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

Jinyong Hahn	Jerry Hausman	Guido Kuersteiner
Brown University	MIT	MIT

January 9, 2001

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 biggest 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 implicitly or explicitly adopted by Chamberlain (1984) and Griliches and Hausman (1986). It has been noted for quite a while that 2SLS 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 bias and mean squared error 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 and contrary to Bekker (1994) 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 and continuous updating GMM estimators in situations that are particularly relevant for the dynamic panel model. We show that standard 2SLS estimators which are asymptotically efficient under first order or standard asymptotic approximations are inadmissible under the alternative asymptotic approximations. We identify a complete class within the class of GMM estimators based on a finite set of instruments or moment conditions.

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. We identify a complete class of GMM estimators and show that a bias minimal estimator within this class can approximately be based on taking long differences of the dynamic panel model. Long differences were introduced by Griliches and Hausman (1986) . Similar problems have been studied by Blundell and Bond (1998) and Moon and Phillips (2000). 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 inadmissible and inference optimally should be based on a smaller than the full set of moment conditions. We also show that it is usually not efficient to focus on original moment conditions. Rather one should consider optimal linear combinations of the moment conditions. Due to the special structure of the panel model the optimal linear combinations are known a priori.

2 Review of the Bias of GMM Estimator

Consider the usual dynamic panel model with fixed effects:

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

It has been common in the literature to consider the case where n is large and T is small. The usual GMM estimator are 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 dramatically simplifies 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 $\operatorname{Var}\left(\varepsilon_{it}\right) / \operatorname{plim} n^{-1} x_t^{*\prime} P_t x_t^*$. 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):

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

Therefore, the GMM estimator $\widehat{b}_{GMM}$ may be understood as a linear combination of the 2SLS estimators $\widehat{b}_{2SLS,1}, \ldots, \widehat{b}_{2SLS,T-1}$. It has long been known that the 2SLS may be subject to substantial finite sample bias. See Nagar (1959), Rothenberg (1983), Bekker (1994), and Donald and Newey (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 Corrected GMM Estimators

In the previous section, we explained the bias of GMM estimator as a result of the biases of the 2SLS estimators. With such understanding, it should be straightforward to apply the standard methods of correcting for biases of 2SLS and eliminate the bias of the GMM estimator itself. Depending on the nature of the higher order asymptotic approximation, we may come up with several strategies of correcting for biases. Below, we discuss two different higher order asymptotic approximations and related methods of bias correction. The first one is the second order Taylor type approximation. Such perspective has been adopted by Nagar (1959), and Rothenberg (1983). The second approximation pretends that the number of parameters increases to infinity as a function of the sample size. Such approximation was originally developed by Bekker (1994), and was adopted by Alvarez and Arellano (1998) and Hahn and Kuersteiner (2000) in dynamic panel context.

3.1 Second Order Biases of GMM Type Estimators

We first present a theory that justifies our second order bias calculation later in this section. Consider a class of estimators solving the minimization problem

$$\min_c g(c)' G(c)^{-1} g(c), \tag{3}$$

where

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

Let b denote the minimizer. First order condition for (3) is given by

$$0 = 2g_1(b)' G(b)^{-1} g(b) - g(b)' G(b)^{-1} G_1(b) G(b)^{-1} g(b), \tag{4}$$

where $g_1(b) \equiv \partial g(b)/\partial b$, and $G_1 \equiv \partial G(b)/\partial b$. By expanding the first order condition, we can obtain the following result:

Theorem 1 *Second order bias of b is equal to*

$$\begin{aligned}
&-\frac{1}{n}\frac{\operatorname{trace}\left(\Lambda^{-1}E\left[\delta_i\frac{\partial\delta_i'}{\partial\beta}\right]\right)}{\lambda_1\Lambda^{-1}\lambda_1} + \frac{1}{n}\frac{\lambda_1'\Lambda^{-1}E\left[\psi_i\psi_i'\Lambda^{-1}\delta_i\right]}{\lambda_1\Lambda^{-1}\lambda_1} + \frac{1}{2n}\frac{\operatorname{trace}\left(\Lambda^{-1}\Lambda_1\Lambda^{-1}E\left[\delta_i\delta_i'\right]\right)}{\lambda_1\Lambda^{-1}\lambda_1} \\
&+2\frac{1}{n}\frac{\lambda_1'\Lambda^{-1}E\left[\delta_i\frac{\partial\delta_i'}{\partial\beta}\right]\Lambda^{-1}\lambda_1}{\left(\lambda_1\Lambda^{-1}\lambda_1\right)^2} - \frac{1}{n}\frac{\lambda_1'\Lambda^{-1}E\left[\delta_i\lambda_1'\Lambda^{-1}\psi_i\psi_i'\right]\Lambda^{-1}\lambda_1}{\left(\lambda_1\Lambda^{-1}\lambda_1\right)^2} \\
&-2\frac{1}{n}\frac{\lambda_1'\Lambda^{-1}E\left[\delta_i\delta_i'\right]\Lambda^{-1}\Lambda_1\Lambda^{-1}\lambda_1}{\left(\lambda_1\Lambda^{-1}\lambda_1\right)^2} + \frac{1}{n}\frac{\lambda_1'\Lambda^{-1}E\left[\delta_i\delta_i'\right]\Lambda^{-1}\lambda_2}{\left(\lambda_1\Lambda^{-1}\lambda_1\right)^2} \\
&-\frac{3}{2n}\frac{\lambda_1'\Lambda^{-1}\lambda_2}{\left(\lambda_1\Lambda^{-1}\lambda_1\right)^3}\lambda_1'\Lambda^{-1}E\left[\delta_i\delta_i'\right]\Lambda^{-1}\lambda_1 + \frac{3}{2n}\frac{\lambda_1'\Lambda^{-1}\Lambda_1\Lambda^{-1}\lambda_1}{\left(\lambda_1\Lambda^{-1}\lambda_1\right)^3}\lambda_1'\Lambda^{-1}E\left[\delta_i\delta_i'\right]\Lambda^{-1}\lambda_1,
\end{aligned} \tag{5}$$

where $G = G(\beta)$, $G_1 = G_1(\beta)$, $G_2 = \partial G_1(\beta)/\partial b$, $G_3 = \partial G_2(\beta)/\partial b$, $\lambda_j = E[g_j]$, and $\Lambda_j = E[G_j]$.

Proof. See Appendix A. ■

Remark 1 *For the particular case where $\psi_i = \delta_i$, i.e. when b is a CUE, the bias formula (5) exactly coincides with Newey and Smith's (2000).*

3.2 Motivation 1: Higher Order Expansion

There are $T-1$ ingredients of the minimum distance estimator $\widehat{b}_{2SLS,1}, \ldots, \widehat{b}_{2SLS,T-1}$. Because all of them are 2SLS, and because 2SLS is known to be biased, it would not be surprising if the resultant minimum distance estimator, i.e., the GMM estimator, is biased. Using Theorem 1, it can be shown that:

Theorem 2 *If the conditional distribution of ε_{it}^* given z_{it} is symmetric, the second order bias of $\widehat{b}_{GMM}$ is equal to*

$$\frac{B_1+B_2+B_3}{n}+o\left(\frac{1}{n}\right), \tag{6}$$

where

$$B_1 \equiv \frac{\sum_{t=1}^{T-1} \operatorname{trace}\left(\Lambda_t^{-1} E\left[\varepsilon_{it}^* x_{it}^* z_{it} z_{it}'\right]\right)}{\sum_{t=1}^{T-1} E\left[z_{it} x_{it}^*\right]' E\left[z_{it} z_{it}'\right]^{-1} E\left[z_{it} x_{it}^*\right]}$$

$$B_2 \equiv -2 \frac{\sum_{t=1}^{T-1} \sum_{s=1}^{T-1} E\left[z_{it} x_{it}^*\right]' E\left[z_{it} z_{it}'\right]^{-1} E\left[\varepsilon_{it}^* x_{is}^* z_{it} z_{is}'\right] E\left[z_{is} z_{is}'\right]^{-1} E\left[z_{is} x_{is}^*\right]}{\left(\sum_{t=1}^{T-1} E\left[z_{it} x_{it}^*\right]' E\left[z_{it} z_{it}'\right]^{-1} E\left[z_{it} x_{it}^*\right]\right)^2}$$

$$B_3 \equiv \frac{\sum_{t=1}^{T-1} \sum_{s=1}^{T-1} E\left[E\left[z_{it} x_{it}^*\right]' E\left[z_{it} z_{it}'\right]^{-1} \varepsilon_{it}^* z_{it} \lambda_s' \Lambda_s^{-1} z_{is} z_{is}' E\left[z_{is} z_{is}'\right]^{-1} E\left[z_{is} x_{is}^*\right]\right]}{\left(\sum_{t=1}^{T-1} E\left[z_{it} x_{it}^*\right]' E\left[z_{it} z_{it}'\right]^{-1} E\left[z_{it} x_{it}^*\right]\right)^2}.$$

Proof. See Appendix B. ■

In Table 1, we compare the actual performance of $\widehat{b}_{GMM}$ and the prediction of its bias based on Theorem 2. Table 1 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 (6).[2] 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 2 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

$$\widehat{b}_{BC1} \equiv \widehat{b}_{GMM} - \frac{1}{n}\left(\widehat{B}_1+\widehat{B}_2+\widehat{B}_3\right) \tag{7}$$

is first order equivalent to $\widehat{b}_{GMM}$, and has second order bias equal to zero. Let

$$\frac{1}{n}\widehat{B}_1 \equiv \frac{\sum_{t=1}^{T-1} \operatorname{trace}\left(\left(\sum_{i=1}^n z_{it} z_{it}'\right)^{-1}\left(\sum_{i=1}^n e_{it}^* x_{it}^* z_{it} z_{it}'\right)\right)}{\sum_{t=1}^{T-1}\left(\sum_{i=1}^n z_{it} x_{it}^*\right)'\left(\sum_{i=1}^n z_{it} z_{it}'\right)^{-1}\left(\sum_{i=1}^n z_{it} x_{it}^*\right)},$$

$$\frac{1}{n}\widehat{B}_2 \equiv -2\frac{\sum_{t=1}^{T-1} \sum_{s=1}^{T-1}\left(\sum_{i=1}^n z_{it} x_{it}^*\right)'\left(\sum_{i=1}^n z_{it} z_{it}'\right)^{-1}\left(\sum_{i=1}^n e_{it}^* x_{is}^* z_{it} z_{is}'\right)\left(\sum_{i=1}^n z_{is} z_{is}'\right)^{-1}\left(\sum_{i=1}^n z_{is} x_{is}^*\right)}{\left(\sum_{t=1}^{T-1}\left(\sum_{i=1}^n z_{it} x_{it}^*\right)'\left(\sum_{i=1}^n z_{it} z_{it}'\right)^{-1}\left(\sum_{i=1}^n z_{it} x_{it}^*\right)\right)^2},$$

$$\frac{1}{n}\widehat{B}_3 \equiv \frac{\sum_{t=1}^{T-1} \sum_{s=1}^{T-1} \widehat{B}_3(t,s)}{\left(\sum_{t=1}^{T-1}\left(\sum_{i=1}^n z_{it} x_{it}^*\right)'\left(\sum_{i=1}^n z_{it} z_{it}'\right)^{-1}\left(\sum_{i=1}^n z_{it} x_{it}^*\right)\right)^2},$$

where

$$\widehat{B}_3(t,s) \equiv \left(\sum_{i=1}^n z_{it} x_{it}^*\right)'\left(\sum_{i=1}^n z_{it} z_{it}'\right)^{-1}$$
$$\left(\sum_{i=1}^n e_{it}^* z_{it}\left(\sum_{i=1}^n z_{is} x_{is}^*\right)'\left(\sum_{i=1}^n z_{is} z_{is}'\right)^{-1} z_{is} z_{is}'\right)\left(\sum_{i=1}^n z_{is} z_{is}'\right)^{-1}\left(\sum_{i=1}^n z_{is} x_{is}^*\right)$$

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

and

$$e_{it}^{*} \equiv y_{it}^{*} - x_{it}^{*}\widehat{b}_{GMM}.$$

Then the $\widehat{B}$s will satisfy the $\sqrt{n}$-consistency requirement, and hence, the estimator (7) 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

$$\widehat{b}_{BC2} \equiv \widehat{b}_{GMM} - \frac{1}{n}\left(\widehat{B}_1 + \widehat{B}_2 + \widehat{\widehat{B}}_3\right) \tag{8}$$

where

$$\frac{1}{n}\widehat{\widehat{B}}_3 \equiv \frac{\sum_{s=t}^{T-1}\sum_{t=1}^{T-1}\widehat{B}_3(t,s)}{\left(\sum_{t=1}^{T-1}\left(\sum_{i=1}^{n} z_{it}x_{it}^{*}\right)'\left(\sum_{i=1}^{n} z_{it}z_{it}'\right)^{-1}\left(\sum_{i=1}^{n} z_{it}x_{it}^{*}\right)\right)^2}.$$

Second order asymptotic theory roughly predicts that $\widehat{b}_{BC2}$ would be relatively free of bias. We examined whether such prediction is reasonably accurate in finite sample by 5000[3] Monte Carlo runs. Table 2 summarizes the properties of $\widehat{b}_{BC2}$. We have seen in Table 1 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 conjecture is verified in Table 2.

3.3 Motivation 2: 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 alternative asymptotics have been adopted by Alvarez and Arellano (1998) and Hahn and Kuersteiner (2000) in analyzing biases of GMM estimators and maximum likelihood estimator for the model (1). 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[4]:

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* $0 < \rho < 1$.[5]

Alvarez and Arellano (1998) show that, under this alternative asymptotic approximation where n and T grow to infinity at the same rate,

$$\sqrt{nT}\left(\widehat{b}_{GMM} - \left(\beta - \frac{1}{n}(1+\beta)\right)\right) \to \mathcal{N}\left(0, 1-\beta^2\right). \tag{9}$$

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

[4] 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.

[5] 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.

By examining the asymptotic distribution (9) under the alternative asymptotics derived by Alvarez and Arellano (1998), we can develop a bias-corrected estimator. This bias-corrected estimator is given by

$$\widetilde{b}_{GMM} \equiv \frac{n}{n-1}\widehat{b}_{GMM} + \frac{1}{n-1}. \tag{10}$$

Combining (9) and (10), we can easily obtain:

Theorem 3 *Suppose that Conditions 1-3 are satisfied. Also suppose that n and T grow to infinity at the same rate. 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 (10) 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 obvious estimator that gets rid of 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.

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

$$\widetilde{b}_{Nagar} \equiv \frac{\sum_t \left(x_t^{*\prime} P_t y_t^* - \lambda_t x_t^{*\prime} M_t y_t^*\right)}{\sum_t \left(x_t^{*\prime} P_t x_t^* - \lambda_t x_t^{*\prime} M_t x_t^*\right)} = \beta + \frac{\sum_t \left(x_t^{*\prime} P_t \varepsilon_t^* - \lambda_t x_t^{*\prime} M_t \varepsilon_t^*\right)}{\sum_t \left(x_t^{*\prime} P_t x_t^* - \lambda_t x_t^{*\prime} M_t x_t^*\right)}. \tag{11}$$

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

Theorem 4 *Suppose that Conditions 1-3 are satisfied. Also suppose that n and T grow to infinity at the same rate. Then, $\sqrt{nT}\left(b_{Nagar} - \beta\right) \to N\left(0, 1-\beta^2\right)$.*

Proof. Lemmas 12, and 13 in Appendix C along with Lemma 2 of Alvarez and Arellano (1998) establish that

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

and

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

from which the conclusion follows. ■

In Table 3, we summarized finite sample properties of $\widetilde{b}_{Nagar}$ and $\widehat{b}_{LIML}$ approximated by 10000 Monte Carlo runs. Here, $\widehat{b}_{LIML}$ is the estimator where λs in (11) are replaced by the corresponding "eigenvalues".

4 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 phenoemenon 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. 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

$$y_{iT} - y_{i1} = \beta\left(y_{iT-1} - y_{i0}\right) + \left(\varepsilon_{iT} - \varepsilon_{i1}\right) \tag{12}$$

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.

4.1 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. I 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:

$$y_4 - y_3 = \beta\left(y_3 - y_2\right) + \varepsilon_4 - \varepsilon_3 \tag{13}$$

For the RHS variables it uses the instrument equation:

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

Now calculate the R2 for equation (13) 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 $\sigma_\varepsilon^2 \frac{-\beta+1}{\beta+1}$. 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:

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

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

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

where $\sigma^2 = \frac{\sigma_\alpha^2}{\sigma_\varepsilon^2}$. 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

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

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

$$-.37408$$

which is much smaller than -19. We therefore conclude that the long difference increases R^2 but decreases the covariance. Further, number of instruments is smaller in the long difference specification so we should expect even smaller bias.

4.2 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 (12) using y_{i0} as instrument. We may then use $\left(y_{i0}, y_{iT-1} - \widehat{b}_{2SLS} y_{iT-2}, \ldots, y_{i2} - \widehat{b}_{2SLS} y_{i1}\right)$ as instrument to the long difference equation (12) 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 $\left(y_{i0}, y_{iT-1} - \widehat{b}_{GMM} y_{iT-2}, \ldots, y_{i2} - \widehat{b}_{GMM} y_{i1}\right)$ as instrument to the long difference equation (12) 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 $\left(y_{i0}, y_{iT-1} - \widehat{b}_{LIML} y_{iT-2}, \ldots, y_{i2} - \widehat{b}_{LIML} y_{i1}\right)$ as

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

instrument to the long difference equation (12) 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.

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 $\widehat{b}_{BB1}, \ldots, \widehat{b}_{BB4}$ 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 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 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 sensitivity of Blundell and Bond's estimator to misspecification, i.e., nonstationary distribution of y_{i0}. Obviously 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.

4.3 Second Order Theory

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

$$b = \left[\left(\sum_{i=1}^n x_i \widehat{z}_i'\right)\left(\sum_{i=1}^n \widehat{z}_i \widehat{z}_i'\right)^{-1}\left(\sum_{i=1}^n \widehat{z}_i x_i\right)\right]^{-1}\left(\sum_{i=1}^n x_i \widehat{z}_i'\right)\left(\sum_{i=1}^n \widehat{z}_i \widehat{z}_i'\right)^{-1}\left(\sum_{i=1}^n \widehat{z}_i y_i\right) \tag{14}$$

applied to the single equation

$$y_i = \beta x_i + \varepsilon_i$$

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

$$\widehat{\theta} = \frac{1}{\sqrt{n}}\sum_{i=1}^n f_i + \frac{1}{\sqrt{n}} Q_n + o_p\left(\frac{1}{\sqrt{n}}\right) \tag{15}$$

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 (15).

Theorem 5 *Let b denote the 2SLS in (14). Under conditional symmetry of ε_i given z_i, the second order* *bias of b is equal to $\frac{1}{n}$ times*

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

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

Proof. See Appendix F. ■

Using Theorem 5, we can characterize the second order bias of iterated 2SLS applied to the long difference equation using LIML like estimator as the initial estimator. For this purpose, we need to have a second order bias of 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.

5 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 such 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 bias and mean squared error 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 such subset of moment restrictions lead to the inference based on "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 the 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.

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

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

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 $\Omega \equiv \lim_n E\left[g_n(\beta_n) g_n(\beta_n)'\right]$. The infeasible GMM of a possibly transformed set of moment conditions $C' g_n(\beta)$ then solves

$$\beta_{2SLS} = \arg\min_\beta g_n(\beta)' C \left(C' \Omega C\right)^+ C' g_n(\beta)$$

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

$$\beta_{2SLS} - \beta_{n0} = \left(f_n' C \left(C'\Omega C\right)^+ C' f_n\right)^{-1} f_n' C \left(C'\Omega C\right)^+ C' g_n(\beta_{n0}). \tag{17}$$

We are now analyzing the behavior of $\beta_{2SLS} - \beta_{n0}$ under local to unity asymptotics. We make the following additional assumptions.

Condition 4 *Let $y_{it} = \alpha_i + \beta_n y_{it-1} + \varepsilon_{it}$ with $\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$, $\alpha_i \sim N(0, \sigma_\alpha^2)$ and $y_{i0} \sim N(\frac{\alpha_i}{1-\beta_n}, \frac{1}{1-\beta_n^2})$ where $\beta_n = \exp(-c/n)$.*

Also note that $\Delta y_{it} = \beta_n^{t-1}\eta_{i0} + \varepsilon_{it} + \frac{c}{\sqrt{n}} \sum_{s=1}^{t-1} \beta_n^{s-1} \varepsilon_{it-s} + o_p(n^{-1})$ where $\eta_{i0} \sim N(0, (\beta_n - 1)^2 / (1 - \beta_n^2))$. We now establish the following Lemma.

Lemma 1 *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$\left[\xi_x', \xi_y'\right] \sim N(0, \Sigma)$ with $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$.with $\Sigma_{11} = \delta I$, $\Sigma_{12} = \delta M_1$ $\Sigma_{22} = \delta M_2$, where $\delta = \frac{\sigma_\alpha^2 \sigma_\varepsilon^2}{c^2}$,

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

and $\Sigma_{12} = \Sigma_{21}'$.

Proof. See Appendix 5. ■

Using Lemma (1) the limiting distribution of $\beta_{2SLS} - \beta_n$ can now be obtained in the next corollary. We define the augmented vectors $\xi_x^\# = \left[0, ..., 0, \xi_x'\right]'$ and $\xi_y^\# = \left[0, ..., 0, \xi_y'\right]'$.

Corollary 1 *Let $\beta_{2SLS} - \beta_n$ be given by (17). If Condition (4) is satisfied then*

$$\beta_{2SLS} - 1 \xrightarrow{d} \frac{\xi_x^{\#\prime} C(C'\Omega C)^+ C'\xi_y^\#}{\xi_x^{\#\prime} C(C'\Omega C)^+ C'\xi_x^\#} = X \tag{18}$$

Unlike the limiting distribution for the standard weak instrument problem, X, as defined in 18, is based on normal vectors that have zero mean. This degeneracy is generated by the presence of the fixed effect in the initial condition, scaled up appropriately to satisfy the stationarity requirement for the process y_{it}. Inspection of the proof shows that the usual concentration parameter appearing in the limit distribution is dominated by a stochastic component related to the fixed effect. This situation seems to be similar to time series models where deterministic trends can dominate the asymptotic distribution.

A problem with analyzing the class of estimators having a limiting distribution X is the fact that $C'\xi_x^\#$ has a potentially degenerate distribution. In order to proceed we therefore have to show first that it is never optimal to choose C such that $C'\xi_x^\#$ is degenerate.

In particular, in the proof of the next theorem we show that it is never optimal to choose W singular where $W = L'C(C'\Omega C)^+ C'L$ with $LL' = \Sigma_{11}$. It then follows by Lemma (14) in the Appendix that $\Omega = \Sigma_{11}$ and $C = C_1$. The latter implies that only moment conditions involving the initial conditions should optimally be picked.

Theorem 6 *Assume Condition 4 holds and* $\beta_{2SLS} - \beta_n$ *is as defined in (17). Then* $E\left[(X^*)^2\right] < E\left[(X)^2\right]$ *where* X *is defined in (18) and* X^* *is obtained by setting* $W^* = L'C_1(C_1'\Sigma_{11}C_1)^+C_1L = I$.

Theorem (6) shows that standard efficient GMM estimators for the dynamic panel model based on exploiting all the available moment conditions are inadmissible under the weak instrument asymptotic approximation. There are two elements that lead to inadmissibility. First, according to Theorem (14) the first order optimal weight matrix Σ_{22} produces an estimator that is dominated in terms of L^2 risk by an estimator based on Σ_{11}. Second, as shown above most of the moment conditions become irrelevant under these asymptotics, the only exception being moment conditions involving initial conditions. This result has been discussed elsewhere by Bond and Blundell (1988) and Hahn (1999). One important consequence of Theorem (6) is that optimal inference for the Panel model is feasible since the matrix Σ_{11} is known up to a scalar which is irrelevant for estimation purposes.

Next we turn to the analysis of bias and mean squared error for the dynamic panel model. We now restrict C_1 to be a $(T-1)\times r$ matrix of full column rank $r \leq T-1$ such that $C_1'C_1 = I$. Restricting $r \leq T-1$ means that moment restrictions can not be used twice which can not be ruled out from Theorem (6). From a practical point of view imposing this restriction is very natural. Since the limit only depends on zero mean normal random vectors we can directly apply the results of Smith (1993).

Theorem 7 *Let* X^* *be as defined in Theorem (6). Let* $\bar{D} = (D + D')/2$ *where* $D = C_1'\Sigma_{12}C_1$. *Then*

$$E[X^*] = \text{trace}\left(\bar{D}/r\right)$$

where $r = rank\,(C)$ *and*

$$E\left[(X^*)^2\right] = \left(2\,\text{trace}(\bar{D}^2) + \left(tr\bar{D}\right)^2\right)/r(r+2) + 2\frac{\left(\frac{r}{2}\right)_{-1}\left(\frac{1}{2}\right)_1}{\left(\frac{r}{2}\right)_1}\,\text{trace}\left(DGD'\right)$$

where $E[X^*]$ *exists for* $r \geq 1$ *and* $E\left[(X^*)^2\right]$ *exists for* $r \geq 3$.

An immediate consequence of Theorem 7 is that both bias and mean squared error are monotonically decreasing in the parameter c. The further β is away from the unit circle the lower both bias and mean squared error are.

We can now consider the problem of choosing an optimal matrix C_1 to minimize bias and L^2 risk. It turns out that an analytical solution for the bias minimal estimator can be found. For the bias term we can write trace $\bar{D}/r = \frac{\delta}{2r}\,\text{trace}\left[C_1'(M_1 + M_1')C_1\right]$ which shows that the bias minimal estimator does not depend on the unknown parameter δ. For the case of L^2 risk the situation is more complicated. Note however that $\text{trace}(\bar{D}^2) = \delta^2\,\text{trace}\left[\left(\frac{1}{2}C_1'(M_1 + M_1')C_1\right)^2\right]$, $\left(\text{trace}\,\bar{D}\right)^2 = \delta^2\left(\text{trace}\left[C_1'(M_1 + M_1)C_1\right]\right)^2$ and

$$\text{trace}\left(DGD'\right) = \delta^3\,\text{trace}\left(C_1'M_1C_1C_1'M_2C_1C_1'M_1'C_1\right) - \delta^4\,\text{trace}\left[\left(C_1'M_1C_1\right)^2\left(C_1'M_1'C_1\right)^2\right]$$

which shows that the optimum only depends in a relatively simple way on the unknown parameter δ. We could in principle use a prior distribution for this parameter to obtain a tractable risk function. Since this will only change the relative weights on the different components we will not explicitly analyze it here. Once the weights are known or assumed to be known the optimal matrix C can in principle be found numerically.

The following theorem describes the bias optimal 2SLS estimator for the dynamic panel model

Theorem 8 *Let X^* be as defined in Theorem (6). Let $\bar{D} = (D + D')/2$ where $D = C_1' \Sigma_{12} C_1$. Let $C^* = \underset{C \ s.t. \ C'C=I}{\arg\min} \ tr\bar{D}/n$. Then $C^* = r_i$ where r_i is the eigenvector corresponding to the smallest eigenvalue l_i of $\bar{D}$. As $T \to \infty$ the smallest eigenvalue of $\bar{D}$, $l_i \to 0$. Let $\mathbf{1} = [1, ..., 1]'$ be a $T-1$ vector. Then for $C = \mathbf{1}/(\mathbf{1}'\mathbf{1})^{1/2}$ it follows that* $\operatorname{trace} \bar{D} \to 0$ *as $T \to \infty$.*

Theorem 8 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 r_i corresponding to the smallest eigenvalue of $(M_1 + M_1')/2$. This eigenvalue can be easily computed for any given T. The Theorem also shows that at least for large T the optimal procedure can be approximated by heuristic method which puts equal weight on all moment conditions. The heuristic procedure turns out be equal to the moment condition $E(u_{iT} - u_{i1}) y_{i0}$ 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 $T \to \infty$.

6 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 $z_i\left(\widehat{\beta}_{(\ell)}\right) \equiv \left(y_{i0}, y_{i2} - \widehat{\beta}_{(\ell)} y_{i1}, \ldots, y_{iT-1} - \widehat{\beta}_{(\ell)} y_{iT-2}\right)$, 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[7], the estimator is a fixed point to the minimization problem

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

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

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

Their actual performance approximated by 5000 Monte Carlo runs along with the biases predicted by second order theory in Theorem 1 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.

[7] 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.

Remark 2 *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 methods. 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 Table* **??**. *For comparison purpose, we reported properties of Arellano and Bover's estimator (1995) as well. We can see that the long difference estimator has a comparable property to Ahn and Schmidt's estimator. We do not know why the version of long difference CUE has such a large median bias at* $\beta = .95$ *whereas the CUE itself does not have such problem.*

7 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 solved the problem. Thus, we propose and investigate a new estimator, the long difference estimator of Griliches and Hausman (1986). We find that 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 are inadmissable. 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 Proof of Theorem 1

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

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

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

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

$$
\begin{aligned}
h_4 &= \frac{1}{2}g_2' G^{-1}G_1G^{-1}g + \frac{1}{2}g'\left(2G^{-1}G_1G^{-1}G_1G^{-1} - G^{-1}G_2G^{-1}\right)G_1G^{-1}g \\
&+ \frac{1}{2}g'G^{-1}G_3G^{-1}g + \frac{1}{2}g'G^{-1}G_1\left(2G^{-1}G_1G^{-1}G_1G^{-1} - G^{-1}G_2G^{-1}\right)g \\
&+ \frac{1}{2}g'G^{-1}G_1G^{-1}g_2 \\
&- g_1'G^{-1}G_1G^{-1}G_1G^{-1}g + g_1'G^{-1}G_2G^{-1}g - g_1'G^{-1}G_1G^{-1}G_1G^{-1}g \\
&+ g_1'G^{-1}G_1G^{-1}g_1 - g'G^{-1}G_1G^{-1}G_2G^{-1}g \\
&+ g'G^{-1}G_1G^{-1}G_1G^{-1}G_1G^{-1}g - g'G^{-1}G_1G^{-1}G_1G^{-1}g_1 - g'G^{-1}G_2G^{-1}G_1G^{-1}g \\
&+ g'G^{-1}G_2G^{-1}g_1 - g'G^{-1}G_1G^{-1}G_1G^{-1}g_1
\end{aligned}
$$

We may therefore rewrite the first order condition (4) as

$$
\begin{aligned}
0 &= \left(2g_1'G^{-1}g - g'G^{-1}G_1G^{-1}g\right) + \frac{1}{\sqrt{n}}\left(2h_1 - h_3\right)\sqrt{n}\left(b-\beta\right) \\
&+ \frac{1}{n}\left(2h_2 - h_4\right)\left(\sqrt{n}\left(b-\beta\right)\right)^2 + o_p\left(\frac{1}{n}\right). \qquad (19)
\end{aligned}
$$

Let

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

Lemma 2

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

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

Lemma 3

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

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

$$
\begin{aligned}
g_1'G^{-1}g_1 &= \left(\lambda_1 + \left(g_1-\lambda_1\right)\right)'\left(\Lambda^{-1} - \Lambda^{-1}\left(G-\Lambda\right)\Lambda^{-1} + o_p\left(\frac{1}{\sqrt{n}}\right)\right)\left(\lambda_1 + \left(g_1-\lambda_1\right)\right) \\
&= \lambda_1\Lambda^{-1}\lambda_1 + 2\left(g_1-\lambda_1\right)'\Lambda^{-1}\lambda_1 - \lambda_1\Lambda^{-1}\left(G-\Lambda\right)\Lambda^{-1}\lambda_1 + o_p\left(\frac{1}{\sqrt{n}}\right), \qquad (20)
\end{aligned}
$$

we obtain

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

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 4

$$2 g_1' G^{-1} g - g' G^{-1} G_1 G^{-1} g = \frac{1}{\sqrt{n}} \Phi + \frac{1}{n} \Gamma + o_p \left(\frac{1}{n} \right).$$

Proof. We have

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

and

$$g' G^{-1} G_1 G^{-1} g = g' \Lambda^{-1} \Lambda_1 \Lambda^{-1} g + o_p \left(\frac{1}{n} \right).$$

from which the conclusion follows. ■

Using Lemmas 2, 3, and 4, we may rewrite the first order condition (19) 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 can conclude that

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

It therefore follows that the approximate mean of $\sqrt{n} (b - \beta)$ is equal to

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

Noting that

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

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

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

and

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

we obtain the desired conclusion.

B Proof of Theorem 2

The second order bias is computed using Theorem 1. 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 1 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 1 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

$$
\begin{aligned}
\lambda_1\Lambda^{-1}\lambda_1 &= \sum_{t=1}^{T-1}E\left[z_{it}x_{it}^*\right]'E\left[z_{it}z_{it}'\right]^{-1}E\left[z_{it}x_{it}^*\right], \\
\operatorname{trace}\left(\Lambda^{-1}E\left[\delta_i\frac{\partial\delta_i'}{\partial\beta}\right]\right) &= -\sum_{t=1}^{T-1}\operatorname{trace}\left(E\left[z_{it}z_{it}'\right]^{-1}E\left[\varepsilon_{it}^*x_{it}^*z_{it}z_{it}'\right]\right), \\
\lambda_1'\Lambda^{-1}E\left[\delta_i\frac{\partial\delta_i'}{\partial\beta}\right]\Lambda^{-1}\lambda_1 &= -\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}E\left[z_{it}x_{it}^*\right]'E\left[z_{it}z_{it}'\right]^{-1}E\left[\varepsilon_{it}^*x_{is}^*z_{it}z_{is}'\right]E\left[z_{is}z_{is}'\right]^{-1}E\left[z_{is}x_{is}^*\right],
\end{aligned}
$$

and

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

C Technical Lemmas for Section 3.3

Lemma 5

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

Proof. We have

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

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

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

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 6

$$\operatorname{Var}(x_t^{*\prime} M_t \varepsilon_t^*) = (n-t)\sigma^2 E\left[v_{it}^{*2}\right] + (n-t)\left(E\left[v_{it}^*\varepsilon_{it}^*\right]\right)^2,$$

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

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 7 *Suppose that* $s < t$. *We have*

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

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

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

and hence

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

It follows that

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

from which we obtain

$$\begin{aligned}
v_{it}^{*} &= -\sqrt{\frac{T-t}{T-t+1}}\left(\frac{1}{1-\beta}-\frac{\beta-\beta^{T-t+1}}{\left(T-t\right)\left(1-\beta\right)^{2}}\right)\left(\alpha_i - E\left[\left.\alpha_i\right|z_{it}\right]\right) \\
&\quad -\sqrt{\frac{T-t}{T-t+1}}\frac{\left(1-\beta\right)\varepsilon_{i,T-1}+\left(1-\beta^{2}\right)\varepsilon_{i,T-2}+\cdots+\left(1-\beta^{T-t}\right)\varepsilon_{i,t}}{\left(T-t\right)\left(1-\beta\right)}.
\end{aligned} \tag{21}$$

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

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

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

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

Because

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

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[8]

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

from which we obtain

$$E\left[(\alpha_i - E[\alpha_i | z_{it}])^2 \right] = \frac{\sigma_\alpha^2}{1 + \frac{\sigma_\alpha^2}{\sigma^2} \frac{2}{1-\beta} + \frac{\sigma_\alpha^2}{\sigma^2}(t-2)}. \tag{22}$$

We now characterize $E\left[v_{it}^{*2}\right]$. Using (21), 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 (22), we obtain the first conclusion.

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

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

Combining with (21), we obtain

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

from which follows the second conclusion.

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

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

and

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

■

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

Lemma 8

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

Proof. Write

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

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

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

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

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

where C is a generic constant. Therefore, we have

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

It can be shown that

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

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

Lemma 9

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

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 8. ■

Lemma 10

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

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

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

But because

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

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 11

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

Proof. Note that

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

Here, the second equality is based on Lemma 6. Lemmas 8, 9, and 10 establish that variances of the three terms on the far right are all of order $o(1)$. ■

Lemma 12

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

Proof. Follows easily by combining Lemma 11 and the proof of Theorem 2 in Alvarez and Arellano (1998). ■

Lemma 13

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

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

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

By conditioning, it can be shown that

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

Therefore,

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

Modifying the proof of Lemma 8, we can establish that the right is $o(1)$.

We now show that

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

We have

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

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

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

Using (21), we can show that

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

Adopting the same argument as in the proofs for Lemmas 8 - 10, we can show that the variance is $o(1)$.
■

D Proofs for Section 5

We need the following auxiliary result to prove admissibility of our bias minimal procedures under asymptotic L^2 risk. Define the random variable Y by

$$Y \equiv \frac{\xi_x' C(C'\Omega C)^+ C'\xi_y}{\xi_x' C(C'\Omega C)^+ C'\xi_x} \tag{23}$$

where $\left[\xi_x, \xi_y\right] \sim N(0, \Sigma)$ and $C(C'\Omega C)^{+}C$ is full rank.

Lemma 14 *Let Y^* be defined in (23) such that $\Omega = \Sigma_{11}$. Then $E\left[(Y^*)^2\right] \leq E\left[(Y)^2\right]$ where Y is as defined in (23) with any positive definite weight matrix Ω.*

Proof. Choose L such that $LL' = \Sigma_{11} = \delta I$. Define $W \equiv L'C\left(C'\Omega C\right)^{+}C'L$ and $D \equiv L'C\left(C'\Omega C\right)^{+}C'\Sigma_{21}'\Sigma_{11}^{-1}L$. Let Γ_1 be an orthogonal matrix of eigenvectors of W such that $\Gamma_1\Gamma_1' = \Gamma_1'\Gamma_1 = I$ and Λ_1 a diagonal matrix with eigenvalues of W. Define $z \equiv L^{-1}\xi_x$ and $z_1 \equiv \Gamma_1 z$. Define $G \equiv \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$ such that

$$E\left[Y^2\right] = E\left[\frac{\left(z_1'\Gamma_1 D\Gamma_1' z_1\right)^2}{\left(z_1'\Lambda_1 z_1\right)^2}\right] + E\left[\frac{z_1'\Gamma_1 DGD'\Gamma_1' z_1}{\left(z_1'\Lambda_1 z_1\right)^2}\right]$$

where $\left(z_1'Dz_1\right)^2 \geq 0$ and $z'DGD'z \geq 0$ since G is positive definite. We define $z_1 \equiv vs^{1/2}$ with $s \equiv z_1'z_1$ and $v \equiv z_1/\left(z_1'z_1\right)^{1/2}$. Without loss of generality, we can assume that the largest eigenvalue of in Λ_1 is 1. Because $v'v = 1$, we have $\left(v'\Lambda_1 v\right)^2 \leq 1$. It therefore follows that $\left(v'\Gamma_1 D\Gamma_1' v\right)^2 \big/ \left(v'\Lambda_1 v\right)^2 \geq \left(v'\Gamma_1 D\Gamma_1' v\right)^2$, from which we obtain

$$\frac{\left(z_1'\Gamma_1 D\Gamma_1' z_1\right)^2}{\left(z_1'\Lambda_1 z_1\right)^2} \geq \frac{\left(z_1'\Gamma_1 D\Gamma_1' z_1\right)^2}{\left(z_1' z_1\right)^2}.$$

The same arguments show that

$$E\left[\frac{z_1'\Gamma_1 DGD'\Gamma_1' z_1}{\left(z_1'\Lambda_1 z_1\right)^2}\right] \geq E\left[\frac{z_1'\Gamma_1 DGD'\Gamma_1' z_1}{\left(z_1' z_1\right)^2}\right]$$

■

Proof of Lemma 1. Note that

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

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

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

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

and

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

such that $\operatorname{Var}\left(n^{-3/2}\sum_{i=1}^n \Delta y_{it} y_{i0}\right)=O(1)$. For $n^{-3/2}\sum_{i=1}^n g_{i,2}(\beta_0)$ we have from the moment conditions that $E[g_{i,2}(\beta_0)]=0$ and

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

The 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

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

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

$$
\Sigma_n=\begin{bmatrix}\Sigma_{11,n} & \Sigma_{12,n}\\ \Sigma_{21,n} & \Sigma_{22,n}\end{bmatrix}
$$

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

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

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

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

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 $n^{-3/2}\sum_{i=1}^{n}\ell' E\left[f_{i,2}\right] = O(n^{-1/2})$ and thus does not affect the limit distribution. ■

Proof of Theorem 6. Define $W \equiv L'C_1\left(C'\Omega C\right)^{+} C'_1 L$ where L satisfies $\Sigma_{11} = LL'$. We first show that it is never optimal to choose W singular. For this purpose partition $C = [C'_0, C'_1]'$ such that $C'\xi_x^{\#} = C'_1\xi_x$. The limiting random variable X can therefore be represented as

$$X = \frac{\xi'_x C_1\left(C'\Omega C\right)^{+} C'_1\xi_y}{\xi'_x C_1\left(C'\Omega C\right)^{+} C'_1\xi_x}.$$

We observe that $\xi_y|\xi_x \sim N(F\xi_x, G)$ where $F = \Sigma_{21}\Sigma_{11}^{-1}$ and $G = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$ and define $D = WL^{-1}FL$ and $z = L^{-1}\xi_x$ such that $z \sim N(0, I)$.

We now consider the case where W is singular. Let Γ be an orthogonal matrix of eigenvectors of W such that $\Gamma\Gamma' = \Gamma'\Gamma = I$ and Λ a diagonal matrix with eigenvalues of W such that Λ_1 contains all nonzero eigenvalues and Λ_2 contains the zero eigenvalues. Partition $\Gamma = [\Gamma_1, \Gamma_2]$ conformably such that $\Gamma'_1\Gamma_2 = 0$ and $\Gamma_1\Gamma'_1 + \Gamma_2\Gamma'_2 = I$. Then $W = \Gamma\Lambda\Gamma' = \Gamma_1\Lambda_1\Gamma'_1$ and $D = \Gamma_1\Lambda_1\Gamma'_1 L^{-1}FL$ such that $\Gamma'_2 D = 0$. Define $z_1 = \Gamma'_1 z$ and $z_2 = \Gamma'_2 z$. Then, using the fact that $E\left[\xi_y\xi'_y|\xi_x\right] = F\xi_x\xi'_x F' + G$ leads to

$$E\left[X^2\right] = E\left[\frac{\left(z'Dz\right)^2}{\left(z'Wz\right)^2}\right] + E\left[\frac{z'DGD'z}{\left(z'Wz\right)^2}\right]$$

where $z'Wz = z'_1\Lambda_1 z_1$ and $z'DGD'z = z'_1\Gamma'_1 DGD'\Gamma_1 z_1$. We therefore only need to consider the first term where $\left(z'Dz\right)^2 = \left(z'_1\Gamma'_1 D\left(\Gamma_1\Gamma'_1 + \Gamma_2\Gamma'_2\right)z\right)^2$. Since z_1 and z_2 are independent we can use a conditioning argument to evaluate the first term

$$\begin{aligned} E\left[\frac{\left(z'Dz\right)^2}{\left(z'Wz\right)^2}\right] &= E\left[\frac{\left(z'_1\Gamma'_1 D\Gamma_1 z_1\right)^2}{\left(z'_1\Lambda_1 z_1\right)^2} + \frac{z'_1\Gamma'_1 D\Gamma_1 z_1 z'_2\Gamma'_2 D\Gamma_1 z_1}{\left(z'_1\Lambda_1 z_1\right)^2}\right] \\ &\quad + E\left[\frac{z'_1\Gamma'_1 D\Gamma_2 z_2 z'_1\Gamma'_1 D\Gamma_1 z_1}{\left(z'_1\Lambda_1 z_1\right)^2} + \frac{\left(z'_1\Gamma'_1 D\Gamma_2 z_2\right)^2}{\left(z'_1\Lambda_1 z_1\right)^2}\right] \end{aligned}$$

where $E\left[\frac{z'_1\Gamma'_1 D\Gamma_2 z_2 z'_1\Gamma'_1 D\Gamma_1 z_1}{\left(z'_1\Lambda_1 z_1\right)^2}|z_1\right] = 0$ because $Ez_2 = 0$ such that the cross terms vanish. It follows that

$$\begin{aligned} E\left[\frac{\left(z'Dz\right)^2}{\left(z'Wz\right)^2}\right] &= E\left[\frac{\left(z'_1\Gamma'_1 D\Gamma_1 z_1\right)^2 + \left(z'_1\Gamma'_1 D\Gamma_2 z_2\right)^2}{\left(z'_1\Lambda_1 z_1\right)^2}\right] \\ &\geq E\left[\frac{\left(z'_1\Gamma'_1 D\Gamma_1 z_1\right)^2}{\left(z'_1\Lambda_1 z_1\right)^2}\right] \end{aligned}$$

We can therefore assume that W is nonsingular. Then by Lemma 14 it follows that the optimal $W = I$ which can only occur if $C = C_1$ and $\Omega = \Sigma_{11}$. ■

Proof of Theorem (8). First note that $\operatorname{trace}\left(\bar{D}\right) = \frac{1}{2}\operatorname{trace}\left(L^{-1}C'_1\left(\Sigma_{12} + \Sigma_{21}\right)C_1 L'^{-1}\right) = \operatorname{trace}\left(C'_1\bar{\Sigma}_{12}C_1\right)$ with $\bar{\Sigma}_{12} = \left(\Sigma_{12} + \Sigma_{21}\right)/2$. It can be checked easily that $\bar{\Sigma}_{12}$ is negative definite symmetric. We can therefore minimize $-\operatorname{trace}\left(C'_1\bar{\Sigma}_{12}C_1\right)$. It is now useful to chose an orthogonal matrix R such that $R'R = RR' = I$ and $-\bar{\Sigma}_{12} = RLR'$ where L is the diagonal matrix of eigenvalues of

$-\bar{\Sigma}_{12} = \sum_{i=1}^{n} l_i r_i r_i'$. Then it follows that $-trC_1'\bar{\Sigma}_{12}C_1 = \sum_{i=1}^{T-1} l_i r_i' C_1 C_1' r_i$. Next note that all the eigenvalues of $C_1 C_1'$ are either zero or one such that $0 \leq r_i' C_1 C_1' r_i \leq 1$. The minimum of $-\operatorname{trace}\left(C_1'\bar{\Sigma}_{12}C_1\right)$ is then found by choosing C_1 such that $C_1' r_i = 0$ except for the eigenvector r_i corresponding to the smallest l_i. It now follows that $E\left[X^*\right] = \operatorname{trace}\left(\bar{D}/n\right)$ is minimized for $C = r_i$ where r_i is the eigenvector corresponding to the smallest eigenvalue. To show this note that if $C = r_i$ then $\operatorname{trace}\left(\bar{D}/n\right) = \operatorname{trace}\left(\bar{D}\right) = l_i$. Now suppose a vector x such that $x'x = 1$ and $r_i'x = 0$ is added to C_1. Then $\operatorname{trace} C_1'\bar{\Sigma}_{12}C_1 = l_i + \sum_{j\neq i}^{T-1} l_j \left(r_j'x\right)^2$. By Parseval's equality $\sum_{j\neq i}^{T-1}\left(r_j'x\right)^2 = 1$. Since $l_j \geq l_i$ we can bound $\operatorname{trace}\left(C_1'\bar{\Sigma}_{12}C_1\right) \geq (T-1)\, l_i$ but then $\operatorname{trace}\left(\bar{D}/n\right) \geq l_i$. This argument can be repeated to more than one orthogonal additions x. Next note that from $x'x = 1$ such that $\min l_i \leq -x'\bar{\Sigma}_{12}x \leq \max l_i$ it follows that $\min l_i \leq \mathbf{1}'\bar{\Sigma}_{12}\mathbf{1}/(\mathbf{1}'\mathbf{1}) = (T-1)^{-1}$ for $\mathbf{1} = [1, ..., 1]'$ which shows that the smallest eigenvalue is bounded by a monotonically decreasing function of the number of moment conditions. Also note that $\bar{\Sigma}_{12}\mathbf{1}/(\mathbf{1}'\mathbf{1})^{1/2} \to \mathbf{0}$ in l^2 norm where $\mathbf{0}$ is an element of the infinite sequence space l^2. ■

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

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

where

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

They suggest a GMM estimation:

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

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 $E\left[q_i\left(\beta\right)q_i\left(\beta\right)\right]$:

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 $\left(\sum_{i=1}^{n} q_i\left(b\right)\right)' A_1^{-1}\left(\sum_{i=1}^{n} q_i\left(b\right)\right)$. 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 Proof of Theorems 5

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

$$\sqrt{n}(b-\beta) = \frac{\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{z}_i x_i\right)'\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{z}_i\widehat{z}_i'\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{z}_i\varepsilon_i\right)}{\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{z}_i x_i\right)'\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{z}_i\widehat{z}_i'\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{z}_i x_i\right)} \tag{24}$$

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

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

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

and

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

Here, ϕ and Δ are defined in Theorem 5. Using arguments similar to the derivation of (20), we obtain

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

and

$$\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{z}_i x_i\right)'\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{z}_i\widehat{z}_i'\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{z}_i x_i\right)$$

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

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

Therefore, we may conclude that

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

$$+ \frac{1}{\sqrt{n}}B_1 + \frac{1}{\sqrt{n}}B_2 + o_p\left(\frac{1}{\sqrt{n}}\right), \tag{25}$$

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

$$
\begin{aligned}
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.
\end{aligned}
$$

The first two terms on the right side of (25) 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

$$
E\left[B_1\right] = \frac{(K-2)\,\sigma_{u\varepsilon}}{\lambda'\Lambda^{-1}\lambda}. \tag{26}
$$

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

$$
\begin{aligned}
E\left[B_2\right] = & -\frac{\lambda'\Lambda^{-1}\varphi}{\lambda'\Lambda^{-1}\lambda}E\left[Q_n\right]-\frac{\lambda'\Lambda^{-1}E\left[f_i w_i \varepsilon_i\right]}{\lambda'\Lambda^{-1}\lambda} \\
& -\frac{E\left[f_i z_i x_i\right]'\Lambda^{-1}\varphi}{\lambda'\Lambda^{-1}\lambda}-\frac{\phi'\Lambda^{-1}E\left[f_i z_i \varepsilon_i\right]}{\lambda'\Lambda^{-1}\lambda} \\
& +\frac{\lambda'\Lambda^{-1}E\left[f_i z_i z_i'\right]\Lambda^{-1}\varphi}{\lambda'\Lambda^{-1}\lambda}+\frac{\lambda'\Lambda^{-1}\Delta\Lambda^{-1}E\left[f_i z_i \varepsilon_i\right]}{\lambda'\Lambda^{-1}\lambda}-E\left[f_i^2\right]\frac{\lambda'\Lambda^{-1}\Delta\Lambda^{-1}\varphi}{\lambda'\Lambda^{-1}\lambda} \\
& +2\frac{\lambda'\Lambda^{-1}\varphi}{\left(\lambda'\Lambda^{-1}\lambda\right)^2}E\left[f_i z_i x_i\right]'\Lambda^{-1}\lambda \\
& +2\frac{\lambda'\Lambda^{-1}E\left[f_i z_i \varepsilon_i\right]}{\left(\lambda'\Lambda^{-1}\lambda\right)^2}\phi'\Lambda^{-1}\lambda-2E\left[f_i^2\right]\frac{\lambda'\Lambda^{-1}\varphi}{\left(\lambda'\Lambda^{-1}\lambda\right)^2}\phi'\Lambda^{-1}\lambda \\
& -\frac{\lambda'\Lambda^{-1}\varphi}{\left(\lambda'\Lambda^{-1}\lambda\right)^2}\lambda'\Lambda^{-1}E\left[f_i z_i z_i'\right]\Lambda^{-1}\lambda \\
& -\frac{\lambda'\Lambda^{-1}E\left[f_i z_i \varepsilon_i\right]}{\left(\lambda'\Lambda^{-1}\lambda\right)^2}\lambda'\Lambda^{-1}\Delta\Lambda^{-1}\lambda+E\left[f_i^2\right]\frac{\lambda'\Lambda^{-1}\varphi}{\left(\lambda'\Lambda^{-1}\lambda\right)^2}\lambda'\Lambda^{-1}\Delta\Lambda^{-1}\lambda. \qquad (27)
\end{aligned}
$$

Using (25), (26), and (27), 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

$$
\sqrt{n}\left(b-\beta\right)=\frac{\frac{1}{\sqrt{n}}\sum_{t=1}^{T-1}\left(x_t^{*\prime}P_t\varepsilon_t^*-\kappa_t x_t^{*\prime}\varepsilon_t^*\right)}{\frac{1}{n}\sum_{t=1}^{T-1}\left(x_t^{*\prime}P_t x_t^*-\kappa_t x_t^{*\prime}x_t^*\right)}, \qquad (28)
$$

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.[9]

G.1 Characterization of Second Order Bias of LIML

Consider a simple simultaneous equations model

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

and examine LIML b that solves

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

where

$$
e\left(c\right)=y-xc
$$

[9] 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.

Here, the first order condition is given by

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

or

$$G_{n}\left(b\right)=0,$$

where

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

Note that

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

and

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

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

$$0=G_{n}\left(\beta\right)+\frac{1}{\sqrt{n}}\frac{\partial G_{n}\left(\beta\right)}{\partial b}\left(\sqrt{n}\left(b-\beta\right)\right)+\frac{1}{n}\frac{\partial^{2}G_{n}\left(\beta\right)}{\partial b^{2}}\left(\sqrt{n}\left(b-\beta\right)\right)^{2}+o_{p}\left(\frac{1}{n}\right).\tag{29}$$

First, note that

$$\begin{aligned}G_{n}\left(\beta\right)&=\left(\frac{1}{n}x'P\varepsilon\right)\left(\frac{1}{n}\varepsilon'\varepsilon\right)-\left(\frac{1}{n}x'\varepsilon\right)\left(\frac{1}{n}\varepsilon'P\varepsilon\right)\\&=\left(\frac{1}{n}\sum_{i=1}^{n}\varepsilon_{i}^{2}\right)\left(\frac{1}{n}\sum_{i=1}^{n}z_{i}x_{i}\right)'\left(\frac{1}{n}\sum_{i=1}^{n}z_{i}z_{i}'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}z_{i}\varepsilon_{i}\right)\\&\quad-\left(\frac{1}{n}\sum_{i=1}^{n}x_{i}\varepsilon_{i}\right)\left(\frac{1}{n}\sum_{i=1}^{n}z_{i}\varepsilon_{i}\right)'\left(\frac{1}{n}\sum_{i=1}^{n}z_{i}z_{i}'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}z_{i}\varepsilon_{i}\right).\end{aligned}$$

Because

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\varepsilon_{i}^{2}&=\sigma_{\varepsilon}^{2}+\frac{1}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\varepsilon_{i}^{2}-\sigma_{\varepsilon}^{2}\right)\right),\\\frac{1}{n}\sum_{i=1}^{n}z_{i}x_{i}&=\lambda+\frac{1}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(z_{i}x_{i}-\lambda\right)\right),\\\frac{1}{n}\sum_{i=1}^{n}x_{i}\varepsilon_{i}&=\sigma_{x\varepsilon}+\frac{1}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(x_{i}\varepsilon_{i}-\sigma_{x\varepsilon}\right)\right),\\\left(\frac{1}{n}\sum_{i=1}^{n}z_{i}z_{i}'\right)^{-1}&=\Lambda^{-1}-\frac{1}{\sqrt{n}}\Lambda^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(z_{i}z_{i}'-\Lambda\right)\right)\Lambda^{-1}+o_{p}\left(\frac{1}{\sqrt{n}}\right),\end{aligned}$$

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

$$G_n(\beta) = \frac{1}{\sqrt{n}}\Phi + \frac{1}{n}\Gamma + o_p\left(\frac{1}{n}\right) \tag{30}$$

where

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

and

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

Now, note that

$$\begin{aligned}
\frac{\partial G_n(\beta)}{\partial b} &= -\left(\frac{1}{n}x'Px\right)\left(\frac{1}{n}\varepsilon'\varepsilon\right) + \left(\frac{1}{n}x'x\right)\left(\frac{1}{n}\varepsilon'P\varepsilon\right) \\
&= -\left(\frac{1}{n}\sum_{i=1}^n \varepsilon_i^2\right)\left(\frac{1}{n}\sum_{i=1}^n z_i x_i\right)'\left(\frac{1}{n}\sum_{i=1}^n z_i z_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n z_i x_i\right) \\
&\quad + \left(\frac{1}{n}\sum_{i=1}^n x_i^2\right)\left(\frac{1}{n}\sum_{i=1}^n z_i \varepsilon_i\right)'\left(\frac{1}{n}\sum_{i=1}^n z_i z_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n z_i \varepsilon_i\right) \\
&= \Upsilon + \frac{1}{\sqrt{n}}\Xi + o_p\left(\frac{1}{\sqrt{n}}\right),
\end{aligned} \tag{31}$$

where

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

and

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

Finally, note that

$$\begin{aligned}
\frac{\partial^2 G_n(\beta)}{\partial b^2} &= 2\left(\frac{1}{n}x'Px\right)\left(\frac{1}{n}x'\varepsilon\right) - 2\left(\frac{1}{n}x'P\varepsilon\right)\left(\frac{1}{n}x'x\right) \\
&= 2\Psi + o_p(1).
\end{aligned} \tag{32}$$

where

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

Combining (29), (30), (31), and (32), we obtain

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

from which we obtain

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

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

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

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

G.2 Higher Order Analysis of the "Eigenvalue"

Let

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

Getting back to the first order condition

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

we can write

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

the usual expression.

Note that

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

The numerator and the denominator may be rewritten as

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

and

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

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 (28). First, note that

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

and

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

It therefore follows that

$$
\begin{aligned}
\sqrt{n}\left(b-\beta\right) &= \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 (14). We have*

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

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

$$\frac{\lambda' \Lambda^{-1}}{\lambda' \Lambda^{-1} \lambda} z_i \varepsilon_i - \frac{\lambda' \Lambda^{-1} \varphi}{\lambda' \Lambda^{-1} \lambda} f_{LIML,i} \tag{33}$$

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

$$\frac{\lambda' \Lambda^{-1}}{\lambda' \Lambda^{-1} \lambda} z_i \varepsilon_i - \frac{\lambda' \Lambda^{-1} \varphi}{\lambda' \Lambda^{-1} \lambda} \left(\frac{\lambda' \Lambda^{-1}}{\lambda' \Lambda^{-1} \lambda} z_i \varepsilon_i - \frac{\lambda' \Lambda^{-1} \varphi}{\lambda' \Lambda^{-1} \lambda} f_{LIML,i} \right). \tag{34}$$

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

$$\frac{\lambda' \Lambda^{-1}}{\lambda' \Lambda^{-1} \lambda} z_i \varepsilon_i - \frac{\lambda' \Lambda^{-1} \varphi}{\lambda' \Lambda^{-1} \lambda} \left[\frac{\lambda' \Lambda^{-1}}{\lambda' \Lambda^{-1} \lambda} z_i \varepsilon_i - \frac{\lambda' \Lambda^{-1} \varphi}{\lambda' \Lambda^{-1} \lambda} \left(\frac{\lambda' \Lambda^{-1}}{\lambda' \Lambda^{-1} \lambda} z_i \varepsilon_i - \frac{\lambda' \Lambda^{-1} \varphi}{\lambda' \Lambda^{-1} \lambda} f_{LIML,i} \right) \right] \tag{35}$$

and

$$\frac{\lambda' \Lambda^{-1}}{\lambda' \Lambda^{-1} \lambda} z_i \varepsilon_i - \frac{\lambda' \Lambda^{-1} \varphi}{\lambda' \Lambda^{-1} \lambda} \left\{ \frac{\lambda' \Lambda^{-1}}{\lambda' \Lambda^{-1} \lambda} z_i \varepsilon_i - \frac{\lambda' \Lambda^{-1} \varphi}{\lambda' \Lambda^{-1} \lambda} \left[\frac{\lambda' \Lambda^{-1}}{\lambda' \Lambda^{-1} \lambda} z_i \varepsilon_i - \frac{\lambda' \Lambda^{-1} \varphi}{\lambda' \Lambda^{-1} \lambda} \left(\frac{\lambda' \Lambda^{-1}}{\lambda' \Lambda^{-1} \lambda} z_i \varepsilon_i - \frac{\lambda' \Lambda^{-1} \varphi}{\lambda' \Lambda^{-1} \lambda} f_{LIML,i} \right) \right] \right\} \tag{36}$$

Using (33) - (36), 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 (33) - (36), 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

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

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

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

[6] 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.

[7] Arellano, M. and O. Bover, 1995, "Another Look at the Instrumental Variable Estimation of Error-Component Models", *Journal of Econometrics* 68, pp. 1 - 268.

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

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

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

[11] 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.*

[12] 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.

[13] Donald, S. and W. Newey, 1998, "Choosing the Number of Instruments", *unpublished manuscript.*

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

[15] Griliches, Z. and J. Hausman, 1986, "Errors in Variables in Panel Data," *Journal of Econometrics* 31, pp. 93 - 118.

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

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

[18] Hahn, J. and J. Hausman, 2000, "A New Specification Test for the Validity of Instrumental Variables", *unpublished manuscript.*

[19] 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.*

[20] Hall, P. and J. Horowitz, 1996, "Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators", *Econometrica* 64, 891-916.

[21] 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.

[22] 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.

[23] Holtz-Eakin, D., W.K. Newey, and H. Rosen, "Estimating Vector Autoregressions with Panel Data", *Econometrica* 56, pp. 1371 - 1395.

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

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

[26] Lancaster, T., 1997, "Orthogonal Parameters and Panel Data", *unpublished manuscript.*

[27] 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.

[28] Moon, H.R. and P.C.B. Phillips, 2000, "GMM Estimation of Autoregressive Roots Near Unity with Panel Data", *unpublished manuscript.*

[29] 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.

[30] 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.

[31] Newey, W.K. and R.J. Smith, 2000, "Asymptotic Equivalence of Empirical Likelihood and the Continuous Updating Estimator", *unpublished manuscript.*

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

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

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

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

[36] 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.

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

Table 1: 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 2: 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 3: 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	0.023	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 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	1.1570
RMSE	0.1699	0.1954	0.2545	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

	$\widehat{b}_{BB1}$	$\widehat{b}_{BB2}$	$\widehat{b}_{BB3}$	$\widehat{b}_{BB4}$	$\widehat{b}_{LIML,1}$	$\widehat{b}_{LIML,2}$	$\widehat{b}_{LIML,3}$
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	0.2391	0.2292

Table 6: Sensitivity of Blundell and Bond's (1998) Estimator

$\beta_F = .5$	$\widehat{b}_{BB1}$	$\widehat{b}_{BB2}$	$\widehat{b}_{BB3}$	$\widehat{b}_{BB4}$	$\widehat{b}_{LIML,1}$	$\widehat{b}_{LIML,2}$	$\widehat{b}_{LIML,3}$
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	-0.2163	-0.2214	-0.2469
RMSE	0.0994	0.1400	0.1899	0.1944	0.0570	0.0611	0.0630
$\beta_F = 0$	$\widehat{b}_{BB1}$	$\widehat{b}_{BB2}$	$\widehat{b}_{BB3}$	$\widehat{b}_{BB4}$	$\widehat{b}_{LIML,1}$	$\widehat{b}_{LIML,2}$	$\widehat{b}_{LIML,3}$
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 $T = 5$

	$N = 100$	$\widehat{b}_{LIML,1}$	$\widehat{b}_{LIML,2}$	$\widehat{b}_{LIML,3}$	$\widehat{b}_{LIML,4}$
$\beta = .75$	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	0.2458	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
	$N = 200$	$\widehat{b}_{LIML,1}$	$\widehat{b}_{LIML,2}$	$\widehat{b}_{LIML,3}$	$\widehat{b}_{LIML,4}$
$\beta = .75$	Actual Mean %bias	1.2054	3.0110	3.8420	5.1421
	Actual Median % bias	-1.6533	-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 $T = 5$

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

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

$N = 200$		$\widehat{\beta}_{CUE,LD}$	$\widehat{\beta}_{CUE2,AS}$	$\widehat{\beta}_{CUE2,LD}$	$\widehat{\beta}_{CUE2,BB}$	$\widehat{\beta}_{CUE,BB}$
$\beta = .75$	Median % Bias	4.2172	3.4952	4.0943	1.2644	0.4242
	Interquartile Range	0.2297	0.1855	0.2034	0.1195	0.1032
	Mean % Bias	8.8638	116.0861	29.5421	4.4327	0.1604
	RMSE	0.1704	85.9117	10.6641	4.5595	0.0719
$\beta = .8$	Median % Bias	5.4765	5.8182	5.3421	0.8893	0.5388
	Interquartile Range	0.2468	0.2105	0.2132	0.1472	0.1063
	Mean % Bias	8.3701	16.9181	-233.6177	-21.1440	-0.1898
	RMSE	0.1674	13.7393	127.6513	9.5825	0.0736
$\beta = .85$	Median % Bias	5.8672	5.1660	3.7226	1.0708	0.6441
	Interquartile Range	0.2341	0.2295	0.2347	0.2619	0.1143
	Mean % Bias	7.3021	688.7913	-50.0828	59.6610	-0.7076
	RMSE	0.1585	455.6137	19.8972	66.8390	0.0779
$\beta = .9$	Median % Bias	5.3657	0.9958	-2.8551	-1.0774	0.9204
	Interquartile Range	0.2101	0.3766	1.3425	0.5870	0.1152
	Mean % Bias	5.4221	479.7193	31.6093	-29.9555	-0.3893
	RMSE	0.1448	381.8271	23.2206	42.2422	0.0913
$\beta = .95$	Median % Bias	2.8984	-11.2026	-125.6884	-12.6677	2.5208
	Interquartile Range	0.1915	2.5978	1.5877	2.6203	0.1099
	Mean % Bias	2.8482	-82.2370	-6464.7709	-177.6883	0.9733
	RMSE	0.1331	39.5396	4315.6181	116.8096	0.1044