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

Jinyong Hahn
Brown University

Jerry Hausman
MIT

Guido Kuersteiner
MIT

June, 2001

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.

Abstract

This paper analyzes the second order bias of instrumental variables estimators for a dynamic panel model with fixed effects. Three different methods of second order bias correction are considered. Simulation experiments show that these methods perform well if the model does not have a root near unity but break down near the unit circle. To remedy the problem near the unit root a weak instrument approximation is used. We show that an estimator based on long differencing the model is approximately achieving the minimal bias in a certain class of instrumental variables (IV) estimators. Simulation experiments document the performance of the proposed procedure in finite samples.

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. Comparisons of information contents of varieties of moment restrictions made by Ahn and Schmidt (1995) and Hahn (1999) suggest that, unless stationarity of the initial level y_{i0} is somehow exploited as in Blundell and Bond (1998), the orthogonality of lagged levels with first differences provide the largest source of information.

Unfortunately, the standard GMM estimator obtained after first differencing has been found to suffer from substantial finite sample biases. 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 (2000). 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.

In this paper, we attempt to eliminate the finite sample bias of the standard GMM estimator obtained after first differencing. We view the standard GMM estimator as a minimum distance estimator that combines $T-1$ instrumental variable estimators (2SLS) applied to first differences. This view has been adopted by Chamberlain (1984) and Griliches and Hausman (1986). It has been noted for quite a while that IV estimators can be quite biased in finite sample. See Nagar (1959), Mariano and Sawa (1972), Rothenberg (1983), Bekker (1994), Donald and Newey (1998) and Kuersteiner (2000). If the ingredients of the minimum distance estimator are all biased, it is natural to expect such bias in the resultant minimum distance estimator, or equivalently, GMM. We propose to eliminate the bias of the GMM estimator by replacing all the ingredients with Nagar type bias corrected instrumental variable estimators. To our knowledge, the idea of applying a minimum distance estimator to bias corrected instrumental variables estimators is new in the literature.

We consider a second order approach to the bias of the GMM estimator using the formula contained in Hahn and Hausman (2000). We find that the standard GMM estimator suffers from significant bias. The bias arises from two primary sources: the correlation of the structural equation error with the reduced form error and the low explanatory power of the instruments. 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 increases the explanatory power of the instruments which further reduces the finite sample bias and also decreases the MSE of the estimator. To increase further the explanatory power of the instruments, we use the technique of using estimated residuals as additional instruments a technique introduced in the simultaneous equations model by Hausman, Newey, and Taylor (1987) and used in the dynamic panel data context by Ahn and Schmidt (1995). 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.

However, the second order bias calculations do not predict well the performance of the estimator for these high values of the coefficient. Simulation evidence shows that our approximations do not work well near the unit circle where the model suffers from a near non-identification problem. In order to analyze the bias of standard GMM procedures under these circumstances we consider a local to non-identification asymptotic approximation.

The alternative asymptotic approximation of Staiger and Stock (1997) and Stock and Wright (2000) is based on letting the correlation between instruments and regressors decrease at a prescribed rate of the sample size. In their work, it is assumed that the number of instruments is held fixed as the sample size increases. Their limit distribution is nonstandard and in special cases corresponds to exact small sample distributions such as the one obtained by Richardson (1968) for the bivariate simultaneous equations model. This approach is related to the work by Phillips (1989) and Choi and Phillips (1992) on the asymptotics of 2SLS in the partially identified case. Dufour (1997), Wang and Zivot (1998) and Nelson, Startz and Zivot (1998) analyze valid inference and tests in the presence of weak instruments. The associated bias and mean squared error of 2SLS under weak instrument assumptions was obtained by Chao and Swanson (2000).

In this paper we use the weak instrument asymptotic approximations to analyze 2SLS for the dynamic panel model. We analyze the impact of stationarity assumptions on the nonstandard limit distribution. Here we let the autoregressive parameter tend to unity in a similar way as in the near unit root literature. Nevertheless we are not considering time series cases since in our approximation the number of time periods T is held constant while 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 inference optimally 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. We are using the weak instrument asymptotic approximation to the distribution of the IV estimator to derive the form of the optimal linear combination.

2 Review of the Bias of GMM Estimator

Consider the usual dynamic panel model with fixed effects:

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

It has been common in the literature to consider the case where n is large and T is small. The usual GMM estimator 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

$$E \left[y_{i,s} \left(\varepsilon_{it} - \varepsilon_{i,t-1} \right) \right] = 0 \qquad s = 0, \ldots, t-2.$$

Instead, 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.[1] 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

$$E\left[z_{it}\varepsilon_{it}^*\right] = 0 \qquad t = 1, \ldots, T-1$$

It can be shown that, with the homoscedasticity 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

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

where $x_t^* \equiv (x_{1t}^*, \cdots, x_{nt}^*)'$, $y_t^* \equiv (y_{1t}^*, \cdots, y_{nt}^*)'$, $Z_t \equiv (z_{1t}, \cdots, z_{nt})'$, and $P_t \equiv Z_t \left(Z_t'Z_t\right)^{-1} Z_t'$. Now, let $\widehat{b}_{2SLS,t}$ denote the 2SLS of y_t^* on x_t^*:

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

If ε_{it} are i.i.d. across t, then under the standard (first order) asymptotics where T is fixed and n grows to infinity, it can be shown that

$$\sqrt{n}\left(\widehat{b}_{2SLS,1} - \beta, \ldots, \widehat{b}_{2SLS,T-1} - \beta\right)' \to \mathcal{N}\left(0, \Psi\right),$$

where Ψ is a diagonal matrix with the t-th diagonal elements equal to $\operatorname{Var}(\varepsilon_{it}) / \left(\operatorname{plim} n^{-1} x_t^{*\prime} P_t x_t^*\right)$. Therefore, we may consider a minimum distance estimator, which solves

$$\min_b \begin{pmatrix} \widehat{b}_{2SLS,1} - b \\ \vdots \\ \widehat{b}_{2SLS,T-1} - b \end{pmatrix}' \begin{bmatrix} \left(x_1^{*\prime} P_1 x_1^*\right)^{-1} & & 0 \\ & \ddots & \\ 0 & & \left(x_{T-1}^{*\prime} P_{T-1} x_{T-1}^*\right)^{-1} \end{bmatrix}^{-1} \begin{pmatrix} \widehat{b}_{2SLS,1} - b \\ \vdots \\ \widehat{b}_{2SLS,T-1} - b \end{pmatrix}$$

The resultant minimum distance estimator is numerically identical to the GMM estimator in (2):

$$\widehat{b}_{GMM} = \frac{\sum_{t=1}^{T-1} x_t^{*\prime} P_t x_{it}^* \cdot \widehat{b}_{2SLS,t}}{\sum_{t=1}^{T-1} x_t^{*\prime} P_t x_t^*} \tag{3}$$

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 (1998) 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.

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

3 Bias Correction using Alternative Asymptotics

In this section, we consider the usual dynamic panel model with fixed effects (1) using the alternative asymptotics where n and T grow to infinity at the same rate. Such approximation was originally developed by Bekker (1994), and was adopted by Alvarez and Arellano (1998) and Hahn and Kuersteiner (2000) in the dynamic panel context. We assume

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

We also assume stationarity on $y_{i,0}$ and normality on α_i[2]:

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

In order to guarantee that $Z_t'Z_t$ is nonsingular, we will assume that

Condition 3 *$\frac{T}{n} \to \rho$, where $0 < \rho < 1$.*[3]

Alvarez and Arellano (1998) show that, under Conditions 1 - 3,

$$\sqrt{nT}\left(\widehat{b}_{GMM} - \left(\beta - \frac{1}{n}\left(1+\beta\right)\right)\right) \to \mathcal{N}\left(0, 1-\beta^2\right), \tag{4}$$

where $\widehat{b}_{GMM}$ is defined in (2) and (3). By examining the asymptotic distribution (4) under such alternative asymptotic approximation where n and T grow to infinity at the same rate, we can develop a bias-corrected estimator. This bias-corrected estimator is given by

$$\widetilde{b}_{GMM} \equiv \frac{n}{n-1}\widehat{b}_{GMM} + \frac{1}{n-1}. \tag{5}$$

Combining (4) and (5), we can easily obtain:

Theorem 1 *Suppose that Conditions 1 - 3 are satisfied. Then, $\sqrt{nT}\left(\widetilde{b}_{GMM} - \beta\right) \to \mathcal{N}\left(0, 1-\beta^2\right)$.*

Hahn and Kuersteiner (2000) establish by a Hajék-type convolution theorem that $\mathcal{N}\left(0, 1-\beta^2\right)$ is the minimal asymptotic distribution. As such, the bias corrected GMM is efficient. Although the bias corrected GMM estimator $\widetilde{b}_{GMM}$ does have a desirable property under the alternative asymptotics, it would not be easy to generalize the development leading to (5) to the model involving other strictly exogenous variables. Such a generalization would require the characterization of the asymptotic distribution of the standard GMM estimator under the alternative asymptotics, which may not be trivial. We therefore consider eliminating biases in $\widehat{b}_{2SLS,t}$ instead. An estimator that removes the higher order bias of $\widehat{b}_{2SLS,t}$ is the Nagar type estimator. Let

$$\widehat{b}_{Nagar,t} = \frac{x_t^{*\prime}P_t y_t^* - \lambda_t x_t^{*\prime}M_t y_t^*}{x_t^{*\prime}P_t x_t^* - \lambda_t x_t^{*\prime}M_t x_t^*},$$

where $M_t \equiv I - P_t$, $\lambda_t \approx \frac{K_t}{n-K_t}$, and K_t denotes the number of instruments for the t-th equation. For example, we may use $\lambda_t = \frac{K_t-2}{n-K_t+2}$ as in Donald and Newey (1998). We may also use LIML for the t-th equation, in which case λ_t would be estimated by the usual minimum eigenvalue search.

[2] This condition allows us to use lots of intermediate results in Alvarez and Arellano (1998). Our results are expected to be robust to violation of this condition.

[3] Alvarez and Arellano (1998) only require $0 \leq \rho < \infty$. We require $\rho < 1$ to guarantee that $Z_t'Z_t$ is singular for every t.

We now examine properties of the corresponding minimum distance estimator. One possible weight matrix for this problem is given by

$$\begin{bmatrix} \left(x_1^{*\prime} P_1 x_1^* - \lambda_1 x_1^{*\prime} M_1 x_1^*\right)^{-1} & & 0 \\ & \ddots & \\ 0 & & \left(x_{T-1}^{*\prime} P_{T-1} x_{T-1}^* - \lambda_{T-1} x_{T-1}^{*\prime} M_{T-1} x_{T-1}^*\right)^{-1} \end{bmatrix}^{-1}$$

With this weight matrix, it can be shown that the minimum distance estimator is given by

$$\widetilde{b}_{Nagar} \equiv \frac{\sum_t \left(x_t^{*\prime} P_t y_t^* - \lambda_t x_t^{*\prime} M_t y_t^*\right)}{\sum_t \left(x_t^{*\prime} P_t x_t^* - \lambda_t x_t^{*\prime} M_t x_t^*\right)} = \beta + \frac{\sum_t \left(x_t^{*\prime} P_t \varepsilon_t^* - \lambda_t x_t^{*\prime} M_t \varepsilon_t^*\right)}{\sum_t \left(x_t^{*\prime} P_t x_t^* - \lambda_t x_t^{*\prime} M_t x_t^*\right)}. \tag{6}$$

One possible way to examine the finite sample property of the new estimator is to use the alternative asymptotics:

Theorem 2 *Suppose that Conditions 1-3 are satisfied. Also suppose that n and T grow to infinity at the same rate. Then,* $\sqrt{nT}\left(b_{Nagar} - \beta\right) \rightarrow N\left(0, 1 - \beta^2\right)$.

Proof. Lemmas 10, and 11 in Appendix A along with Lemma 2 of Alvarez and Arellano (1998) establish that

$$\frac{1}{\sqrt{nT}} \sum_t \left(x_t^{*\prime} P_t \varepsilon_t^* - \frac{K_t}{n - K_t} x_t^{*\prime} M_t \varepsilon_t^*\right) \rightarrow N\left(0, \frac{\sigma^4}{1 - \beta^2}\right)$$

and

$$\frac{1}{nT} \sum_t \left(x_t^{*\prime} P_t x_t^* - \frac{K_t}{n - K_t} x_t^{*\prime} M_t x_t^*\right) \rightarrow \frac{\sigma^2}{1 - \beta^2},$$

from which the conclusion follows. ■

In Table 1, we summarized finite sample properties of $\widetilde{b}_{Nagar}$ and $\widehat{b}_{LIML}$ approximated by 10000 Monte Carlo runs.[4] Here, $\widehat{b}_{LIML}$ is the estimator where λs in (6) are replaced by the corresponding "eigenvalues". In general, we find that $\widetilde{b}_{Nagar}$ and $\widehat{b}_{LIML}$ successfully remove bias accurate unless β is close to one *and* the sample size is small.

4 Bias Correction using Higher Order Expansions

In the previous section, we explained the bias of the GMM estimator as a result of the biases of the 2SLS estimators. In this section, we consider elimination of the bias by adopting the second order Taylor type approximation. Such perspective has been adopted by Nagar (1959), and Rothenberg (1983).

For this purpose we first analyze the second order bias of a general minimization estimator b of a single parameter $\beta \in \mathbb{R}$ defined by

$$b = \underset{c \in C}{\operatorname{argmin}}\, Q_n(c) \tag{7}$$

for some $C \subset \mathbb{R}$. The score for the minimization problem is denoted by $S_n(c) \equiv \partial Q_n(c) / \partial c$. The criterion function is assumed to be of the form $Q_n(c) = g(c)' G(c)^{-1} g(c)$ where $g(c)$ and $G(c)$ are defined in Condition 7. The criterion function depends on primitive functions $\delta(w_i, c)$ and $\psi(w_i, c)$ mapping $\mathbb{R}^k \times \mathbb{R}$ into $\mathbb{R}^d$ for $d \geq 1$, where w_i are i.i.d. observations. We assume that $E[\delta(w_i, \beta)] = 0$. We impose the following additional conditions on w_i, δ, and ψ.

[4] In our Monte Carlo experiment, we let $\varepsilon_{it} \sim N(0,1)$, $\alpha_i \sim N(0,1)$, and $y_{i0} \sim N\left(\frac{\alpha_i}{1-\beta}, \frac{1}{1-\beta^2}\right)$.

Condition 4 *The random variables w_i are i.i.d.*

Condition 5 *The functions $\delta(w,c)$ and $\psi(w,c)$ are three times differentiable in c for $c \in C$ where $C \subset \mathbb{R}$ is a compact set such that $\beta \in \operatorname{int} C$. Assume that $\delta(w_i,c)$ and $\psi(w_i,c)$ satisfy a Lipschitz condition $\|\delta(w_i,c_1) - \delta(w_i,c_2)\| \leq M_\delta(w_i)|c_1 - c_2|$ for some function $M_\delta(.) : \mathbb{R}^k \to \mathbb{R}$ and $c_1, c_2 \in C$ with the same statement holding for ψ. The functions $M_\delta(.)$ and $M_\psi(.)$ satisfy $E[M_\delta(w_i)] < \infty$ and $E\left[|M_\psi(w_i)|^2\right] < \infty$.*

Condition 6 *Let $\delta_j(w_i,c) \equiv \partial^j \delta(w_i,c)/\partial c^j$, $\Psi(w_i,c) \equiv \psi(w_i,c)\psi(w_i,c)'$ and $\Psi_j(w_i,c) \equiv \partial^j \Psi(w_i,c)/\partial c^j$. Then, $\lambda_j(c) \equiv E[\delta_j(w_i,c)]$, and $\Lambda_j(c) \equiv E[\Psi_j(w_i,c)]$ all exist and are finite for $j = 0, ..., 3$. For simplicity, we use the notation $\lambda_j \equiv \lambda_j(\beta)$, $\Lambda_j \equiv \Lambda_j(\beta)$, $\lambda(c) \equiv \lambda_0(c)$ and $\Lambda(c) \equiv \Lambda_0(c)$.*

Condition 7 *Let $g(c) \equiv \frac{1}{n}\sum_{i=1}^n \delta(w_i,c)$, $g_j(c) \equiv \frac{1}{n}\sum_{i=1}^n \delta_j(w_i,c)$, $G(c) \equiv \frac{1}{n}\sum_{i=1}^n \psi(w_i,c)\psi(w_i,c)'$ and $G_j(c) \equiv \frac{1}{n}\sum_{i=1}^n \Psi_j(w_i,c)$. Then $g(c) \xrightarrow{p} E[\delta(w_i,c)]$, $g_j(c) \to E[\delta_j(w_i,c)]$, $G(c) \xrightarrow{p} E[\Psi(w_i,c)]$, and $G_j(c) \xrightarrow{p} E[\Psi_j(w_i,c)]$ for all $c \in C$.*

Our asymptotic approximation of the second order bias of b is based on an approximate estimator $\tilde{b}$ such that $b - \tilde{b} = o_p(n^{-1})$. The approximate bias of b is then defined as $E\left[\tilde{b}\right] - \beta$ while the original estimator b need not necessarily possess moments of any order. In order to justify our approximation we need to establish that b is $\sqrt{n}$-consistent and that $S_n(b) = 0$ with probability tending to one. For this purpose we introduce the following additional conditions.

Condition 8 *(i) There exists some finite $0 < \mathrm{M} < \infty$ such that the eigenvalues of $E[\Psi(w_i,c)]$ are contained in the compact interval $[\mathrm{M}^{-1}, \mathrm{M}]$ for all $c \in C$; (ii) the vector $E[\delta(w_i,c)] = 0$ if and only if $c = \beta$; (iii) $\lambda_1 \neq 0$.*

Condition 9 *There exists some $\eta > 0$ such that $E\left[M_\delta(w_i)^{2+\eta}\right] < \infty$, $E\left[M_\psi(w_i)^{2+\eta}\right] < \infty$, $E\left[\sup_{c\in C}\|\delta(w_i,c)\|^{2+\eta}\right] < \infty$, and $E\left[\sup_{c\in C}\|\psi(w_i,c)\|^{4+\eta}\right] < \infty$.*

Condition 8 is an identification condition that guarantees the existence of a unique interior minimum of the limiting criterion function. Condition 9 corresponds to Assumption B of Andrews (1994) and is used to establish a stochastic equicontinuity property of the criterion function.

Lemma 1 *Under conditions 4 - 9, b defined in (7) satisfies $\sqrt{n}(b - \beta) = O_p(1)$ and $S_n(b) = 0$ with probability tending to 1.*

Proof. See Appendix B. ■

Based on Lemma 1 the first order condition for (7) can be characterized by

$$0 = 2g_1(b)' G(b)^{-1} g(b) - g(b)' G(b)^{-1} G_1(b) G(b)^{-1} g(b) \text{ wp } \to 1. \tag{8}$$

A second order Taylor expansion of (8) around β leads to a representation of $b - \beta$ up to terms of order $o_p(n^{-1})$. In Appendix B, it is shown that

$$\sqrt{n}(b - \beta) = -\frac{1}{\Upsilon}\Phi + \frac{1}{\sqrt{n}}\left(-\frac{1}{\Upsilon}\Gamma + \frac{1}{\Upsilon^2}\Phi\Xi - \frac{\Psi}{\Upsilon^3}\Phi^2\right) + o_p\left(\frac{1}{\sqrt{n}}\right). \tag{9}$$

(See Definition 2 in Appendix B for definition of $\Psi, \Upsilon, \Phi, \Xi$, and Γ.) Ignoring the $o_p\left(\frac{1}{\sqrt{n}}\right)$ term in (9), and taking expectations, we obtain the "approximate mean" of $\sqrt{n}(b - \beta)$. We present the second order bias of b in the next Theorem.

Theorem 3 *Under Conditions 4-9, the second order bias of b is equal to*

$$-\frac{1}{\sqrt{n}}\frac{1}{\Upsilon}E\left[\Phi\right]-\frac{1}{n}\frac{1}{\Upsilon}E\left[\Gamma\right]+\frac{1}{n}\frac{1}{\Upsilon^{2}}E\left[\Phi\Xi\right]-\frac{\Psi}{n\Upsilon^{3}}E\left[\Phi^{2}\right]. \quad (10)$$

where

$$E\left[\Phi\right]=0,$$

$$E\left[\Gamma\right]=2\,\mathrm{trace}\left(\Lambda^{-1}E\left[\delta_{i}\frac{\partial\delta_{i}'}{\partial\beta}\right]\right)-2\lambda_{1}'\Lambda^{-1}E\left[\psi_{i}\psi_{i}'\Lambda^{-1}\delta_{i}\right]-\mathrm{trace}\left(\Lambda^{-1}\Lambda_{1}\Lambda^{-1}E\left[\delta_{i}\delta_{i}'\right]\right),$$

and

$$\begin{aligned}E\left[\Phi\Xi\right] \;=\; & 8\lambda_{1}'\Lambda^{-1}E\left[\delta_{i}\frac{\partial\delta_{i}'}{\partial\beta}\right]\Lambda^{-1}\lambda_{1}-4\lambda_{1}'\Lambda^{-1}E\left[\delta_{i}\lambda_{1}'\Lambda^{-1}\psi_{i}\psi_{i}'\right]\Lambda^{-1}\lambda_{1}\\ & -8\lambda_{1}'\Lambda^{-1}E\left[\delta_{i}\delta_{i}'\right]\Lambda^{-1}\Lambda_{1}\Lambda^{-1}\lambda_{1}+4\lambda_{1}'\Lambda^{-1}E\left[\delta_{i}\delta_{i}'\right]\Lambda^{-1}\lambda_{2},\end{aligned}$$

and

$$E\left[\Phi^{2}\right]=4\lambda_{1}'\Lambda^{-1}E\left[\delta_{i}\delta_{i}'\right]\Lambda^{-1}\lambda_{1},$$

Proof. See Appendix B. ■

Remark 1 *For the particular case where $\psi_i = \delta_i$, i.e. when b is a CUE, the bias formula (10) exactly coincides with Newey and Smith's (2000).*

We now apply these general results to the GMM estimator of the dynamic panel model. The GMM estimator $\widehat{b}_{GMM}$ can be understood to be a solution to the minimization problem

$$\min_{c}\left(\frac{1}{n}\sum_{i=1}^{n}m_{i}\left(c\right)\right)'V_{n}^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}m_{i}\left(c\right)\right)$$

for $c \in C$ where C is some closed interval on the real line containing the true parameter value and

$$m_{i}\left(c\right)=\begin{pmatrix}z_{i1}\left(y_{i1}^{*}-c\cdot x_{i1}^{*}\right)\\ \vdots\\ z_{i,T-1}\left(y_{i,T-1}^{*}-c\cdot x_{i,T-1}^{*}\right)\end{pmatrix},\qquad V_{n}=\frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix}z_{i1}z_{i1}' & & 0\\ & \ddots & \\ 0 & & z_{i,T-1}z_{i,T-1}'\end{bmatrix}.$$

We now characterize the finite sample bias of the GMM estimator $\widehat{b}_{GMM}$ of the dynamic panel model using Theorem 4. It can be shown that:

Theorem 4 *Under Conditions 1-3 the second order bias of $\widehat{b}_{GMM}$ is equal to*

$$\frac{B_{1}+B_{2}+B_{3}}{n}+o\left(\frac{1}{n}\right), \quad (11)$$

where

$$\begin{aligned}B_{1} &\equiv \Upsilon_{1}^{-1}\textstyle\sum_{t=1}^{T-1}\mathrm{trace}\left(\left(\Gamma_{t}^{zz}\right)^{-1}\Gamma_{t,t}^{\varepsilon xzz}\right)\\ B_{2} &\equiv -2\Upsilon_{1}^{-2}\textstyle\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}\Gamma_{t}^{zx\prime}\left(\Gamma_{t}^{zz}\right)^{-1}\Gamma_{t,s}^{\varepsilon xzz}\left(\Gamma_{s}^{zz}\right)^{-1}\Gamma_{s}^{zx}\\ B_{3} &\equiv \Upsilon_{1}^{-2}\textstyle\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}\Gamma_{t}^{zx\prime}\left(\Gamma_{t}^{zz}\right)^{-1}B_{3,1}(t,s)\left(\Gamma_{s}^{zz}\right)^{-1}\Gamma_{s}^{zx}.\end{aligned}$$

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

Proof. See Appendix C. ■

In Table 2, we compare the actual performance of $\widehat{b}_{GMM}$ and the prediction of its bias based on Theorem 4. Table 2 tabulates the actual bias of the estimator approximated by 10000 Monte Carlo runs, and compares it with the second order bias based on the formula (11). 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 4 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

$$\widehat{b}_{BC1} \equiv \widehat{b}_{GMM} - \frac{1}{n}\left(\widehat{B}_1 + \widehat{B}_2 + \widehat{B}_3\right) \tag{12}$$

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.

Then the $\widehat{B}$s will satisfy the $\sqrt{n}$-consistency requirement, and hence, the estimator (12) will be first order equivalent to $\widehat{b}_{GMM}$ and will have zero second order bias. Because the summand

$$E\left[E\left[z_{it}x_{it}^*\right]' E\left[z_{it}z_{it}'\right]^{-1} \varepsilon_{it}^* z_{it}\lambda_s'\Lambda_s^{-1} z_{is}z_{is}' E\left[z_{is}z_{is}'\right]^{-1} E\left[z_{is}x_{is}^*\right]\right]$$

in the numerator of B_3 is equal to zero for $s < t$, we may instead consider

$$\widehat{b}_{BC2} \equiv \widehat{b}_{GMM} - \frac{1}{n}\left(\widehat{B}_1 + \widehat{B}_2 + \widehat{\widehat{B}}_3\right) \tag{13}$$

where

$$\frac{1}{n}\widehat{\widehat{B}}_3 \equiv \hat{\Upsilon}_1^{-2}\sum_{s=t}^{T-1}\sum_{t=1}^{T-1}\hat{\Gamma}_t^{zx\prime}\left(\hat{\Gamma}_t^{zz}\right)^{-1}\hat{B}_{3,1}(t,s)\left(\hat{\Gamma}_s^{zz}\right)^{-1}\hat{\Gamma}_s^{zx}.$$

Second order asymptotic theory predicts approximately that $\widehat{b}_{BC2}$ would be relatively free of bias. We examined whether such prediction is reasonably accurate in finite sample by 5000 Monte Carlo runs.[5] Table 3 summarizes the properties of $\widehat{b}_{BC2}$. We have seen in Table 2 that the second order theory is reasonably accurate unless β is close to one. It is therefore sensible to conjecture that $\widehat{b}_{BC2}$ would have a reasonable finite sample bias property as long as β is not too close to one. Such a conjecture is verified in Table 3.

5 Long Difference Specification: Finite Iteration

In previous sections, we noted that even 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}. Such stationarity condition may or may not be appropriate for particular applications. Further, stationarity assumption turns out to be a predominant source of information around $\beta \approx 1$ as noted by Hahn (1999). We therefore turn to some other method to overcome the weak instrument problem around the unit circle avoiding the stationarity assumption.

[5] The difference of Monte Carlo runs here induced some minor numerical difference (in properties of b_{GMM}) across Tables 1 - 3.

We argue that some of the difficulties of inference around the unit circle would be alleviated by taking a long difference. To be specific, we focus on a single equation based on the long difference

$$y_{iT} - y_{i1} = \beta \left(y_{iT-1} - y_{i0} \right) + \left(\varepsilon_{iT} - \varepsilon_{i1} \right) \tag{14}$$

It is easy to see that the initial observation y_{i0} would serve as a valid instrument. Using intuition as in Hausman and Taylor (1983) or Ahn and Schmidt (1995), we can see that $y_{iT-1} - \beta y_{iT-2}, \ldots, y_{i2} - \beta y_{i1}$ would be valid instruments as well.

5.1 Intuition

In Hahn-Hausman (HH) (1999) we found that the bias of 2SLS (GMM) 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:

$$\frac{1}{n} \frac{\text{(number of instruments)} \times (\text{"covariance"})}{\text{"Explained" variance of the first stage reduced form equation}}$$

Similarly, the Donald-Newey (DN) (1999) MSE formula depends on the same 4 factors. We now consider first differences (FD) and long differences (LD) to see why LD does so much better in our Monte-Carlo experiments.

Assume that $T = 4$. The first difference set up is:

$$y_4 - y_3 = \beta \left(y_3 - y_2 \right) + \varepsilon_4 - \varepsilon_3 \tag{15}$$

For the RHS variables it uses the instrument equation:

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

Now calculate the R2 for equation (15) using Ahn-Schmidt (AS) moments under "ideal conditions" where you know β in the sense that the nonlinear restrictions become linear restrictions: We would then use $(y_2, y_1, y_0, \alpha + \varepsilon_1, \alpha + \varepsilon_2)$ as instruments. Assuming stationarity for symbols, but not using it as additional moment information, we can write

$$y_0 = \frac{\alpha}{1 - \beta} + \xi_0,$$

where $\xi_0 \sim \left(0, \frac{\sigma_\varepsilon^2}{1 - \beta^2} \right)$. It can be shown that the covariance between the structure error and the first stage error is $-\sigma_\varepsilon^2$, and the "explained variance" in the first stage is equal to $\sigma_\varepsilon^2 \frac{1-\beta}{\beta+1}$. Therefore,the ratio that determines the bias of 2SLS is equal to

$$\frac{-\sigma_\varepsilon^2}{\sigma_\varepsilon^2 \frac{-\beta+1}{\beta+1}} = -\frac{1+\beta}{1-\beta},$$

which is equal to -19 for $\beta = .9$. For $n = 100$, this implies the percentage bias of

$$\frac{\text{Number of Instruments}}{\text{Sample Size}} \frac{-19}{\beta} \times 100 = \frac{5}{100} \frac{-19}{0.9} \times 100 = -105.56$$

We now turn to the LD setup:

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

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

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

which is equal to

$$-.374\,08 + \frac{2.570\,3 \times 10^{-4}}{\sigma^2 + 4.830\,6 \times 10^{-2}}$$

for $\beta = .9$. Note that the maximum value that this ratio can take in absolute terms is

$$-.374\,08$$

which is much smaller than -19. We therefore conclude that the long difference increases R^2 but decreases the covariance. Further, the number of instruments is smaller in the long difference specification so we should expect even smaller bias. Thus, all of the factors in the HH equation, except sample size, cause the LD estimator to have smaller bias.

5.2 Monte Carlo

For the long difference specification, we can use y_{i0} as well as the "residuals" $y_{iT-1} - \beta y_{iT-2}, \ldots, y_{i2} - \beta y_{i1}$ as valid instruments.[6] We may estimate β by applying 2SLS to the long difference equation (14) using y_{i0} as instrument. We may then use $\left(y_{i0}, y_{iT-1} - \widehat{b}_{2SLS} y_{iT-2}, \cdots, y_{i2} - \widehat{b}_{2SLS} y_{i1}\right)$ as instrument to the long difference equation (14) to estimate β. Call the estimator $\widehat{b}_{2SLS,1}$. By iterating this procedure, we can define $\widehat{b}_{2SLS,2}$, $\widehat{b}_{2SLS,3}$, ... Similarly, we may first estimate β by Arellano and Bover, and use $\left(y_{i0}, y_{iT-1} - \widehat{b}_{GMM} y_{iT-2}, \cdots, y_{i2} - \widehat{b}_{GMM} y_{i1}\right)$ as instrument to the long difference equation (14) to estimate β. Call the estimator $\widehat{b}_{2SLS,1}$. By iterating this procedure, we can define $\widehat{b}_{GMM,2}$, $\widehat{b}_{GMM,3}$, ... Likewise, we may first estimate β by $\widehat{b}_{LIML}$, and use $\left(y_{i0}, y_{iT-1} - \widehat{b}_{LIML} y_{iT-2}, \cdots, y_{i2} - \widehat{b}_{LIML} y_{i1}\right)$ as instrument to the long difference equation (14) to estimate β. Call the estimator $\widehat{b}_{LIML,1}$. By iterating this procedure, we can define $\widehat{b}_{LIML,2}$, $\widehat{b}_{LIML,3}$, ... We found that such iteration of the long difference estimator works quite well. We implemented these procedures for $T = 5$, $n = 100$, $\beta = 0.9$ and $\sigma_\alpha^2 = \sigma_\varepsilon^2 = 1$. Our finding with 5000 monte carlo runs is summarized in Table 4. In general, we found that the iteration of the long difference estimator works quite well.[7]

We compared performances of our estimator with Blundell and Bond's (1998) estimator, which uses additional information, i.e., stationarity. We compared four versions of their estimators $\widehat{b}_{BB1}, \cdots, \widehat{b}_{BB4}$ with the long difference estimators $\widehat{b}_{LIML,1}, \widehat{b}_{LIML,2}, \widehat{b}_{LIML,3}$. For exact definition of $\widehat{b}_{BB1}, \ldots, \widehat{b}_{BB4}$, see Appendix E. Of the four versions, $\widehat{b}_{BB3}$ and $\widehat{b}_{BB4}$ are the ones reported in their Monte Carlo section. In our Monte Carlo exercise, we set $\beta = 0.9$, $\sigma_\varepsilon^2 = 1$, $\alpha_i \sim N(0,1)$. Our finding based on 5000 Monte

[6] We acknowledge that the residual instruments are irrelevant under the near unity asymptotics.

[7] Second order theory does not seem to explain the behavior of long difference estimator. In Table 7, we compare the actual performance of the long difference based estimators with the second order theory developed in Appendix F.

Carlo runs is contained in Table 5. In terms of bias, we find that Blundell and Bond's estimators $\widehat{b}_{BB3}$ and $\widehat{b}_{BB4}$ have similar properties as the long difference estimator(s), although the former dominates the latter in terms of variability. (We note, however, that $\widehat{b}_{BB1}$ and $\widehat{b}_{BB2}$ are seriously biased. This indicates that the choice of weight matrix matters in implementing Blundell and Bond's procedure.) This result is not surprising because the long difference estimator does not use the information contained in the initial condition. See Hahn (1999) for related discussion. We also wanted to examine the sensitivity of Blundell and Bond's estimator to misspecification, i.e., nonstationary distribution of y_{i0}. For this situation the estimator will be inconsistent. In order to assess the finite sample sensitivity, we considered the cases where $y_{i0} \sim \left(\frac{\alpha_i}{1-\beta_F}, \frac{\sigma_\varepsilon^2}{1-\beta_F^2}\right)$. Our Monte Carlo results based on 5000 runs are contained in Table 6, which contains results for $\beta_F = .5$ and $\beta_F = 0$. We find that the long difference estimator is quite robust, whereas $\widehat{b}_{BB3}$ and $\widehat{b}_{BB4}$ become quite biased as predicted by the first order theory. (We note that $\widehat{b}_{BB1}$ and $\widehat{b}_{BB2}$ are less sensitive to misspecification. Such robustness consideration suggests that choice of weight matrix is not straightforward in implementing Blundell and Bond's procedure.) We conclude that the long difference estimator works quite well even compared to Blundell and Bond's (1998) estimator.[8]

6 Near Unit Root Approximation

Our Monte Carlo simulation results summarized in Tables 1, 2, and 3 indicate that the previously discussed approximations and the bias corrections that are based on them do not work well near the unit circle. This is because the identification of the model becomes "weak" near the unit circle. See Blundell and Bond (1998), who related the problem to the analysis by Staiger and Stock (1997). In this Section, we formally adopt approximations local to the points in the parameter space that are not identified. To be specific, we consider model (1) for T fixed and $n \to \infty$ when also β_n tends to unity. 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) moment conditions and show that a strict subset of the full set of moment restrictions should be used in estimation in order to minimize bias. We argue that this subset of moment restrictions leads to the inference based on the "long difference" specification.

Following Ahn and Schmidt we exploit the moment conditions

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

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

$$b = \begin{bmatrix} \operatorname{vech} E\left[u_i u_i'\right] \\ E\left[u_i y_{i0}\right] \end{bmatrix} = \sigma_\varepsilon^2 \begin{bmatrix} \operatorname{vech} I \\ 0 \end{bmatrix} + \sigma_\alpha^2 \begin{bmatrix} \operatorname{vech}(I + \mathbf{1}\mathbf{1}') \\ 0 \end{bmatrix} + \sigma_{\alpha y_0} \begin{bmatrix} 0 \\ \mathbf{1} \end{bmatrix} \tag{16}$$

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 (16) makes it clear that the vector $b \in \mathbb{R}^{T(T+1)/2+T}$ 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 b. This statement is equivalent to Ahn and Schmidt's (1997) analysis of the number of moment conditions.

[8] We did try to compare the sensitivity of the two moment restrictions by implementing CUE, but we experienced some numerical problem. Numerical problem seems to be an issue with CUE in general. Windmeijer (2000) report similar problems with CUE.

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 b. This matrix A satisfies

$$b'A = 0.$$

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

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

where $\Delta u_i = [u_{i2} - u_{i1}, ..., u_{iT} - u_{iT-1}]'$. It becomes transparent that any other representation of the moment conditions can be obtained by applying a corresponding nonsingular linear operator C to the matrix A. It can be checked that there exists a nonsingular matrix C such that $b'AC = 0$ 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. 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 $g_{i2}(\beta) \equiv [y_{i0}\Delta u_i(\beta)]$. 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

$$\beta_{2SLS} = \underset{\beta}{\operatorname{argmin}}\, g_n(\beta)' C\left(C'\Omega_n C\right)^+ C'g_n(\beta) \tag{17}$$

where C is a $G \times r$ matrix for $1 \le r \le G$ such that $C'C = I_r$ and $\operatorname{rank}\left(C\left(C'\Omega C\right)^+ C'\right) \ge 1$. We use $(C'\Omega C)^+$ 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

$$\beta_{2SLS} - \beta_n = \left(f_n' C\left(C'\Omega_n C\right)^+ C' f_n\right)^{-1} f_n' C\left(C'\Omega_n C\right)^+ C' g_n(\beta_n). \tag{18}$$

We are now analyzing the behavior of $\beta_{2SLS} - \beta_n$ under local to unity asymptotics. We make the following additional assumptions.[9]

Condition 10 *Let* $y_{it} = \alpha_i + \beta_n y_{it-1} + \varepsilon_{it}$ *with* $\varepsilon_{it} \sim N\left(0, \sigma_\varepsilon^2\right)$, $\alpha_i \sim N\left(0, \sigma_\alpha^2\right)$ *and* $y_{i0} \sim N\left(\frac{\alpha_i}{1-\beta_n}, \frac{1}{1-\beta_n^2}\right)$, *where* $\beta_n = \exp(-c/n)$ *for some* $c > 0$.

Also note that $\Delta y_{it} = \beta_n^{t-1}\eta_{i0} + \varepsilon_{it} + \frac{c}{\sqrt{n}}\sum_{s=1}^{t-1}\beta_n^{s-1}\varepsilon_{it-s} + o_p\left(n^{-1}\right)$ where $\eta_{i0} \sim N\left(0, (\beta_n - 1)^2 \big/ \left(1-\beta_n^2\right)\right)$. Under the generating mechanism described in the previous definition the following Lemma can be established.

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

Lemma 2 *Assume $\beta_n = \exp(-c/n)$ for some $c > 0$. 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

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

where $[\xi_x', \xi_y'] \sim N(0, \Sigma)$ *with* $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$ *and* $\Sigma_{11} = \delta I$, $\Sigma_{12} = \delta M_1$ $\Sigma_{22} = \delta M_2$, *where*

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

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

and $\Sigma_{12} = \Sigma_{21}'$. *We also have*

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

Proof. See Appendix D. ■

Using Lemma (2) the limiting distribution of $\beta_{2SLS} - \beta_n$ 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 $C = [C_0', C_1']'$ such that $C'\xi_x^{\#} = C_1'\xi_x$. Let r_1 denote the rank of C_1.

Corollary 1 *Let $\beta_{2SLS} - \beta_n$ be given by (18). If Condition 10 is satisfied then*

$$\beta_{2SLS} - 1 \xrightarrow{d} \frac{\xi_x' C_1 \left(C_1'\Sigma_{22}C_1\right)^{+} C_1'\xi_y}{\xi_x' C_1 \left(C_1'\Sigma_{22}C_1\right)^{+} C_1'\xi_x} = X\left(C, \Sigma_{22}\right) \tag{19}$$

Unlike the limiting distribution for the standard weak instrument problem, $X\left(C, \Sigma_{22}\right)$, as defined in (19), 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[10] for the process y_{it}. Inspection of the proof shows that the usual concentration parameter appearing in the

[10] An alternative way to parametrize the stationarity condition is to set $y_{it} = (1 - \beta_n)\alpha_i + \beta_n y_{it-1} + \varepsilon_{it}$, $\varepsilon_{it} \sim N\left(0, \sigma_\varepsilon^2\right)$, $\alpha_i \sim N\left(0, \sigma_\alpha^2\right)$ and $y_{i0} \sim N\left(\alpha_i, \frac{\sigma_y^2}{1-\beta_n^2}\right)$ with $\beta_n = \exp(-c/n)$. It can be shown that an estimator soley based on the moment condition $E\left[\overline{u}_i \Delta u_{ik}(\beta_0)\right] = 0$ is consistent. Restricting attention to estimators that are based on all moment conditions except the condition $E\left[\overline{u}_i \Delta u_{ik}(\beta_0)\right] = 0$, one can show that

$$\beta_{2SLS} - 1 \xrightarrow{d} \frac{(\mu + \xi_x)' C_1 (C_1'\Sigma_{22}C_1)^{-1} C_1'\xi_y}{(\mu + \xi_x)' C_1 (C_1'\Sigma_{22}C_1)^{-1} C_1' (\mu + \xi_x)} = X$$

where $\mu = 1_{T-1}\sigma_y^2/2$. Estimators in the class β_{2SLS}^* defined in Definition 1 can be analyzed by the same methods as under the stationarity assumptions discussed in the paper. Their limiting distributions are denoted by $X^*\left(C_1, \tilde{\Omega}\right)$. Moments of limit distribution is dominated by a stochastic component related to the fixed effect. This situation seems to be similar to time series models where deterministic trends can dominate the asymptotic distribution.

Based on Corollary 1, we define the following class of 2SLS estimators for the dynamic panel model. The class contains estimators that only use a nonredundant set of moment conditions involving the initial conditions y_{i0}.

Definition 1 *Let β^*_{2SLS} be defined as $\beta^*_{2SLS} \equiv \operatorname{argmin}_\beta g_{2,n}(\beta)' C_1 \left(C_1'\tilde{\Omega}C_1\right)^{-1} C_1' g_{2,n}(\beta)$, 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 $r_1 \le T-1$ such that $C_1'C_1 = I$.*

Next we turn to the analysis of the asymptotic bias for the estimator β^*_{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 5 *Let $X^*\left(C_1,\tilde{\Omega}\right)$ be the limiting distribution of $\beta^*_{2SLS} - 1$ in Definition 1 under Condition 10. Let $\bar{D} \equiv (D + D')/2$, where $D = \left(C_1'\tilde{\Omega}C_1\right)^{-1} C_1'M_1'C_1$. Then*

$$E\left[X\left(C_1,\tilde{\Omega}\right)\right] = \bar{\lambda}^{-1}\sum_{k=0}^{\infty}\frac{(1)_k\left(\frac{1}{2}\right)_{1+k}}{\left(\frac{r_1}{2}\right)_{1+k} k!} C^{1,k}_{1+k}\left(\bar{D}, I_{r_1} - \bar{\lambda}^{-1}\left(C_1'\tilde{\Omega}C_1\right)^{-1}\right)$$

where $r_1 = \operatorname{rank}(C_1)$, $(a)_b$ is the Pochammer 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 $\left(C_1'\tilde{\Omega}C_1\right)^{-1}$. The mean $E\left[X\left(C_1,\tilde{\Omega}\right)\right]$ exists for $r_1 \geq 1$.

Proof. See Appendix D. ∎

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

The problem of minimizing the bias 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. We show however that for two particular choices of $\tilde{\Omega}$ where $\tilde{\Omega} = I_{T-1}$ or $\tilde{\Omega} = \Sigma_{22}$ and subsequent minimization over C_1 an analytical solution for the bias minimal estimator can be found. It turns out that the optimum is the same for both weight matrices. The theorem also shows that the optimum can be reasonably well approximated by a procedure that is very easy to implement.

X^* are intractably complicated unless $\tilde{\Omega} = I_{T-1}$. Then the mean of $X^*(C_1, I_{T-1})$ is given by

$$\begin{aligned} E[X^*] &= B(\bar{D},\bar{\mu}) = e^{-\bar{\mu}'\bar{\mu}/2}\left(\frac{tr\bar{D}}{r}\,{}_1F_1(\frac{r}{2},\frac{r}{2}+1,\frac{\bar{\mu}'\bar{\mu}}{2}) + \frac{\bar{\mu}'\bar{D}\bar{\mu}}{r+2}\,{}_1F_1(\frac{r}{2}+1,\frac{r}{2}+2,\frac{\bar{\mu}'\bar{\mu}}{2})\right. \\ &\left. -\frac{\bar{\mu}'\bar{D}\bar{\mu}}{r+1}\,{}_1F_1(\frac{r-1}{2},\frac{r-1}{2}+1,\frac{\bar{\mu}'\bar{\mu}}{2}) - \frac{\bar{\mu}'\bar{D}\bar{\mu}}{r+1}\bar{\mu}'\bar{\mu}\,{}_1F_1(\frac{r-1}{2},\frac{r-1}{2}+1,\frac{\bar{\mu}'\bar{\mu}}{2})\right) \end{aligned}$$

with $\bar{D} = (D+D')/2$, $D = C_1'M_1C_1$ and $\bar{\mu} = C_1'\mu$. While finding an exact analytical minimum of the above bias expression as a function of the weight matrix is probably infeasible due to the complexity of the formula, one sees relatively easily that for large T the minimum must approximately be achieved for $C = 1_{T-1}$, thus leading approximately to the long difference estimator.

Theorem 6 *Let $X\left(C_1,\tilde{\Omega}\right)$ be as defined in Definition 1. Let $\bar{D}=(D+D')/2$ where $D=\left(C_1'\tilde{\Omega}C_1\right)^{-1}C_1'M_1'C_1$. Then*

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

Moreover,

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

Let $C_1^*=\operatorname{argmin}_{C_1}\left|E\left[X(C_1,I_{T-1})\right]\right|$ 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 $\bar{D}$. Thus, $\min_{C_1}\operatorname{tr}\bar{D}/r_1=\min l_i/2$. As $T\to\infty$ the smallest eigenvalue of $\bar{D}$, $\min l_i\to 0$. Let $\mathbf{1}=[1,...,1]'$ be a $T-1$ vector. Then for $C_1=\mathbf{1}/(\mathbf{1}'\mathbf{1})^{1/2}$ it follows that $\operatorname{tr}\bar{D}\to 0$ as $T\to\infty$.

Theorem 6 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 $\left(M_1+M_1'\right)/2$. This eigenvalue can be easily computed for any given T. The Theorem also shows that at least for large T a heuristic method which puts equal weight on all moment conditions leads to essentially the same bias reduction as the optimal procedure. The heuristic procedure turns out be equal to the moment condition $E\left[\left(u_{iT}-u_{i1}\right)y_{i0}\right]$ which can be motivated by taking "long differences" of the model equation $y_{it}=\alpha_i+\beta_n y_{it-1}+\varepsilon_{it}$ i.e. by considering

$$y_{iT}-y_{i1}=\alpha_i+\beta_n(y_{iT-1}-y_{i0})+\varepsilon_{iT}-\varepsilon_{i1}.$$

It can also be shown that a 2SLS estimator that uses all moment conditions involving y_{i0} remains biased even as $T\to\infty$.

7 Long Difference Specification: Infinite Iteration

We found that the iteration of the long difference estimator works quite well. In the $(\ell+1)$-th iteration, our iterated estimator estimates the model

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

based on 2SLS using instruments $z_i\left(\widehat{\beta}_{(\ell)}\right)\equiv\left(y_{i0},y_{i2}-\widehat{\beta}_{(\ell)}y_{i1},\dots,y_{iT-1}-\widehat{\beta}_{(\ell)}y_{iT-2}\right)$, where $\widehat{\beta}_{(\ell)}$ is the estimator obtained in the previous iteration. We might want to examine properties of an estimator based on an infinite iteration, and see if it improves bias property. If we continue the iteration and it converges[11], the estimator is a fixed point to the minimization problem

$$\min_b\left(\sum_{i=1}^N\xi_i(b)\right)'\left(\sum_{i=1}^N z_i(b)\,z_i(b)'\right)^{-1}\left(\sum_{i=1}^N\xi_i(b)\right)$$

[11] There is no a priori reason to believe that the iterations converge to the fixed point. To show that, one would have to prove that the iterations are a contraction mapping.

where $\xi_i(b) \equiv z_i(b)((y_{iT} - y_{i1}) - b(y_{iT-1} - y_{i0}))$. Call the minimizer the infinitely iterated 2SLS and denote it $\widehat{\beta}_{I2SLS}$. Another estimator which resembles $\widehat{\beta}_{I2SLS}$ is CUE, which solves

$$\widehat{\beta}_{CUE} \equiv \underset{b}{\operatorname{argmin}} L(b) = \underset{b}{\operatorname{argmin}} \left(\sum_{i=1}^{N} \xi_i(b)\right)' \left(\sum_{i=1}^{N} \xi_i(b)\xi_i(b)'\right)^{-1} \left(\sum_{i=1}^{N} \xi_i(b)\right).$$

Their actual performance approximated by 5000 Monte Carlo runs along with the biases predicted by second order theory in Theorem 4 are summarized in Tables 8 and 9. We find that the long difference based estimators have quite reasonable finite sample properties even when β is close to 1. Similar to the finite iteration in the previous section, the second order theory seem to be next to irrelevant for β close to 1.

We compared performances of our estimators with Ahn and Schmidt's (1995) estimator as well as Blundell and Bond's (1998) estimator. Both estimators are defined in two-step GMM procedures. In order to make a accurate comparison with our long difference strategy, for which there is no ambiguity of weight matrix, we decided to apply the continuous updating estimator to their moment restrictions. We had difficulty of finding global minimum for Ahn and Schmidt's (1995) moment restrictions. We therefore used Rothenberg type two step iteration, which would have the same second order property as the CUE itself. (See Appendix I.) Again, in order to make a accurate comparison, we applied the two step iteration idea to our long difference and Blundell and Bond (1998) as well. We call these estimators $\widehat{\beta}_{CUE2,AS}$, $\widehat{\beta}_{CUE2,LD}$, and $\widehat{\beta}_{CUE2,BB}$. We set $n = 100$ and $T = 5$. Again the number of monte carlo runs was set equal to 5000. Our results are reported in Tables 10 and 11. We can see that the long difference estimator has a comparable property to Ahn and Schmidt's estimator. We do not know why $\widehat{\beta}_{CUE2,LD}$ has such a large median bias at $\beta = .95$ whereas $\widehat{\beta}_{CUE,LD}$ does not have such problem.

8 Conclusion

We have investigated the bias of the dynamic panel effects estimators using second order approximations and Monte Carlo simulations. The second order approximations confirm the presence of significant bias as the parameter becomes large, as has previously been found in Monte Carlo investigations. Use of the second order asymptotics to define a second order unbiased estimator using the Nagar approach improve matters, but unfortunately does not solve the problem. Thus, we propose and investigate a new estimator, the long difference estimator of Griliches and Hausman (1986). We find in Monte Carlo experiments that this estimator works quite well, removing most of the bias even for quite high values of the parameter. Indeed, the long differences estimator does considerably better than "standard" second order asymptotics would predict. Thus, we consider alternative asymptotics with a near unit circle approximation. These asymptotics indicate that the previously proposed estimators for the dynamic fixed effects problem suffer from larger biases. The calculations also demonstrate that the long difference estimator should work in eliminating the finite sample bias previously found. Thus, the alternative asymptotics explain our Monte Carlo finding of the excellent performance of the long differences estimator.

Technical Appendix

A Technical Details for Section 3

Lemma 3

$$E\left[\sum_t \left(x_t^{*\prime} P_t \varepsilon_t^* - \frac{K_t}{n-K_t} x_t^{*\prime} M_t \varepsilon_t^*\right)\right] = 0.$$

Proof. We have

$$E\left[x_t^{*\prime} P_t \varepsilon_t^* - \frac{K_t}{n} x_t^{*\prime} M_t \varepsilon_t^*\right] = E\left[\operatorname{trace}\left(P_t E_t\left[\varepsilon_t^* x_t^{*\prime}\right]\right)\right] - \frac{K_t}{n-K_t} E\left[\operatorname{trace}\left(M_t E_t\left[\varepsilon_t^* x_t^{*\prime}\right]\right)\right],$$

where $E_t[\cdot]$ denotes the conditional expectation given Z_t. Because $E_t[\varepsilon_t^*] = 0$, $E_t[\varepsilon_t^* x_t^{*\prime}]$ is the conditional covariance between ε_t^* and $y_{t-1}^{*\prime}$, which does not depend on Z_t due to joint normality. Moreover, by cross-sectional independence, we have

$$E_t\left[\varepsilon_t^* x_t^{*\prime}\right] = E_t\left[\varepsilon_{i,t}^* x_{i,t}^*\right] I_n.$$

Hence, using the fact that $\operatorname{trace}(P_t) = K_t$ and $\operatorname{trace}(M_t) = n - K_t$, we have

$$E\left[x_t^{*\prime} P_t \varepsilon_t^* - \frac{K_t}{n} x_t^{*\prime} M_t \varepsilon_t^*\right] = E_t\left[\varepsilon_{i,t}^* x_{i,t}^*\right] \cdot \left(K_t - \frac{K_t}{n-K_t}(n-K_t)\right) = 0,$$

from which the conclusion follows. ■

Lemma 4

$$\operatorname{Var}\left(x_t^{*\prime} M_t \varepsilon_t^*\right) = (n-t)\,\sigma^2 E\left[v_{it}^{*2}\right] + (n-t)\left(E\left[v_{it}^* \varepsilon_{it}^*\right]\right)^2,$$
$$\operatorname{Cov}\left(x_t^{*\prime} M_t \varepsilon_t^*, x_s^{*\prime} M_s \varepsilon_s^*\right) = (n-s)\, E\left[v_{is}^* \varepsilon_{it}^*\right] E\left[v_{it}^* \varepsilon_{is}^*\right], \qquad s < t$$

where $v_{it}^* \equiv x_{it}^* - E\left[x_{it}^* \mid z_{it}\right]$.

Proof. Follows by modifying the developments from (A23) to (A30) and from (A31) to A(34) in Alvarez and Arellano (1998). ■

Lemma 5 *Suppose that* $s < t$. *We have*

$$E\left[v_{it}^{*2}\right] = \frac{T-t}{T-t+1}\left(\frac{1}{1-\beta} - \frac{\beta-\beta^{T-t+1}}{(T-t)(1-\beta)^2}\right)^2 \frac{\sigma_\alpha^2}{1+\frac{\sigma_\alpha^2}{\sigma^2}\frac{2}{1-\beta}+\frac{\sigma_\alpha^2}{\sigma^2}(t-2)}$$
$$-\sigma^2 \frac{T-t}{T-t+1}\frac{1}{(T-t)^2(1-\beta)^2}$$
$$\times\left((T-t) + \frac{\beta^2 - 2\beta^{T-t+2} + \beta^{2(T-t)+2} - 2\beta^{T-t+1} + 2\beta}{\beta^2-1}\right),$$

$$E\left[v_{it}^* \varepsilon_{it}^*\right] = -\sigma^2 \sqrt{\frac{T-t}{T-t+1}}\frac{\left(1-\beta^{T-t}\right)}{(T-t)(1-\beta)}$$
$$+\sigma^2 \sqrt{\frac{T-t}{T-t+1}}\frac{1}{(T-t)^2(1-\beta)}\left((T-t) - \frac{\beta-\beta^{T-t}}{1-\beta}\right),$$

$$E\left[v_{is}^{*}\varepsilon_{it}^{*}\right]=-\sigma^{2}\sqrt{\frac{T-s}{T-s+1}}\frac{\left(1-\beta^{T-t}\right)}{(T-s)(1-\beta)}$$
$$+\sigma^{2}\sqrt{\frac{T-s}{T-s+1}}\frac{1}{(T-s)(T-t)(1-\beta)}\left((T-t)-\frac{1-\beta^{T-t}}{1-\beta}\right),$$

$$E\left[v_{it}^{*}\varepsilon_{is}^{*}\right]=\sigma^{2}\sqrt{\frac{T-t}{T-t+1}}\frac{1}{(T-s)(T-t)(1-\beta)}\left(T-t-\frac{\beta-\beta^{T-t+1}}{1-\beta}\right).$$

Proof. We first characterize v_{it}^{*}. We have

$$x_{i,t}=y_{i,t-1}$$
$$x_{i,t+1}=y_{i,t}=\alpha_i+\beta y_{i,t-1}+\varepsilon_{i,t}$$
$$\vdots$$
$$x_{i,T}=y_{i,T-1}=\frac{1-\beta^{T-t}}{1-\beta}\alpha_i+\beta^{T-t}y_{i,t-1}+\left(\varepsilon_{i,T-1}+\beta\varepsilon_{i,T-2}+\cdots+\beta^{T-t-1}\varepsilon_{i,t}\right)$$

and hence

$$\sqrt{\frac{T-t+1}{T-t}}x_{it}^{*}=x_{it}-\frac{1}{T-t}\left(x_{it+1}+\cdots+x_{iT}\right)$$
$$=y_{i,t-1}-\frac{1}{T-t}\left(x_{it+1}+\cdots+x_{iT}\right)$$
$$=\left(1-\frac{\beta-\beta^{T-t+1}}{(T-t)(1-\beta)}\right)y_{i,t-1}-\left(\frac{1}{1-\beta}-\frac{\beta-\beta^{T-t+1}}{(T-t)(1-\beta)^{2}}\right)\alpha_i$$
$$-\frac{(1-\beta)\varepsilon_{i,T-1}+\left(1-\beta^{2}\right)\varepsilon_{i,T-2}+\cdots+\left(1-\beta^{T-t}\right)\varepsilon_{i,t}}{(T-t)(1-\beta)}.$$

It follows that

$$\sqrt{\frac{T-t+1}{T-t}}E\left[x_{it}^{*}\middle|z_{it}\right]=\left(1-\frac{\beta-\beta^{T-t+1}}{(T-t)(1-\beta)}\right)y_{i,t-1}-\left(\frac{1}{1-\beta}-\frac{\beta-\beta^{T-t+1}}{(T-t)(1-\beta)^{2}}\right)E\left[\alpha_i\middle|z_{it}\right],$$

from which we obtain

$$v_{it}^{*}=-\sqrt{\frac{T-t}{T-t+1}}\left(\frac{1}{1-\beta}-\frac{\beta-\beta^{T-t+1}}{(T-t)(1-\beta)^{2}}\right)\left(\alpha_i-E\left[\alpha_i\middle|z_{it}\right]\right)$$
$$-\sqrt{\frac{T-t}{T-t+1}}\frac{(1-\beta)\varepsilon_{i,T-1}+\left(1-\beta^{2}\right)\varepsilon_{i,T-2}+\cdots+\left(1-\beta^{T-t}\right)\varepsilon_{i,t}}{(T-t)(1-\beta)}.\qquad(20)$$

We now compute $E\left[\left(\alpha_i-E\left[\alpha_i\middle|z_{it}\right]\right)^{2}\right]=\mathrm{Var}\left[\alpha_i\middle|z_{it}\right]$. It can be shown that

$$\mathrm{Cov}\left(\alpha_i,\left(y_{i0},\ldots,y_{it-1}\right)'\right)=\frac{\sigma_\alpha^{2}}{1-\beta}\ell,\qquad\text{and}\qquad\mathrm{Var}\left(\left(y_{i0},\ldots,y_{it-1}\right)'\right)=\frac{\sigma_\alpha^{2}}{(1-\beta)^{2}}\ell\ell'+Q$$

where ℓ is a t-dimensioanl column vector of ones, and

$$Q=\frac{\sigma^{2}}{1-\beta^{2}}\begin{bmatrix}1&\beta&&\beta^{t-1}\\\beta&1&&\beta^{t-2}\\&&\ddots&\\\beta^{t-1}&&&1\end{bmatrix}$$

Therefore, the conditional variance is given by

$$\sigma_\alpha^2 - \sigma_\alpha^2 \ell' \left[\ell\ell' + \frac{(1-\beta)^2}{\sigma_\alpha^2} Q \right]^{-1} \ell$$

Because

$$\begin{aligned}
\left[\ell\ell' + \frac{(1-\beta)^2}{\sigma_\alpha^2} Q \right]^{-1} &= \left(\frac{(1-\beta)^2}{\sigma_\alpha^2} Q \right)^{-1} \\
&\quad - \frac{1}{1 + \ell' \left(\frac{(1-\beta)^2}{\sigma_\alpha^2} Q \right)^{-1} \ell} \left(\frac{(1-\beta)^2}{\sigma_\alpha^2} Q \right)^{-1} \ell\ell' \left(\frac{(1-\beta)^2}{\sigma_\alpha^2} Q \right)^{-1} \\
&= \frac{\sigma_\alpha^2}{(1-\beta)^2} Q^{-1} - \frac{1}{1 + \frac{\sigma_\alpha^2}{(1-\beta)^2} \ell' Q^{-1} \ell} \left(\frac{\sigma_\alpha^2}{(1-\beta)^2} \right)^2 Q^{-1} \ell\ell' Q^{-1},
\end{aligned}$$

we obtain

$$\ell' \left[\ell\ell' + \frac{(1-\beta)^2}{\sigma_\alpha^2} Q \right]^{-1} \ell = \frac{\sigma_\alpha^2}{(1-\beta)^2} \ell' Q^{-1} \ell - \frac{\left(\frac{\sigma_\alpha^2}{(1-\beta)^2} \ell' Q^{-1} \ell \right)^2}{1 + \frac{\sigma_\alpha^2}{(1-\beta)^2} \ell' Q^{-1} \ell} = \frac{\frac{\sigma_\alpha^2}{(1-\beta)^2} \ell' Q^{-1} \ell}{1 + \frac{\sigma_\alpha^2}{(1-\beta)^2} \ell' Q^{-1} \ell}$$

and hence,

$$\sigma_\alpha^2 - \sigma_\alpha^2 \ell' \left[\ell\ell' + \frac{(1-\beta)^2}{\sigma_\alpha^2} Q \right]^{-1} \ell = \frac{\sigma_\alpha^2}{1 + \frac{\sigma_\alpha^2}{(1-\beta)^2} \ell' Q^{-1} \ell}$$

Now, it can be shown that[12]

$$\ell' Q^{-1} \ell = \frac{1}{\sigma^2} \left(2(1-\beta) + (t-2)(1-\beta)^2 \right)$$

from which we obtain

$$E\left[(\alpha_i - E[\alpha_i | z_{it}])^2 \right] = \frac{\sigma_\alpha^2}{1 + \frac{\sigma_\alpha^2}{\sigma^2} \frac{2}{1-\beta} + \frac{\sigma_\alpha^2}{\sigma^2} (t-2)}. \tag{21}$$

We now characterize $E\left[v_{it}^{*2}\right]$. Using (20), and the independence of the first and second term there, we can see that

$$\begin{aligned}
E\left[v_{it}^{*2}\right] &= \frac{T-t}{T-t+1} \left(\frac{1}{1-\beta} - \frac{\beta - \beta^{T-t+1}}{(T-t)(1-\beta)^2} \right)^2 E\left[(\alpha_i - E[\alpha_i | z_{it}])^2 \right] \\
&\quad - \sigma^2 \frac{T-t}{T-t+1} \frac{1}{(T-t)^2 (1-\beta)^2} \left((T-t) + \frac{\beta^2 - 2\beta^{T-t+2} + \beta^{2(T-t)+2} - 2\beta^{T-t+1} + 2\beta}{\beta^2 - 1} \right).
\end{aligned}$$

With (21), we obtain the first conclusion.

As for $E[v_{it}^* \varepsilon_{it}^*]$, we note that

$$\varepsilon_{it}^* = \varepsilon_{it} - \frac{1}{T-t} (\varepsilon_{iT} + \cdots + \varepsilon_{it+1}).$$

[12] See Amemiya (1985, p. 164), for example.

Combining with (20), we obtain

$$E\left[v_{it}^{*}\varepsilon_{it}^{*}\right] = -\sqrt{\frac{T-t}{T-t+1}}\frac{\left(1-\beta^{T-t}\right)}{(T-t)(1-\beta)}\sigma^2 + \sqrt{\frac{T-t}{T-t+1}}\frac{(1-\beta)+\cdots+\left(1-\beta^{T-t-1}\right)}{(T-t)^2(1-\beta)}\sigma^2,$$

from which follows the second conclusion.

As for $E\left[v_{is}^{*}\varepsilon_{it}^{*}\right]$ and $E\left[v_{it}^{*}\varepsilon_{is}^{*}\right]$ $s<t$, we note that

$$\begin{aligned} E\left[v_{is}^{*}\varepsilon_{it}^{*}\right] &= -\sqrt{\frac{T-s}{T-s+1}}\frac{\left(1-\beta^{T-t}\right)}{(T-s)(1-\beta)}\sigma^2 \\ &\quad + \sqrt{\frac{T-s}{T-s+1}}\frac{(1-\beta)+\left(1-\beta^2\right)+\cdots+\left(1-\beta^{T-t-1}\right)}{(T-s)(T-t)(1-\beta)}\sigma^2 \end{aligned}$$

and

$$E\left[v_{it}^{*}\varepsilon_{is}^{*}\right] = \sqrt{\frac{T-t}{T-t+1}}\frac{(1-\beta)+\left(1-\beta^2\right)+\cdots+\left(1-\beta^{T-t}\right)}{(T-s)(T-t)(1-\beta)}\sigma^2.$$

■

Lemma 6

$$\frac{1}{nT}\sum_t \frac{t^2}{n-t}E\left[v_{it}^{*2}\right] = o(1)$$

Proof. Write

$$\begin{aligned} E\left[v_{it}^{*2}\right] &= \frac{T-t}{T-t+1}\left(\frac{1}{1-\beta}-\frac{\beta-\beta^{T-t+1}}{(T-t)(1-\beta)^2}\right)^2 \frac{\sigma_\alpha^2}{1+\frac{\sigma_\alpha^2}{\sigma^2}\frac{2}{1-\beta}+\frac{\sigma_\alpha^2}{\sigma^2}(t-2)} \\ &\quad - \sigma^2\frac{T-t}{T-t+1}\frac{1}{(T-t)^2(1-\beta)^2}\left((T-t)+\frac{\beta^2+2\beta}{\beta^2-1}\right) \\ &\quad - \sigma^2\frac{T-t}{T-t+1}\frac{1}{(T-t)^2(1-\beta)^2}\left(\frac{-2\beta^{T-t+2}+\beta^{2(T-t)+2}-2\beta^{T-t+1}}{\beta^2-1}\right) \end{aligned}$$

Sum of the first two terms on the right can be bounded above by

$$C\frac{\sigma_\alpha^2}{1+\frac{\sigma_\alpha^2}{\sigma^2}\frac{2}{1-\beta}+\frac{\sigma_\alpha^2}{\sigma^2}(t-2)},$$

and the third term can be bounded above in absolute value by

$$C\frac{1}{(T-t)^2}$$

where C is a generic constant. Therefore, we have

$$\begin{aligned} \left|\frac{1}{nT}\sum_t \frac{t^2}{n-t}E\left[v_{it}^{*2}\right]\right| &\le \frac{C}{nT}\sum_t \frac{t^2}{n-t}\frac{\sigma_\alpha^2}{1+\frac{\sigma_\alpha^2}{\sigma^2}\frac{2}{1-\beta}+\frac{\sigma_\alpha^2}{\sigma^2}(t-2)} + \frac{C}{nT}\sum_t \frac{t^2}{n-t}\frac{1}{(T-t)^2} \\ &\le \frac{C}{nT}\sum_t \frac{T^2}{n-T}\frac{\sigma_\alpha^2}{1+\frac{\sigma_\alpha^2}{\sigma^2}\frac{2}{1-\beta}+\frac{\sigma_\alpha^2}{\sigma^2}(t-2)} + \frac{C}{nT}\sum_t \frac{T^2}{n-T}\frac{1}{(T-t)^2} \end{aligned}$$

It can be shown that

$$\sum_t \frac{\sigma_\alpha^2}{1+\frac{\sigma_\alpha^2}{\sigma^2}\frac{2}{1-\beta}+\frac{\sigma_\alpha^2}{\sigma^2}(t-2)} = O(\log T), \quad \sum_t \frac{1}{(T-t)^2} = O(1)$$

Using the assumption that $T/n = O(1)$, we obtain the desired conclusion. ■

Lemma 7

$$\frac{1}{nT}\sum_t \frac{t^2}{n-t}\left(E\left[v_{it}^* \varepsilon_{it}^*\right]\right)^2 = o(1)$$

Proof. We can bound $\left(E\left[v_{it}^* \varepsilon_{it}^*\right]\right)^2$ by $\frac{C}{(T-t)^2}$, where C is a generic constant. Conclusion easily follows by adopting the same proof as in Lemma 6. ■

Lemma 8

$$\frac{1}{nT}\sum_{s<t} \frac{st}{n-t} E\left[v_{is}^* \varepsilon_{it}^*\right] E\left[v_{it}^* \varepsilon_{is}^*\right] = o(1)$$

Proof. We can bound $\left|E\left[v_{is}^* \varepsilon_{it}^*\right] E\left[v_{it}^* \varepsilon_{is}^*\right]\right|$ by $\frac{C}{(T-s)^2}$. Therefore, we have

$$\left|\frac{1}{nT}\sum_{s<t} \frac{st}{n-t} E\left[v_{is}^* \varepsilon_{it}^*\right] E\left[v_{it}^* \varepsilon_{is}^*\right]\right| \le \frac{C}{nT}\sum_{t=1}^{T-1}\sum_{s=1}^{t-1}\frac{st}{n-t}\frac{1}{(T-s)^2} \le \frac{C}{nT}\sum_{t=1}^{T-1}\frac{t}{n-t}\left(\sum_{s=1}^{t-1}\frac{s}{(T-s)^2}\right)$$

But because

$$\frac{s}{(T-s)^2} = \frac{T}{(T-s)^2} - \frac{1}{T-s} \le \frac{T}{(T-s)^2}$$

we can bound $\left|\frac{1}{nT}\sum_{s<t}\frac{st}{n-t}E\left[v_{is}^* \varepsilon_{it}^*\right] E\left[v_{it}^* \varepsilon_{is}^*\right]\right|$ further by

$$\frac{C}{n}\sum_{t=1}^{T-1}\frac{t}{n-t}\sum_{s=1}^{t-1}\frac{1}{(T-s)^2}$$

Because

$$\sum_{s=1}^{t-1}\frac{1}{(T-s)^2} = O\left(\int_t^T \frac{1}{s^2}ds\right) = O\left(\frac{T-t}{Tt}\right) = O\left(\frac{1}{t}\right)$$

we have

$$\frac{C}{n}\sum_{t=1}^{T-1}\frac{t}{n-t}\left(\frac{1}{t}-\frac{1}{T}\right) = \frac{C}{n}\sum_{t=1}^{T-1}\frac{1}{n-t} - \frac{C}{nT}\sum_{t=1}^{T-1}\frac{t}{n-t}.$$

Conclusion follows from

$$\left|\frac{1}{nT}\sum_{s<t}\frac{st}{n-t}E\left[v_{is}^* \varepsilon_{it}^*\right] E\left[v_{it}^* \varepsilon_{is}^*\right]\right| = O\left(\frac{C}{n}\sum_{t=1}^{T-1}\frac{1}{n-t}\right) = O\left(\frac{\log n - \log T}{n}\right) = o(1).$$

■

Lemma 9

$$\mathrm{Var}\left(\frac{1}{\sqrt{nT}}\sum_t \frac{t}{n-t} x_t^{*\prime} M_t \varepsilon_t^*\right) = o(1)$$

Proof. Note that

$$\begin{aligned}\operatorname{Var}\left(\frac{1}{\sqrt{nT}}\sum_t \frac{t}{n-t}x_t^{*\prime}M_t\varepsilon_t^*\right) &= \frac{1}{nT}\sum_t\left(\frac{t}{n-t}\right)^2\operatorname{Var}\left(x_t^{*\prime}M_t\varepsilon_t^*\right)\\ &+\frac{2}{nT}\sum_{s<t}\left(\frac{t}{n-t}\right)\left(\frac{s}{n-s}\right)\operatorname{Cov}\left(x_t^{*\prime}M_t\varepsilon_t^*, x_s^{*\prime}M_s\varepsilon_s^*\right)\\ &=\frac{\sigma^2}{nT}\sum_t\frac{t^2}{n-t}E\left[v_{it}^{*2}\right]\\ &+\frac{1}{nT}\sum_t\frac{t^2}{n-t}\left(E\left[v_{it}^*\varepsilon_{it}^*\right]\right)^2+\frac{2}{nT}\sum_{s<t}\frac{st}{n-t}E\left[v_{is}^*\varepsilon_{it}^*\right]E\left[v_{it}^*\varepsilon_{is}^*\right]\end{aligned}$$

Here, the second equality is based on Lemma 4. Lemmas 6, 7, and 8 establish that variances of the three terms on the far right are all of order $o(1)$. ■

Lemma 10

$$\frac{1}{\sqrt{nT}}\sum_t\left(x_t^{*\prime}P_t\varepsilon_t^*-\frac{K_t}{n-K_t}x_t^{*\prime}M_t\varepsilon_t^*\right)\to N\left(0,\frac{\sigma^4}{1-\beta^2}\right)$$

Proof. Follows easily by combining Lemma 9 and the proof of Theorem 2 in Alvarez and Arellano (1998). ■

Lemma 11

$$\frac{1}{nT}\sum_t\frac{K_t}{n-K_t}x_t^{*\prime}M_tx_t^*=o_p(1)$$

Proof. First, note that $x_t^{*\prime}M_tx_t^*=v_t^{*\prime}M_tv_t^*$ by normality. We therefore have

$$E\left(\frac{1}{nT}\sum_t\frac{K_t}{n-K_t}x_t^{*\prime}M_tx_t^*\right)=\frac{1}{nT}\sum_t\frac{t}{n-t}E\left[v_t^{*\prime}M_tv_t^*\right]$$

By conditioning, it can be shown that

$$E\left[v_t^{*\prime}M_tv_t^*\right]=(n-t)E\left[v_{it}^{*2}\right]$$

Therefore,

$$E\left(\frac{1}{nT}\sum_t\frac{K_t}{n-K_t}x_t^{*\prime}M_tx_t^*\right)=\frac{1}{nT}\sum_t tE\left[v_{it}^{*2}\right]$$

Modifying the proof of Lemma 6, we can establish that the right is $o(1)$.

We now show that

$$\operatorname{Var}\left(\frac{1}{nT}\sum_t\frac{K_t}{n-K_t}x_t^{*\prime}M_tx_t^*\right)=o(1).$$

We have

$$\begin{aligned}\operatorname{Var}\left(\frac{1}{nT}\sum_t\frac{K_t}{n-K_t}x_t^{*\prime}M_tx_t^*\right)&=\frac{1}{n^2T^2}\sum_t\left(\frac{t}{n-t}\right)^2\operatorname{Var}\left(v_t^{*\prime}M_tv_t^*\right)\\ &+\frac{2}{n^2T^2}\sum_{s<t}\frac{t}{n-t}\frac{s}{n-s}\operatorname{Cov}\left(v_s^{*\prime}M_sv_s^*, v_t^{*\prime}M_tv_t^*\right)\end{aligned}$$

Modifying the development from (A53) to (A58) in Alvarez and Arellano (1998) and using normality, we can show that

$$\operatorname{Var}\left(v_t^{*\prime} M_t v_t^*\right) = 2(n-t) E\left[v_{it}^{*4}\right] = 6(n-t)\left(E\left[v_{it}^{*2}\right]\right)^2,$$
$$\operatorname{Cov}\left(v_s^{*\prime} M_s v_s^*, v_t^{*\prime} M_t v_t^*\right) = 2(n-t)\left(E\left[v_{it}^* v_{is}^*\right]\right)^2.$$

Using (20), we can show that

$$\begin{aligned}
E\left[v_{it}^* v_{is}^*\right] &= \sqrt{\frac{T-t}{T-t+1}}\sqrt{\frac{T-s}{T-s+1}}\left(\frac{1}{1-\beta} - \frac{\beta-\beta^{T-t+1}}{(T-t)(1-\beta)^2}\right) \\
&\quad \times \left(\frac{1}{1-\beta} - \frac{\beta-\beta^{T-s+1}}{(T-s)(1-\beta)^2}\right)\frac{\sigma_\alpha^2}{1+\frac{\sigma_\alpha^2}{\sigma^2}\frac{2}{1-\beta}+\frac{\sigma_\alpha^2}{\sigma^2}(t-2)} \\
&\quad + \sigma^2\sqrt{\frac{T-t}{T-t+1}}\sqrt{\frac{T-s}{T-s+1}}\frac{1}{(T-t)(T-s)(1-\beta)^2} \\
&\quad \times \left((T-t) + \frac{\beta^2-2\beta^{T-t+2}+\beta^{2(T-t)+2}-2\beta^{T-t+1}+2\beta}{\beta^2-1}\right).
\end{aligned}$$

Adopting the same argument as in the proofs for Lemmas 6 - 8, we can show that the variance is $o(1)$. ■

B Technical Details for Section 4

Definition 2

$$\Psi \equiv 3\lambda_1'\Lambda^{-1}\lambda_2 - 3\lambda_1'\Lambda^{-1}\Lambda_1\Lambda^{-1}\lambda_1,$$
$$\Upsilon \equiv 2\lambda_1\Lambda^{-1}\lambda_1, \quad \frac{1}{\sqrt{n}}\Phi \equiv 2\lambda_1'\Lambda^{-1}g,$$
$$\frac{1}{\sqrt{n}}\Xi \equiv 4(g_1-\lambda_1)'\Lambda^{-1}\lambda_1 - 2\lambda_1'\Lambda^{-1}(G-\Lambda)\Lambda^{-1}\lambda_1 - 4\lambda_1'\Lambda^{-1}\Lambda_1\Lambda^{-1}g + 2\lambda_2'\Lambda^{-1}g,$$
$$\frac{1}{n}\Gamma \equiv 2(g_1-\lambda_1)'\Lambda^{-1}g - 2\lambda_1'\Lambda^{-1}(G-\Lambda)\Lambda^{-1}g - g'\Lambda^{-1}\Lambda_1\Lambda^{-1}g.$$

Proof of Lemma 1. By Lemma 2 (a) and Theorem 1(a) of Andrews (1992) and Conditions 4-7 it follows that $\sup_{c\in C}|Q_n(c) - Q(c)| = o_p(1)$, where $Q(c) \equiv \lambda(c)'\Lambda(c)^{-1}\lambda(c)$. Let $B(\beta,\epsilon)$ be an open interval of length ϵ centered at β. By Condition 8 it follows that $\inf_{c\notin B(\beta,\epsilon)} Q(c) > Q(\beta) = 0$ for all $\epsilon > 0$. It then follows from standard arguments that $b - \beta = o_p(1)$. It therefore follows that $\Pr(S_n(b) \neq 0) \leq \Pr(b \in \partial C) = 1 - \Pr(b \in \operatorname{int} C) \leq 1 - \Pr(b \in B(\beta,\epsilon)) \to 0$ for any $\epsilon > 0$ where ∂C denotes the boundary of C.

Using similar arguments as in Pakes and Pollard (1989) we write

$$\begin{aligned}
|Q(b)| &= \left|\lambda(b)'\Lambda(b)^{-1}\lambda(b)\right| \\
&\leq \left|g(b)'G(b)^{-1}g(b) - \lambda(b)'\Lambda(b)^{-1}\lambda(b) - g(\beta)'G(b)^{-1}g(\beta)\right| \\
&\quad + \left|g(b)'G(b)^{-1}g(b)\right| + \left|g(\beta)'G(b)^{-1}g(\beta)\right| \\
&\leq \left|g(b)'G(b)^{-1}g(b) - \lambda(b)'G(b)^{-1}\lambda(b) - g(\beta)'G(b)^{-1}g(\beta)\right| \\
&\quad + \left|\lambda(b)'G(b)^{-1}\lambda(b) - \lambda(b)'\Lambda(b)^{-1}\lambda(b)\right| \\
&\quad + \left|g(b)'G(b)^{-1}g(b)\right| + \left|g(\beta)'G(b)^{-1}g(\beta)\right|,
\end{aligned}$$

where $\left|g(b)'G(b)^{-1}g(b)\right| \leq \left|g(\beta)'G(\beta)^{-1}g(\beta)\right| = O_p(n^{-1})$ by the definition of b and Condition 9. We have $G(b)^{-1} = O_p(1)$ by consistency of b and the uniform law of large numbers, from which we obtain $\left|g(\beta)'G(b)^{-1}g(\beta)\right| = O_p(n^{-1})$. We also have

$$\begin{aligned}
& g(b)'G(b)^{-1}g(b) - \lambda(b)'G(b)^{-1}\lambda(b) - g(\beta)'G(b)^{-1}g(\beta) \\
= \; & g(b)'G(b)^{-1}(g(b) - g(\beta) - \lambda(b)) + (g(b) - g(\beta) - \lambda(b))'G(b)^{-1}\lambda(b) \\
& + 2g(\beta)'G(b)^{-1}\lambda(b) + (g(b) - g(\beta) - \lambda(b))'G(b)^{-1}g(\beta).
\end{aligned}$$

Therefore, we obtain

$$\begin{aligned}
|Q(b)| &\leq \left|\lambda(b)'G(b)^{-1}\lambda(b) - \lambda(b)'\Lambda(b)^{-1}\lambda(b)\right| \\
&\quad + \left|g(b)'G(b)^{-1}(g(b) - g(\beta) - \lambda(b))\right| \\
&\quad + \left|(g(b) - g(\beta) - \lambda(b))'G(b)^{-1}\lambda(b)\right| \\
&\quad + 2\left|g(\beta)'G(b)^{-1}\lambda(b)\right| + \left|(g(b) - g(\beta) - \lambda(b))'G(b)^{-1}g(\beta)\right| \\
&\quad + O_p(n^{-1}).
\end{aligned}$$

The terms $G(\beta)^{-1}$, and $\Lambda(b)^{-1}$ are $O_p(1)$ by consistency of b and the uniform law of large numbers. Also, the terms $g(b)$ and $\lambda(b)$ are $o_p(1)$ by consistency of b and the uniform law of large numbers. From Theorems 1 and 2 in Andrews (1994) and Conditions 4-9 it follows that $g(b) - \lambda(b) - g(\beta) = o_p(n^{-1/2})$. From a standard CLT and consistency of b it follows that $(g(b) - g(\beta) - \lambda(b))'G(b)^{-1}g(\beta) = o_p(n^{-1})$, and $g(\beta)'G(b)^{-1} = O_p(n^{-1/2})$. These results show that

$$\begin{aligned}
|Q(b)| &\leq \left\|G(b)^{-1} - \Lambda(b)^{-1}\right\| \|\lambda(b)\|^2 + O_p\left(n^{-1/2}\right)\|\lambda(b)\| + o_p(n^{-1}) \\
&= o_p(1)\|\lambda(b)\|^2 + O_p\left(n^{-1/2}\right)\|\lambda(b)\| + o_p(n^{-1}).
\end{aligned}$$

Because $|Q(b)| = \left|\lambda(b)'\Lambda(b)^{-1}\lambda(b)\right| \geq \frac{1}{\mathrm{M}}\|\lambda(b)\|^2$, we conclude that

$$\left(\frac{1}{\mathrm{M}} - o_p(1)\right)\|\lambda(b)\|^2 - O_p\left(n^{-1/2}\right)\|\lambda(b)\| \leq o_p\left(n^{-1}\right)$$

or

$$\|\lambda(b)\| = O_p\left(n^{-1/2}\right),$$

which implies that $b-\beta = O_p\left(n^{-1/2}\right)$. ■

Proof of Theorem 3. Note that we have

$$g_1(b) = g_1 + \frac{1}{\sqrt{n}} g_2 \cdot \sqrt{n}(b-\beta) + \frac{1}{2n} g_3 \cdot \left(\sqrt{n}(b-\beta)\right)^2 + o_p\left(\frac{1}{n}\right),$$
$$g(b) = g + \frac{1}{\sqrt{n}} g_1 \cdot \sqrt{n}(b-\beta) + \frac{1}{2n} g_2 \cdot \left(\sqrt{n}(b-\beta)\right)^2 + o_p\left(\frac{1}{n}\right),$$

and

$$\begin{aligned} G(b)^{-1} &= G^{-1} - \frac{1}{\sqrt{n}} G^{-1} G_1 G^{-1} \cdot \sqrt{n}(b-\beta) \\ &\quad + \frac{1}{2}\frac{1}{n}\left(2G^{-1}G_1G^{-1}G_1G^{-1} - G^{-1}G_2G^{-1}\right)\left(\sqrt{n}(b-\beta)\right)^2 + o_p\left(\frac{1}{n}\right), \\ G_1(b) &= G_1 + \frac{1}{\sqrt{n}} G_2 \cdot \sqrt{n}(b-\beta) + \frac{1}{2}\frac{1}{n} G_3 \cdot \left(\sqrt{n}(b-\beta)\right)^2 + o_p\left(\frac{1}{n}\right) \end{aligned}$$

where g, g_j, G, G_j and $\sqrt{n}(b-\beta)$ are $O_p(1)$ by Conditions 6 and 7 and Lemma 1. Therefore, we have

$$g_1(b)' G(b)^{-1} g(b) = g_1' G^{-1} g + \frac{1}{\sqrt{n}} h_1 \cdot \sqrt{n}(b-\beta) + \frac{1}{n} h_2 \cdot \left(\sqrt{n}(b-\beta)\right)^2 + o_p\left(\frac{1}{n}\right),$$

and

$$\begin{aligned} g(b)' G(b)^{-1} G_1(b) G(b)^{-1} g(b) &= g' G^{-1} G_1 G^{-1} g + \frac{1}{\sqrt{n}} h_3 \cdot \sqrt{n}(b-\beta) \\ &\quad + \frac{1}{n} h_4 \cdot \left(\sqrt{n}(b-\beta)\right)^2 + o_p\left(\frac{1}{n}\right), \end{aligned}$$

where

$$\begin{aligned} h_1 &= g_2' G^{-1} g - g_1' G^{-1} G_1 G^{-1} g + g_1' G^{-1} g_1, \\ h_2 &= \frac{1}{2} g_3' G^{-1} g + \frac{1}{2} g_1' \left(2G^{-1}G_1G^{-1}G_1G^{-1} - G^{-1}G_2G^{-1}\right) g + \frac{1}{2} g_1' G^{-1} g_2 \\ &\quad - g_2' G^{-1} G_1 G^{-1} g - g_1' G^{-1} G_1 G^{-1} g_1 + g_2' G^{-1} g_1 \\ &= \frac{1}{2} g_3' G^{-1} g + g_1' G^{-1} G_1 G^{-1} G_1 G^{-1} g - \frac{1}{2} g_1' G^{-1} G_2 G^{-1} g + \frac{3}{2} g_1' G^{-1} g_2 \\ &\quad - g_2' G^{-1} G_1 G^{-1} g - g_1' G^{-1} G_1 G^{-1} g_1, \end{aligned}$$

and

$$\begin{aligned} h_3 &= g_1' G^{-1} G_1 G^{-1} g - g' G^{-1} G_1 G^{-1} G_1 G^{-1} g + g' G^{-1} G_2 G^{-1} g \\ &\quad - g' G^{-1} G_1 G^{-1} G_1 G^{-1} g + g' G^{-1} G_1 G^{-1} g_1 \\ &= 2 g_1' G^{-1} G_1 G^{-1} g - 2 g' G^{-1} G_1 G^{-1} G_1 G^{-1} g + g' G^{-1} G_2 G^{-1} g, \end{aligned}$$

$$
\begin{aligned}
h_4 = {} & \frac{1}{2} g_2' G^{-1} G_1 G^{-1} g + \frac{1}{2} g' \left(2G^{-1} G_1 G^{-1} G_1 G^{-1} - G^{-1} G_2 G^{-1} \right) G_1 G^{-1} g \\
& + \frac{1}{2} g' G^{-1} G_3 G^{-1} g + \frac{1}{2} g' G^{-1} G_1 \left(2G^{-1} G_1 G^{-1} G_1 G^{-1} - G^{-1} G_2 G^{-1} \right) g \\
& + \frac{1}{2} g' G^{-1} G_1 G^{-1} g_2 \\
& - g_1' G^{-1} G_1 G^{-1} G_1 G^{-1} g + g_1' G^{-1} G_2 G^{-1} g - g_1' G^{-1} G_1 G^{-1} G_1 G^{-1} g \\
& + g_1' G^{-1} G_1 G^{-1} g_1 - g' G^{-1} G_1 G^{-1} G_2 G^{-1} g \\
& + g' G^{-1} G_1 G^{-1} G_1 G^{-1} G_1 G^{-1} g - g' G^{-1} G_1 G^{-1} G_1 G^{-1} g_1 - g' G^{-1} G_2 G^{-1} G_1 G^{-1} g \\
& + g' G^{-1} G_2 G^{-1} g_1 - g' G^{-1} G_1 G^{-1} G_1 G^{-1} g_1
\end{aligned}
$$

We may therefore rewrite the score $S_n(b)$ as

$$
\begin{aligned}
S_n(b) = {} & \left(2g_1' G^{-1} g - g' G^{-1} G_1 G^{-1} g \right) + \frac{1}{\sqrt{n}} (2h_1 - h_3) \sqrt{n} (b - \beta) \\
& + \frac{1}{n} (2h_2 - h_4) \left(\sqrt{n} (b - \beta) \right)^2 + o_p\left(\frac{1}{n} \right).
\end{aligned}
\tag{22}
$$

Next note that $P(|S_n(b)| > \epsilon) = P(|S_n(b)| > \epsilon / n^{-1})$ for any $\epsilon > 0$ because of Lemma 1. Thus $S_n(b) = o_p(n^{-1})$ and we can subsume this error into the $o_p(n^{-1})$ term of 22. Using these arguments and Lemmas 12, 13, and 14 below, we may rewrite the first order condition (22) as

$$
0 = \frac{1}{\sqrt{n}} \Phi + \frac{1}{n} \Gamma + \frac{1}{\sqrt{n}} \left(\Upsilon + \frac{1}{\sqrt{n}} \Xi \right) \sqrt{n} (b - \beta) + \frac{1}{n} \Psi \left(\sqrt{n} (b - \beta) \right)^2 + o_p\left(\frac{1}{n} \right)
$$

or

$$
0 = \Phi + \frac{1}{\sqrt{n}} \Gamma + \left(\Upsilon + \frac{1}{\sqrt{n}} \Xi \right) \sqrt{n} (b - \beta) + \frac{1}{\sqrt{n}} \Psi \left(\sqrt{n} (b - \beta) \right)^2 + o_p\left(\frac{1}{\sqrt{n}} \right),
$$

based on which we obtain (9). Noting that

$$
E[\Phi] = 2\sqrt{n} \lambda_1' \Lambda^{-1} E[g] = 0, \tag{23}
$$

$$
\begin{aligned}
E[\Gamma] & = 2nE\left[(g_1 - \lambda_1)' \Lambda^{-1} g \right] - 2n\lambda_1' \Lambda^{-1} E\left[(G - \Lambda) \Lambda^{-1} g \right] - nE\left[g' \Lambda^{-1} \Lambda_1 \Lambda^{-1} g \right] \\
& = 2 \operatorname{trace}\left(\Lambda^{-1} E\left[\delta_i \frac{\partial \delta_i'}{\partial \beta} \right] \right) - 2\lambda_1' \Lambda^{-1} E\left[\psi_i \psi_i' \Lambda^{-1} \delta_i \right] - \operatorname{trace}\left(\Lambda^{-1} \Lambda_1 \Lambda^{-1} E\left[\delta_i \delta_i' \right] \right),
\end{aligned}
\tag{24}
$$

$$
\begin{aligned}
E[\Phi \Xi] & = 8n\lambda_1' \Lambda^{-1} E\left[g (g_1 - \lambda_1)' \right] \Lambda^{-1} \lambda_1 - 4nE\left[\lambda_1' \Lambda^{-1} g \lambda_1' \Lambda^{-1} (G - \Lambda) \Lambda^{-1} \lambda_1 \right] \\
& \quad - 8n\lambda_1' \Lambda^{-1} E\left[g g' \right] \Lambda^{-1} \Lambda_1 \Lambda^{-1} \lambda_1 + 4n\lambda_1' \Lambda^{-1} E\left[g g' \right] \Lambda^{-1} \lambda_2 \\
& = 8\lambda_1' \Lambda^{-1} E\left[\delta_i \frac{\partial \delta_i'}{\partial \beta} \right] \Lambda^{-1} \lambda_1 - 4\lambda_1' \Lambda^{-1} E\left[\delta_i \lambda_1' \Lambda^{-1} \psi_i \psi_i' \right] \Lambda^{-1} \lambda_1 \\
& \quad - 8\lambda_1' \Lambda^{-1} E\left[\delta_i \delta_i' \right] \Lambda^{-1} \Lambda_1 \Lambda^{-1} \lambda_1 + 4\lambda_1' \Lambda^{-1} E\left[\delta_i \delta_i' \right] \Lambda^{-1} \lambda_2,
\end{aligned}
\tag{25}
$$

and

$$
E\left[\Phi^2 \right] = 4\lambda_1' \Lambda^{-1} E\left[\delta_i \delta_i' \right] \Lambda^{-1} \lambda_1, \tag{26}
$$

we obtain the desired conclusion. ■

Lemma 12 *Under Conditions 6 and 7*

$$h_2 = \frac{3}{2}\lambda_1'\Lambda^{-1}\lambda_2 - \lambda_1'\Lambda^{-1}\Lambda_1\Lambda^{-1}\lambda_1 + o_p(1),$$
$$h_4 = \lambda_1'\Lambda^{-1}\Lambda_1\Lambda^{-1}\lambda_1 + o_p(1).$$

Proof. Follows from $\operatorname{plim} g = 0$. ■

Lemma 13 *Under Conditions 6 and 7*

$$2h_1 - h_3 = \Upsilon + \frac{1}{\sqrt{n}}\Xi + o_p\left(\frac{1}{\sqrt{n}}\right).$$

Proof. Because

$$g_2'G^{-1}g = \lambda_2'\Lambda^{-1}g + o_p\left(\frac{1}{\sqrt{n}}\right),$$

$$g_1'G^{-1}G_1G^{-1}g = \lambda_1'\Lambda^{-1}\Lambda_1\Lambda^{-1}g + o_p\left(\frac{1}{\sqrt{n}}\right),$$

and

$$\begin{aligned} g_1'G^{-1}g_1 &= (\lambda_1 + (g_1 - \lambda_1))'\left(\Lambda^{-1} - \Lambda^{-1}(G-\Lambda)\Lambda^{-1} + o_p\left(\frac{1}{\sqrt{n}}\right)\right)(\lambda_1 + (g_1 - \lambda_1)) \\ &= \lambda_1\Lambda^{-1}\lambda_1 + 2(g_1 - \lambda_1)'\Lambda^{-1}\lambda_1 - \lambda_1\Lambda^{-1}(G-\Lambda)\Lambda^{-1}\lambda_1 + o_p\left(\frac{1}{\sqrt{n}}\right), \end{aligned} \tag{27}$$

we obtain

$$\begin{aligned} h_1 &= \lambda_2'\Lambda^{-1}g - \lambda_1'\Lambda^{-1}\Lambda_1\Lambda^{-1}g \\ &\quad + \lambda_1\Lambda^{-1}\lambda_1 + 2(g_1 - \lambda_1)'\Lambda^{-1}\lambda_1 - \lambda_1\Lambda^{-1}(G-\Lambda)\Lambda^{-1}\lambda_1 + o_p\left(\frac{1}{\sqrt{n}}\right). \end{aligned}$$

Similarly, we obtain

$$h_3 = 2\lambda_1'\Lambda^{-1}\Lambda_1\Lambda^{-1}g + o_p\left(\frac{1}{\sqrt{n}}\right).$$

The conclusion follows. ■

Lemma 14 *Under Conditions 6 and 7*

$$2g_1'G^{-1}g - g'G^{-1}G_1G^{-1}g = \frac{1}{\sqrt{n}}\Phi + \frac{1}{n}\Gamma + o_p\left(\frac{1}{n}\right).$$

Proof. We have

$$\begin{aligned} g_1'G^{-1}g &= (\lambda_1 + (g_1 - \lambda_1))'\left(\Lambda^{-1} - \Lambda^{-1}(G-\Lambda)\Lambda^{-1} + o_p\left(\frac{1}{\sqrt{n}}\right)\right)g \\ &= \lambda_1'\Lambda^{-1}g + (g_1 - \lambda_1)'\Lambda^{-1}g - \lambda_1'\Lambda^{-1}(G-\Lambda)\Lambda^{-1}g + o_p\left(\frac{1}{n}\right) \end{aligned}$$

and

$$g'G^{-1}G_1G^{-1}g = g'\Lambda^{-1}\Lambda_1\Lambda^{-1}g + o_p\left(\frac{1}{n}\right),$$

from which the conclusion follows. ■

C Proof of Theorem 4

The second order bias is computed using Theorem 4. Because the "weight matrix" here does not involve the parameter of interest, we have $\Lambda_1 = 0$, which renders the third, sixth, and last terms in Theorem 4 equal to zero. Also, because the moment restriction is linear in the parameter of interest, we have $\lambda_2 = 0$, which renders the seventh and eight terms in Theorem 4 equal to zero. Furthermore, because $E\left[z_{it}z_{it}'E\left[z_{it}z_{it}'\right]^{-1}z_{it}\varepsilon_{it}^*\right] = 0$ under conditional symmetry of ε_{it}^*, the numerator in the second term $\lambda_1'\Lambda^{-1}E\left[\psi_i\psi_i'\Lambda^{-1}\delta_i\right] = -\sum_{t=1}^{T-1}E\left[z_{it}x_{it}^*\right]'E\left[z_{it}z_{it}'\right]^{-1}E\left[z_{it}z_{it}'E\left[z_{it}z_{it}'\right]^{-1}z_{it}\varepsilon_{it}^*\right]$ should be equal to zero, and therefore, the second term should be equal to zero. We obtain the desired conclusion by noting that

$$\lambda_1\Lambda^{-1}\lambda_1 = \sum_{t=1}^{T-1}E\left[z_{it}x_{it}^*\right]'E\left[z_{it}z_{it}'\right]^{-1}E\left[z_{it}x_{it}^*\right],$$

$$\text{trace}\left(\Lambda^{-1}E\left[\delta_i\frac{\partial\delta_i'}{\partial\beta}\right]\right) = -\sum_{t=1}^{T-1}\text{trace}\left(E\left[z_{it}z_{it}'\right]^{-1}E\left[\varepsilon_{it}^*x_{it}^*z_{it}z_{it}'\right]\right),$$

$$\lambda_1'\Lambda^{-1}E\left[\delta_i\frac{\partial\delta_i'}{\partial\beta}\right]\Lambda^{-1}\lambda_1 = -\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}E\left[z_{it}x_{it}^*\right]'E\left[z_{it}z_{it}'\right]^{-1}E\left[\varepsilon_{it}^*x_{is}^*z_{it}z_{is}'\right]E\left[z_{is}z_{is}'\right]^{-1}E\left[z_{is}x_{is}^*\right],$$

and

$$\begin{aligned}&\lambda_1'\Lambda^{-1}E\left[\delta_i\lambda_1'\Lambda^{-1}\psi_i\psi_i'\right]\Lambda^{-1}\lambda_1\\&\quad= -\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}E\left[z_{it}x_{it}^*\right]'E\left[z_{it}z_{it}'\right]^{-1}E\left[\varepsilon_{it}^*z_{it}E\left[z_{is}x_{is}^*\right]'E\left[z_{is}z_{is}'\right]^{-1}z_{is}z_{is}'\right]E\left[z_{is}z_{is}'\right]^{-1}E\left[z_{is}x_{is}^*\right].\end{aligned}$$

D Proofs for Section 6

Proof of Lemma 2. Note that

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

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 $n^{-3/2}\sum_{i=1}^n g_{i,1} = o_p(1)$.

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

$$\begin{aligned}E\left[\Delta y_{it}y_{i0}\right] &= E\left[y_{i0}\left(\beta_n^{t-2}\left(\beta_n-1\right)\xi_{i0}+\varepsilon_{is-1}+\left(\beta_n-1\right)\sum_{r=1}^{s-2}\beta_n^{r-1}\varepsilon_{is-1-r}\right)\right]\\&= \beta_n^{t-2}\sigma_\varepsilon^2\frac{\beta_n-1}{1-\beta_n^2} = O(1),\end{aligned}$$

and

$$\begin{aligned}
E\left[\left(\Delta y_{it}y_{i0}\right)^2\right] &= E\left[y_{i0}^2\left(\beta_n^{t-2}\left(\beta_n-1\right)\xi_{i0}+\varepsilon_{is-1}+\left(\beta_n-1\right)\sum_{r=1}^{s-2}\beta_n^{r-1}\varepsilon_{is-1-r}\right)^2\right] \\
&= \beta_n^{2(t-2)}\frac{\left(\beta_n-1\right)^2\sigma_\varepsilon^2}{1-\beta_n^2}\frac{\sigma_\alpha^2}{\left(1-\beta_n\right)^2}+3\beta_n^{2(t-2)}\frac{\sigma_\varepsilon^4(\beta_n-1)^2}{\left(1-\beta_n^2\right)^2} \\
&\quad+\left(\sigma_\varepsilon^2+\left(\beta_n-1\right)^2\sigma_\varepsilon^2\sum_{r=1}^{s-2}\beta_n^{2(r-1)}\right)\left(\frac{\sigma_\alpha^2}{\left(1-\beta_n\right)^2}+\frac{\sigma_\varepsilon^2}{\left(1-\beta_n^2\right)}\right) \\
&= \frac{\sigma_\varepsilon^2\sigma_\alpha^2}{c^2}n^2+O(n).
\end{aligned}$$

such that $\operatorname{Var}\left(n^{-3/2}\sum_{i=1}^n \Delta y_{it}y_{i0}\right)=O(1)$. 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

$$\operatorname{Var}\left(\Delta u_i(\beta_0)y_{i0}\right)=2\sigma_\varepsilon^2\sigma_\alpha^2\left(1-\beta_n\right)^{-2}+O(n)=\frac{2\sigma_\varepsilon^2\sigma_\alpha^2}{c^2}n^2+O(n).$$

The joint limiting distribution of $n^{-3/2}\sum_{i=1}^n\left[f_{i,2}'-Ef_{i,2}',g_{i,2}(\beta_0)'\right]'$ 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 $\mu=\sigma_y^2/2\iota+O(n^{-1})$ 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 $\frac{\sigma_\varepsilon^2\sigma_\alpha^2}{c^2}n^2$ and $\frac{2\sigma_\varepsilon^2\sigma_\alpha^2}{c^2}n^2$. The off-diagonal elements of $\Sigma_{11,n}$ are found to be

$$\begin{aligned}
E\left[\Delta y_{it}\Delta y_{is}y_{i0}^2\right] &= E\left[y_{i0}^2\left(\beta_n^{s-2}\left(\beta_n-1\right)\xi_{i0}+\varepsilon_{is-1}+\left(\beta_n-1\right)\sum_{r=1}^{s-2}\beta_n^{r-1}\varepsilon_{is-1-r}\right)\right. \\
&\quad\left.\times\left(\beta_n^{t-2}\left(\beta_n-1\right)\xi_{i0}+\varepsilon_{it-1}+\left(\beta_n-1\right)\sum_{r=1}^{t-2}\beta_n^{r-1}\varepsilon_{it-1-r}\right)\right] \\
&= \beta_n^{t-2}\beta_n^{s-2}\frac{(\beta_n-1)^2}{(1-\beta_n^2)}\left(\frac{\sigma_\alpha^2}{(1-\beta_n)^2}+3\frac{\sigma_\varepsilon^4}{(1-\beta_n^2)}\right)+O(1)=\frac{\sigma_\alpha^2}{2c}n+O(1)
\end{aligned}$$

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

$$E\left[\Delta u_{it}\Delta u_{is}y_{i0}^2\right]=\begin{cases}-\sigma_\varepsilon^2\sigma_\alpha^2\left(1-\beta_n\right)^{-2}+O(n) & t=s+1\text{ or }t=s-1\\ 0 & \text{otherwise}\end{cases}$$

For $\Sigma_{12,n}$, we consider

$$E\left[\Delta y_{it}\Delta u_{is}y_{i0}^2\right]=\begin{cases}\frac{\sigma_\varepsilon^2\sigma_\alpha^2}{c^2}n^2+O(n) & \text{if }t=s\\ -\frac{\sigma_\varepsilon^2\sigma_\alpha^2}{c^2}n^2+O(n) & \text{if }t=s-1\\ 0 & \text{otherwise}\end{cases}$$

It then follows that for $\ell \in \mathbb{R}^{T(T+1)/+2T-6}$ such that $\ell'\ell = 1$ $n^{-3/2} \sum_{i=1}^{n} \ell' \Sigma_n^{-1/2} \left[f_{i,2}' - Ef_{i,2}, g_{i,2}(\beta_0)' \right]' \xrightarrow{d} N(0,1)$ 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 $\left[\xi_x', \xi_y' \right]' \sim N(0, \Sigma)$. Note that $n^{-3/2} \sum_{i=1}^{n} \ell' E[f_{i,2}] = O(n^{-1/2})$ and thus does not affect the limit distribution.

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

$$\operatorname{Var}(\Delta u_i(\beta_0) y_{i0}) = 2\sigma_\varepsilon^2 \sigma_\alpha^2 (1 - \beta_n)^{-2} + O(n) = \frac{2\sigma_\varepsilon^2 \sigma_\alpha^2}{c^2} n^2 + O(n).$$

The off-diagonal elements of $E\left[g_{i,2} g_{i,2}' \right]$ are obtained from

$$E\left[\Delta u_{it} \Delta u_{is} y_{i0}^2 \right] = \begin{cases} -\sigma_\varepsilon^2 \sigma_\alpha^2 (1 - \beta_n)^{-2} + O(n) & t = s+1 \text{ or } t = s-1 \\ 0 & \text{otherwise} \end{cases}$$

It therefore follows that

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

■

Proof of Theorem 5. Use the transformation $z = L^{-1} C_1' \xi_x$ with $C_1' \Sigma_{11} C_1 = LL'$. Note that we can take $L = \sqrt{\delta} I_{r_1}$. Define $W = \left(C_1' \tilde{\Omega} C_1 \right)^{-1}$. Then $X\left(C_1, \tilde{\Omega} \right) = z'L'WC_1'\xi_y / z'L'WLz$. Next use the fact that $E\left[C_1'\xi_y \middle| C_1 \xi_x \right] = FC_1'\xi_x = FLz$ with $F = C_1' \Sigma_{21} C_1 \left(C_1' \Sigma_{11} C_1 \right)^{-1} = C_1' M_1' C_1$, where the second equality is based on $\left(C_1' \Sigma_{11} C_1 \right)^{-1} = \delta^{-1} I_{r_1}$. Using a conditioning argument it then follows that

$$E\left[X\left(C_1, \tilde{\Omega} \right) \right] = E\left[\frac{z'L'WFLz}{z'L'WLz} \right] = E\left[\frac{z'Dz}{z'Wz} \right].$$

Note that $z'Dz = z'D'z = z'\bar{D}z$ where $\bar{D} = \frac{1}{2}(D + D')$ is symmetric. Also, W is symmetric positive definite. The result then follows from Smith (1993, Eq. 2.4, p. 273). ■

Proof of Theorem 6. We first analyze $E\left[X\left(C_1, \Sigma_{22} \right) \right]$. Note that in this case $W = \left(C_1' \Sigma_{22} C_1 \right)^{-1} = \delta^{-1} \left(C_1' M_2 C_1 \right)^{-1}$ such that

$$E\left[X\left(C_1, \Sigma_{22} \right) \right] = E\left[\frac{z'\delta\bar{D}z}{z'\left(C_1' M_2 C_1 \right)^{-1} z} \right]$$

and $\delta \operatorname{tr} \bar{D} = \delta/2 \operatorname{tr} WC_1' \left(M_1' + M_1 \right) C_1 = -r_1/2$ since $M_1' + M_1 = -M_2$. Let Γ_1 be an orthogonal matrix of eigenvectors of $\left(C_1' M_2 C_1 \right)^{-1}$ with corresponding diagonal matrix of eigenvalues Λ_1 such that $\left(C_1' M_2 C_1 \right)^{-1} = \Gamma_1 \Lambda_1 \Gamma_1'$ and $z_1 = \Gamma_1' z$. Let $\bar{\lambda}$ be the largest element of Λ_1. Then it follows from Smith (1993, p. 273 and Appendix A)

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

where for any two real symmetric matrices Y_1, Y_2

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

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 symmetric then $\operatorname{tr} Y_3 Y_1 Y_2^i = \operatorname{tr} Y_1 Y_3 Y_2^i$ such that

$$\begin{aligned}
\operatorname{tr}\Gamma_1'\bar{D}\Gamma_1(I_{r_1} - \bar{\lambda}^{-1}\Lambda_1)^i &= \operatorname{tr}\Gamma_1'\left(WC_1'M_1'C_1 + C_1'M_1C_1W\right)\Gamma_1\left(I_{r_1} - \bar{\lambda}^{-1}\Lambda_1\right)^i \\
&= \delta^{-1}\operatorname{tr}\left(\Lambda_1\Gamma_1'C_1'M_1'C_1\Gamma_1 + \Gamma_1'C_1'M_1C_1\Gamma_1\Lambda_1\right)\left(I_{r_1} - \bar{\lambda}^{-1}\Lambda_1\right)^i \\
&= \delta^{-1}\operatorname{tr}\Gamma_1'C_1'\left(M_1' + M_1\right)C_1\Gamma_1\Lambda_1\left(I_{r_1} - \bar{\lambda}^{-1}\Lambda_1\right)^i \\
&= -\delta^{-1}\operatorname{tr}\Gamma_1'C_1'M_2C_1\Gamma_1\Lambda_1\left(I_{r_1} - \bar{\lambda}^{-1}\Lambda_1\right)^i \\
&= -\delta^{-1}\operatorname{tr}\Gamma_1\Lambda_1\Gamma_1'C_1'M_2C_1\left(I_{r_1} - \bar{\lambda}^{-1}\Lambda_1\right)^i \\
&= -\delta^{-1}\operatorname{tr}(I_{r_1} - \bar{\lambda}^{-1}\Lambda_1)^i.
\end{aligned}$$

This shows that all the terms $\operatorname{tr}\left(\Gamma_1'\bar{D}\Gamma_1\left(I_{r_1} - \bar{\lambda}^{-1}\Lambda_1\right)^i\right)C_{k-i}\left(I_{r_1} - \bar{\lambda}^{-1}\Lambda_1\right)$ have the same sign and therefore all the terms $C_{1+k}^{1,k}\left(\Gamma_1'\bar{D}\Gamma_1, I_{r_1} - \bar{\lambda}^{-1}\Lambda_1\right)$ have the same sign. Therefore, we have

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

For $k = 0$, we have

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

and $(1)_0\left(\frac{1}{2}\right)_1/\left(\frac{r_1}{2}\right)_1 = 1/r_1$. This shows that $|E\left[X\left(C_1,\Sigma_{22}\right)\right]| \geq \bar{\lambda}^{-1}/2$ for all C_1 such that $C_1'C_1 = I_{r_1}$ and all $r_1 \in \{1, 2, .., T-1\}$. Then

$$\min_{\substack{C_1 \text{ s.t. } C_1'C_1 = I_{r_1} \\ r_1 \in \{1,2,..,T-1\}}} |E\left[X\left(C_1,\Sigma_{22}\right)\right]| \geq \min_{\substack{C_1 \text{ s.t. } C_1'C_1 = I_{r_1} \\ r_1 \in \{1,2,..,T-1\}}} \frac{\bar{\lambda}^{-1}}{2} \geq \frac{\min l_j}{2}, \tag{28}$$

where $\min l_j$ is the smallest eigenvalue of M_2 and the last inequality follows from Magnus and Neudecker (1988, Theorem 10, p. 209). Now let $r_1 = 1$ and $C_1 = \rho_i$, where ρ_i is the eigenvector corresponding to $\min l_j$. Then $I_{r_1} - \bar{\lambda}^{-1}\Lambda_1 = 1 - 1 = 0$ such that $|E\left[X\left(C_1,\Sigma_{22}\right)\right]| = \bar{\lambda}^{-1}\left|\operatorname{tr}\Gamma_1'\bar{D}\Gamma_1\right| = \left(1/\min l_j\right)^{-1}/2 = \min l_j/2$. Inequality (28) therefore holds with equality.

Next consider $E\left[X\left(C_1, I_{T-1}\right)\right] = \operatorname{tr} C_1'\bar{M}_1C_1/r_1$ with $\bar{M}_1 = \left(M_1' + M_1\right)/2$. We analyze

$$\min_{\substack{C_1 \text{ s.t. } C_1'C_1 = I_{r_1} \\ r_1 \in \{1,2,..,T-1\}}} \left|\operatorname{tr} C_1'\bar{M}_1C_1/r_1\right|.$$

It can be checked easily that $\bar{M}_1$ is negative definite symmetric. We can therefore minimize $-\operatorname{tr}\left(C_1'\bar{M}_1C_1\right)$. It is now useful to choose an orthogonal matrix R with j-th row ρ_j such that $R'R = RR' = I$ and $-\bar{M}_1 = R\mathbb{L}R'$ where $\mathbb{L}$ is the diagonal matrix of eigenvalues of $-\bar{M}_1 = \sum_{j=1}^{n} l_j\rho_j\rho_j'$. Then it follows that $-\operatorname{tr}\left(C_1'\bar{M}_1C_1\right) = \sum_{j=1}^{T-1} l_j\rho_j'C_1C_1'\rho_j$. Next note that all the eigenvalues of C_1C_1' are either zero or one such that $0 \le \rho_j'C_1C_1'\rho_j \le 1$. The minimum of $-\operatorname{tr}\left(C_1'\bar{M}_1C_1\right)$ 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 $\operatorname{tr}\left(\bar{D}/r_1\right) = \min l_j$, 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 $\operatorname{tr} C_1'\bar{M}_1C_1 = l_i + \sum_{j\neq i}^{T-1} l_j\left(\rho_j'x\right)^2$. By Parseval's equality $\sum_{j\neq i}^{T-1}\left(\rho_j'x\right)^2 = 1$. Since $l_j \ge l_i$ we can bound $\operatorname{tr}\left(C_1'\bar{M}_1C_1\right) \ge 2l_i$ but then $\operatorname{tr}\left(\bar{D}/2\right) \ge l_i$. This argument can be repeated to more than one orthogonal additions x. It now follows that $E\left[X\left(C_1, I_{T-1}\right)\right] = \operatorname{tr}\left(\bar{D}/r_1\right)$ 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 $\min l_i \le -x'\bar{M}_1x \le \max l_i$ it follows that

$$\min l_i \le -\mathbf{1}'\bar{M}_1\mathbf{1}/(\mathbf{1}'\mathbf{1}) = (T-1)^{-1}$$

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

The last part of the result follows from $\mathbf{1}'\bar{M}_1\mathbf{1}/(\mathbf{1}'\mathbf{1}) \to 0$. ■

E Blundell and Bond's (1998) Estimator and Weight Matrix

Bludell and Bond (1998) suggest a new set of moment restrictions. If $T = 5$, they can be written as

$$E\left[g_i(\beta)\right] = 0$$

where

$$g_i(b) \equiv \begin{bmatrix} y_{i0}\cdot\left((y_{i2}-y_{i1}) - b\,(y_{i1}-y_{i0})\right) \\ y_{i0}\cdot\left((y_{i3}-y_{i2}) - b\,(y_{i2}-y_{i1})\right) \\ y_{i1}\cdot\left((y_{i3}-y_{i2}) - b\,(y_{i2}-y_{i1})\right) \\ y_{i0}\cdot\left((y_{i4}-y_{i3}) - b\,(y_{i3}-y_{i2})\right) \\ y_{i1}\cdot\left((y_{i4}-y_{i3}) - b\,(y_{i3}-y_{i2})\right) \\ y_{i2}\cdot\left((y_{i4}-y_{i3}) - b\,(y_{i3}-y_{i2})\right) \\ y_{i0}\cdot\left((y_{i5}-y_{i4}) - b\,(y_{i4}-y_{i3})\right) \\ y_{i1}\cdot\left((y_{i5}-y_{i4}) - b\,(y_{i4}-y_{i3})\right) \\ y_{i2}\cdot\left((y_{i5}-y_{i4}) - b\,(y_{i4}-y_{i3})\right) \\ y_{i3}\cdot\left((y_{i5}-y_{i4}) - b\,(y_{i4}-y_{i3})\right) \\ (y_{i1}-y_{i0})\cdot(y_{i2}-by_{i1}) \\ (y_{i2}-y_{i1})\cdot(y_{i3}-by_{i2}) \\ (y_{i3}-y_{i2})\cdot(y_{i4}-by_{i3}) \\ (y_{i4}-y_{i3})\cdot(y_{i5}-by_{i4}) \end{bmatrix}$$

They suggest a GMM estimation:

$$\min_b \left(\sum_{i=1}^{n} g_i(b)\right)' A^{-1} \left(\sum_{i=1}^{n} g_i(b)\right)$$

We examine properties of Blundell and Bond's moment restriction for β near unity. We consider four methods of computing A, which in principle is a consistent estimator of $E\left[q_i(\beta)\, q_i(\beta)'\right]$:

1. We can use $\widehat{b}_{LIML}$ as our consistent estimator and use
$$A_1 = \frac{1}{n}\sum_{i=1}^{n} q_i\left(\widehat{b}_{LIML}\right) q_i\left(\widehat{b}_{LIML}\right)'$$
This gives us a GMM estimator that minimizes $\left(\sum_{i=1}^{n} q_i(b)\right)' A_1^{-1} \left(\sum_{i=1}^{n} q_i(b)\right)$. We call it $\widehat{b}_{BB1}$.

2. We can compute
$$A_2 = \frac{1}{n}\sum_{i=1}^{n} q_i\left(\widehat{b}_{BB1}\right) q_i\left(\widehat{b}_{BB1}\right)'$$
and obtain a GMM estimator that minimizes $\left(\sum_{i=1}^{n} q_i(b)\right)' A_2^{-1} \left(\sum_{i=1}^{n} q_i(b)\right)$. We call it $\widehat{b}_{BB2}$.

3. We can compute
$$A_1 = \frac{1}{n}\sum_{i=1}^{n} Z_i' Z_i$$
where
$$Z_i = \begin{bmatrix} y_{i,0} & 0 & 0 & & & \cdots & & & & 0 \\ 0 & y_{i,0} & y_{i,1} & & & & & & & \\ & & & \ddots & & & & & & \\ & & & & y_{i,0} & y_{i,1} & \cdots & y_{i,T-2} & & \\ \vdots & & & & & & & & \Delta y_{i1} & \\ & & & & & & & & & \Delta y_{i,2} \\ & & & & & & & & & & \ddots \\ 0 & & & & \cdots & & & & 0 & & \Delta y_{i,T-1} \end{bmatrix}$$
and obtain a GMM estimator that minimizes $\left(\sum_{i=1}^{n} q_i(b)\right)' A_3^{-1} \left(\sum_{i=1}^{n} q_i(b)\right)$. We call it $\widehat{b}_{BB3}$. This is one of the estimators considered by Blundell and Bond (1998) in thier Monte Carlo.

4. We can compute
$$A_4 = \frac{1}{n}\sum_{i=1}^{n} q_i\left(\widehat{b}_{BB3}\right) q_i\left(\widehat{b}_{BB3}\right)'$$
and obtain a GMM estimator that minimizes $\left(\sum_{i=1}^{n} q_i(b)\right)' A_4^{-1} \left(\sum_{i=1}^{n} q_i(b)\right)$. We call it $\widehat{b}_{BB4}$. Again, this is one of the estimators considered by Blundell and Bond (1998) in thier Monte Carlo.

F Second Order Theory for Finitely Iterated Long Difference Estimator

We examine second order bias of finitely iterated 2SLS. For this purpose, we consider 2SLS

$$b = \left[\left(\sum_{i=1}^{n} x_i \widehat{z}_i'\right)\left(\sum_{i=1}^{n} \widehat{z}_i \widehat{z}_i'\right)^{-1}\left(\sum_{i=1}^{n} \widehat{z}_i x_i\right)\right]^{-1}\left(\sum_{i=1}^{n} x_i \widehat{z}_i'\right)\left(\sum_{i=1}^{n} \widehat{z}_i \widehat{z}_i'\right)^{-1}\left(\sum_{i=1}^{n} \widehat{z}_i y_i\right) \tag{29}$$

applied to the single equation

$$y_i = \beta x_i + \varepsilon_i \tag{30}$$

using instrument $\widehat{z}_i = z_i - \frac{1}{\sqrt{n}}\widehat{\theta} w_i$, where $\widehat{\theta} = \sqrt{n}\left(\widehat{\beta} - \beta\right)$. Here, z_i is the "proper" instrument. We assume that

$$\widehat{\theta} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} f_i + \frac{1}{\sqrt{n}} Q_n + o_p\left(\frac{1}{\sqrt{n}}\right) \tag{31}$$

where f_i is i.i.d. and has mean zero, and $Q_n = O_p(1)$. It can be seen that $\frac{E[Q_n]}{n}$ is equal to the second order bias of $\widehat{\beta}$ under our assumption (31). If ε_i in (30) is symmetrically distributed given z_i, then the second order bias of b is equal to $\frac{1}{n}$ times

$$\begin{aligned}
&\frac{(K-2)\,\sigma_{u\varepsilon}}{\lambda'\Lambda^{-1}\lambda} - \frac{\lambda'\Lambda^{-1}\varphi}{\lambda'\Lambda^{-1}\lambda} E[Q_n] - \frac{\lambda'\Lambda^{-1}E[f_i w_i \varepsilon_i]}{\lambda'\Lambda^{-1}\lambda} \\
&- \frac{E[f_i z_i x_i]'\,\Lambda^{-1}\varphi}{\lambda'\Lambda^{-1}\lambda} - \frac{\phi'\Lambda^{-1}E[f_i z_i \varepsilon_i]}{\lambda'\Lambda^{-1}\lambda} \\
&+ \frac{\lambda'\Lambda^{-1}E[f_i z_i z_i']\,\Lambda^{-1}\varphi}{\lambda'\Lambda^{-1}\lambda} + \frac{\lambda'\Lambda^{-1}\Delta\Lambda^{-1}E[f_i z_i \varepsilon_i]}{\lambda'\Lambda^{-1}\lambda} - E\left[f_i^2\right]\frac{\lambda'\Lambda^{-1}\Delta\Lambda^{-1}\varphi}{\lambda'\Lambda^{-1}\lambda} \\
&+ 2\frac{\lambda'\Lambda^{-1}\varphi}{\left(\lambda'\Lambda^{-1}\lambda\right)^2} E[f_i z_i x_i]'\,\Lambda^{-1}\lambda \\
&+ 2\frac{\lambda'\Lambda^{-1}E[f_i z_i \varepsilon_i]}{\left(\lambda'\Lambda^{-1}\lambda\right)^2}\phi'\Lambda^{-1}\lambda - 2E\left[f_i^2\right]\frac{\lambda'\Lambda^{-1}\varphi}{\left(\lambda'\Lambda^{-1}\lambda\right)^2}\phi'\Lambda^{-1}\lambda \\
&- \frac{\lambda'\Lambda^{-1}\varphi}{\left(\lambda'\Lambda^{-1}\lambda\right)^2}\lambda'\Lambda^{-1}E[f_i z_i z_i']\,\Lambda^{-1}\lambda \\
&- \frac{\lambda'\Lambda^{-1}E[f_i z_i \varepsilon_i]}{\left(\lambda'\Lambda^{-1}\lambda\right)^2}\lambda'\Lambda^{-1}\Delta\Lambda^{-1}\lambda + E\left[f_i^2\right]\frac{\lambda'\Lambda^{-1}\varphi}{\left(\lambda'\Lambda^{-1}\lambda\right)^2}\lambda'\Lambda^{-1}\Delta\Lambda^{-1}\lambda.
\end{aligned} \tag{32}$$

where $\lambda = E[z_i x_i]$, $\Lambda = E[z_i z_i']$, $\phi = E[w_i x_i]$, $\Delta = E[w_i z_i' + z_i w_i']$, and $\varphi = E[w_i \varepsilon_i]$.

Using (32), we can characterize the second order bias of iterated 2SLS applied to the long difference equation using a LIML like estimator as the initial estimator. For this purpose, we need to have the second order bias of the LIML like estimator. In Appendix G, we present a second order bias of the LIML like estimator. In fact, based on 5000 runs, we found in our Monte Carlo experiments that the biases of $\widehat{b}_{LIML,1}$ and $\widehat{b}_{LIML,2}$ are smaller than predicted by the second order theory. In Table 7, we compare the actual performance of the long difference based estimators with the second order theory.

It is sometimes of interest to construct a consistent estimator for the asymptotic variance. Although such exercise may appear to be related only to first order asymptotics, a consistent estimator of the asymptotic variance could be useful in practice for refinement of confidence interval as well: Pivoted bootstrap as considered by Hall and Horowitz (1996) require such consistent estimator for second order refinement. In Appendix H, we present a first order asymptotic result as well as a consistent estimator for the asymptotic variance.

Proof of (32). We first present an expansion for 2SLS using instrument $\widehat{z}_i = z_i - \frac{1}{\sqrt{n}}\widehat{\theta} w_i$. We have

$$\sqrt{n}\,(b - \beta) = \frac{\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{z}_i x_i\right)'\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{z}_i\widehat{z}_i'\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{z}_i\varepsilon_i\right)}{\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{z}_i x_i\right)'\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{z}_i\widehat{z}_i'\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{z}_i x_i\right)} \tag{33}$$

Write

$$\frac{1}{n}\sum_{i=1}^{n} z_i x_i = \lambda + \frac{1}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(z_i x_i - \lambda)\right),$$
$$\frac{1}{n}\sum_{i=1}^{n} z_i z_i' = \Lambda + \frac{1}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(z_i z_i' - \Lambda)\right).$$

Recalling that

$$\widehat{\theta} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} f_i + \frac{1}{\sqrt{n}}Q_n + o_p\left(\frac{1}{\sqrt{n}}\right),$$

we can derive that

$$\frac{1}{n}\sum_{i=1}^{n}\left(z_i - \frac{1}{\sqrt{n}}\widehat{\theta}w_i\right)x_i = \lambda + \frac{1}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(z_i x_i - \lambda)\right) - \frac{1}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f_i\right)\phi + o_p\left(\frac{1}{\sqrt{n}}\right),$$

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\left(z_i - \frac{1}{\sqrt{n}}\widehat{\theta}w_i\right)\left(z_i - \frac{1}{\sqrt{n}}\widehat{\theta}w_i\right)' = \Lambda &+ \frac{1}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(z_i z_i' - \Lambda)\right)\\ &- \frac{1}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f_i\right)\Delta + o_p\left(\frac{1}{\sqrt{n}}\right),\end{aligned}$$

and

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(z_i - \frac{1}{\sqrt{n}}\widehat{\theta}w_i\right)\varepsilon_i = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i\varepsilon_i &- \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f_i\right)\varphi - \frac{1}{\sqrt{n}}Q_n\varphi\\ &- \frac{1}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f_i\right)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(w_i\varepsilon_i - \varphi)\right) + o_p\left(\frac{1}{\sqrt{n}}\right).\end{aligned}$$

Here, ϕ and Δ are defined in Theorem 32. Using arguments similar to the derivation of (27), we obtain

$$\begin{aligned}
&\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{z}_i x_i\right)'\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{z}_i\widehat{z}_i'\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{z}_i\varepsilon_i\right)\\
&=\lambda'\Lambda^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}z_i\varepsilon_i\right)-\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}f_i\right)\lambda'\Lambda^{-1}\varphi\\
&-\frac{1}{\sqrt{n}}\lambda'\Lambda^{-1}\varphi Q_n-\frac{1}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}f_i\right)\lambda'\Lambda^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(v_i\varepsilon_i-\varphi)\right)\\
&+\frac{1}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(z_ix_i-\lambda)\right)'\Lambda^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}z_i\varepsilon_i\right)\\
&-\frac{1}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}f_i\right)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(z_ix_i-\lambda)\right)'\Lambda^{-1}\varphi\\
&-\frac{1}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}f_i\right)\phi'\Lambda^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}z_i\varepsilon_i\right)+\frac{1}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}f_i\right)\phi'\Lambda^{-1}\varphi\\
&-\frac{1}{\sqrt{n}}\lambda'\Lambda^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(z_iz_i'-\Lambda)\right)\Lambda^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}z_i\varepsilon_i\right)\\
&+\frac{1}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}f_i\right)\lambda'\Lambda^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(z_iz_i'-\Lambda)\right)\Lambda^{-1}\varphi\\
&+\frac{1}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}f_i\right)\lambda'\Lambda^{-1}\Delta\Lambda^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}z_i\varepsilon_i\right)-\frac{1}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}f_i\right)^2\lambda'\Lambda^{-1}\Delta\Lambda^{-1}\varphi\\
&+o_p\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}$$

and

$$\begin{aligned}
&\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{z}_i x_i\right)'\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{z}_i\widehat{z}_i'\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{z}_i x_i\right)\\
&=\lambda'\Lambda^{-1}\lambda+\frac{2}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(z_ix_i-\lambda)\right)'\Lambda^{-1}\lambda-\frac{2}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}f_i\right)\phi'\Lambda^{-1}\lambda\\
&-\frac{1}{\sqrt{n}}\lambda'\Lambda^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(z_iz_i'-\Lambda)\right)\Lambda^{-1}\lambda+\frac{1}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}f_i\right)\lambda'\Lambda^{-1}\Delta\Lambda^{-1}\lambda+o_p\left(\frac{1}{\sqrt{n}}\right)
\end{aligned}$$

Therefore, we may conclude that

$$\begin{aligned}
\sqrt{n}(b-\beta)&=\frac{\lambda'\Lambda^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}z_i\varepsilon_i\right)}{\lambda'\Lambda^{-1}\lambda}-\frac{\lambda'\Lambda^{-1}\varphi}{\lambda'\Lambda^{-1}\lambda}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}f_i\right)\\
&+\frac{1}{\sqrt{n}}B_1+\frac{1}{\sqrt{n}}B_2+o_p\left(\frac{1}{\sqrt{n}}\right), \qquad (34)
\end{aligned}$$

where

$$B_1 = \frac{\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n (z_i x_i - \lambda)\right)' \Lambda^{-1} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n z_i \varepsilon_i\right)}{\lambda' \Lambda^{-1} \lambda}$$

$$- \frac{\lambda' \Lambda^{-1} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n (z_i z_i' - \Lambda)\right) \Lambda^{-1} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n z_i \varepsilon_i\right)}{\lambda' \Lambda^{-1} \lambda}$$

$$- 2\frac{\lambda' \Lambda^{-1} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n z_i \varepsilon_i\right)}{\left(\lambda' \Lambda^{-1} \lambda\right)^2} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n (z_i x_i - \lambda)\right)' \Lambda^{-1} \lambda$$

$$+ \frac{\lambda' \Lambda^{-1} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n z_i \varepsilon_i\right)}{\left(\lambda' \Lambda^{-1} \lambda\right)^2} \lambda' \Lambda^{-1} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n (z_i z_i' - \Lambda)\right) \Lambda^{-1} \lambda,$$

and

$$B_2 = -\frac{\lambda' \Lambda^{-1} \varphi}{\lambda' \Lambda^{-1} \lambda} Q_n - \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n f_i\right) \frac{\lambda' \Lambda^{-1} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n (w_i \varepsilon_i - \varphi)\right)}{\lambda' \Lambda^{-1} \lambda}$$

$$- \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n f_i\right) \frac{\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n (z_i x_i - \lambda)\right)' \Lambda^{-1} \varphi}{\lambda' \Lambda^{-1} \lambda}$$

$$- \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n f_i\right) \frac{\phi' \Lambda^{-1} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n z_i \varepsilon_i\right)}{\lambda' \Lambda^{-1} \lambda} + \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n f_i\right) \frac{\phi' \Lambda^{-1} \varphi}{\lambda' \Lambda^{-1} \lambda}$$

$$+ \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n f_i\right) \frac{\lambda' \Lambda^{-1} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n (z_i z_i' - \Lambda)\right) \Lambda^{-1} \varphi}{\lambda' \Lambda^{-1} \lambda}$$

$$+ \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n f_i\right) \frac{\lambda' \Lambda^{-1} \Delta \Lambda^{-1} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n z_i \varepsilon_i\right)}{\lambda' \Lambda^{-1} \lambda} - \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n f_i\right)^2 \frac{\lambda' \Lambda^{-1} \Delta \Lambda^{-1} \varphi}{\lambda' \Lambda^{-1} \lambda}$$

$$+ 2\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n f_i\right) \frac{\lambda' \Lambda^{-1} \varphi}{\left(\lambda' \Lambda^{-1} \lambda\right)^2} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n (z_i x_i - \lambda)\right)' \Lambda^{-1} \lambda$$

$$+ 2\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n f_i\right) \frac{\lambda' \Lambda^{-1} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n z_i \varepsilon_i\right)}{\left(\lambda' \Lambda^{-1} \lambda\right)^2} \phi' \Lambda^{-1} \lambda - 2\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n f_i\right)^2 \frac{\lambda' \Lambda^{-1} \varphi}{\left(\lambda' \Lambda^{-1} \lambda\right)^2} \phi' \Lambda^{-1} \lambda$$

$$- \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n f_i\right) \frac{\lambda' \Lambda^{-1} \varphi}{\left(\lambda' \Lambda^{-1} \lambda\right)^2} \lambda' \Lambda^{-1} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n (z_i z_i' - \Lambda)\right) \Lambda^{-1} \lambda$$

$$- \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n f_i\right) \frac{\lambda' \Lambda^{-1} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n z_i \varepsilon_i\right)}{\left(\lambda' \Lambda^{-1} \lambda\right)^2} \lambda' \Lambda^{-1} \Delta \Lambda^{-1} \lambda$$

$$+ \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n f_i\right)^2 \frac{\lambda' \Lambda^{-1} \varphi}{\left(\lambda' \Lambda^{-1} \lambda\right)^2} \lambda' \Lambda^{-1} \Delta \Lambda^{-1} \lambda.$$

The first two terms on the right side of (34) capture the standard first order asymptotics of the plug in estimator, which establishes Lemma 15. Obviously, they have mean equal to zero. The third term $\frac{1}{\sqrt{n}} B_1$ is the standard second order expansion term when $\widehat{\theta} = 0$, i.e., when the proper instrument is known exactly. Therefore, under conditional symmetry of ε_i, it can be shown that

$$E\left[B_1\right] = \frac{\left(K-2\right)\sigma_{u\varepsilon}}{\lambda'\Lambda^{-1}\lambda}. \tag{35}$$

The third term $\frac{1}{\sqrt{n}}B_2$ is the correction to the second order expansion to accommodate the plug-in nature of the estimation. It is not difficult to see that

$$\begin{aligned}
E\left[B_2\right] = & -\frac{\lambda'\Lambda^{-1}\varphi}{\lambda'\Lambda^{-1}\lambda}E\left[Q_n\right] - \frac{\lambda'\Lambda^{-1}E\left[f_i w_i \varepsilon_i\right]}{\lambda'\Lambda^{-1}\lambda} \\
& - \frac{E\left[f_i z_i x_i\right]'\Lambda^{-1}\varphi}{\lambda'\Lambda^{-1}\lambda} - \frac{\phi'\Lambda^{-1}E\left[f_i z_i \varepsilon_i\right]}{\lambda'\Lambda^{-1}\lambda} \\
& + \frac{\lambda'\Lambda^{-1}E\left[f_i z_i z_i'\right]\Lambda^{-1}\varphi}{\lambda'\Lambda^{-1}\lambda} + \frac{\lambda'\Lambda^{-1}\Delta\Lambda^{-1}E\left[f_i z_i \varepsilon_i\right]}{\lambda'\Lambda^{-1}\lambda} - E\left[f_i^2\right]\frac{\lambda'\Lambda^{-1}\Delta\Lambda^{-1}\varphi}{\lambda'\Lambda^{-1}\lambda} \\
& + 2\frac{\lambda'\Lambda^{-1}\varphi}{\left(\lambda'\Lambda^{-1}\lambda\right)^2}E\left[f_i z_i x_i\right]'\Lambda^{-1}\lambda \\
& + 2\frac{\lambda'\Lambda^{-1}E\left[f_i z_i \varepsilon_i\right]}{\left(\lambda'\Lambda^{-1}\lambda\right)^2}\phi'\Lambda^{-1}\lambda - 2E\left[f_i^2\right]\frac{\lambda'\Lambda^{-1}\varphi}{\left(\lambda'\Lambda^{-1}\lambda\right)^2}\phi'\Lambda^{-1}\lambda \\
& - \frac{\lambda'\Lambda^{-1}\varphi}{\left(\lambda'\Lambda^{-1}\lambda\right)^2}\lambda'\Lambda^{-1}E\left[f_i z_i z_i'\right]\Lambda^{-1}\lambda \\
& - \frac{\lambda'\Lambda^{-1}E\left[f_i z_i \varepsilon_i\right]}{\left(\lambda'\Lambda^{-1}\lambda\right)^2}\lambda'\Lambda^{-1}\Delta\Lambda^{-1}\lambda + E\left[f_i^2\right]\frac{\lambda'\Lambda^{-1}\varphi}{\left(\lambda'\Lambda^{-1}\lambda\right)^2}\lambda'\Lambda^{-1}\Delta\Lambda^{-1}\lambda.
\end{aligned} \tag{36}$$

Using (34), (35), and (36), we can obtain the desired conclusion. ■

G Second Order Bias of $\widehat{b}_{LIML}$

Our $\widehat{b}_{LIML}$ modifies Arellano and Bover's estimator. It is given by

$$\sqrt{n}\left(b-\beta\right) = \frac{\frac{1}{\sqrt{n}}\sum_{t=1}^{T-1}\left(x_t^{*\prime}P_t\varepsilon_t^* - \kappa_t x_t^{*\prime}\varepsilon_t^*\right)}{\frac{1}{n}\sum_{t=1}^{T-1}\left(x_t^{*\prime}P_t x_t^* - \kappa_t x_t^{*\prime}x_t^*\right)}, \tag{37}$$

where

$$\kappa_t = \min_c \frac{\left(\frac{1}{n}\sum_{i=1}^n z_{it}\left(y_{it}^* - x_{it}^* c\right)\right)'\left(\frac{1}{n}\sum_{i=1}^n z_{it}z_{it}'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n z_{it}\left(y_{it}^* - x_{it}^* c\right)\right)}{\frac{1}{n}\sum_{i=1}^n\left(y_{it}^* - x_{it}^* c\right)^2}$$

We make the second order expansion of $\sqrt{n}\left(b-\beta\right)$. We make a digression to the discussion of single equation model.[13]

G.1 Characterization of Second Order Bias of LIML

Consider a simple simultaneous equations model

$$y_i = \beta x_i + \varepsilon_i, \quad x_i = z_i'\pi + u_i$$

[13] The digression mostly confirms the usual higher order analysis of LIML readily available in the literature. The only reason we consider such analysis is because all the analysis we found in the literature are conditional analysis given instruments: They all assume that the instruments are nonstochastic. Our purpose is to make a marginal second order analysis, which is more natural in the dynamic panel model context.

and examine LIML b that solves

$$\min_c \frac{e(c)' Pe(c)}{e(c)' e(c)} = \min_c \frac{\left(\frac{1}{n}\sum_{i=1}^n z_i (y_i - x_i c)\right)' \left(\frac{1}{n}\sum_{i=1}^n z_i z_i'\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^n z_i (y_i - x_i c)\right)}{\frac{1}{n}\sum_{i=1}^n (y_i - x_i c)^2}$$

where

$$e(c) = y - xc$$

Here, the first order condition is given by

$$\frac{-2x' Pe(b)}{e(b)' e(b)} - \frac{e(b)' Pe(b)}{(e(b)' e(b))^2} (-2x' e(b)) = 0$$

or

$$G_n(b) = 0,$$

where

$$G_n(b) = \left(\frac{1}{n} x' Pe(b)\right) \left(\frac{1}{n} e(b)' e(b)\right) - \left(\frac{1}{n} x' e(b)\right) \left(\frac{1}{n} e(b)' Pe(b)\right).$$

Note that

$$\begin{aligned} \frac{\partial G_n(b)}{\partial b} &= \left(-\frac{1}{n} x' Px\right) \left(\frac{1}{n} e(b)' e(b)\right) + \left(\frac{1}{n} x' Pe(b)\right) \left(-2\frac{1}{n} x' e(b)\right) \\ &\quad - \left(-\frac{1}{n} x' x\right) \left(\frac{1}{n} e(b)' Pe(b)\right) - \left(\frac{1}{n} x' e(b)\right) \left(-2\frac{1}{n} x' Pe(b)\right) \end{aligned}$$

and

$$\begin{aligned} \frac{\partial^2 G_n(b)}{\partial b^2} &= \left(-\frac{1}{n} x' Px\right) \left(-2\frac{1}{n} x' e(b)\right) + \left(-\frac{1}{n} x' Px\right) \left(-2\frac{1}{n} x' e(b)\right) + \left(\frac{1}{n} x' Pe(b)\right) \left(2\frac{1}{n} x' x\right) \\ &\quad - \left(-\frac{1}{n} x' x\right) \left(-2\frac{1}{n} x' Pe(b)\right) - \left(-\frac{1}{n} x' x\right) \left(-2\frac{1}{n} x' Pe(b)\right) - \left(\frac{1}{n} x' e(b)\right) \left(2\frac{1}{n} x' Px\right). \end{aligned}$$

We now expand $G_n(\beta)$, $\frac{\partial G_n(\beta)}{\partial b}$, and $\frac{\partial^2 G_n(\beta)}{\partial b^2}$ using $\sqrt{n}$-consistency of b:

$$0 = G_n(\beta) + \frac{1}{\sqrt{n}} \frac{\partial G_n(\beta)}{\partial b} \left(\sqrt{n}(b - \beta)\right) + \frac{1}{n} \frac{\partial^2 G_n(\beta)}{\partial b^2} \left(\sqrt{n}(b - \beta)\right)^2 + o_p\left(\frac{1}{n}\right). \qquad (38)$$

First, note that

$$\begin{aligned} G_n(\beta) &= \left(\frac{1}{n} x' P\varepsilon\right) \left(\frac{1}{n} \varepsilon' \varepsilon\right) - \left(\frac{1}{n} x' \varepsilon\right) \left(\frac{1}{n} \varepsilon' P\varepsilon\right) \\ &= \left(\frac{1}{n} \sum_{i=1}^n \varepsilon_i^2\right) \left(\frac{1}{n} \sum_{i=1}^n z_i x_i\right)' \left(\frac{1}{n} \sum_{i=1}^n z_i z_i'\right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n z_i \varepsilon_i\right) \\ &\quad - \left(\frac{1}{n} \sum_{i=1}^n x_i \varepsilon_i\right) \left(\frac{1}{n} \sum_{i=1}^n z_i \varepsilon_i\right)' \left(\frac{1}{n} \sum_{i=1}^n z_i z_i'\right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n z_i \varepsilon_i\right). \end{aligned}$$

Because

$$\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2=\sigma_\varepsilon^2+\frac{1}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\varepsilon_i^2-\sigma_\varepsilon^2\right)\right),$$

$$\frac{1}{n}\sum_{i=1}^{n}z_i x_i=\lambda+\frac{1}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(z_i x_i-\lambda\right)\right),$$

$$\frac{1}{n}\sum_{i=1}^{n}x_i\varepsilon_i=\sigma_{x\varepsilon}+\frac{1}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(x_i\varepsilon_i-\sigma_{x\varepsilon}\right)\right),$$

$$\left(\frac{1}{n}\sum_{i=1}^{n}z_i z_i'\right)^{-1}=\Lambda^{-1}-\frac{1}{\sqrt{n}}\Lambda^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(z_i z_i'-\Lambda\right)\right)\Lambda^{-1}+o_p\left(\frac{1}{\sqrt{n}}\right),$$

where $\lambda=E[z_i x_i]$, and $\Lambda=E[z_i z_i']$. Therefore, we have

$$G_n(\beta)=\frac{1}{\sqrt{n}}\Phi+\frac{1}{n}\Gamma+o_p\left(\frac{1}{n}\right)\tag{39}$$

where

$$\Phi=\sigma_\varepsilon^2\lambda'\Lambda^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}z_i\varepsilon_i\right)$$

and

$$\begin{aligned}\Gamma=&\ \sigma_\varepsilon^2\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(z_i x_i-\lambda\right)\right)'\Lambda^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}z_i\varepsilon_i\right)\\&-\sigma_\varepsilon^2\lambda'\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(z_i z_i'-\Lambda\right)\right)\Lambda^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}z_i\varepsilon_i\right)\\&+\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\varepsilon_i^2-\sigma_\varepsilon^2\right)\right)\lambda'\Lambda^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}z_i\varepsilon_i\right)\\&-\sigma_{x\varepsilon}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}z_i\varepsilon_i\right)'\Lambda^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}z_i\varepsilon_i\right).\end{aligned}$$

Now, note that

$$\begin{aligned}\frac{\partial G_n(\beta)}{\partial b}&=-\left(\frac{1}{n}x'Px\right)\left(\frac{1}{n}\varepsilon'\varepsilon\right)+\left(\frac{1}{n}x'x\right)\left(\frac{1}{n}\varepsilon'P\varepsilon\right)\\&=-\left(\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2\right)\left(\frac{1}{n}\sum_{i=1}^{n}z_i x_i\right)'\left(\frac{1}{n}\sum_{i=1}^{n}z_i z_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}z_i x_i\right)\\&\quad+\left(\frac{1}{n}\sum_{i=1}^{n}x_i^2\right)\left(\frac{1}{n}\sum_{i=1}^{n}z_i\varepsilon_i\right)'\left(\frac{1}{n}\sum_{i=1}^{n}z_i z_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}z_i\varepsilon_i\right)\\&=\Upsilon+\frac{1}{\sqrt{n}}\Xi+o_p\left(\frac{1}{\sqrt{n}}\right),\end{aligned}\tag{40}$$

where

$$\Upsilon=-\sigma_\varepsilon^2\lambda'\Lambda^{-1}\lambda$$

and

$$\Xi = -\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\varepsilon_i^2 - \sigma_\varepsilon^2\right)\right)\lambda'\Lambda^{-1}\lambda - 2\sigma_\varepsilon^2\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(z_i x_i - \lambda\right)\right)'\Lambda^{-1}\lambda$$
$$+\sigma_\varepsilon^2\lambda'\Lambda^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(z_i z_i' - \Lambda\right)\right)\Lambda^{-1}\lambda.$$

Finally, note that

$$\begin{aligned}\frac{\partial^2 G_n(\beta)}{\partial b^2} &= 2\left(\frac{1}{n}x'Px\right)\left(\frac{1}{n}x'\varepsilon\right) - 2\left(\frac{1}{n}x'P\varepsilon\right)\left(\frac{1}{n}x'x\right)\\ &= 2\Psi + o_p(1).\end{aligned} \tag{41}$$

where

$$\Psi = \sigma_{x\varepsilon}\lambda'\Lambda^{-1}\lambda.$$

Combining (38), (39), (40), and (41), we obtain

$$0 = \frac{1}{\sqrt{n}}\Phi + \frac{1}{n}\Gamma + \frac{1}{\sqrt{n}}\left(\Upsilon + \frac{1}{\sqrt{n}}\Xi\right)\sqrt{n}(b-\beta) + \frac{1}{n}\Psi\left(\sqrt{n}(b-\beta)\right)^2 + o_p\left(\frac{1}{n}\right),$$

from which we obtain

$$\sqrt{n}(b-\beta) = -\frac{1}{\Upsilon}\Phi + \frac{1}{\sqrt{n}}\left(-\frac{1}{\Upsilon}\Gamma + \frac{1}{\Upsilon^2}\Phi\Xi - \frac{\Psi}{\Upsilon^3}\Phi^2\right) + o_p\left(\frac{1}{\sqrt{n}}\right).$$

Note that Φ has a mean equal to zero. Therefore, under symmetry, the second order bias of b is given by

$$E\left[\frac{1}{n}\left(-\frac{1}{\Upsilon}\Gamma + \frac{1}{\Upsilon^2}\Phi\Xi - \frac{\Psi}{\Upsilon^3}\Phi^2\right)\right] = \frac{-\sigma_{\varepsilon v}}{\lambda'\Lambda^{-1}\lambda},$$

which is qualitatively of the same form as Rothenberg's mean.

G.2 Higher Order Analysis of the "Eigenvalue"

Let

$$\kappa = \frac{e(b)'Pe(b)}{e(b)'e(b)}$$

Getting back to the first order condition

$$0 = x'Pe(b) - \frac{e(b)'Pe(b)}{e(b)'e(b)}x'e(b) = x'Py - \kappa x'y - (x'Px - \kappa x'x)b,$$

we can write

$$b = \frac{x'Py - \kappa x'y}{x'Px - \kappa x'x},$$

the usual expression.

Note that

$$\kappa = \frac{\frac{1}{n}\varepsilon'P\varepsilon - 2\frac{1}{n}(b-\beta)\varepsilon'Px + \frac{1}{n}(b-\beta)^2 x'Px}{\frac{1}{n}\varepsilon'\varepsilon - 2\frac{1}{n}(b-\beta)\varepsilon'x + \frac{1}{n}(b-\beta)^2 x'x}$$

The numerator and the denominator may be rewritten as

$$\frac{1}{n}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}z_i\varepsilon_i\right)'\Lambda^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}z_i\varepsilon_i\right)-2\frac{1}{n}\left(\sqrt{n}\left(b-\beta\right)\right)\lambda'\Lambda^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}z_i\varepsilon_i\right)$$
$$+\frac{1}{n}\left(\sqrt{n}\left(b-\beta\right)\right)^2\lambda'\Lambda^{-1}\lambda+o_p\left(\frac{1}{n}\right)$$
$$=\frac{1}{n}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}z_i\varepsilon_i\right)'\Lambda^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}z_i\varepsilon_i\right)-\frac{1}{n}\frac{\left(\lambda'\Lambda^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}z_i\varepsilon_i\right)\right)^2}{\lambda'\Lambda^{-1}\lambda}+o_p\left(\frac{1}{n}\right),$$

and

$$\frac{1}{n}\varepsilon'\varepsilon-2\frac{1}{n}\left(b-\beta\right)\varepsilon'x+\frac{1}{n}\left(b-\beta\right)^2x'x=\sigma_\varepsilon^2+o_p\left(1\right).$$

We may therefore write

$$\kappa=\frac{1}{n\sigma_\varepsilon^2}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}z_i\varepsilon_i\right)'\Lambda^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}z_i\varepsilon_i\right)-\frac{1}{n\sigma_\varepsilon^2}\frac{\left(\lambda'\Lambda^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}z_i\varepsilon_i\right)\right)^2}{\lambda'\Lambda^{-1}\lambda}+o_p\left(\frac{1}{n}\right).$$

G.3 Application to Dynamic Panel Model

We now adopt obvious notations, and make a second order analysis of the right side of (37). First, note that

$$\frac{1}{\sqrt{n}}\left(x_t^{*\prime}P_t\varepsilon_t^*-\kappa_t x_t^{*\prime}\varepsilon_t^*\right)$$
$$=\left(\frac{1}{n}\sum_{i=1}^{n}z_{it}x_{it}^*\right)'\left(\frac{1}{n}\sum_{i=1}^{n}z_{it}z_{it}'\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}z_{it}\varepsilon_{it}^*\right)-\kappa_t\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}x_{it}^*\varepsilon_{it}^*\right)$$
$$=\lambda_t'\Lambda_t^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}z_{it}\varepsilon_{it}^*\right)+\frac{1}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(z_{it}x_{it}^*-\lambda_t\right)\right)'\Lambda_t^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}z_{it}\varepsilon_{it}^*\right)$$
$$-\frac{1}{\sqrt{n}}\lambda_t'\Lambda_t^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(z_{it}z_{it}'-\Lambda_t\right)\right)\Lambda_t^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}z_{it}\varepsilon_{it}^*\right)$$
$$-\frac{1}{\sqrt{n}\sigma_{\varepsilon,t}^2}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}z_{it}\varepsilon_{it}^*\right)'\Lambda_t^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}z_{it}\varepsilon_{it}^*\right)\sigma_{u\varepsilon,t}$$
$$+\frac{1}{\sqrt{n}\sigma_{\varepsilon,t}^2}\frac{\left(\lambda_t'\Lambda_t^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}z_{it}\varepsilon_{it}^*\right)\right)^2}{\lambda_t'\Lambda_t^{-1}\lambda_t}\sigma_{u\varepsilon,t}+o_p\left(\frac{1}{\sqrt{n}}\right)$$

and

$$\frac{1}{n}\left(x_t^{*\prime}P_tx_t^*-\kappa_t x_t^{*\prime}x_t^*\right)$$
$$=\left(\frac{1}{n}\sum_{i=1}^{n}z_{it}x_{it}^*\right)'\left(\frac{1}{n}\sum_{i=1}^{n}z_{it}z_{it}'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}z_{it}x_{it}^*\right)-\kappa_t\left(\frac{1}{n}\sum_{i=1}^{n}\left(x_{it}^*\right)^2\right)$$
$$=\lambda_t'\Lambda_t^{-1}\lambda_t+\frac{2}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(z_{it}x_{it}^*-\lambda_t\right)\right)'\Lambda_t^{-1}\lambda_t$$
$$-\frac{1}{\sqrt{n}}\lambda_t'\Lambda_t^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(z_{it}z_{it}'-\Lambda_t\right)\right)\Lambda_t^{-1}\lambda_t+o_p\left(\frac{1}{\sqrt{n}}\right).$$

It therefore follows that

$$
\begin{aligned}
\sqrt{n}\,(b-\beta) =& \frac{\sum_{t=1}^{T-1} \lambda_t' \Lambda_t^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{it}\varepsilon_{it}^{*} \right)}{\sum_{t=1}^{T-1} \lambda_t' \Lambda_t^{-1} \lambda_t} \\
&+ \frac{1}{\sqrt{n}} \frac{\sum_{t=1}^{T-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(z_{it} x_{it}^{*} - \lambda_t \right) \right)' \Lambda_t^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{it}\varepsilon_{it}^{*} \right)}{\sum_{t=1}^{T-1} \lambda_t' \Lambda_t^{-1} \lambda_t} \\
&- \frac{1}{\sqrt{n}} \frac{\sum_{t=1}^{T-1} \lambda_t' \Lambda_t^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(z_{it} z_{it}' - \Lambda_t \right) \right) \Lambda_t^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{it}\varepsilon_{it}^{*} \right)}{\sum_{t=1}^{T-1} \lambda_t' \Lambda_t^{-1} \lambda_t} \\
&- \frac{1}{\sqrt{n}} \frac{\sum_{t=1}^{T-1} \frac{\sigma_{u\varepsilon,t}}{\sigma_{\varepsilon,t}^{2}} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{it}\varepsilon_{it}^{*} \right)' \Lambda_t^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{it}\varepsilon_{it}^{*} \right)}{\sum_{t=1}^{T-1} \lambda_t' \Lambda_t^{-1} \lambda_t} \\
&+ \frac{1}{\sqrt{n}} \frac{\sum_{t=1}^{T-1} \frac{\sigma_{u\varepsilon,t}}{\sigma_{\varepsilon,t}^{2}} \frac{\left(\lambda_t' \Lambda_t^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{it}\varepsilon_{it}^{*} \right) \right)^2}{\lambda_t' \Lambda_t^{-1} \lambda_t}}{\sum_{t=1}^{T-1} \lambda_t' \Lambda_t^{-1} \lambda_t} \\
&- \frac{2}{\sqrt{n}} \frac{\sum_{t=1}^{T-1} \lambda_t' \Lambda_t^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{it}\varepsilon_{it}^{*} \right)}{\left(\sum_{t=1}^{T-1} \lambda_t' \Lambda_t^{-1} \lambda_t \right)^2} \left(\sum_{t=1}^{T-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(z_{it} x_{it}^{*} - \lambda_t \right) \right)' \Lambda_t^{-1} \lambda_t \right) \\
&+ \frac{1}{\sqrt{n}} \frac{\sum_{t=1}^{T-1} \lambda_t' \Lambda_t^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_{it}\varepsilon_{it}^{*} \right)}{\left(\sum_{t=1}^{T-1} \lambda_t' \Lambda_t^{-1} \lambda_t \right)^2} \left(\sum_{t=1}^{T-1} \lambda_t' \Lambda_t^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(z_{it} z_{it}' - \Lambda_t \right) \right) \Lambda_t^{-1} \lambda_t \right) \\
&+ o_p\left(\frac{1}{\sqrt{n}} \right).
\end{aligned}
$$

Therefore, under symmetry, the second order bias of the LIML-like estimator is given by

$$
\begin{aligned}
& \frac{1}{n} \frac{\sum_{t=1}^{T-1} \sigma_{u\varepsilon,t}}{\sum_{t=1}^{T-1} \lambda_t' \Lambda_t^{-1} \lambda_t} \\
&- \frac{2}{n} \frac{\sum_{t=1}^{T-1} \sum_{s=1}^{T-1} \lambda_t' \Lambda_t^{-1} E\left[\left(z_{it}\varepsilon_{it}^{*} \right) \left(z_{is} x_{is}^{*} - \lambda_s \right)' \right] \Lambda_s^{-1} \lambda_s}{\left(\sum_{t=1}^{T-1} \lambda_t' \Lambda_t^{-1} \lambda_t \right)^2} \\
&+ \frac{1}{n} \frac{\sum_{t=1}^{T-1} \sum_{s=1}^{T-1} \lambda_t' \Lambda_t^{-1} E\left[\left(z_{it}\varepsilon_{it}^{*} \right) \lambda_s' \Lambda_s^{-1} \left(z_{is} z_{is}' - \Lambda_s \right) \right] \Lambda_s^{-1} \lambda_s}{\left(\sum_{t=1}^{T-1} \lambda_t' \Lambda_t^{-1} \lambda_t \right)^2} \\
&+ o_p\left(\frac{1}{\sqrt{n}} \right).
\end{aligned}
$$

H First Order Asymptotic Theory for Finitely Iterated Long Difference Estimator

Lemma 15 *Let b denote the 2SLS in (29). We have*

$$
\sqrt{n}\,(b-\beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{\lambda' \Lambda^{-1}}{\lambda' \Lambda^{-1} \lambda} z_i \varepsilon_i - \frac{\lambda' \Lambda^{-1} \varphi}{\lambda' \Lambda^{-1} \lambda} f_i \right) + o_p(1) \to N\left(0, \frac{\lambda' \Lambda^{-1} \Sigma \Lambda^{-1} \lambda}{\left(\lambda' \Lambda^{-1} \lambda \right)^2} \right),
$$

where

$$
\Sigma \equiv E\left[\left(z_i \varepsilon_i - f_i \varphi \right) \left(z_i \varepsilon_i - f_i \varphi \right)' \right].
$$

Proof. See Appendix F. ■

Lemma 15 can be used to establish the influence function of iterated 2SLS estimators $\widehat{b}_{LIML,1}, \ldots, \widehat{b}_{LIML,4}$ applied to the long difference. We first note that the influence function of $\widehat{b}_{LIML}$ is given by

$$f_{LIML,i} = \frac{\sum_{t=1}^{T-1} \lambda_t' \Lambda_t^{-1} z_{it} \varepsilon_{it}^*}{\sum_{t=1}^{T-1} \lambda_t' \Lambda_t^{-1} \lambda_t}$$

where $\lambda_t = E\left[z_{it} x_{it}^*\right]$, and $\Lambda_t = E\left[z_{it} z_{it}'\right]$. We also note that $y_i = y_{iT} - y_{i1}$, $x_i = y_{iT-1} - y_{i0}$, and $w_i = (0, y_{iT-2}, \ldots, y_{i1})'$. This is because we use the instrument of the form $\left(y_{i0}, y_{iT-1} - \widehat{\beta} y_{iT-2}, \ldots, y_{i2} - \widehat{\beta} y_{i1}\right)$ at each iteration, where $\widehat{\beta}$ is some preliminary estimator of β. By Lemma 15, the influence function of $\widehat{b}_{LIML,1}$ is equal to

$$\frac{\lambda' \Lambda^{-1}}{\lambda' \Lambda^{-1} \lambda} z_i \varepsilon_i - \frac{\lambda' \Lambda^{-1} \varphi}{\lambda' \Lambda^{-1} \lambda} f_{LIML,i} \tag{42}$$

Using Lemma 15 again, we can see that the influence function of b_2 is equal to

$$\frac{\lambda' \Lambda^{-1}}{\lambda' \Lambda^{-1} \lambda} z_i \varepsilon_i - \frac{\lambda' \Lambda^{-1} \varphi}{\lambda' \Lambda^{-1} \lambda} \left(\frac{\lambda' \Lambda^{-1}}{\lambda' \Lambda^{-1} \lambda} z_i \varepsilon_i - \frac{\lambda' \Lambda^{-1} \varphi}{\lambda' \Lambda^{-1} \lambda} f_{LIML,i} \right). \tag{43}$$

Likewise, we can see that the influence functions of b_3 and b_4 are equal to

$$\frac{\lambda' \Lambda^{-1}}{\lambda' \Lambda^{-1} \lambda} z_i \varepsilon_i - \frac{\lambda' \Lambda^{-1} \varphi}{\lambda' \Lambda^{-1} \lambda} \left[\frac{\lambda' \Lambda^{-1}}{\lambda' \Lambda^{-1} \lambda} z_i \varepsilon_i - \frac{\lambda' \Lambda^{-1} \varphi}{\lambda' \Lambda^{-1} \lambda} \left(\frac{\lambda' \Lambda^{-1}}{\lambda' \Lambda^{-1} \lambda} z_i \varepsilon_i - \frac{\lambda' \Lambda^{-1} \varphi}{\lambda' \Lambda^{-1} \lambda} f_{LIML,i} \right) \right] \tag{44}$$

and

$$\frac{\lambda' \Lambda^{-1}}{\lambda' \Lambda^{-1} \lambda} z_i \varepsilon_i - \frac{\lambda' \Lambda^{-1} \varphi}{\lambda' \Lambda^{-1} \lambda} \left\{ \frac{\lambda' \Lambda^{-1}}{\lambda' \Lambda^{-1} \lambda} z_i \varepsilon_i - \frac{\lambda' \Lambda^{-1} \varphi}{\lambda' \Lambda^{-1} \lambda} \left[\frac{\lambda' \Lambda^{-1}}{\lambda' \Lambda^{-1} \lambda} z_i \varepsilon_i - \frac{\lambda' \Lambda^{-1} \varphi}{\lambda' \Lambda^{-1} \lambda} \left(\frac{\lambda' \Lambda^{-1}}{\lambda' \Lambda^{-1} \lambda} z_i \varepsilon_i - \frac{\lambda' \Lambda^{-1} \varphi}{\lambda' \Lambda^{-1} \lambda} f_{LIML,i} \right) \right] \right\} \tag{45}$$

Using (42) - (45), we can easily construct consistent estimators of asymptotic variances of $\sqrt{n}\left(\widehat{b}_{LIML,1} - \beta\right)$, $\sqrt{n}\left(\widehat{b}_{LIML,2} - \beta\right)$, $\sqrt{n}\left(\widehat{b}_{LIML,3} - \beta\right)$, and $\sqrt{n}\left(\widehat{b}_{LIML,4} - \beta\right)$. Suppose that $\widehat{\Lambda}$, $\widehat{\lambda}$, and $\widehat{\varphi}$ are consistent estimators of Λ, λ, and φ. Likewise, let $\widehat{\Lambda}_t$, and $\widehat{\lambda}_t$ denote some consistent estimators of Λ_t, and λ_t. For example,

$$\widehat{\Lambda} = \frac{1}{n} \sum_{i=1}^{n} \widehat{z}_i \widehat{z}_i', \qquad \widehat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} \widehat{z}_i x_i, \qquad \widehat{\varphi} = \frac{1}{n} \sum_{i=1}^{n} w_i \left(y_i - b x_i\right)$$

$$\widehat{z}_i = \left(y_{i0}, y_{iT-1} - \widehat{\beta} y_{iT-2}, \ldots, y_{i2} - \widehat{\beta} y_{i1}\right),$$

$$\widehat{\Lambda}_t = \frac{1}{n} \sum_{i=1}^{n} z_{it} z_{it}', \qquad \widehat{\lambda}_t = \frac{1}{n} \sum_{i=1}^{n} z_{it} x_{it}^*.$$

where $\widehat{\beta}$ is any $\sqrt{n}$-consistent estimator of β. Also, let

$$\widehat{\varepsilon}_i = y_i - \widehat{\beta} x_i, \qquad \widehat{f}_{LIML,i} = \frac{\sum_{t=1}^{T-1} \widehat{\lambda}_t' \widehat{\Lambda}_t^{-1} z_{it} \left(y_{it}^* - \widehat{\beta} x_{it}^*\right)}{\sum_{t=1}^{T-1} \widehat{\lambda}_t' \widehat{\Lambda}_t^{-1} \widehat{\lambda}_t}$$

From (42) - (45), it then follows that

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\widehat{\lambda}'\widehat{\Lambda}^{-1}}{\widehat{\lambda}'\widehat{\Lambda}^{-1}\widehat{\lambda}}\widehat{z}_i\widehat{\varepsilon}_i-\frac{\widehat{\lambda}'\widehat{\Lambda}^{-1}\widehat{\varphi}}{\widehat{\lambda}'\widehat{\Lambda}^{-1}\widehat{\lambda}}\widehat{f}_{LIML,i}\right)^2,$$

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\widehat{\lambda}'\widehat{\Lambda}^{-1}}{\widehat{\lambda}'\widehat{\Lambda}^{-1}\widehat{\lambda}}\widehat{z}_i\widehat{\varepsilon}_i-\frac{\widehat{\lambda}'\widehat{\Lambda}^{-1}\widehat{\varphi}}{\widehat{\lambda}'\widehat{\Lambda}^{-1}\widehat{\lambda}}\left(\frac{\widehat{\lambda}'\widehat{\Lambda}^{-1}}{\widehat{\lambda}'\widehat{\Lambda}^{-1}\widehat{\lambda}}\widehat{z}_i\widehat{\varepsilon}_i-\frac{\widehat{\lambda}'\widehat{\Lambda}^{-1}\widehat{\varphi}}{\widehat{\lambda}'\widehat{\Lambda}^{-1}\widehat{\lambda}}\widehat{f}_{LIML,i}\right)\right)^2,$$

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\widehat{\lambda}'\widehat{\Lambda}^{-1}}{\widehat{\lambda}'\widehat{\Lambda}^{-1}\widehat{\lambda}}\widehat{z}_i\widehat{\varepsilon}_i-\frac{\widehat{\lambda}'\widehat{\Lambda}^{-1}\widehat{\varphi}}{\widehat{\lambda}'\widehat{\Lambda}^{-1}\widehat{\lambda}}\left[\frac{\widehat{\lambda}'\widehat{\Lambda}^{-1}}{\widehat{\lambda}'\widehat{\Lambda}^{-1}\widehat{\lambda}}\widehat{z}_i\widehat{\varepsilon}_i-\frac{\widehat{\lambda}'\widehat{\Lambda}^{-1}\widehat{\varphi}}{\widehat{\lambda}'\widehat{\Lambda}^{-1}\widehat{\lambda}}\left(\frac{\widehat{\lambda}'\widehat{\Lambda}^{-1}}{\widehat{\lambda}'\widehat{\Lambda}^{-1}\widehat{\lambda}}\widehat{z}_i\widehat{\varepsilon}_i-\frac{\widehat{\lambda}'\widehat{\Lambda}^{-1}\widehat{\varphi}}{\widehat{\lambda}'\widehat{\Lambda}^{-1}\widehat{\lambda}}\widehat{f}_{LIML,i}\right)\right]\right)^2,$$

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\widehat{\lambda}'\widehat{\Lambda}^{-1}}{\widehat{\lambda}'\widehat{\Lambda}^{-1}\widehat{\lambda}}\widehat{z}_i\widehat{\varepsilon}_i-\frac{\widehat{\lambda}'\widehat{\Lambda}^{-1}\widehat{\varphi}}{\widehat{\lambda}'\widehat{\Lambda}^{-1}\widehat{\lambda}}\left\{\frac{\widehat{\lambda}'\widehat{\Lambda}^{-1}}{\widehat{\lambda}'\widehat{\Lambda}^{-1}\widehat{\lambda}}\widehat{z}_i\widehat{\varepsilon}_i-\frac{\widehat{\lambda}'\widehat{\Lambda}^{-1}\widehat{\varphi}}{\widehat{\lambda}'\widehat{\Lambda}^{-1}\widehat{\lambda}}\left[\frac{\widehat{\lambda}'\widehat{\Lambda}^{-1}}{\widehat{\lambda}'\widehat{\Lambda}^{-1}\widehat{\lambda}}\widehat{z}_i\widehat{\varepsilon}_i-\frac{\widehat{\lambda}'\widehat{\Lambda}^{-1}\widehat{\varphi}}{\widehat{\lambda}'\widehat{\Lambda}^{-1}\widehat{\lambda}}\left(\frac{\widehat{\lambda}'\widehat{\Lambda}^{-1}}{\widehat{\lambda}'\widehat{\Lambda}^{-1}\widehat{\lambda}}\widehat{z}_i\widehat{\varepsilon}_i-\frac{\widehat{\lambda}'\widehat{\Lambda}^{-1}\widehat{\varphi}}{\widehat{\lambda}'\widehat{\Lambda}^{-1}\widehat{\lambda}}\widehat{f}_{LIML,i}\right)\right]\right\}\right)^2$$

are consistent estimators of asymptotic variances of $\sqrt{n}\left(\widehat{b}_{LIML,1}-\beta\right)$, $\sqrt{n}\left(\widehat{b}_{LIML,2}-\beta\right)$, $\sqrt{n}\left(\widehat{b}_{LIML,3}-\beta\right)$, and $\sqrt{n}\left(\widehat{b}_{LIML,4}-\beta\right)$.

I Approximation of CUE

We examine an easier method of calculating an estimator that is equivalent to CUE up to the second order adapting Rothenberg's (1984) argument, who was concerned about properties of linearized version of MLE. We basically argue that two Newton iterations suffice for second order bias removal. The CUE b_{CUE} solves

$$\min_c L(c)=\min_c g(c)' G(c)^{-1} g(c),$$

where

$$g(c)\equiv\frac{1}{n}\sum_{i=1}^{n}\delta_i(c),\quad G(c)\equiv\frac{1}{n}\sum_{i=1}^{n}\delta_i(c)\,\delta_i(c)'$$

Let b denote the minimizer, and let $L_j(c)\equiv\frac{\partial^j L(c)}{\partial c^j}$. We consider an iterated version of CUE. Suppose that we have a $\sqrt{n}$-consistent estimator b_0. Such estimator can be easily found by the usual GMM estimation method. Note that we would have $b_0-b_{CUE}=O_p\left(\frac{1}{\sqrt{n}}\right)$. Assume that

$$L_2\left(\widehat{b}\right)=O_p(1),\quad L_3\left(\widehat{b}\right)=O_p(1)$$

for any $\sqrt{n}$−consistent estimator $\widehat{b}$. (This condition is expected to be satisfied for most estimators.) Let

$$b_{r+1}\equiv b_r-\frac{L_1(b_r)}{L_2(b_r)}.$$

Expanding around b_{CUE}, and noting that $L_1(b_{CUE}) = 0$, we can obtain

$$\begin{aligned}
b_1 - b_{CUE} &= b_0 - b_{CUE} - \frac{L_1(b_0)}{L_2(b_0)} \\
&= b_0 - b_{CUE} \\
&\quad - \frac{L_2(b_{CUE}) \cdot (b_0 - b_{CUE}) + \frac{1}{2} L_3(b_{CUE}) \cdot (b_0 - b_{CUE})^2 + o_p\left((b_0 - b_{CUE})^2\right)}{L_2(b_{CUE}) + L_3(b_{CUE}) \cdot (b_0 - b_{CUE}) + o_p(b_0 - b_{CUE})} \\
&= b_0 - b_{CUE} \\
&\quad - \left(L_2(b_{CUE}) \cdot (b_0 - b_{CUE}) + \frac{1}{2} L_3(b_{CUE}) \cdot (b_0 - b_{CUE})^2\right) \\
&\qquad \times \left(\frac{1}{L_2(b_{CUE})} - \frac{L_3(b_{CUE}) \cdot (b_0 - b_{CUE})}{L_2(b_{CUE})^2}\right) \\
&\quad + o_p\left((b_0 - b_{CUE})^2\right) \\
&= \frac{L_3(b_{CUE})}{2L_2(b_{CUE})} \cdot (b_0 - b_{CUE})^2 + o_p\left((b_0 - b_{CUE})^2\right) \\
&= \frac{L_3(b_{CUE})}{2L_2(b_{CUE})} \cdot (b_0 - b_{CUE})^2 + o_p\left(\frac{1}{n}\right).
\end{aligned}$$

It follows that

$$b_1 - b_{CUE} = O_p\left(\frac{1}{n}\right).$$

We can similarly show that

$$b_2 - b_{CUE} = \frac{L_3(b_{CUE})}{2L_2(b_{CUE})} \cdot (b_1 - b_{CUE})^2 + o_p\left((b_1 - b_{CUE})^2\right) = O_p\left(\frac{1}{n^2}\right).$$

or

$$\sqrt{n}\,(b_2 - b_{CUE}) = O_p\left(n^{-3/2}\right).$$

This implies that b_2 has very similar properties as b_{CUE}: Its (approximate) mean and variance up to $O\left(n^{-1}\right)$ coincide with those of b_{CUE}.

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] ANDREWS, D.W.K., 1992, "Generic Uniform Convergence", *Econometric Theory,* 8, pp. 241-257.

[5] 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.

[6] ALVAREZ, J. AND M. ARELLANO, 1998, "The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators", *unpublished manuscript.*

[7] ANDERSON, T.W. AND C. HSIAO, 1982, "Formulation and Estimation of Dynamic Models Using Panel Data", *Journal of Econometrics* 18, pp. 47 - 82.

[8] 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.

[9] ARELLANO, M. AND O. BOVER, 1995, "Another Look at the Instrumental Variable Estimation of Error-Component Models", *Journal of Econometrics* 68, pp. 29 - 51.

[10] BEKKER, P., 1994, "Alternative Approximations to the Distributions of Instrumental Variable Estimators", *Econometrica* 62, pp. 657 - 681.

[11] BLUNDELL, R. AND S. BOND, 1998, "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models", *Journal of Econometrics* 87, pp. 115 - 143.

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

[13] 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.*

[14] 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.

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

[16] DONALD, S. AND W. NEWEY, 1998, "Choosing the Number of Instruments", *unpublished manuscript.*

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

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

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

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

[21] HAHN, J. AND J. HAUSMAN, 2000, "A New Specification Test for the Validity of Instrumental Variables", *unpublished manuscript.*

[22] HAHN, J. AND G. KUERSTEINER, 2000, "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both n and T are Large", *unpublished manuscript.*

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

[24] 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.

[25] 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.

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

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

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

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

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

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

[32] 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.

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

[34] 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.

[35] 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.

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

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

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

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

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

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

[42] 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.

[43] 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.

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

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

Table 1: Performance of $\widehat{b}_{Nagar}$ and $\widehat{b}_{LIML}$

			%bias			RMSE		
T	n	β	$\widehat{b}_{GMM}$	$\widehat{b}_{Nagar}$	$\widehat{b}_{LIML}$	$\widehat{b}_{GMM}$	$\widehat{b}_{Nagar}$	$\widehat{b}_{LIML}$
5	100	0.1	-16	3	-3	0.081	0.084	0.082
10	100	0.1	-14	1	-1	0.046	0.046	0.045
5	500	0.1	-4	0	-1	0.036	0.036	0.036
10	500	0.1	-3	0	-1	0.020	0.020	0.020
5	100	0.3	-9	1	-3	0.099	0.103	0.099
10	100	0.3	-7	0	-1	0.053	0.051	0.050
5	500	0.3	-2	0	-1	0.044	0.044	0.044
10	500	0.3	-2	0	0	0.023	0.025	0.023
5	100	0.5	-10	1	-3	0.132	0.140	0.130
10	100	0.5	-7	0	-1	0.064	0.059	0.058
5	500	0.5	-2	0	-1	0.057	0.057	0.057
10	500	0.5	-2	0	0	0.027	0.026	0.026
5	100	0.8	-28	-129	-15	0.321	102.156	0.327
10	100	0.8	-14	0	-5	0.136	0.128	0.109
5	500	0.8	-7	1	-3	0.130	0.141	0.127
10	500	0.8	-3	0	-1	0.050	0.044	0.044
5	100	0.9	-51	-70	-41	0.555	26.984	0.604
10	100	0.9	-24	-4	-15	0.250	4.712	0.229
5	500	0.9	-20	-41	-10	0.278	46.933	0.277
10	500	0.9	-9	0	-2	0.102	0.087	0.080

Table 2: Performance of Second Order Theory in Predicting Properties of $\widehat{\beta}_{GMM}$

T	n	β	Actual Bias	Actual %Bias	Second Order Bias	Second Order %Bias
5	100	0.1	-0.016	-16.00	-0.018	-17.71
10	100	0.1	-0.014	-14.26	-0.016	-15.78
5	500	0.1	-0.004	-3.72	-0.004	-3.54
10	500	0.1	-0.003	-3.20	-0.003	-3.16
5	100	0.3	-0.028	-9.23	-0.032	-10.60
10	100	0.3	-0.021	-7.11	-0.024	-8.13
5	500	0.3	-0.006	-2.08	-0.006	-2.12
10	500	0.3	-0.005	-1.58	-0.005	-1.63
5	100	0.5	-0.052	-10.32	-0.060	-12.09
10	100	0.5	-0.034	-6.78	-0.040	-8.00
5	500	0.5	-0.011	-2.29	-0.012	-2.42
10	500	0.5	-0.008	-1.51	-0.008	-1.60
5	100	0.8	-0.224	-28.06	-0.302	-37.81
10	100	0.8	-0.108	-13.53	-0.152	-18.98
5	500	0.8	-0.056	-7.02	-0.060	-7.56
10	500	0.8	-0.027	-3.44	-0.030	-3.80
5	100	0.9	-0.455	-50.56	-1.068	-118.64
10	100	0.9	-0.220	-24.47	-0.474	-52.66
5	500	0.9	-0.184	-20.48	-0.214	-23.73
10	500	0.9	-0.078	-8.64	-0.095	-10.53

Table 3: Performance of $\widehat{b}_{BC2}$

T	n	β	%bias$\left(\widehat{b}_{GMM}\right)$	%bias$\left(\widehat{b}_{BC2}\right)$	RMSE$\left(\widehat{b}_{GMM}\right)$	RMSE$\left(\widehat{b}_{BC2}\right)$
5	100	0.1	-14.96	0.25	0.08	0.08
10	100	0.1	-14.06	-0.77	0.05	0.05
5	500	0.1	-3.68	-0.38	0.04	0.04
10	500	0.1	-3.15	-0.16	0.02	0.02
5	100	0.3	-8.86	-0.47	0.10	0.10
10	100	0.3	-7.06	-0.66	0.05	0.05
5	500	0.3	-2.03	-0.16	0.04	0.04
10	500	0.3	-1.58	-0.10	0.02	0.02
5	100	0.5	-10.05	-1.14	0.13	0.13
10	100	0.5	-6.76	-0.93	0.06	0.06
5	500	0.5	-2.25	-0.15	0.06	0.06
10	500	0.5	-1.53	-0.11	0.03	0.03
5	100	0.8	-27.65	-11.33	0.32	0.34
10	100	0.8	-13.45	-4.55	0.14	0.11
5	500	0.8	-6.98	-0.72	0.13	0.13
10	500	0.8	-3.48	-0.37	0.05	0.04
5	100	0.9	-50.22	-42.10	0.55	0.78
10	100	0.9	-24.27	-15.82	0.25	0.23
5	500	0.9	-20.50	-6.23	0.28	0.30
10	500	0.9	-8.74	-2.02	0.10	0.08

Table 4: Performance of Iterated Long Difference Estimator

	$\widehat{b}_{2SLS,1}$	$\widehat{b}_{2SLS,2}$	$\widehat{b}_{2SLS,3}$	$\widehat{b}_{2SLS,4}$
Bias	-0.0813	-0.0471	-0.0235	-0.0033
%Bias	-9.0316	-5.2316	-2.6072	-0.3644
RMSE	0.3802	0.2863	0.2479	0.2536
	$\widehat{b}_{GMM,1}$	$\widehat{b}_{GMM,2}$	$\widehat{b}_{GMM,3}$	$\widehat{b}_{GMM,4}$
Bias	-0.0770	-0.0374	0.0006	0.0104
%Bias	-8.5505	-4.1599	0.0622	1.1670
RMSE	0.1699	0.1964	0.2546	0.2851
	$\widehat{b}_{LIML,1}$	$\widehat{b}_{LIML,2}$	$\widehat{b}_{LIML,3}$	$\widehat{b}_{LIML,4}$
Bias	-0.0878	-0.0475	-0.0186	0.0074
%Bias	-9.7571	-5.2756	-2.0698	0.8251
RMSE	0.2458	0.2391	0.2292	0.2638

Table 5: Comparison with Blundell and Bond's (1998) Estimator: $\beta = .9$, $T = 5$, $N = 100$

	$\hat{b}_{BB1}$	$\hat{b}_{BB2}$	$\hat{b}_{BB3}$	$\hat{b}_{BB4}$	$\hat{b}_{LIML,1}$	$\hat{b}_{LIML,2}$	$\hat{b}_{LIML,3}$
Mean % Bias	-33.8148	-29.4131	4.7432	4.2551	-9.7571	-5.2755	-2.0697
Median % Bias	-31.1881	-25.9085	5.9111	5.6280	-15.3878	-9.0639	-6.9573
RMSE	0.4796	0.4257	0.0823	0.0882	0.2458	0.2591	0.2292

Table 6: Sensitivity of Blundell and Bond's (1998) Estimator: $\beta = .9$, $T = 5$, $N = 100$

$\beta_F = .5$	$\hat{b}_{BB1}$	$\hat{b}_{BB2}$	$\hat{b}_{BB3}$	$\hat{b}_{BB4}$	$\hat{b}_{LIML,1}$	$\hat{b}_{LIML,2}$	$\hat{b}_{LIML,3}$
Mean % Bias	8.9525	14.4790	20.9971	21.5154	0.0252	0.1691	0.2334
Median % Bias	9.5207	15.4609	21.1202	21.6144	-0.2168	-0.2214	-0.2469
RMSE	0.0994	0.1400	0.1899	0.1944	0.0570	0.0611	0.0630
$\beta_F = 0$	$\hat{b}_{BB1}$	$\hat{b}_{BB2}$	$\hat{b}_{BB3}$	$\hat{b}_{BB4}$	$\hat{b}_{LIML,1}$	$\hat{b}_{LIML,2}$	$\hat{b}_{LIML,3}$
Mean % Bias	10.8819	17.3840	24.8534	25.4517	0.0429	0.1455	0.1860
Median % Bias	11.4178	18.2542	24.9990	25.5079	-0.1621	-0.1890	-0.2168
RMSE	0.1156	0.1654	0.2246	0.2299	0.0521	0.0543	0.0555

Table 7: Performance of Iterated Long Difference Estimator for $T = 5$

	$N = 100$	$\hat{b}_{LIML,1}$	$\hat{b}_{LIML,2}$	$\hat{b}_{LIML,3}$	$\hat{b}_{LIML,4}$
$\beta = .75$	Actual Mean % bias	1.2977	4.6584	5.3703	7.6702
	Actual Median % bias	-3.0867	-0.4467	-0.0800	-0.4800
	2nd order Mean % bias	-.1358	2.8043	4.6720	6.5872
	RMSE	0.1806	0.2278	0.2465	0.2857
$\beta = .80$	Actual Mean %bias	-0.1119	2.5878	4.2732	6.3443
	Actual Median % bias	-5.7250	-2.3438	-1.4188	-1.4000
	2nd order Mean % bias	-.4020	4.6019	7.3205	9.8596
	RMSE	0.2128	0.2452	0.2523	0.3032
$\beta = .85$	Actual Mean %bias	-3.8994	-0.5921	1.6201	3.6981
	Actual Median % bias	-10.1176	-5.4235	-4.0059	-3.5471
	2nd order Mean % bias	-.8477	9.2416	14.3129	18.0912
	RMSE	0.2333	0.2494	0.2532	0.2848
$\beta = .90$	Actual Mean %bias	-9.7571	-5.2756	-2.0698	0.8251
	Actual Median % bias	-15.3889	-9.0667	-6.9556	-5.6444
	2nd order Mean % bias	-1.7413	25.3502	40.2274	49.3254
	RMSE	0.2453	0.2391	0.2292	0.2638
$\beta = .95$	Actual Mean %bias	-15.2028	-9.5738	-6.0855	-2.8321
	Actual Median % bias	-19.6368	-12.4895	-9.6105	-8.0684
	2nd order Mean % bias	-4.4189	132.5028	229.9208	298.9023
	RMSE	0.2518	0.2191	0.2124	0.2397
	$N = 200$	$\hat{b}_{LIML,1}$	$\hat{b}_{LIML,2}$	$\hat{b}_{LIML,3}$	$\hat{b}_{LIML,4}$
$\beta = .75$	Actual Mean %bias	1.2054	3.0110	3.8420	5.1421
	Actual Median % bias	-1.6553	-0.2333	-0.1000	-0.4733
	2nd order Mean % bias	-.0679	1.4022	2.3360	3.2936
	RMSE	0.1336	0.1630	0.1906	0.2189
$\beta = .80$	Actual Mean %bias	1.4085	3.7041	4.3488	4.8453
	Actual Median % bias	-3.3125	-1.1813	-0.5938	-1.1500
	2nd order Mean % bias	-.2010	2.3010	3.6602	4.9210
	RMSE	0.1740	0.2071	0.2245	0.2435
$\beta = .85$	Actual Mean %bias	0.0299	1.7783	1.7835	3.6882
	Actual Median % bias	-6.8412	-3.7059	-2.8588	-2.6000
	2nd order Mean % bias	-.4238	4.6208	7.1565	9.0456
	RMSE	0.2239	0.2363	0.2288	0.2513
$\beta = .90$	Actual Mean %bias	-5.8274	-2.7803	-1.3073	0.1639
	Actual Median % bias	-13.1000	-7.9111	-5.9278	-5.2333
	2nd order Mean % bias	-.8706	12.6751	20.1137	24.6627
	RMSE	0.2406	0.2257	0.2252	0.2396
$\beta = .95$	Actual Mean %bias	-13.3638	-8.7034	-6.2646	-4.1416
	Actual Median % bias	-19.3737	-12.3211	-9.5579	-8.0526
	2nd order Mean % bias	-2.2094	66.2514	114.9604	149.4511
	RMSE	0.2515	0.2156	0.1991	0.2020

Table 8: Performance of $\widehat{\beta}_{I2SLS}$ and $\widehat{\beta}_{CUE}$ for $T = 5$

$N = 100$		$\widehat{\beta}_{I2SLS,LD}$	$\widehat{\beta}_{CUE,LD}$
$\beta = 0.75$	Actual Mean % Bias	5.5331	11.5527
	Second Order Mean % Bias	5.4224	7.6105
	Actual Median %Bias	1.3811	7.4700
	RMSE	0.1761	0.2132
	InterQuartile Range	0.2434	0.3067
$\beta = 0.8$	Actual Mean % Bias	4.3037	10.4126
	Second Order Mean % Bias	9.6240	13.0702
	Actual Median % Bias	1.4569	8.6510
	RMSE	0.1727	0.2048
	InterQuartile Range	0.2422	0.3031
$\beta = 0.85$	Actual Mean % Bias	1.9659	7.9833
	Second Order Mean % Bias	20.9080	27.0025
	Actual Median % Bias	0.0656	7.5588
	RMSE	0.1604	0.1947
	InterQuartile Range	0.2270	0.2900
$\beta = 0.9$	Actual Mean % Bias	-0.7710	6.1387
	Second Order Mean % Bias	65.0609	78.5269
	Actual Median % Bias	-2.3467	6.1147
	RMSE	0.1534	0.1803
	InterQuartile Range	0.2115	0.2668
$\beta = 0.95$	Actual Mean % Bias	-3.3676	3.1244
	Second Order Mean % Bias	481.1993	533.4268
	Actual Median % Bias	-4.7764	3.1365
	RMSE	0.1494	0.1655
	InterQuartile Range	0.2002	0.2512

Table 9: Performance of $\widehat{\beta}_{I2SLS}$ and $\widehat{\beta}_{CUE}$ for $T = 5$

$N = 200$		$\widehat{\beta}_{I2SLS,LD}$	$\widehat{\beta}_{CUE,LD}$
$\beta = 0.75$	Actual Mean % Bias	5.9078	8.8638
	Second Order Mean % Bias	2.7112	3.8052
	Actual Median %Bias	1.5982	4.2172
	RMSE	0.1519	0.1704
	InterQuartile Range	0.1896	0.2297
$\beta = 0.8$	Actual Mean % Bias	4.9410	8.3701
	Second Order Mean % Bias	4.8120	6.5351
	Actual Median % Bias	1.8273	5.4765
	RMSE	0.1447	0.1674
	InterQuartile Range	0.1997	0.2468
$\beta = 0.85$	Actual Mean % Bias	2.7966	7.3021
	Second Order Mean % Bias	10.4540	13.5012
	Actual Median % Bias	1.0718	5.8672
	RMSE	0.1373	0.1585
	InterQuartile Range	0.1909	0.2341
$\beta = 0.9$	Actual Mean % Bias	0.8948	5.4221
	Second Order Mean % Bias	32.5304	39.2635
	Actual Median % Bias	-0.0204	5.3657
	RMSE	0.1271	0.1448
	InterQuartile Range	0.1750	0.2101
$\beta = 0.95$	Actual Mean % Bias	-2.0943	2.8482
	Second Order Mean % Bias	240.5997	266.7134
	Actual Median % Bias	-2.5881	2.8984
	RMSE	0.1216	0.1331
	InterQuartile Range	0.1594	0.1915

Table 10: Performance of $\widehat{\beta}_{CUE,LD}$, $\widehat{\beta}_{CUE2,AS}$, $\widehat{\beta}_{CUE2,LD}$, and $\widehat{\beta}_{CUE2,BB}$ for $T = 5$

$N = 100$		$\widehat{\beta}_{CUE,LD}$	$\widehat{\beta}_{CUE,BB}$	$\widehat{\beta}_{CUE2,AS}$	$\widehat{\beta}_{CUE2,LD}$	$\widehat{\beta}_{CUE2,BB}$
$\beta = .75$	Median % Bias	7.4700	1.2705	6.6814	4.2643	2.0471
	Interquartile Range	0.3067	0.1480	0.2864	0.2911	0.2456
	Mean % Bias	11.5527	0.4852	-296.6631	1250.1149	-136.4730
	RMSE	0.2132	0.1050	152.2249	676.8912	117.9397
$\beta = .8$	Median % Bias	8.6510	1.2629	4.7391	1.6364	0.6595
	Interquartile Range	0.3031	0.1540	0.3206	0.3410	0.3676
	Mean % Bias	10.4126	-0.0913	33.6498	-15.0393	-125.6554
	RMSE	0.2048	0.1092	29.2436	12.6828	74.6934
$\beta = .85$	Median % Bias	7.5588	1.9808	0.9468	-2.2980	-1.0482
	Interquartile Range	0.2900	0.1645	0.4253	1.2817	0.4902
	Mean % Bias	7.9833	0.3824	-100.7981	-161.9267	6.4686
	RMSE	0.1947	0.1225	28.4546	25.6932	23.8489
$\beta = .9$	Median % Bias	6.1147	3.0423	-4.2248	-16.4693	-6.9530
	Interquartile Range	0.2668	0.1637	2.3282	1.5503	2.3169
	Mean % Bias	6.1387	1.2087	-30.2898	-177.0842	495.0171
	RMSE	0.1803	0.1344	24.6733	131.8341	193.9465
$\beta = .95$	Median % Bias	3.1365	3.4897	-17.7102	-129.4765	-21.5058
	Interquartile Range	0.2512	0.1452	2.5936	1.6277	2.5714
	Mean % Bias	3.1244	1.0877	-290.6542	-42.6293	-32.0973
	RMSE	0.1655	0.1347	166.1415	67.0635	98.0361

Table 11: Performance of $\widehat{\beta}_{CUE,LD}$, $\widehat{\beta}_{CUE2,AS}$, $\widehat{\beta}_{CUE2,LD}$, and $\widehat{\beta}_{CUE2,BB}$ for $T = 5$

$N = 200$		$\widehat{\beta}_{CUE,LD}$	$\widehat{\beta}_{CUE,BB}$	$\widehat{\beta}_{CUE2,AS}$	$\widehat{\beta}_{CUE2,LD}$	$\widehat{\beta}_{CUE2,BB}$
$\beta = .75$	Median % Bias	4.2172	0.4242	3.4952	4.0943	1.2644
	Interquartile Range	0.2297	0.1032	0.1855	0.2034	0.1195
	Mean % Bias	8.8638	0.1604	116.0861	29.5421	4.4327
	RMSE	0.1704	0.0719	85.9117	10.6641	4.5595
$\beta = .8$	Median % Bias	5.4765	0.5388	5.8182	5.3421	0.8895
	Interquartile Range	0.2468	0.1063	0.2105	0.2132	0.1472
	Mean % Bias	8.3701	-0.1898	16.9181	-233.6177	-21.1440
	RMSE	0.1674	0.0736	13.7393	127.6513	9.5825
$\beta = .85$	Median % Bias	5.8672	0.6441	5.1660	3.7226	1.0708
	Interquartile Range	0.2341	0.1143	0.2295	0.2347	0.2619
	Mean % Bias	7.3021	-0.7076	688.7913	-50.0828	59.6610
	RMSE	0.1585	0.0779	455.6137	19.8972	66.8390
$\beta = .9$	Median % Bias	5.3657	0.9204	0.9958	-2.8551	-1.0774
	Interquartile Range	0.2101	0.1152	0.3766	1.3425	0.5870
	Mean % Bias	5.4221	-0.3893	479.7193	31.6093	-29.9555
	RMSE	0.1448	0.0913	381.8271	23.2206	42.2422
$\beta = .95$	Median % Bias	2.8984	2.5208	-11.2026	-125.6884	-12.6677
	Interquartile Range	0.1915	0.1099	2.5978	1.5877	2.6203
	Mean % Bias	2.8482	0.9733	-82.2370	-6464.7709	-177.6883
	RMSE	0.1331	0.1044	59.5396	4315.6181	116.8096