u_{ik}, \tilde{y}_{iT}\Delta u_{ij}\right]$ for $s = 2, ..., T$; $t = 1, ...s-2$; $j = 2, ..., T-1$; $k = 2, ..., T$. It can be checked easily that $E\left(\tilde{y}_{it-1}\Delta u_{is}\right) = O(n^{-1})$ for $s = 2, ..., T$; $t = 1, ...s-2$, $E\tilde{y}_{iT}\Delta u_{ij} = O(n^{-1})$ for $j = 2, ..., T-1$ and $E\left[T^{-1}\left(\sum_{t=1}^{T-1}\tilde{y}_{it} + y_{i0}\right)\Delta u_{ik}\right] = -T^{-1}\sigma_\varepsilon^2 + O(n^{-1})$. This shows that $E\tilde{g}_{i1} = -T^{-1}\sigma_\varepsilon^2\left[0, ..., 0, \mathbf{1}_{T-1}', 0, ..., 0\right]'$. Note that from arguments in the proof of Lemma 3 the estimator $\hat{b}_{GMM}$ is a continunous function of $n^{-3/2}\sum_{i=1}^n \left[f_{i2}', g_{i2}'\right]'$ such that convergence of $n^{-3/2}\sum_{i=1}^n \left[f_{i,1}', f_{i,2}', g_{i,1}' g_{i,2}'\right]'$ and $\hat{b}_{GMM}$ is joint and the continuous mapping theorem can be invoked to establish the limit of $n^{-3/2}\sum_{i=1}^n \left[\hat{f}_{i,1}', \hat{g}_{i,1}'\right]'$. It then follows from previous results that

$$\begin{aligned} n^{-3/2}\sum_{i=1}^n \hat{g}_{i,1} &= \left(\hat{b}_{GMM} - \beta_n\right) B\left(I_{T-1} \otimes \bar{b}_n\right) n^{-3/2}\sum_{i=1}^n g_{i2} + o_p(1) \\ &\xrightarrow{d} \xi_{GMM} B\left(I_{T-1} \otimes \mathbf{1}_T\right) H\xi_u \end{aligned}$$

and by the same arguments as before it follows that

$$\begin{aligned} n^{-3/2}\sum_{i=1}^n \hat{f}_{i,1} &= \left(\hat{b}_{GMM} - \beta_n\right) B\left(I_{T-1} \otimes \bar{b}_n\right) n^{-3/2}\sum_{i=1}^n f_{i2} + o_p(1) \\ &\xrightarrow{d} \xi_{GMM} B\left(I_{T-1} \otimes \mathbf{1}_T\right) \xi_x. \end{aligned}$$

Based on these results one concludes that

$$\begin{aligned} & n^{-3/2}\sum_{i=1}^n \left[\hat{f}_{i,1}', f_{i,2}', \hat{g}_{i,1}' g_{i,2}'\right]' \\ = \; & n^{-3/2}\sum_{i=1}^n \begin{bmatrix} \left(\hat{b}_{GMM} - \beta_n\right) B\left(I_{T-1} \otimes \bar{b}_n\right) & 0 \\ I_{T-1} & 0 \\ 0 & \left(\hat{b}_{GMM} - \beta_n\right) B\left(I_{T-1} \otimes \bar{b}_n\right) \\ 0 & I_{T-1} \end{bmatrix} \begin{bmatrix} f_{i,2} \\ g_{i,2} \end{bmatrix} + o_p(1). \end{aligned}$$

and the the statement of the Lemma follows from previous arguments. ■

Proof of Lemma 5. One needs to consider $\hat{g}_{i,1}'\left(\hat{b}_{GMM}\right)$ and $g_{i,2}'\left(\hat{b}_{GMM}\right)$. First for $\Delta y_i = [y_{iT} - y_{iT-1}, ..., y_{i2} - y_{i1}]'$

$$\begin{aligned} g_{i,2}'\left(\hat{b}_{GMM}\right) &= y_{i0}\left(\Delta y_i - \hat{b}_{GMM}\Delta y_{i,-1}\right) = y_{i0}\Delta u_i - \left(\hat{b}_{GMM} - \beta_n\right) y_{i0}\Delta y_{i,-1} \\ &= g_{i,2} - \left(\hat{b}_{GMM} - \beta_n\right) f_{i,2}. \end{aligned}$$

Second,

$$\begin{aligned} \hat{g}_{i1}\left(\hat{b}_{GMM}\right) &= B\left(I_{T(T-1)} + \left(\hat{b}_{GMM} - \beta_n\right)\left(I_{T-1}\otimes \bar{B}_n\right)\right)\left(u_i \otimes \left(\Delta y_i - \hat{b}_{GMM}\Delta y_{i,-1}\right)\right) \\ &\quad + \left(\hat{b}_{GMM} - \beta_n\right) B\left(I_{T-1}\otimes \bar{b}_n\right) g_{i,2}'\left(\hat{b}_{GMM}\right) \\ &= B\left(I_{T(T-1)} + \left(\hat{b}_{GMM} - \beta_n\right)\left(I_{T-1}\otimes \bar{B}_n\right)\right)\left(\left(\Delta u_i \otimes u_i\right) - \left(\hat{b}_{GMM} - \beta_n\right)\left(\Delta y_{i,-1}\otimes u_i\right)\right) \\ &\quad + \left(\hat{b}_{GMM} - \beta_n\right) B\left(I_{T-1}\otimes \bar{b}_n\right)\left(g_{i,2} - \left(\hat{b}_{GMM} - \beta_n\right) f_{i,2}\right) \end{aligned}$$

We have seen before that $E\left(\Delta u_i \Delta u_i' \otimes u_i u_i'\right) = O(1)$, $E\left(\Delta y_{i,-1}\Delta y_{i,-1}' \otimes u_i u_i'\right) = O(1)$ and $E\left(\Delta u_i \Delta y_{i,-1}' \otimes u_i u_i'\right) = O(1)$. Furthermore,

$$\begin{aligned} E\left(\Delta u_i \Delta u_i' \otimes u_i y_{i0}\right) &= E\left(\Delta u_i \Delta u_i'\right) \otimes E\left(u_i y_{i0}\right) + E\left(\Delta u_i y_{i0}\right) \otimes E\left(u_i \Delta u_i'\right) + E\left(u_i \otimes \Delta u_i\right) E\left(y_{i0}\Delta u_i'\right) \\ &= \sigma_\varepsilon^2 M_2 \otimes \frac{\sigma_\alpha^2 n}{c}\mathbf{1}_T + O(1), \end{aligned}$$

and in the same way

$$E\left(\Delta u_i \Delta y_{i,-1}' \otimes u_i y_{i0}\right) = \sigma_\varepsilon^2 M_1 \otimes \frac{\sigma_\alpha^2 n}{c}\mathbf{1}_T + O(1),$$

and

$$E\left(\Delta y_{i,-1}\Delta y_{i,-1}' \otimes u_i y_{i0}\right) = \sigma_\varepsilon^2 I_{T-1} \otimes \frac{\sigma_\alpha^2 n}{c}\mathbf{1}_T + O(1).$$

These results imply that

$$\hat{\Omega}_n = R_n n^{-3}\sum_{i=1}^{n}\begin{bmatrix} f_{i,2} \\ g_{i,2}\end{bmatrix}\begin{bmatrix} f_{i,2}' & g_{i,2}'\end{bmatrix} R_n' + o_p(1)$$

where

$$R_n = \begin{bmatrix} \left(\hat{b}_{GMM} - \beta_n\right) B\left(I_{T-1}\otimes \bar{b}_n\right) & -\left(\hat{b}_{GMM} - \beta_n\right)^2 B\left(I_{T-1}\otimes \bar{b}_n\right) \\ -\left(\hat{b}_{GMM} - \beta_n\right) I_{T-1} & I_{T-1}\end{bmatrix}$$

and

$$R_n \xrightarrow{d} R_\xi\left(\xi_{GMM}\right) = \begin{bmatrix} \xi_{GMM} B\left(I_{T-1}\otimes \mathbf{1}_T\right) & -\xi_{GMM}^2 B\left(I_{T-1}\otimes \mathbf{1}_T\right) \\ -\xi_{GMM} I_{T-1} & I_{T-1}\end{bmatrix}.$$

■

Proof of Theorem 5. As before unobserved innovations are replaced with $\hat{u}_i$ such that

$\hat{g}_{i,1}^{LD} = B^{LD}\left(I_{T(T-1)} + \left(\hat{b}_{GMM} - \beta_n\right) B^{LD}\left(I_{T-1} \otimes \bar{B}_n\right)\right)(\Delta u_i \otimes u_i) + \left(\hat{b}_{GMM} - \beta_n\right) B^{LD}\left(I_{T-1} \otimes \bar{b}_n\right) g_{i2}$

where B^{LD} is defined in (9) and $n^{-3/2}\sum_{i=1}^n \hat{g}_{i,1}^{LD} \xrightarrow{d} \xi_{GMM} B^{LD}\left(I_{T-1} \otimes \mathbf{1}_T\right) H\xi_u$. Similarly,

$\hat{f}_{i,1}^{LD} = B^{LD}\left(I_{T(T-1)} + \left(\hat{b}_{GMM} - \beta_n\right)\left(I_{T-1} \otimes \bar{B}_n\right)\right)(\Delta y_{i,-1} \otimes u_i) + \left(\hat{b}_{GMM} - \beta_n\right) B^{LD}\left(I_{T-1} \otimes \bar{b}_n\right) f_{i2}$

such that

$$n^{-3/2}\sum_{i=1}^n \left[\hat{f}_{i,1}^{LD\prime}, f_{i,2}^{\prime}, \hat{g}_{i,1}^{LD\prime} g_{i,2}^{\prime}\right]^{\prime} \xrightarrow{d} \begin{bmatrix} \xi_{GMM} B^{LD}\left(I_{T-1} \otimes \mathbf{1}_T\right) & 0 \\ I_{T-1} & 0 \\ 0 & \xi_{GMM} B^{LD}\left(I_{T-1} \otimes \mathbf{1}_T\right) \\ 0 & I_{T-1} \end{bmatrix} \begin{bmatrix} \xi_x \\ H\xi_u \end{bmatrix}.$$

By the same arguments as before it also follows that for

$$\hat{\Omega}_n^{LD} = n^{-3}\sum_{i=1}^n \left[\hat{g}_{i,1}^{LD\prime}\left(\hat{b}_{GMM}\right), g_{i,2}^{\prime}\left(\hat{b}_{GMM}\right)\right]^{\prime}\left[\hat{g}_{i,1}^{LD\prime}\left(\hat{b}_{GMM}\right), g_{i,2}^{\prime}\left(\hat{b}_{GMM}\right)\right]$$

one obtains

$$\hat{\Omega}_n^{LD} = R_n^{LD} n^{-3}\sum_{i=1}^n \begin{bmatrix} f_{i,2} \\ g_{i,2} \end{bmatrix}\begin{bmatrix} f_{i,2}^{\prime} & g_{i,2}^{\prime} \end{bmatrix} R_n^{LD\prime} + o_p(1)$$

where

$$R_n = \begin{bmatrix} \left(\hat{b}_{GMM} - \beta_n\right) B^{LD}\left(I_{T-1} \otimes \bar{b}_n\right) & -\left(\hat{b}_{GMM} - \beta_n\right)^2 B^{LD}\left(I_{T-1} \otimes \bar{b}_n\right) \\ -\left(\hat{b}_{GMM} - \beta_n\right) I_{T-1} & I_{T-1} \end{bmatrix}$$

the weight matrix converges to the random limit

$$\hat{\Omega}_n^{LD} \xrightarrow{d} R_\xi^{LD}\left(\xi_{GMM}\right)\Sigma R_\xi^{LD}\left(\xi_{GMM}\right)^{\prime}$$

where

$$R_n^{LD} \xrightarrow{d} R_\xi^{LD}\left(\xi_{GMM}\right) = \begin{bmatrix} \xi_{GMM} B^{LD}\left(I_{T-1} \otimes \mathbf{1}_T\right) & -\xi_{GMM}^2 B^{LD}\left(I_{T-1} \otimes \mathbf{1}_T\right) \\ -\xi_{GMM} I_{T-1} & I_{T-1} \end{bmatrix}.$$

Next note that $B^{LD}\left(I_{T-1} \otimes \mathbf{1}_T\right) = \mathbf{1}_{T-1}^{\prime} \otimes \mathbf{1}_{T-2}$, $\mathbf{1}_{T-1}^{\prime}\Sigma_{11}\mathbf{1}_{T-1} = (T-1)\,\delta$, $\mathbf{1}_{T-1}^{\prime}\Sigma_{12}\mathbf{1}_{T-1} = -\delta$ and $\mathbf{1}_{T-1}^{\prime}\Sigma_{22}\mathbf{1}_{T-1} = 2\delta$. Use the short hand notation $\xi_{GMM} = \xi_G$. It then follows that

$$C^{LD\prime} R_\xi^{LD}\left(\xi_G\right)\Sigma R_\xi^{LD}\left(\xi_G\right)^{\prime} C^{LD} = \begin{bmatrix} a\mathbf{1}_{T-2}\mathbf{1}_{T-2}^{\prime} & b\mathbf{1}_{T-2} \\ b\mathbf{1}_{T-2}^{\prime} & d \end{bmatrix}$$

where $a=\xi_G^2(T-1)\delta+2\xi_G^3\delta+2\xi_G^4\delta$, $b=-\xi_G^2(T-1)\delta-\xi_G^3\delta-\xi_G\delta-2\xi_G^2\delta$ and $d=\xi_G^2(T-1)\delta+2\xi\delta+2\delta$. It then follows that

$$\left(C^{LD\prime}R_\xi^{LD}(\xi_G)\,\Sigma R_\xi^{LD}(\xi_G)'\,C^{LD}\right)^+=\begin{bmatrix} e\mathbf{1}_{T-2}\mathbf{1}_{T-2}' & f\mathbf{1}_{T-2} \\ f\mathbf{1}_{T-2}' & g \end{bmatrix}$$

for some coefficients e,f,g as well as

$$C^{LD}\left(C^{LD\prime}R_\xi^{LD}(\xi_G)\,\Sigma R_\xi^{LD}(\xi_G)'\,C^{LD}\right)^+C^{LD\prime}=\begin{bmatrix} e\mathbf{1}_{T-2}\mathbf{1}_{T-2}' & f\mathbf{1}_{T-2}\mathbf{1}_{T-1}' \\ f\mathbf{1}_{T-1}\mathbf{1}_{T-2}' & g\mathbf{1}_{T-1}\mathbf{1}_{T-1}' \end{bmatrix}.$$

Finally, noting that $B^{LD}(\xi_x\otimes\mathbf{1}_T)=\left(\xi_x'\mathbf{1}_{T-1}\right)\mathbf{1}_{T-2}$ one obtains

$$\begin{aligned}
& \hat{\xi}_x(\xi_G)'\,C^{LD}\left(C^{LD\prime}R_\xi^{LD}(\xi_G)\,\Sigma R_\xi^{LD}(\xi_G)'\,C^{LD}\right)^+C^{LD\prime}\hat{\xi}_x(\xi_G) \\
= \; & e\xi_G^2\left(\mathbf{1}_{T-2}'\mathbf{1}_{T-2}\right)^2\left(\xi_x'\mathbf{1}_{T-1}\right)^2+2f\xi_G\left(\mathbf{1}_{T-2}'\mathbf{1}_{T-2}\right)\left(\xi_x'\mathbf{1}_{T-1}\right)^2+g\left(\xi_x'\mathbf{1}_{T-1}\right)^2
\end{aligned}$$

as well as

$$\begin{aligned}
& \hat{\xi}_x(\xi_G)'\,C^{LD}\left(C^{LD\prime}R_\xi^{LD}(\xi_G)\,\Sigma R_\xi^{LD}(\xi_G)'\,C^{LD}\right)^+C^{LD\prime}\hat{\xi}_y(\xi_G) \\
= \; & \left(e\xi_G^2\left(\mathbf{1}_{T-2}'\mathbf{1}_{T-2}\right)^2+2f\xi_G\left(\mathbf{1}_{T-2}'\mathbf{1}_{T-2}\right)+g\right)\left(\xi_x'\mathbf{1}_{T-1}\right)\left(\xi_u'H'\mathbf{1}_{T-1}\right)
\end{aligned}$$

This now implies that

$$\xi_{FLD}=\frac{\left(\xi_x'\mathbf{1}_{T-1}\right)\left(\xi_u'H'\mathbf{1}_{T-1}\right)}{\left(\xi_x'\mathbf{1}_{T-1}\right)^2}=\xi_{LD}.$$

■

C Definitions

C.1 Definitions for Limiting Distributions

Definitions for Lemma (1)

$$\mathbb{K}_1=\begin{bmatrix} c_{1,1,\alpha,\alpha} & \cdots & c_{1,T-1,\alpha,\alpha} \\ \vdots & \ddots & \vdots \\ c_{T-1,1,\alpha,\alpha} & \cdots & c_{T-1,T-1,\alpha,\alpha} \end{bmatrix},$$

$$\mathbb{K}_2=\begin{bmatrix} c_{2,2,\alpha,\alpha}-c_{2,1,\alpha,\alpha}-c_{1,2,\alpha,\alpha}+c_{1,1,\alpha,\alpha} & \cdots & c_{2,T,\alpha,\alpha}-c_{2,T-1,\alpha,\alpha}-c_{1,T,\alpha,\alpha}+c_{1,T-1,\alpha,\alpha} \\ \vdots & \ddots & \vdots \\ c_{T,2,\alpha,\alpha}-c_{T,1,\alpha,\alpha}-c_{T-1,2,\alpha,\alpha}+c_{T-1,T-1,\alpha,\alpha} & \cdots & c_{T,T,\alpha,\alpha}-c_{T,T-1,\alpha,\alpha}-c_{T,T,\alpha,\alpha}+c_{T-1,T-1,\alpha,\alpha} \end{bmatrix},$$

$$\mathbb{K}_3 = \begin{bmatrix} c_{1,2,\alpha,\alpha} - c_{1,1,\alpha,\alpha} & \cdots & c_{1,T,\alpha,\alpha} - c_{1,T-1,\alpha,\alpha} \\ \vdots & \ddots & \vdots \\ c_{T-1,2,\alpha,\alpha} - c_{T-1,1,\alpha,\alpha} & \cdots & c_{T-1,T,\alpha,\alpha} - c_{T-1,T-1,\alpha,\alpha} \end{bmatrix}.$$

Define

$$H = \begin{bmatrix} -1 & 1 & & 0 & 0 \\ & \ddots & \ddots & & \vdots \\ & 0 & -1 & 1 & 0 \\ 0 & \cdots & & 0 & 1 \end{bmatrix}. \tag{19}$$

C.2 Selector Matrix B

Let B be given as

$$B = \left[S'_{\Delta,T-2}, ..., S'_{\Delta,1}, S'_{\Delta,0}, S'_{\Delta,-1}\right]' \tag{20}$$

where

$$S_{\Delta,j} = \begin{cases} \left[0_{T-1-j,(T-1-j)(T-1)}, I_{T-1-j}, 0_{T-1-j,j}, 0_{T-1-j,j(T-1)}\right] & \text{for } j > 0 \\ I_{T-1} \otimes T^{-1}\mathbf{1}'_T & \text{for } j = 0 \\ \left[0_{T-2,(T-1)(T-1)}, I_{T-2}, 0_{T-2,1}\right] & \text{for } j = -1 \end{cases}$$

with I_i the $i \times i$ identity matrix and $0_{j,i}$ a $j \times i$ matrix of zeros. The convention is used that when the index i or j of $0_{i,j}$ is zero then the corresponding submatrix is absent.

C.3 Definitions for the Arellano and Bond Estimator

Define the matrices

$$S_{0,j}^{AB,n} = \left[\beta_n^j, ..., \beta_n^0\right],$$

$$S_{1,k}^{AB,n} = \left[0_{k+1,(k-1)k/2}, \begin{matrix} 0_{1,k} \\ \left[S_{0,0}^{AB,n}, 0_{1,k-1}\right] \\ \vdots \\ \left[S_{0,k-1}^{AB,n}\right] \end{matrix}, 0_{k+1,q_3-k(k-1)/2-k} \right]$$

where $q_3 = T\,(T+1)\,/2 - 2$ and

$$C^{AB,n} = \begin{bmatrix} C_0^{AB,n} \\ C_1^{AB,n} \end{bmatrix}$$

with

$$C_0^{AB,n} = \left[\begin{array}{cccc} 0_{q_2,1} & S_{1,1}^{AB,n\prime} & \cdots & S_{1,T-2}^{AB,n\prime} \end{array} \right], \; C_1^{AB,n} = \left[\begin{array}{cccccccccc} 1 & 0 & \cdots & & & & & & & \\ 0 & 1 & \beta_n & 0 & \cdots & & & & & \\ & & & 1 & \beta_n & \beta_n^2 & 0 & \cdots & & \\ & & & & & & \ddots & & & \\ & & & & & & & 1 & \cdots & \beta_n^{T-2} \end{array} \right].$$

Then $\lim_{n\to\infty} C_0^{AB,n} = C_0^{AB}$ and $\lim_{n\to\infty} C_1^{AB,n} = C_1^{AB}$ are given by

$$C_0^{AB} = \left[\begin{array}{cccc} 0_{q_2,1} & S_{1,1}^{AB\prime} & \cdots & S_{1,T-2}^{AB\prime} \end{array} \right], \; C_1^{AB} = \left[\begin{array}{cccccc} 1 & 0 & \cdots & & & \\ 0 & \mathbf{1}_2' & 0 & \cdots & & \\ & & \mathbf{1}_3' & 0 & & \cdots \\ & & & & \ddots & \\ & & & & & \mathbf{1}_{T-1}' \end{array} \right]$$

where

$$S_{1,k}^{AB} = \left[\begin{array}{c} 0_{1,k} \\ [1, 0_{1,k-1}] \\ 0_{k+1,(k-1)k/2}, \;\; [1,1,0_{1,k-2}] \;\; , 0_{k+1,q_3-k(k-1)/2-k} \\ \vdots \\ [1,...,1] \end{array} \right]$$

C.4 Definitions for the Anderson and Hsiao Estimator

Define the matrices

$$S_{0,j}^{AH,n} = \left[\beta_n^j, ..., \beta_n^0\right],$$

$$S_{1,k}^{AH,n} = \left[0_{1,(k-1)k/2}, S_{0,k-1}^{AB,n}, 0_{1,q_3-k(k-1)/2-k}\right]$$

where $q_3 = T\,(T+1)\,/2 - 2$ and

$$C^{AH,n} = \left[\begin{array}{c} C_0^{AH,n} \\ C_1^{AH,n} \end{array} \right]$$

with

$$C_0^{AH,n} = \begin{bmatrix} 0_{q_2,1} & S_{1,1}^{AB,n\prime} & \cdots & S_{1,T-2}^{AB,n\prime} \end{bmatrix} \mathbf{1}_{T-1}, C_1^{AH,n} = \begin{bmatrix} 1 & \cdots & & & 0 \\ 0 & \beta_n & & & \\ & & \beta_n^2 & & \\ & & & \ddots & \\ & & & & \beta_n^{T-2} \end{bmatrix} \mathbf{1}_{T-1}$$

$$C_1^{AH} = \mathbf{1}_{T-1}.$$

C.5 Definitions for the Arellano and Bover Estimator

$$B^* = \begin{bmatrix} \sqrt{\frac{T-1}{T}} & \sqrt{\frac{T-1}{T}}\frac{T-2}{T-1} & \sqrt{\frac{T-1}{T}}\frac{T-3}{T-1} & \cdots & \sqrt{\frac{T-1}{T}}\frac{1}{T-1} \\ 0 & \sqrt{\frac{T-2}{T-1}} & \sqrt{\frac{T-2}{T-1}}\frac{T-3}{T-2} & \cdots & \sqrt{\frac{T-2}{T-1}}\frac{1}{T-2} \\ \vdots & & \ddots & & \vdots \\ & & \cdots & \sqrt{\frac{2}{3}} & \sqrt{\frac{2}{3}}\frac{1}{2} \\ 0 & & \cdots & 0 & \sqrt{\frac{1}{2}} \end{bmatrix} \tag{21}$$

C.6 Definitions for the Ahn and Schmidt Estimator

Define $C^{AS,n}$ as

$$C^{AS,n} = \begin{bmatrix} C_0^{AS,n} \\ C_1^{AS,n} \end{bmatrix}$$

where

$$C_0^{AS,n\prime} = \begin{bmatrix} C_0^{AB,n\prime} \\ \left[0_{T-2,(T-2)(T-1)/2+T-1}, I_{T-2}\right] \end{bmatrix}, \ C_1^{AS,n\prime} = \begin{bmatrix} C_1^{AB,n\prime} \\ 0_{T-2,T-1} \end{bmatrix}$$

References

[1] Ahn, S., and P. Schmidt, 1995, “Efficient Estimation of Models for Dynamic Panel Data”, *Journal of Econometrics* 68, 5 - 28.

[2] Ahn, S., and P. Schmidt, 1997, “Efficient Estimation of Dynamic Panel Data Models: Alternative Assumptions and Simplified Estimation”, *Journal of Econometrics* 76, 309 - 321.

[3] Alonso-Borrego, C. and M. Arellano, 1999, “Symmetrically Normalized Instrumental-Variable Estimation Using Panel Data”, *Journal of Business and Economic Statistics* 17, pp. 36-49.

[4] Alvarez, J. and M. Arellano, 1998, “The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators”, *unpublished manuscript*.

[5] Andrews, D.W.K., 1992, ”Generic Uniform Convergence”, *Econometric Theory,* 8, pp. 241-257.

[6] Andrews, D.W.K., 1994, ”Empirical Process Methods in Econometrics.” In *Handbook of Econometrics,* Vol IV (R.F. Engle and D.L. McFadden, eds.). North Holland, Amsterdam.

[7] Anderson, T.W. and C. Hsiao, 1981, “Estimation of Dynamic Models with Error Components”, *Journal of the American Statistical Association* 76, No. 375, pp. 598 - 606.

[8] Anderson, T.W. and C. Hsiao, 1982, “Formulation and Estimation of Dynamic Models Using Panel Data”, *Journal of Econometrics* 18, pp. 47 - 82.

[9] Arellano, M. and S.R. Bond, 1991, “Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations”, *Review of Economic Studies* 58, pp. 277 - 297.

[10] Arellano, M. and O. Bover, 1995, “Another Look at the Instrumental Variable Estimation of Error-Component Models”, *Journal of Econometrics* 68, pp. 29 - 51.

[11] Bekker, P., 1994, “Alternative Approximations to the Distributions of Instrumental Variable Estimators”, *Econometrica* 62, pp. 657 - 681.

[12] Blundell, R. and S. Bond, 1998, “Initial Conditions and Moment Restrictions in Dynamic Panel Data Models”, *Journal of Econometrics* 87, pp. 115 - 143.

[13] CHAMBERLAIN, G., 1984, "Panel Data", in Griliches, Z. and M. D. Intriligator, eds., *Handbook of Econometrics* Vol. 2. Amsterdam: North-Holland.

[14] CHAO, J. AND N. SWANSON, 2000, "An Analysis of the Bias and MSE of the IV Estimator under Weak Identification with Application to Bias Correction", *unpublished manuscript.*

[15] CHOI, I. AND P.C.B. PHILLIPS, 1992, "Asymptotic and Finite Sample Distribution Theory for IV Estimators and Tests in Partially Identified Structural Equations", *Journal of Econometrics* 51, 113 - 150.

[16] DAVIS, A.W, 1980, "Invariant polynomials with two matrix arguments, extending the zonal polynomials", in P.R. Krishnaiah, ed, *Mulitivariate Analysis V*, 287-299. Amsterdam: North Holland.

[17] DONALD, S. AND W. NEWEY, 2001, "Choosing the Number of Instruments", *Econometrica,* 69, 1161-1191.

[18] DUFOUR, J-M, 1997, "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models", *Econometrica* 65, 1365 - 1388.

[19] GRILICHES, Z. AND J. HAUSMAN, 1986, "Errors in Variables in Panel Data," *Journal of Econometrics* 31, pp. 93 - 118.

[20] HAHN, J., 1997, "Efficient Estimation of Panel Data Models with Sequential Moment Restrictions", *Journal of Econometrics* 79, pp. 1 - 21.

[21] HAHN, J., 1999, "How Informative Is the Initial Condition in the Dynamic Panel Model with Fixed Effects?", *Journal of Econometrics* 93, pp. 309 - 326.

[22] HAHN, J. AND J. HAUSMAN, 2002, "A New Specification Test for the Validity of Instrumental Variables", *Econometrica*, 70, 163-189.

[23] HAHN, J., J. HAUSMAN AND G. KUERSTEINER, 2004, "Estimation with Weak Instruments: Accuracy of Higher Order Bias and MSE Approximations", *Econometrics Journal*, 7, 272-306.

[24] HAHN, J. AND G. KUERSTEINER, 2002, "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both n and T are Large", *Econometrica,* 70, 1639-1657.

[25] HAHN, J. AND G. KUERSTEINER, 2002, "Discontinuities of Weak Instrument Limiting Distributions," *Economics Letters,* 75, 325-331.

[26] HAHN, J., G. KUERSTEINER, AND W.K. NEWEY, 2002, "Higher Order Properties of Bootstrap and Jackknife Bias Corrected Maximum Likelihood Estimators", *unpublished manuscript.*

[27] HALL, P. AND J. HOROWITZ, 1996, "Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators", *Econometrica* 64, 891-916.

[28] HAUSMAN, J.A., AND W.E. TAYLOR, 1983, "Identification in Linear Simultaneous Equations Models with Covariance Restrictions: An Instrumental Variables Interpretation", *Econometrica* 51, pp. 1527-1550.

[29] HAUSMAN, J.A., W.K. NEWEY, AND W.E. TAYLOR, 1987, "Efficient Estimation and Identification of Simultaneous Equation Models with Covariance Restrictions", *Econometrica* 55, pp. 849-874.

[30] HOLTZ-EAKIN, D., W.K. NEWEY, AND H. ROSEN, "Estimating Vector Autoregressions with Panel Data", *Econometrica* 56, pp. 1371 - 1395.

[31] KIVIET, J.F., 1995, "On Bias, Inconsistency, and Efficiency of Various Estimators in Dynamic Panel Data Models", *Journal of Econometrics* 68, pp. 1 - 268.

[32] KRUINIGER, H., 2000, "GMM Estimation of Dynamic Panel Data Models with Persistent Data," *unpublished manuscript.*

[33] KUERSTEINER, G.M., 2000, "Mean Squared Error Reduction for GMM Estimators of Linear Time Series Models" *unpulished manuscript.*

[34] LANCASTER, T., 1997, "Orthogonal Parameters and Panel Data", *unpublished manuscript.*

[35] MAGNUS J., AND H. NEUDECKER (1988), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley & Sons Ltd.

[36] MARIANO, R.S. AND T. SAWA, 1972, "The Exact Finite-Sample Distribution of the Limited-Information Maximum Likelihood Estimator in the Case of Two Included Endogenous Variables", *Journal of the American Statistical Association* 67, 159 - 163.

[37] MOON, H.R. AND P.C.B. PHILLIPS, 2000, "GMM Estimation of Autoregressive Roots Near Unity with Panel Data", *unpublished manuscript.*

[38] NAGAR, A.L., 1959, "The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations", *Econometrica* 27, pp. 575 - 595.

[39] NELSON, C., R. STARTZ, AND E. ZIVOT, 1998, "Valid Confidence Intervals and Inference in the Presence of Weak Instruments", *International Economic Review* 39, 1119 - 1146.

[40] NEWEY, W.K. AND R.J. SMITH, 2000, "Asymptotic Equivalence of Empirical Likelihood and the Continuous Updating Estimator", *unpublished manuscript.*

[41] NICKELL, S., 1981, "Biases in Dynamic Models with Fixed Effects", *Econometrica* 49, pp, 1417 - 1426.

[42] PAKES, A AND D. POLLARD, 1989, "Simulation and the Asymptotics of Optimazation Estimators," *Econometrica* 57, pp. 1027-1057.

[43] PHILLIPS, P.C.B., 1989, "Partially Identified Econometric Models", *Econometric Theory* 5, 181 - 240.

[44] RILSTONE, P., V.K. SRIVASTAVA, A. ULLAH, 1996, "The Second-Order Bias and Mean Squared Error of Nonlinear Estimators", *Journal of Econometrics* 75, pp. 369 - 395.

[45] RICHARDSON, D.J., 1968, "The Exact Distribution of a Structural Coefficient Estimator", *Journal of the American Statistical Association* 63, 1214 - 1226.

[46] ROTHENBERG, T.J., 1983, "Asymptotic Properties of Some Estimators In Structural Models", in *Studies in Econometrics, Time Series, and Multivariate Statistics.*

[47] ROTHENBERG, T. J., 1984, "Approximating the Distributions of Econometric Estimators and Test Statistics", in Griliches, Z. and M. D. Intriligator, eds., *Handbook of Econometrics*, Vol. 2. Amsterdam: North-Holland.

[48] SMITH, M.D., 1993, "Expectations of Ratios of Quadratic Forms in Normal Variables: Evaluating Some Top-Order Invariant Polynomials", *Australian Journal of Statistics* 35, 271-282.

[49] STAIGER, D., AND J. H. STOCK (1997): "Instrumental Variables Regressions with Weak Instruments," *Econometrica*, 65, 557 - 586.

[50] STOCK, J. H. AND J. H. WRIGHT (2000): "GMM with Weak Identification," *Econometrica*, 68, 1055 - 1096.

[51] WANG, J. AND E. ZIVOT, 1998, "Inference on Structural Parameters in Instrumental Variables Regression with Weak Instruments", *Econometrica* 66,1389 – 1404.

[52] WINDMEIJER, F., 2000, "GMM Estimation of Linear Dynamic Panel Data Models using Nonlinear Moment Conditions", *unpublished manuscript.*

[53] ZHANG, F., 1999, "Matrix Theory", Springer Verlag.

Table 1: Monte Carlo Comparison of Estimators

Estimator	$\widehat{b}_{BB}$	$\widehat{b}_{LDGMM}$	$\widehat{b}_{AS}$	$\widehat{b}_{LD-AB}$	$\widehat{b}_{LD1}$	$\widehat{b}_{LD2}$	$\widehat{b}_{LD3}$
$\beta = 0.1, T = 5, n = 100$							
Mean	0.105	0.101	0.098	0.103	0.100	0.102	0.100
RMSE	0.072	0.092	0.070	0.101	0.106	0.104	0.103
Median	0.104	0.101	0.098	0.102	0.101	0.102	0.099
1st percentile	-0.054	-0.113	-0.057	-0.127	-0.144	-0.145	-0.136
25th percentile	0.057	0.040	0.050	0.033	0.031	0.034	0.032
75th percentile	0.152	0.160	0.143	0.169	0.169	0.169	0.165
99th percentile	0.276	0.318	0.268	0.344	0.352	0.352	0.349
$\beta = 0.5, T = 5, n = 100$							
Mean	0.508	0.526	0.504	0.507	0.512	0.519	0.520
RMSE	0.092	0.150	0.113	0.102	0.123	0.146	0.170
Median	0.507	0.503	0.491	0.504	0.503	0.503	0.497
1st percentile	0.296	0.273	0.295	0.282	0.255	0.250	0.245
25th percentile	0.448	0.429	0.430	0.437	0.429	0.421	0.418
75th percentile	0.569	0.588	0.559	0.573	0.588	0.598	0.592
99th percentile	0.725	1.033	0.922	0.759	0.826	0.953	1.073
$\beta = 0.8, T = 5, n = 100$							
Mean	0.753	0.816	0.766	0.797	0.819	0.835	0.854
RMSE	0.158	0.182	0.184	0.214	0.254	0.274	0.294
Median	0.767	0.786	0.736	0.751	0.775	0.783	0.783
1st percentile	0.331	0.493	0.405	0.513	0.369	0.423	0.436
25th percentile	0.662	0.682	0.641	0.672	0.685	0.685	0.682
75th percentile	0.861	0.936	0.871	0.856	0.892	0.911	0.938
99th percentile	1.054	1.280	1.208	1.602	1.763	1.849	1.855
$\beta = 0.9, T = 5, n = 100$							
Mean	0.664	0.826	0.709	0.816	0.853	0.880	0.902
RMSE	0.388	0.221	0.317	0.249	0.240	0.234	0.264
Median	0.693	0.798	0.697	0.762	0.814	0.838	0.844
1st percentile	-0.221	0.380	0.077	0.467	0.384	0.427	0.495
25th percentile	0.499	0.689	0.558	0.691	0.734	0.747	0.750
75th percentile	0.882	0.954	0.854	0.861	0.911	0.961	0.981
99th percentile	1.273	1.349	1.295	1.752	1.764	1.722	1.879

Table 2: Median Bias of Near Unit Root Approximations

T	AS/AB/GMM	LD/AH	Optimal
5	-0.6619	-0.2445	-0.1910
6	-0.5694	-0.1822	-0.1340
7	-0.5006	-0.1677	-0.0990
8	-0.4510	-0.1523	-0.0761
9	-0.4121	-0.1279	-0.0603
10	-0.3722	-0.1083	-0.0489

Table 3: Median Bias of Near Unit Root Approximations

T	FGMM	FAS	FLD	FLD(4)
5	-0.6619	-0.8250	-0.2445	-0.2445
6	-0.5694	-0.7903	-0.1822	-0.1822
7	-0.5006	-0.7704	-0.1677	-0.1677
8	-0.4510	-0.7173	-0.1523	-0.1523
9	-0.4121	-0.6905	-0.1279	-0.1279
10	-0.3722	-0.6586	-0.1083	-0.1083

Table 4: MSE of Near Unit Root Approximations

T	Optimal	Moments	AS/GMM	FGMM	FAS	Moments (AS)
5	0.6506	3	0.7411	0.7411	1.2152	17
6	0.4687	4	0.5584	0.5584	1.1012	24
7	0.3159	4	0.4640	0.4640	1.0026	32
8	0.2107	4	0.3702	0.3702	0.9311	41
9	0.1513	4	0.3149	0.3149	0.8823	51
10	0.1079	4	0.2636	0.2636	0.8466	62

Table 5: IQR of Near Unit Root Approximations

T	AS/AB/GMM	LD/AH	FGMM	FAS	FLD
5	0.6798	1.3415	0.6798	1.0007	1.3415
6	0.6110	1.1827	0.6110	0.9973	1.1827
7	0.5403	1.1381	0.5403	1.0285	1.1381
8	0.4761	1.0182	0.4761	1.0291	1.0182
9	0.4424	0.9694	0.4424	1.0280	0.9694
10	0.4071	0.9436	0.4071	1.0339	0.9436

Table 6: Finite Sample Properties of $\widehat{b}_{GMM}$ and $\widehat{b}_{BC}$

T	n	β	%bias$\left(\widehat{b}_{GMM}\right)$		%bias$\left(\widehat{b}_{BC}\right)$	RMSE$\left(\widehat{b}_{GMM}\right)$	RMSE$\left(\widehat{b}_{BC}\right)$
			Actual	Second Order Theory	Actual	Actual	Actual
5	100	0.1	-14.96	-17.71	0.25	0.08	0.08
10	100	0.1	-14.06	-15.78	-0.77	0.05	0.05
5	500	0.1	-3.68	-3.54	-0.38	0.04	0.04
10	500	0.1	-3.15	-3.16	-0.16	0.02	0.02
5	100	0.5	-10.05	-12.09	-1.14	0.13	0.13
10	100	0.5	-6.76	-8.00	-0.93	0.06	0.06
5	500	0.5	-2.25	-2.42	-0.15	0.06	0.06
10	500	0.5	-1.53	-1.60	-0.11	0.03	0.03
5	100	0.8	-27.65	-37.81	-11.33	0.32	0.34
10	100	0.8	-13.45	-18.98	-4.55	0.14	0.11
5	500	0.8	-6.98	-7.56	-0.72	0.13	0.13
10	500	0.8	-3.48	-3.80	-0.37	0.05	0.04
5	100	0.9	-50.22	-118.64	-42.10	0.55	0.78
10	100	0.9	-24.27	-52.66	-15.82	0.25	0.23
5	500	0.9	-20.50	-23.73	-6.23	0.28	0.30
10	500	0.9	-8.74	-10.53	-2.02	0.10	0.08

Table 7: Inconsistency of $\widehat{b}_{BB}$ under Misspecification

Weight Matrix I				
	β_F			
β	0.3	0.6	0.8	0.9
0.3	0.000	0.071	0.074	0.035
0.6	0.267	0.000	0.116	0.054
0.8	0.253	0.205	0.000	0.088
0.9	0.220	0.181	0.129	0.000
Weight Matrix II				
	β_F			
β	0.3	0.6	0.8	0.9
0.3	0.000	0.077	0.149	0.084
0.6	0.267	0.000	0.268	0.149
0.8	0.258	0.208	0.000	0.126
0.9	0.235	0.197	0.137	0.000

Table 8: Higher Order Theoretical Properties of Various Estimators

β	n	T	$\widehat{b}_{GMM}$	$\widehat{b}_{LD-AB}$	$\widehat{b}_{LD1}$	$\widehat{b}_{LD2}$	$\widehat{b}_{LD3}$	$\widehat{b}_{LD4}$	$\widehat{b}_{AS}$	$\widehat{b}_{BB}$	$\widehat{b}_{LDGMM}$
Bias Predicted by Second Order Theory											
0.1	100	5	-0.017	0.000	0.001	0.001	0.001	0.001	-0.001	0.005	0.003
0.3	100	5	-0.029	0.000	0.003	0.005	0.005	0.005	-0.001	0.008	0.005
0.5	100	5	-0.054	-0.007	0.003	0.009	0.013	0.014	-0.002	0.012	0.012
0.7	100	5	-0.135	-0.040	-0.007	0.014	0.030	0.043	0.002	0.014	0.046
0.9	100	5	-1.019	-0.449	-0.119	0.081	0.214	0.317	0.340	-0.078	1.055
RMSE Predicted by Higher Order Theory											
0.1	100	5	0.081	0.099	0.102	0.102	0.102	0.102	0.069	0.070	0.091
0.3	100	5	0.100	0.097	0.109	0.112	0.113	0.113	0.077	0.079	0.100
0.5	100	5	0.133	0.097	0.120	0.134	0.142	0.147	0.094	0.093	0.124
0.7	100	5	0.220	0.121	0.141	0.173	0.212	0.250	0.158	0.123	0.242
0.9	100	5	0.965	0.609	1.181	1.449	1.561	1.633	2.099	0.422	3.558

Table 9: Properties of Selected Estimators: Higher Order Theory and Practice

β	n	T	$\widehat{b}_{GMM}$	$\widehat{b}_{LD-AB}$	$\widehat{b}_{LD1}$	$\widehat{b}_{LD2}$	$\widehat{b}_{LD3}$	$\widehat{b}_{LD4}$
Bias Predicted by Higher Order Theory								
0.10	100	5	-0.017	0.000	0.001	0.001	0.001	0.001
0.50	100	5	-0.054	-0.007	0.003	0.009	0.013	0.014
0.80	100	5	-0.281	-0.105	-0.029	0.015	0.051	0.083
0.90	100	5	-1.019	-0.449	-0.119	0.081	0.214	0.317
Actual Bias								
0.10	100	5	-0.015	0.001	0.002	0.002	0.002	0.002
0.50	100	5	-0.050	-0.005	0.005	0.012	0.019	0.024
0.80	100	5	-0.221	-0.068	-0.024	0.005	0.036	0.044
0.90	100	5	-0.452	-0.142	-0.077	-0.037	0.001	0.010
RMSE Predicted by Higher Order Theory								
0.10	100	5	0.081	0.099	0.102	0.102	0.102	0.102
0.50	100	5	0.133	0.097	0.120	0.134	0.142	0.147
0.80	100	5	0.354	0.189	0.228	0.261	0.318	0.399
0.90	100	5	0.965	0.609	1.181	1.449	1.561	1.633
Actual RMSE								
0.10	100	5	0.081	0.099	0.102	0.102	0.102	0.102
0.50	100	5	0.131	0.097	0.120	0.137	0.158	0.172
0.80	100	5	0.318	0.144	0.161	0.199	0.265	0.298
0.90	100	5	0.552	0.188	0.170	0.195	0.255	0.285