NBER WORKING PAPER SERIES

A MAXIMUM LIKELIHOOD METHOD FOR THE INCIDENTAL PARAMETER PROBLEM

Marcelo Moreira

Working Paper 13787
http://www.nber.org/papers/w13787

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2008

The author thanks Gary Chamberlain, Rustam Ibragimov, Humberto Moreira, Whitney Newey, Thomas Rothenberg, and Tiemen Woutersen for helpful comments. José Miguel Torres provided outstanding research assistance. E-mail: moreira@fas.harvard.edu. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

A Maximum Likelihood Method for the Incidental Parameter Problem
Marcelo Moreira
NBER Working Paper No. 13787
February 2008
JEL No. C13,C23,C30

ABSTRACT

This paper uses the invariance principle to solve the incidental parameter problem. We seek group actions that preserve the structural parameter and yield a maximal invariant in the parameter space with fixed dimension. M-estimation from the likelihood of the maximal invariant statistic yields the maximum invariant likelihood estimator (MILE). We apply our method to (i) a stationary autoregressive model with fixed effects; (ii) an agent-specific monotonic transformation model; (iii) an instrumental variable (IV) model; and (iv) a dynamic panel data model with fixed effects. In the first two examples, there exist group actions that completely discard the incidental parameters. In a stationary autoregressive model with fixed effects, MILE coincides with existing conditional and integrated likelihood methods. The invariance principle also gives a new perspective to the marginal likelihood approach. In an agent-specific monotonic transformation model, our approach yields an estimator that is consistent and asymptotically normal when errors are Gaussian. In an instrumental variable (IV) model, this paper unifies asymptotic results under strong instruments (SIV) and many weak instruments (MWIV) frameworks. We obtain consistency, asymptotic normality, and optimality results for the limited information maximum likelihood estimator directly from the invariant likelihood. Our approach is parallel to M-estimation in problems in which the number of parameters does not change with the sample size. In a dynamic panel data model with N individuals and T time periods, MILE is consistent as long as NT goes to infinity. We obtain a large N, fixed T bound; this bound coincides with Hahn and Kuersteiner's (2002) bound when T goes to infinity. MILE reaches (i) our bound when N is large and T is fixed; and (ii) Hahn and Kuersteiner's (2002) bound when both N and T are large.

Marcelo Moreira
Littauer Center M-6
Harvard University
Cambridge, MA 02138
and NBER
moreira@fas.harvard.edu

1 Introduction

The maximum likelihood estimator (MLE) is a commonly used procedure to estimate a parameter in stochastic models. Under regularity conditions, the MLE is not only consistent but also has asymptotic optimality properties (e.g., Le Cam and Yang (2000)). In the presence of incidental parameters, however, the MLE of structural parameters may not even be consistent. This failure occurs because the dimension of incidental parameters increases with the sample size, affecting the ability of MLE to consistently estimate the structural parameters. This is the so-called incidental parameter problem after the seminal paper by Neyman and Scott (1948). Lancaster (2000) and Arellano and Honoré (1991) provide excellent overviews of the subject.

This paper appeals to the invariance principle to solve the incidental parameter problem. We propose to find a group action that preserves the model and the structural parameter. This yields a maximal invariant statistic. Its distribution depends on the parameters only through the maximal invariant in the parameter space. Maximization of the invariant likelihood yields the maximum invariant likelihood estimator (MILE). Distinct group actions in general yield different estimators. We seek group actions whose maximal invariant in the parameter space has fixed dimension regardless of the sample size.

As is customary in the literature, we illustrate our approach with a series of examples.

Section 3 considers two groups of transformations that completely discard the incidental parameters. The first example is the stationary autoregressive model with fixed effects. For a particular group action, our solution coincides with Andersen's (1970) conditional and Lancaster's (2002) integrated likelihood approaches. The invariance principle also provides a new perspective on the marginal likelihood approach, e.g., Arellano (2003, Section 2.4.3). The second example is the monotonic transformation model. The proposed transformation is agent-specific and has infinite dimension. The conditional and integrated likelihood approaches do not seem to be applicable here. The invariant principle provides an estimator that is consistent and asymptotically normal under the assumption of normal errors.

We then proceed to the two main sections of the paper.

Section 4 considers an instrumental variable (IV) model with N observations and K instruments. In this section, we provide a likelihood maximization approach. It unifies asymptotic results under both the strong instruments (SIV) and many weak instruments (MWIV) asymptotics, e.g., Kunitomo (1980), Morimune (1983), and

Bekker (1994). This framework parallels standard M-estimation in problems in which the number of parameters does not change with the sample size. In particular, we are able to (i) show consistency of the MLE in the IV setup even under MWIV asymptotics from the perspective of likelihood maximization; (ii) derive the asymptotic distribution of the MLE directly from the objective function under SIV and MWIV asymptotics; and (iii) provide an explanation for optimality of MLE within the class of regular invariant estimators.

Section 5 presents a simple dynamic panel data model with N individuals and T time periods. We propose to use MILE based on the orthogonal group of transformations. This estimator is consistent as long as NT goes to infinity (regardless of the relative rate of N and T) and asymptotically normal under (i) large N, fixed T; and (ii) large N, large T asymptotics when the autoregressive parameter is smaller than one. We derive a bound for large N, fixed T asymptotics when errors are normal; our bound coincides with Hahn and Kuersteiner's (2002) bound when $T \to \infty$. MILE reaches (i) our bound when N is large and T is fixed; and (ii) Hahn and Kuersteiner's (2002) bound when both N and T are large. Finally, this paper provides further support to work by Arellano and Bond (1991) and Ahn and Schmidt (1995) from a maximal invariant perspective. Together with Chamberlain and Moreira (2006), we establish a connection between the GMM/MD and the integrated likelihood approaches in the dynamic panel data model.

Section 6 compares MILE with existing fixed-effects estimators for the dynamic panel data model.

Section 7 provides proofs for our results.

2 The Maximum Invariant Likelihood Estimator

Let $P_{\gamma,\eta}$ denote the distribution of the data set $Y \in \mathbf{Y}$ when the structural parameter is $\gamma \in \mathbf{\Gamma}$ and the incidental parameter is $\eta \in \mathbf{N}$: $\mathcal{L}(Y) = P_{\gamma,\eta} \in \mathbf{P}$.

We seek a group $\mathbf{G}$ and actions $\mathcal{A}_1(\cdot, Y)$ and $\mathcal{A}_2(\cdot, (\gamma, \eta))$ in the sample and parameter spaces that preserve the model $\mathbf{P}$:

$$\mathcal{L}(Y) = P_{\gamma,\eta} \Rightarrow \mathcal{L}(\mathcal{A}_1(g, Y)) = P_{\mathcal{A}_2(g,(\gamma,\eta))}, \text{ for any } P_{\gamma,\eta} \in \mathbf{P}.$$

We are interested in γ. This yields the following definition.

Definition 1 *Suppose that* $\mathcal{A}_2 : \mathbf{G} \times \mathbf{\Gamma} \times \mathbf{N} \to \mathbf{\Gamma} \times \mathbf{N}$ *induces an action* $\mathcal{A}_3 : \mathbf{G} \times \mathbf{N} \to \mathbf{N}$ *such that*

$$\mathcal{A}_2(g, (\gamma, \eta)) = (\gamma, \mathcal{A}_3(g, \eta)).$$

Then the parameter γ *is said to be preserved. The incidental parameter space* $\mathbf{N}$ *is preserved if*

$$\mathbf{N} = \{\eta \in \mathbf{N}; \eta = \mathcal{A}_3(g, \widetilde{\eta}) \text{ for some } \widetilde{\eta} \in \mathbf{N}\}.$$

Suppose that both γ and $\mathbf{N}$ are preserved. We then can appeal to the *invariance principle* and focus on invariant statistics $\phi(Y)$ in which $\phi(\mathcal{A}_1(g, Y)) = \phi(Y)$ for every $Y \in \mathcal{Y}$ and $g \in \mathbf{G}$. Any invariant statistic can be written as a function of a maximal invariant statistic defined below.

Definition 2 *A statistic* $M \equiv M(Y)$ *is a maximal invariant in the sample space if*

$$M(\widetilde{Y}) = M(Y) \text{ if and only if } \widetilde{Y} = \mathcal{A}_1(g, Y) \text{ for some } g \in G.$$

Comment: If M is a maximal invariant then $\widetilde{c} \cdot M$ is also a maximal invariant statistic (for any scalar $\widetilde{c} \neq 0$). This shows that the maximal invariant statistic is not unique.

An orbit of $\mathbf{G}$ is an equivalence class of elements Y, where $\widetilde{Y} \sim Y$ (mod $\mathbf{G}$) if there exists $g \in \mathbf{G}$ such that $\widetilde{Y} = \mathcal{A}_1(g, Y)$. By definition, M is a *maximal invariant* statistic if it is invariant and takes distinct values on different *orbits* of $\mathbf{G}$. Every invariant procedure can be written as a function of a maximal invariant. Hence, we restrict our attention to the class of decision rules that depend only on the maximal invariant statistic. An analogous definition holds for the parameter space.

Definition 3 *A parameter* $\theta \equiv \theta(\gamma, \eta)$ *is a maximal invariant in the parameter space if* $\theta(\gamma, \eta)$ *is invariant and takes different values on different* orbits *of* $\mathbf{G}$*:* $O_{\gamma,\eta} = \{\mathcal{A}_2(g, (\gamma, \eta)) \in \mathbf{\Gamma} \times \mathbf{N}; \text{for some } g \in \mathbf{G}\}$.

The distribution of a maximal invariant M depends on (γ, η) only through θ. If $\mathcal{A}_2 : \mathbf{G} \times \mathbf{\Gamma} \times \mathbf{N} \to \mathbf{\Gamma} \times \mathbf{N}$ induces a group action $\mathcal{A}_3 : \mathbf{G} \times \mathbf{N} \to \mathbf{N}$, then $\theta \equiv (\gamma, \lambda)$, where $\lambda \in \mathbf{\Lambda}$ is the maximal invariant in the nuisance parameter space $\mathbf{N}$. The parameter set $\mathbf{\Lambda}$ is allowed to be the empty set.

Definition 4 *Let* $f(M; \theta)$ *be the pdf/pmf of a maximal invariant statistic (we shall abbreviate* $f(M; \theta)$ *as the invariant likelihood). The maximum invariant likelihood estimator (MILE) is defined as*

$$\widehat{\theta} \equiv \arg\max_{\theta \in \Theta} f(M; \theta).$$

Comments: 1. Hereinafter, we assume the set Θ to be compact.

2. In general, different group actions $\mathcal{A}_1(\cdot, Y)$ and $\mathcal{A}_2(\cdot, (\gamma, \eta))$ yield different estimators. Hence, a better notation for $\widehat{\theta}$ would indicate its dependence on the choice of group actions. For brevity, we omit its dependence here.

3. Suppose that $G = \{1\}$, $\mathcal{A}_1(g, Y) = Y$, and $\mathcal{A}_3(g, \eta) = \eta$. Then $M = Y$ is a maximal invariant statistic and $\theta = (\gamma, \eta)$ is a maximal invariant parameter. This shows that MILE is a generalization of the MLE concept.

4. In general we seek group actions $\mathcal{A}_1(\cdot, Y)$ and $\mathcal{A}_2(\cdot, (\gamma, \eta))$ that preserve the model $\mathbf{P}$ and the structural parameter γ, and yield a maximal invariant λ in $\mathbf{N}$ which has fixed dimension with the sample size.

5. MILE is a marginal approach. The use of invariance suggests which likelihoods we should maximize.

We introduce some additional notation. The superscript $*$ indicates the true value of a parameter, e.g., γ^* is the true value of the structural parameter γ. The subscript N denotes dependence on the sample size N, e.g., λ_N^* is the true value of the maximal invariant λ when the sample size is N. In addition, let 1_T be a T-dimensional vector of ones, $O_{j\times k}$ be a $j \times k$ matrix with entries zero, e_j be a vector with entry j equals one and other entries zero.

Hereinafter, additional notation is specific to each example.

3 Transformations Within Individuals

In this section, we present three examples of transformations within individuals. Instead of $P_{\gamma,\eta}$, we work with P^i_{γ,η_i}, the probability of the model for agent i. This clarifies our exposition and highlights the fact that the likelihood of each maximal invariant $M = (M_1, ..., M_N)$ is the sum of marginal likelihoods. In all examples below, the maximal invariant in the parameter space is $\theta = \gamma$, with the objective function simplifying to

$$Q_N(\theta) = \frac{1}{N} \sum_{i=1}^{N} \ln f_i(m_i; \theta), \tag{1}$$

where $f_i(m_i; \theta)$ is the marginal density of the maximal invariant M_i for each individual i. Because the MILE $\widehat{\theta}_N$ maximizes $Q_N(\theta)$, consistency, asymptotic normality, and optimality of $\widehat{\theta}_N$ follow from standard results.

Lemma 1 *Let $Q_N(\theta)$ be defined as in (1) and take all limits as $N \to \infty$.*
(a) *Suppose that* (i) $\sup_{\theta\in\Theta}|Q_N(\theta) - Q(\theta)| \to_p 0$ *for a fixed, nonstochastic function $Q(\theta)$, and* (ii) $\forall\epsilon > 0$, $\inf_{\theta\notin B(\theta^*,\epsilon)} Q(\theta) > Q(\theta^*)$. *Then*

$$\widehat{\theta}_N \to_p \theta^*.$$

(b) *Suppose that* (i) $\widehat{\theta}_N \to_p \theta^*$, (ii) $\theta^* \in int(\Theta)$, (iii) $Q_N(\theta)$ *is twice continuously differentiable in some neighborhood of θ^*,* (iv) $\sqrt{N}\partial Q_N(\theta^*)/\partial\theta \to_d N(0, I(\theta^*))$, *and* (v) $\sup_{\theta\in\Theta}|\partial^2 Q_N(\theta^*)/\partial\theta\partial\theta' + I(\theta)| \to_p 0$ *for some nonstochastic matrix that is continuous at θ^* where $I(\theta^*)$ is nonsingular. Then*

$$\sqrt{N}(\widehat{\theta}_N - \theta^*) \to_d N(0, I(\theta^*)^{-1}).$$

(c) *Suppose that* (i) $\{Q_N(\theta); \theta \in \Theta\}$ *is differentiable in quadratic mean at θ^* with nonsingular information matrix $I(\theta^*)$, and* (ii) $\sqrt{N}(\widehat{\theta}_N - \theta^*) = I(\theta^*)^{-1}\sqrt{N}\partial Q_N(\theta^*)/\partial\theta + o_{Q_N(\theta^*)}(1)$. *Then*

$$\ln \frac{Q_N(\theta + h \cdot N^{-1/2})}{Q_N(\theta)} = h'S_N - \frac{1}{2}h'I(\theta^*)h + o_{Q_N(\theta^*)}(1),$$

where $S_N \to_d N(0, I(\theta^*))$ under $Q_N(\theta^*)$, and $\widehat{\theta}_N$ is the best regular invariant estimator of θ^*.

3.1 A Linear Stationary Panel Data Model

As an introductory example, consider a linear stationary panel data model with exogenous regressors and fixed effects:

$$y_{it} = \eta_i + x_{it}'\beta + u_{it},$$

where $y_{it} \in \mathbb{R}$ and $x_{it} \in \mathbb{R}^K$ are observable variables; u_{it} are unobservable (possibly autocorrelated) errors, $i = 1, ..., N$, $t = 1, ..., T$; $\eta_i \in \mathbb{R}$ are incidental parameters, $i = 1, ..., N$; and $\gamma = (\beta, \sigma^2) \in \mathbb{R}^K \times \mathbb{R}$ are the structural parameters.

The model for $y_{i\cdot} = [y_{i1}, ..., y_{iT}]' \in \mathbb{R}^T$ conditional on $x_{i\cdot} = [x_{i1}, ..., x_{iT}]' \in \mathbb{R}^{T\times K}$ is

$$y_{i\cdot} \overset{iid}{\sim} N\left(\eta_i 1_T + x_{i\cdot}\beta, \sigma^2\Sigma_T\right), \text{ where } \Sigma_T = \frac{1}{1-\rho^2}\begin{bmatrix} 1 & \rho & \cdots & \rho^{T-1} \\ \rho & 1 & & \\ \vdots & & \ddots & \\ \rho^{T-1} & & & 1 \end{bmatrix} \quad (2)$$

This is Example 3 of Lancaster (2002).

Both the model and the structural parameter $\gamma = (\beta, \sigma^2, \rho)$ are preserved by translations $g \cdot 1_T$ (where g is a scalar):

$$y_{i\cdot} + g \cdot 1_T \overset{iid}{\sim} N\left((\eta_i + g)1_T + x_{i\cdot}\beta, \sigma^2 \Sigma_T\right).$$

Proposition 1 *Let g be elements of the real line with $g_1 \circ g_2 = g_1 + g_2$. If the actions on the sample and parameter spaces are, respectively, $\mathcal{A}_1(g, y_{i\cdot}) = (y_{i\cdot} + g \cdot 1_T)$ and $\mathcal{A}_2(g, (\beta, \sigma^2, \rho, \eta_i)) = (\beta, \sigma^2, \rho, \eta_i + g)$, then*
(a) *the vector $M_i = Dy_{i\cdot}$ is a maximal invariant in the sample space, where D is a $T-1 \times T$ differencing matrix with typical row $(0, ..., 0, 1, -1, 0, ..., 0)$,*
(b) *γ is a maximal invariant in the parameter space, and*
(c) *$M_i \equiv M(y_{i\cdot}) \overset{iid}{\sim} N(Dx_{i\cdot}\beta, \sigma^2 D\Sigma_T D')$ with density at $m_i = Dy_{i\cdot}$ given by*

$$f_i\left(m_i; \beta, \rho, \sigma^2\right) = \left(2\pi\sigma^2\right)^{-\frac{(T-1)}{2}} \left|D\Sigma_T D'\right|^{-1/2}$$
$$\times \exp\left\{-\frac{1}{2\sigma^2}(y_{i\cdot} - x_{i\cdot}\beta)' D' \left(D\Sigma_T D'\right)^{-1} D(y_{i\cdot} - x_{i\cdot}\beta)\right\}.$$

Comments: 1. The density $f_i(m_i|\beta, \rho, \sigma^2)$ is free of the incidental parameter η_i (as it should be).

2. Under the assumption that $\frac{1}{N}\sum_{i=1}^{N} vec(x_{i\cdot})vec(x_{i\cdot})' \to_p \Omega_{XX}$ p.d., we can use Lemma 1 to show that $\widehat{\theta}_N$ is consistent and asymptotically normal.

3. Maximization of the invariant likelihood coincides with maximization of the integrated likelihood if the prior on η_i is left unrestricted, e.g., Arellano (2003, Section 2.4). The use of invariance gives an additional result, with $\widehat{\theta}_N = (\widehat{\beta}_N, \widehat{\sigma}_N^2, \widehat{\gamma}_N)$ being asymptotically optimal within the class of invariant regular estimators.

Finally, we give an example in which MILE may not be admissible. Suppose that ρ is known to be equal to zero. The model given by (2) simplifies to

$$y_{i\cdot} \overset{iid}{\sim} N\left(\eta_i 1_T + x_{i\cdot}\beta, \sigma^2 I_T\right), \tag{3}$$

which is Example 2 of Lancaster (2002). The Proposition 1(a),(b) still holds true. The density of M_i at $m_i = Dy_{i\cdot}$ is given by

$$f_i\left(m_i; \beta, \sigma^2\right) = \left(2\pi\sigma^2\right)^{-\frac{(T-1)}{2}} \left|DD'\right|^{-1/2}$$
$$\times \exp\left\{-\frac{1}{2\sigma^2}(y_{i\cdot} - x_{i\cdot}\beta)' D' \left(DD'\right)^{-1} D(y_{i\cdot} - x_{i\cdot}\beta)\right\}.$$

The MILE estimator $\widehat{\theta}_N = (\widehat{\beta}_N, \widehat{\sigma}^2_N)$ is given by

$$\widehat{\beta}_N = \frac{\sum_{i=1}^N x'_{i\cdot} D' \left(DD'\right)^{-1} D y_{i\cdot}}{\sum_{i=1}^N x'_{i\cdot} D' \left(DD'\right)^{-1} D x_{i\cdot}} \text{ and}$$

$$\widehat{\sigma}^2_N = \frac{1}{N\left(T-1\right)} \sum_{i=1}^N (y_{i\cdot} - x_{i\cdot}\widehat{\beta}_N)' D' \left(DD'\right)^{-1} D (y_{i\cdot} - x_{i\cdot}\widehat{\beta}_N).$$

The estimator $\widehat{\beta}_N$ is unbiased, but the estimator $\widehat{\sigma}^2_N$ is biased and not even admissible for a quadratic loss function. This example shows that the MILE method yields consistent, but not necessarily admissible estimators of structural parameters.[1]

3.2 A Linear Transformation Model

Consider a simple panel data transformation model:

$$\eta_i\left(y_{it}\right) = x'_{it}\beta + u_{it},$$

where $y_{it} \in \mathbb{R}$ and $x_{it} \in \mathbb{R}^K$ are observable variables; $u_{it} \in \mathbb{R}$ are unobservable errors, $i = 1, ..., N$, $t = 1, ..., T$, with $T > K$; $\eta_i : \mathbb{R} \to \mathbb{R}$ is an unknown, continuous, strictly increasing incidental function; and $\beta \in \mathbb{R}^K$ is the structural parameter. Unlike Abrevaya (2000), we shall parameterize the distribution of the errors: $u_{it} \overset{iid}{\sim} N\left(\alpha_i, \sigma^2\right)$. Because of location and scale normalizations, we shall assume without loss of generality that $u_{it} \overset{iid}{\sim} N\left(0, 1\right)$.

The model for $y_{i\cdot} = (y_{i1}, y_{i2}, ..., y_{iT}) \in \mathbb{R}^T$ is then given by

$$P\left(y_{i\cdot} \leq v\right) = \prod_{t=1}^T \Phi\left(\eta_i\left(v_t\right) - x'_{it}\beta\right), \text{ where } v = \left[v_1, v_2, ..., v_T\right]'.$$

Both the model and the structural parameter $\gamma \equiv \beta$ are preserved by continuous, strictly increasing transformations.

Proposition 2 *Let g be elements of the group of continuous, strictly increasing transformations, with $g_1 \circ g_2 = g_1(g_2)$. If the actions on the sample and parameter spaces are, respectively, $\mathcal{A}_1\left(g, (y_{i1}, y_{i2}, ..., y_{iT})\right) = (g(y_{i1}), g(y_{i2}), ..., g(y_{iT}))$ and $\mathcal{A}_2\left(g, (\beta, \eta_i)\right) = (\beta, \eta_i(g))$, then*

[1] We can of course fix this problem by finding the model for $Y = vec(y_{1\cdot}, ..., y_{N\cdot})$ and considering an action group that eliminates both the structural parameter β and the incidental parameters η_i, e.g., Harville (1974). This yields a likelihood whose maximum likelihood estimator of σ^2 is unbiased and consistent as $N \to \infty$.

(a) *the statistic* $M_i = (M_{i1}, ..., M_{iT})$ *is the maximal invariant in the sample space, where* M_{it} *is the rank of* y_{it} *in the collection* $y_{i1}, ..., y_{iT}$*,*
(b) *the vector* β *is the maximal invariant in the parameter space, and*
(c) M_i*,* $i = 1, ..., N$*, are independent with marginal probability mass function of* M_i *at* m_i *given by*

$$f_i\left(m_{i1}, ..., m_{iT}; \beta\right) = \frac{1}{T!} E\left[\exp\left\{\left(\sum_{t=1}^{T} V_{(m_{it})} x_{it}'\right)\beta\right\}\right] \exp\left\{-\frac{1}{2}\beta'\left(\sum_{t=1}^{T} x_{it} x_{it}'\right)\beta\right\},$$

where $V_{(1)}, ..., V_{(T)}$ *is an ordered sample from a* $N(0, 1)$ *distribution.*

The likelihood of the maximal invariant also yields semiparametric methods. For example, consider the case in which $T = 2$. If $x_{i2}'\beta > x_{i1}'\beta$, then it is likely that $y_{i2} > y_{i1}$. This yields the semiparametric estimator of Abrevaya (2000). This estimator maximizes

$$Q_n(\beta) = \frac{1}{N}\sum_{i=1}^{N}\left\{H\left(y_{i2}, y_{i1}\right) I\left(\triangle x_i'\beta > 0\right) + H\left(y_{i1}, y_{i2}\right) I\left(\triangle x_i'\beta < 0\right)\right\}$$

where H is an arbitrary function increasing in the first and decreasing in the second argument. This estimator is very appealing as it is consistent under more general error distributions. For asymptotic normality, Abrevaya (2000) proposes to smoothen the objective function to obtain asymptotic normality whose convergence rate can be made arbitrarily close to $N^{-1/2}$. In contrast, the MILE estimator suggested here does not require arbitrary choices of H or smoothing.

4 An Instrumental Variables Model

Consider a simple simultaneous equations model with one endogenous variable, multiple instrumental variables (IVs), and errors that are normal with known covariance matrix. The model consists of a structural equation and a reduced-form equation:

$$\begin{aligned} y_1 &= y_2\beta + u, \\ y_2 &= Z\pi + v_2, \end{aligned}$$

where $y_1, y_2 \in R^N$ and $Z \in R^{N\times K}$ are observed variables; $u, v_2 \in R^N$ are unobserved errors; and $\beta \in R$ and $\pi \in R^K$ are unknown parameters. The matrix Z has full column rank K; the $N \times 2$ matrix of errors $[u : v_2]$ is assumed to be iid across rows with each row having a mean zero bivariate normal distribution with a nonsingular covariance matrix; π is the incidental parameter; and β is the parameter of interest.

The two equation reduced-form model can be written in matrix notation as

$$Y = Z\pi a' + V,$$

where $Y = [y_1 : y_2]$, $V = [v_1 : v_2]$, and $a = (\beta, 1)'$. The distribution of $Y \in R^{N\times 2}$ is multivariate normal with mean matrix $Z\pi a'$, independence across rows, and covariance matrix Σ for each row.

Because the multivariate normal is a member of the exponential family of distributions, Moreira (2001) shows that low dimensional sufficient statistics are available for the parameter $(\beta, \pi')'$. Andrews, Moreira, and Stock (2006) and Chamberlain (2007) propose to use orthogonal transformations applied to the sufficient statistic $(Z'Z)^{-1/2} Z'Y$. The maximal invariant is $Y'N_Z Y$, where $N_Z = Z(Z'Z)^{-1}Z'$.

Reducing the data to a sufficient statistic before applying invariance is a delicate argument. For example, suppose that there is a (nearly) optimal invariant decision rule based on a sufficient statistic. This does not imply that it is (nearly) optimal within invariant decision rules based on the initial data. This problem arises because there may exist invariant decision rules whose equivalent procedures based on the sufficient statistic are not invariant. See for example Hall, Wijsman, and Ghosh (1965) and Lehmann and Romano (2005). To avoid this issue, we shall use an invariance argument without reducing the data to a sufficient statistic.

For convenience, it is useful to write the model in a canonical form. The matrix Z has the polar decomposition $Z = \omega(\rho', 0_{K\times(N-K)})'$, where ω is an $N \times N$ orthogonal matrix, and ρ is the unique symmetric, positive definite square root of $Z'Z$. Define $R = \omega' Y$ and let $\eta = \rho\pi$. Then the canonical model is

$$R \overset{d}{=} \begin{pmatrix} \eta a' \\ 0 \end{pmatrix} + V, \ \mathcal{L}(V) = N(0, I_N \otimes \Sigma).$$

Both model and structural parameters β and Σ are preserved by transformations $O(K)$ in the first K rows of R. The next proposition obtains the maximal invariants in the sample and parameter spaces.

Proposition 3 *Let g be elements of the orthogonal group of transformations $O(K)$ and partition the sample space $R = (R_1', R_2')'$, where R_1 is $K \times 2$ and R_2 is $(N-K) \times 2$. If the actions on the sample and parameter spaces are, respectively, $\mathcal{A}_1(g, R) =$*

$((gR_1)', R_2')'$ *and* $\mathcal{A}_2(g, (\beta, \Sigma, \eta)) = (\beta, \Sigma, g\eta)$, *then*
(a) *the maximal invariant in the sample space is* $M = (R_1'R_1, R_2)$, *and*
(b) *the maximal invariant in the parameter space is* $\theta_N = (\beta, \Sigma, \lambda_N)$, *where* $\lambda_N \equiv \eta'\eta/N$.

To illustrate the approach we assume for simplicity that Σ is known. Hence, we omit Σ from now on, e.g., $\theta_N = (\beta, \lambda_N)$.

The density of M is the product of the marginal densities of $R_1'R_1$ and R_2. Since R_2 is an ancillary statistic, we can focus on the marginal density of $R_1'R_1 \equiv Y'N_ZY$ in the maximization of the log-likelihood. As the density of $Y'N_ZY$ is not well-behaved as N goes to infinity, we work with the density of $W_N \equiv N^{-1}Y'N_ZY$ instead.

Theorem 1 *The density of* $W_N \equiv N^{-1}Y'N_ZY$ *evaluated at* w *is*

$$g(w; \beta, \lambda_N) = C_{1,K} \cdot N^K \cdot \exp\left(-\frac{N\lambda_N}{2}a'\Sigma^{-1}a\right)|\Sigma|^{-K/2}|w|^{\frac{K-3}{2}}\exp\left(-\frac{N}{2}tr(\Sigma^{-1}w)\right)$$
$$\times\left(N\sqrt{\lambda_N \cdot a'\Sigma^{-1}w\Sigma^{-1}a}\right)^{-\frac{K-2}{2}} I_{\frac{K-2}{2}}\left(N\sqrt{\lambda_N \cdot a'\Sigma^{-1}w\Sigma^{-1}a}\right), \tag{4}$$

where $C_{1,K}^{-1} = 2^{\frac{K+2}{2}}\pi^{\frac{1}{2}}\Gamma\left(\frac{K-1}{2}\right)$, $I_\nu(\cdot)$ *denotes the modified Bessel function of the first kind of order* ν, *and* $\Gamma(\cdot)$ *is the gamma function.*

Define MILE as

$$\widehat{\theta}_N \equiv \arg\max_{\theta \in \Theta} Q_N(\theta),$$

where $Q_N(\theta) \equiv N^{-1}\ln g(W_N; \theta_N)$ and $\theta_N = (\beta, \lambda_N)$.[2] The next result shows that $\widehat{\theta}_N = \theta_N^* + o_p(1)$ under general conditions.

Theorem 2 (a) *Under the assumption that* $N \to \infty$ *with* K *fixed or* $K/N \to 0$, (i) *if* λ_N^* *is fixed at* $\lambda^* > 0$, *then* $\widehat{\theta}_N \to_p \theta^* = (\beta^*, \lambda^*)$, (ii) *if* $\lambda_N^* \to_p \lambda^* > 0$, *then* $\widehat{\theta}_N \to_p \theta^* = (\beta^*, \lambda^*)$, *and* (iii) *if* $0 < \liminf \lambda_N^* \leq \limsup \lambda_N^* < \infty$, *then* $\widehat{\theta}_N = \theta_N^* + o_p(1)$.
(b) *Under the assumption that* $N \to \infty$ *with* $K/N \to \alpha > 0$, (i) *if* λ_N^* *is fixed at* $\lambda^* > 0$, *then* $\widehat{\theta}_N \to_p \theta^* = (\beta^*, \lambda^*)$, (ii) *if* $\lambda_N^* \to_p \lambda^* > 0$, *then* $\widehat{\theta}_N \to_p \theta^* = (\beta^*, \lambda^*)$, *and* (iii) *if* $0 < \liminf \lambda_N^* \leq \limsup \lambda_N^* < \infty$, *then* $\widehat{\theta}_N = \theta_N^* + o_p(1)$, *where* $\theta_N^* = (\beta^*, \lambda_N^*)$.

[2] The objective function $Q_N(\theta)$ is not defined if W_N is not positive definite (due to the term $\ln|W_N|$). To avoid this technical issue, we can instead maximize only the terms of $Q_N(\theta)$ that depend on θ.

Comments: 1. Parts (a),(b)(i) yield consistency results conditional on λ_N^*; the remaining results of the theorem are unconditional on λ_N^*. Parts (a),(b)(ii) yield consistency results for β^* under SIV and MWIV asymptotics when $\lambda_N^* \to_p \lambda^*$. The assumption of $\lambda_N^* \to_p \lambda^*$ is standard in the literature, but parts (a),(b)(iii) show that $\widehat{\beta}_N \to_p \beta_N^*$ without imposing convergence of λ_N^*.

2. This result also holds under nonnormal errors as long as $V(W_N) \to 0$.

Proposition 4 *MILE of β is the limited information maximum likelihood (LIMLK) estimator.*

Proposition 4 together with Theorem 2 explain why the LIMLK estimator is consistent when the number of instruments increases. The MILE estimator maximizes a log-likelihood function that is well-behaved as it depends on a finite number of parameters. Because MILE is consistent and LIMLK is equivalent to MILE in the instrumental variable problem, LIMLK is consistent as well.

The next result derives the limiting distribution of LIMLK.

Theorem 3 *Let the score statistic and the Hessian matrix be*

$$S_N(\theta) = \frac{\partial \ln Q_N(\theta)}{\partial \theta} \text{ and } H_N(\theta) = \frac{\partial^2 \ln Q_N(\theta)}{\partial \theta \partial \theta'},$$

respectively, and define the matrix

$$I_\alpha(\theta^*) = \begin{bmatrix} \lambda^{*2} \frac{a^{*\prime}\Sigma^{-1}a^* \cdot e_1'\Sigma^{-1}e_1(\alpha + 2\lambda^* a^{*\prime}\Sigma^{-1}a^*) + \alpha(a^{*\prime}\Sigma^{-1}e_1)^2}{(\alpha + \lambda^* a^{*\prime}\Sigma^{-1}a^*)(\alpha + 2\lambda^* a^{*\prime}\Sigma^{-1}a^*)} & \lambda^* \frac{a^{*\prime}\Sigma^{-1}e_1 \cdot a^{*\prime}\Sigma^{-1}a^*}{\alpha + 2\lambda^* a^{*\prime}\Sigma^{-1}a^*} \\ \lambda^* \frac{a^{*\prime}\Sigma^{-1}e_1 \cdot a^{*\prime}\Sigma^{-1}a^*}{\alpha + 2\lambda^* a^{*\prime}\Sigma^{-1}a^*} & \frac{(a^{*\prime}\Sigma^{-1}a^*)^2}{2(\alpha + 2\lambda^* a^{*\prime}\Sigma^{-1}a^*)} \end{bmatrix}.$$

(a) *Suppose that λ_N^* is fixed at $\lambda^* > 0$ and $N \to \infty$ with K fixed. Then* (i) $\sqrt{N} S_N(\theta^*) \to_d N(0, I_0(\theta^*))$, (ii) $H_N(\theta^*) \to_p -I_0(\theta^*)$, *and* (iii) $\sqrt{N}(\widehat{\theta}_N - \theta^*) \to_d N(0, I_0(\theta^*)^{-1})$.
(b) *Suppose that λ_N^* is fixed at $\lambda^* > 0$ and $N \to \infty$ with $K/N \to \alpha$. Then* (i) $\sqrt{N} S_N(\theta^*) \to_d N(0, I_\alpha(\theta^*))$, (ii) $H_N(\theta^*) \to_p -I_\alpha(\theta^*)$, *and* (iii) $\sqrt{N}(\widehat{\theta}_N - \theta^*) \to_d N(0, I_\alpha(\theta^*)^{-1})$.

Comments: 1. For convenience we provide asymptotic results only for the case in which λ_N^* is fixed at $\lambda^* > 0$. Small changes in the proofs also yield asymptotic results for $\lambda_N^* \to_p \lambda^*$. Alternatively, if the convergence for $\sqrt{N}(\widehat{\theta}_N - \theta^*)$ is uniform in a compact set containing λ^*, we can use Theorem 3 and Sweeting (1989) to show that $\sqrt{N}(\widehat{\theta}_N - \theta_N^*)$ converges to $N(0, I_\alpha(\theta^*))$.

2. If λ_N^* does not converge, then $\sqrt{N}(\widehat{\theta}_N - \theta_N^*)$ does not converge. However, if the conditional convergence for $\sqrt{N}(\widehat{\theta}_N - \theta_N^*)$ is uniform on a compact set that eventually contains λ_N^*, then $N(0, I_\alpha(\theta_N^*))$ provides an approximation to the finite sample distribution in the sense that $\sqrt{N}(\widehat{\theta}_N - \theta_N^*)$ conditioned on $\lambda_N^* = \lambda^*$ converges in distribution to $N(0, I_\alpha(\theta^*))$.

3. It is possible to extend the asymptotic distribution to nonnormal errors, e.g., Bekker and der Ploeg (2005), Hansen, Hausman, and Newey (2006), and van Hasselt (2007). Our approach entails finding the asymptotic distribution $\sqrt{N}S_N(\theta^*)$ for nonnormal errors.

As a corollary, we find the limiting distribution of LIMLK. This result of course coincides with those obtained by Bekker (1994).

Corollary 1 *Define $\sigma_u^2 = b'\Sigma b$. Under SIV asymptotics (or under MWIV asymptotics with $\alpha = 0$), conditional on $\lambda_N^* = \lambda^* > 0$,*

$$\sqrt{N}(\widehat{\beta}_N - \beta^*) \to_d N\left(0, \frac{\sigma_u^2}{\lambda^*}\right). \tag{5}$$

Under MWIV asymptotics, conditional on $\lambda_N^* = \lambda^* > 0$,

$$b\sqrt{N}(\widehat{\beta}_N - \beta^*) \to_d N\left(0, \frac{\sigma_u^2}{\lambda^{*2}}\left\{\lambda^* + \alpha\frac{1}{a^{*\prime}\Sigma^{-1}a^*}\right\}\right). \tag{6}$$

Comments: 1. The limiting distribution given in (6) simplifies to the one given in (5) as $\alpha \to 0$.

2. Instead of using the invariant likelihood to obtain an estimator, we could instead use only its first moment. Define

$$\overline{m}(W_N; \theta_N) = vech\left(\frac{R_1'R_1}{N}\right) - vech\left(aa' \cdot \lambda_N + \frac{K}{N}\Sigma\right). \tag{7}$$

If $\lambda_N^* > 0$, then the following holds (for possibly nonnormal errors):

$$E_{\theta_N^*}(\overline{m}(W_N; \theta)) = 0 \text{ if and only if } \theta_N = \theta_N^*. \tag{8}$$

Because the number of moment conditions does not increase under SIV or MWIV asymptotics, we can show that the MD estimator based on (7) and (8) is consistent and asymptotically normal.

Finally, Chioda and Jansson (2007) derive limits of experiments from the maximal invariant's likelihood. In our setup, we obtain the following result under SIV and MWIV asymptotics.

Theorem 4 *Define the log-likelihood ratio*

$$\Lambda_N\left(\theta^* + h \cdot N^{-1/2}, \theta^*\right) = N(Q_N\left(\theta^* + h \cdot N^{-1/2}\right) - Q_N\left(\theta^*\right)).$$

(a) *Under SIV asymptotics,*

$$\Lambda_N\left(\theta^* + h \cdot N^{-1/2}, \theta^*\right) = h'\sqrt{N}S_N\left(\theta^*\right) - \frac{1}{2}h'I_0\left(\theta^*\right)h + o_{Q_N(\theta^*)}(1), \tag{9}$$

where $\sqrt{N}S_N\left(\theta^*\right) \to_d N(0, I_0\left(\theta^*\right))$ *under* $Q_N\left(\theta^*\right)$.

(b) *Under MWIV asymptotics,*

$$\Lambda_N\left(\theta^* + h \cdot N^{-1/2}, \theta^*\right) = h'\sqrt{N}S_N\left(\theta^*\right) - \frac{1}{2}h'I_\alpha\left(\theta^*\right)h + o_{Q_N(\theta^*)}(1), \tag{10}$$

where $\sqrt{N}S_N\left(\theta^*\right) \to_d N(0, I_\alpha\left(\theta^*\right))$ *under* $Q_N\left(\theta^*\right)$.

Furthermore, the LIMLK estimator is asymptotically efficient within the class of regular invariant estimators under both SIV and MWIV asymptotics.

Comments: 1. Chioda and Jansson's (2007) proof uses Johnson and Kotz's (1970) asymptotic results for Wishart distributions. The standard literature on limit of experiments instead typically provides expansions around the score, e.g., Lehmann and Romano (2005). Theorem 3 shows that the score is asymptotically normal with variance given by the reciprocal of the inverse of the limit of the Hessian matrix. As the remainder terms are asymptotically negligible, (9) and (10) hold true.

2. Theorem 4 requires the assumption of normal errors. Anderson, Kunitomo, and Matsushita (2006) exploit the fact that W_N involves double sums (in terms of N and K) to obtain optimality results for nonnormal errors.

Under SIV asymptotics, the bound $(I_0\left(\theta^*\right)^{-1})_{11}$ for regular invariant estimators of β is the same as the one achieved by limit of experiments applied to the likelihood of Y. Hence, there is no loss of efficiency in focusing on the class of invariant procedures under SIV asymptotics.

The LIMLK achieves the bound $(I_\alpha\left(\theta^*\right)^{-1})_{11}$ under MWIV asymptotics. Proposition 4 and Theorem 4(b) explain why. Standard optimality results apply to an estimator that maximizes a (marginal) likelihood function that is locally asymptotically normal (LAN). Applying this principle to invariant likelihood delivers optimality of MILE (within the class of regular invariant estimators). Because LIMLK coincides with MILE, the LIMLK estimator must be optimal as well.

5 A Nonstationary Dynamic Panel Data Model

Consider a simple dynamic panel data model with fixed effects:

$$y_{i,t} = \rho y_{i,t-1} + \eta_i + u_{it},$$

where $y_{it} \in \mathbb{R}$ are observable variables and $u_{it} \overset{iid}{\sim} N(0, \sigma^2)$ are unobservable errors, $i = 1, ..., N$, $t = 1, ..., T$; $\eta_i \in \mathbb{R}$ are incidental parameters, $i = 1, ..., N$; $\theta = (\rho, \sigma^2) \in \mathbb{R}^K \times \mathbb{R}$ are structural parameters; and $y_{i,0}$ are the initial values of the stochastic process. We follow Lancaster (2002) and seek inference conditional on the initial values $y_{i,0}$. Writing the model as

$$(y_{i,t} - y_{i,0}) = \rho(y_{i,t-1} - y_{i,0}) + (\eta_i - y_{i,0}(1-\rho)) + u_{it},$$

we can assume that $y_{i,0} = 0$ without loss of generality.

In its matrix form, we have

$$[y_{\cdot 1}, y_{\cdot 2}, ..., y_{\cdot T}] = \rho\,[y_{\cdot 0}, y_{\cdot 1}, ..., y_{\cdot T-1}] + \eta 1_T' + [u_{\cdot 1}, u_{\cdot 2}, ..., u_{\cdot T}]\,, \tag{11}$$

where $y_{\cdot t} = [y_{1,t}, y_{2,t}, ..., y_{N,t}]' \in \mathbb{R}^N$, $u_{\cdot t} = [u_{1,t}, u_{2,t}, ..., u_{N,t}]' \in \mathbb{R}^N$, and $\eta = [\eta_1, ..., \eta_N]' \in \mathbb{R}^N$. Solving (11) recursively yields

$$[y_{\cdot 1}, y_{\cdot 2}, ..., y_{\cdot T}] = \eta(B1_T)' + [u_{\cdot 1}, u_{\cdot 2}, ..., u_{\cdot T}]\,B', \text{ where} \tag{12}$$

$$B = \begin{bmatrix} 1 & & \\ \vdots & \ddots & \\ \rho^{T-1} & \cdots & 1 \end{bmatrix}.$$

The inverse of B has a simple form:

$$B^{-1} \equiv D = I_T - \rho \cdot J_T, \text{ where } J_T = \begin{bmatrix} 0_{T-1}' & 0 \\ I_{T-1} & 0_{T-1} \end{bmatrix}$$

and 0_{T-1} is a $T-1$-dimensional column vector with zero entries.

If individuals i are treated equally, the coordinate system used to specify the vectors $y_{\cdot t}$ should not affect inference based on them. In consequence, it is reasonable to restrict attention to coordinate-free functions of $y_{\cdot t}$. Indeed, we find that orthogonal transformations preserve both the model given in (12) and the structural parameter $\gamma = (\rho, \sigma^2)$.

Proposition 5 *Let g be elements of the orthogonal group of transformations $O(N)$. If the actions on the sample and parameter spaces are, respectively, $\mathcal{A}_1(g,R) = ((gR_1)', R_2')'$ and $\mathcal{A}_2(g,(\rho,\sigma^2,\eta)) = (\rho,\sigma^2,g\eta)$, then*
(a) *the maximal invariant in the sample space is $M = Y'Y$, and*
(b) *the maximal invariant in the parameter space is $\theta_N = (\gamma, \lambda_N)$, where $\lambda_N = \eta'\eta/(N\sigma^2)$.*

Comments: 1. The dimension of the maximal invariant M is $T(T+1)/2$. For example, if $T = 2$, the maximal invariant has dimension three.

2. The maximal invariant M has a noncentral Wishart distribution with T degrees of freedom, covariance matrix $\Sigma = \sigma^2 BB'$, and noncentrality matrix $\Omega = \Sigma^{-1}\overline{M}'\overline{M}$ where $\overline{M} = \eta(B1_T)'$. We write that M is $W_T(K,\Sigma,\Omega)$. If there is autocorrelation Σ_T that is homogeneous across individuals, the maximal invariant M remains the same. The covariance matrix however changes to $\Sigma = \sigma^2 B\Sigma_T B'$.

For convenience, we standardize the distribution of $M = Y'Y$.

Theorem 5 *If $N \geq T$, the density of $W_N \equiv N^{-1}Y'N_Z Y$ evaluated at w is*

$$g\left(w;\rho,\sigma^2,\lambda_N\right) = C_{2,N}\cdot\left(\sigma^2\right)^{-\frac{NT}{2}}|w|^{\frac{N-T-1}{2}}\exp\left(-\frac{1}{2\sigma^2}tr(DwD')\right)\exp\left(-\frac{NT}{2}\lambda_N\right)$$
$$\times\left(N\sqrt{\lambda_N\frac{1_T'DwD'1_T}{\sigma^2}}\right)^{-\frac{K-2}{2}} I_{\frac{K-2}{2}}\left(N\sqrt{\lambda_N\frac{1_T'DwD'1_T}{\sigma^2}}\right)\cdot N^{\frac{NT}{2}}, \qquad (13)$$

where $C_{2,N}^{-1} = 2^{\frac{NT}{2}-\frac{N-2}{2}}\pi^{\frac{T(T-1)}{4}}\prod_{i=1}^{T-1}\Gamma\left(\frac{N-i}{2}\right)$.

Define MILE as

$$\widehat{\theta}_N \equiv \arg\max_{\theta\in\Theta} Q_N(\theta),$$

where $Q_N(\theta) \equiv (NT)^{-1}\ln g(W_N;\beta,\lambda)$ and $\theta_N = (\rho,\sigma^2,\lambda_N)$.[3] The next result shows that $\widehat{\theta}_N = \theta_N^* + o_p(1)$ under general conditions.

Theorem 6 (a) *Under the assumption that $N \to \infty$ with T fixed,* (i) *if λ_N^* is fixed at $\lambda^* > 0$, then $\widehat{\theta}_N \to_p \theta^* = (\rho^*,\sigma^{*2},\lambda^*)$,* (ii) *if $\lambda_N^* \to_p \lambda^* > 0$, then $\widehat{\theta}_N \to_p \theta^* = (\rho^*,\sigma^{*2},\lambda^*)$, and* (iii) *if $\limsup\lambda_N^* < \infty$, then $\widehat{\theta}_N = \theta_N^* + o_p(1)$, where*

[3] If $N < T$, W_N is not absolutely continuous with respect to the Lebesgue measure. We will still maximize the pseudo-likelihood to find $\widehat{\theta}_N$.

$\theta_N^* = (\rho^*, \sigma^{*2}, \lambda_N^*)$.
(b) *Under the assumption that* $T \to \infty$ *and* $|\rho^*| < 1$, (i) *if* λ_N^* *is fixed at* $\lambda^* > 0$, *then* $\widehat{\theta}_N \to_p \theta^* = (\beta^*, \lambda^*)$, (ii) *if* $\lambda_N^* \to_p \lambda^* > 0$, *then* $\widehat{\theta}_N \to_p \theta^* = (\beta^*, \lambda^*)$, *and* (iii) *if* $\limsup \lambda_N^* < \infty$, *then* $\widehat{\theta}_N = \theta_N^* + o_p(1)$, *where* $\theta_N^* = (\rho^*, \sigma^{*2}, \lambda_N^*)$.

Comments: 1. This result also holds under nonnormal errors.

2. This theorem implies that $\widehat{\rho}_N \to_p \rho^*$ under the assumption that $NT \to \infty$ (regardless of the growing rate of N and T).

The next result derives the limiting distribution of MILE when $N \to \infty$.

Theorem 7 *Suppose that* λ_N^* *is fixed at* $\lambda^* > 0$, *and let the score statistic and the Hessian matrix be*

$$S_N(\theta) = \frac{\partial \ln Q_N(\theta)}{\partial \theta} \text{ and } H_N(\theta) = \frac{\partial^2 \ln Q_N(\theta)}{\partial \theta \partial \theta'},$$

respectively, and define the matrix

$$I_T(\theta^*) = \begin{bmatrix} h_{1,T} + h_{2,T} + h_{3,T} & \frac{\lambda^*}{2\sigma^{*2}} \frac{1_T' F 1_T}{T} & \frac{1+\lambda^* T}{1+2\lambda^* T} \frac{1_T' F 1_T}{T} \\ \frac{\lambda^*}{2\sigma^{*2}} \frac{1_T' F 1_T}{T} & \frac{1}{2(\sigma^{*2})^2} + \frac{\lambda^*}{4\sigma^{*2}} \frac{2\lambda^* T}{1+2\lambda^* T} & \frac{1}{4\sigma^{*2}} \\ \frac{1+\lambda^* T}{1+2\lambda^* T} \frac{1_T' F 1_T}{T} & \frac{1}{4\sigma^{*2}} & \frac{1}{4\lambda^*} \end{bmatrix},$$

where $DB^* \equiv I_T + (\rho^* - \rho) F$ *and the three terms in the* $(1,1)$ *entry of* $H_T(\theta^*)$ *are*

$$h_{1,T} = \frac{tr(FF')}{T} + \lambda^* \frac{1_T' F' F 1_T}{T}, \; h_{2,T} = \frac{2\lambda^{*2}}{(1+2\lambda^* T)} \frac{(1_T' F 1_T)^2}{T}, \text{ and}$$
$$h_{3,T} = -\frac{\lambda^*}{1+\lambda^* T} \left\{ \frac{1_T' F' F 1_T}{T} + \lambda^* \frac{(1_T' F 1_T)^2}{T} \right\}.$$

As $N \to \infty$ *with* T *fixed,*
(a) (i) $\sqrt{NT} S_N(\theta) \to_d N(0, I_T(\theta^*))$, (ii) $H_N(\theta^*) \to_p -I_T(\theta^*)$, *and* (iii) $\sqrt{NT}(\widehat{\theta}_N - \theta^*) \to_d N(0, I_T(\theta^*)^{-1})$, *and*
(b) *the log-likelihood ratio is*

$$\begin{aligned} \Lambda_N\left(\theta^* + h \cdot (NT)^{-1/2}, \theta^*\right) &= NT(Q_N\left(\theta^* + h \cdot (NT)^{-1/2}\right) - Q_N(\theta^*)) \\ &= h'\sqrt{NT} S_N(\theta^*) - \frac{1}{2} h' I_T(\theta^*) h + o_{Q_N(\theta^*)}(1), \end{aligned} \quad (14)$$

$\sqrt{NT} S_N(\theta^*) \to_d N(0, I_T(\theta^*))$ *under* $Q_N(\theta^*)$. *Furthermore,* $\widehat{\theta}_N$ *is asymptotically efficient within the class of regular invariant estimators under large* N, *fixed* T *asymptotics.*

Comments: 1. If the convergence is uniform on a compact set that eventually contains λ_N^*, $N(0, H_\alpha(\theta_N^*))$ provides an approximation of the finite sample distribution of $\sqrt{NT}(\widehat{\theta}_N - \theta_N^*)$ in the sense of Sweeting (1989). Because $\widehat{\theta}_N = \theta_N^* + o_p(1)$ and $I_T(\cdot)$ is continuous, $N(0, I_T(\widehat{\theta}_N))$ also provides a valid asymptotic approximation to the distribution of $\sqrt{NT}(\widehat{\theta}_N - \theta_N^*)$.

2. It is possible to extend parts (a)(i),(iii) to nonnormal errors by finding the appropriate asymptotic distribution of $\sqrt{NT}S_N(\theta^*)$.

3. The MILE estimator $\widehat{\rho}_N$ achieves the bound $\left(I_T(\theta^*)^{-1}\right)_{11}$ as $N \to \infty$, whereas the bias-corrected OLS estimator does not.

The next proposition considers minimum distance (MD) estimation based on the expectation of W_N; standard semiparametric efficiency arguments (e.g., Chamberlain (1987)) show that the MD estimator is optimal. This proposition also provides a connection between the GMM and integrated likelihood approaches for the dynamic panel data model. It shows that Arellano and Bond's (1991) and Ahn and Schmidt's (1995) moment conditions are transformations of the expectation of the maximal invariant. This result connects and builds on work by Chamberlain and Moreira (2006) who show that the likelihood integrated with respect to the Haar measure (for orthogonal groups) coincides with the marginal likelihood of the maximal invariant.

Proposition 6 *Let $w_i = y_{i\cdot}y_{i\cdot}'$, where $y_{i\cdot} = [y_{i,1}, y_{i,2}, ..., y_{i,T}]' \in \mathbb{R}^T$, and define*

$$\overline{m}(W_N; \theta_N) = \frac{1}{N}\sum_{i=1}^{N} m\left(w_i; \rho, \sigma^2, (\eta_i/\sigma)^2\right) \text{ where} \tag{15}$$

$$m\left(w_i; \rho, \sigma^2, (\eta_i/\sigma)^2\right) = vech\left(w_i - \sigma^2 B\left\{I_T + \left(\frac{\eta_i}{\sigma}\right)^2 \cdot 1_T 1_T'\right\} B\right).$$

(a) *Arellano and Bond's (1991) and Ahn and Schmidt's (1995) moment conditions are subsets of the $T(T+1)/2$ moment conditions given by*

$$\begin{aligned} E_{\theta_N^*}\left(\overline{m}(W_N; \theta_N)\right) &= E_{\theta_N^*}\left(vech\left(W_N - \sigma^2 B\left\{I_T + \lambda_N \cdot 1_T 1_T'\right\} B\right)\right) \\ &= 0 \text{ if and only if } \theta_N = \theta_N^*. \end{aligned} \tag{16}$$

(b) *Consider the minimum distance (MD) estimator $\widetilde{\theta}_N$ that minimizes*

$$Q(\theta_N) = \overline{m}(W_N; \theta_N)' A_N \overline{m}(W_N; \theta_N). \tag{17}$$

Under the assumptions $N \to \infty$ with T fixed, $A_N \to_p A$ p.d., and λ_N^* is fixed at λ^*, $\widetilde{\theta}_N \to_p \theta^* = (\rho^*, \sigma^{*2}, \lambda^*)$ and $\sqrt{N}(\widetilde{\theta}_N - \theta^*) \to_d N(0, (\zeta' A\zeta)^{-1}\zeta' A \Xi A\zeta(\zeta' A\zeta)^{-1})$, where Ξ and ζ are defined as

$$\sqrt{N}\overline{m}(W_N; \theta^*) \to_d N(0, \Xi) \text{ and } \frac{\partial \overline{m}(W_N; \theta^*)}{\partial \theta} \to_p \zeta.$$

(c) *The optimal MD estimator* $\widetilde{\theta}_N$ *achieves the semiparametric efficiency bound derived under the assumption that* $(\eta_i^*/\sigma^*)^2$, $i = 1, ..., N$, *are fixed at* λ^*.

Comments: 1. The additional random effects assumption $\eta_i \overset{iid}{\sim} N\left(0, \sigma_\eta^2\right)$ specifies a data covariance structure that depends on a finite number of parameters. Specifically, $y_{i\cdot} \overset{iid}{\sim} N\left(0, \Psi\left(\rho, \sigma^2, \sigma_\eta^2\right)\right)$ for some covariance matrix Ψ that depends on ρ, σ^2, and σ_η^2, and we can proceed as in Arellano (2003, Section 5.4) to make inference on ρ. This approach differs from ours. We do not impose additional distribution assumptions. As a result, the distribution of $y_{i\cdot}$ depends on ρ, σ^2, and η_i^2, $i = 1, ..., N$. Use of invariance, however, shows that the expectation of sample averages of $w_i = y_{i\cdot} y_{i\cdot}'$ depends on only three parameters: ρ, σ^2, and λ_N.

2. For $T = 2$, the number of nonredundant moments given by (16) equals the dimension of θ_N, and the parameter θ_N is said to be just-identified.

3. The MD estimator dominates MILE under nonnormal errors with large N and small T. For large T, the MD estimator does not perform well. If T grows sufficiently fast with the sample size, the MD estimator is no longer consistent. Consistency of MILE does not depend on particular rates at which both N and T grow with the sample size.

4. If there is autocorrelation Σ_T that is homogeneous across individuals, the maximal invariant remains the same, but (16) changes to

$$E_{\theta_N^*}\left(\overline{m}\left(W_N; \theta_N\right)\right) = E_{\theta_N^*}\left(vech\left(W_N - B\left\{\Sigma_T + \frac{\eta'\eta}{N} \cdot 1_T 1_T'\right\} B\right)\right).$$

In the IV model, the number of moment conditions does not increase with N or K ; see Comment 2 to Corollary 1. In the panel data model, the number of moment conditions increases (too quickly) with T. As a result, semiparametric efficiency results (e.g., Newey (2004)) do not apply to (16) as $T \to \infty$. Instead, Hahn and Kuersteiner (2002) cleverly use Hájek's convolution theorem to obtain an efficiency bound for normal errors as $T \to \infty$ for the stationary case $|\rho^*| < 1$. The bias-corrected OLS estimator of ρ achieves Hahn and Kuersteiner's (2002) bound for large N, large T asymptotics.

Our efficiency bound $\left(I_T\left(\theta^*\right)^{-1}\right)_{11}$ reduces to Hahn and Kuersteiner's (2002) bound when $T \to \infty$. This shows that there is no loss of efficiency in focusing on the class of invariant procedures under large N, large T asymptotics.

Corollary 2 *Under the assumption that* $|\rho^*| < 1$*, the efficiency bound given by the* $(1,1)$ *coordinate of the inverse of* $I_\infty(\theta^*)^{-1} \equiv \left(\lim_{T\to\infty} I_T(\theta^*)\right)^{-1}$ *converges to Hahn and Kuersteiner's (2002) efficiency bound of* $(1-\rho^{*2})$ *as* $T \to \infty$*.*

As a final result, the MILE estimator $\widehat{\rho}_N$ also achieves the bound $\left(I_T(\theta^*)^{-1}\right)_{11}$ for large N, large T asymptotics.

Theorem 8 *Under the assumption that* $N \geq T \to \infty$, $|\rho^*| < 1$, *and* λ_N^* *is fixed at* $\lambda^* > 0$, (i)$\sqrt{NT}S_N(\theta) \to_d N(0, I_\infty(\theta^*))$, (ii) $H_N(\theta^*) \to_p -I_\infty(\theta^*)$, *and* (iii) $\sqrt{NT}(\widehat{\theta}_N - \theta^*) \to_d N(0, I_\infty(\theta^*)^{-1})$.

6 Numerical Results

This section illustrates the MILE approach for estimation of the autoregressive parameter ρ in the dynamic panel data model described in Section 5. The numerical results are presented as means and mean squared errors (MSEs) based on 1,000 Monte Carlo simulations. These results are also available for other fixed-effects estimators: Arellano-Bond (AB), Ahn-Schmidt (AS), and bias-corrected OLS (BCOLS) estimators.

We consider different combinations between short and large panels: $N = 5$, 10, 25, 100, and $T = 2$, 3, 5, 10, 25, 100.

Table I presents the initial design from which several variations are drawn.[4] This design assumes that $\eta_i^* \overset{iid}{\sim} N(0,4)$ (random effects), $u_{it} \overset{iid}{\sim} N(0,1)$ (normal errors), and $\rho^* = 0.5$ (positive autocorrelation). The value σ^* is fixed at one for all designs.

MILE seems to be correctly centered around 0.5. Even in a very short panel with $N = 5$ and $T = 2$, its bias of 0.0408 seems quite small. As N and/or T increases, its mean approaches 0.5. For example, for $N = 5$ and $T = 25$, the bias is around 0.0129; for $N = 25$ and $T = 2$, the simulation mean is around 0.0040. BCOLS estimator seems to have smaller bias than the AB and AS estimators for small N and large T. The AB and AS estimators have large bias with small N and T, but their performance improves with large N and small T.

MILE also seems to have smaller MSE than the other estimators. The AS estimator outperforms the AB estimator in terms of MSE. The BCOLS estimator has

[4] The full set of results for ρ, σ^2, and λ_N using different designs will be available at http://www.economics.harvard.edu/faculty/moreira/softwaresimulations.html.

smaller MSE than AS. The MSE of the BCOLS estimator, however, does not decrease if N increases but T is held constant. For $T \geq 25$, its performance is comparable to that of MILE. This provides numerical support for the theoretical finding that both MILE and BCOLS reach our large N, large T bound.

Table II reports results for $\lambda_N^* = N$ (nonconvergent effects), normal errors, and $\rho^* = 0.5$. Table III presents results for random effects, $u_{it} \overset{iid}{\sim} (\chi^2(1) - 1)/\sqrt{2}$ (nonnormal errors), and $\rho^* = 0.5$. In both cases, MILE continues to have smaller bias and MSE than the other estimators. This result is surprising with nonnormal errors as the AB and AS estimators could potentially dominate MILE when N is large and T is small.

Tables IV and V differ from Table I only in the autoregressive parameter; respectively, $\rho^* = -0.5$ (negative autocorrelation) and $\rho^* = 1.0$ (integrated model). Most —but not all— conclusions drawn from Table I hold here. MILE continues to outperform the AB and AS estimators in terms of mean and MSE. If $\rho^* = -0.5$, MILE and BCOLS seem to perform similarly. If $\rho^* = 1.0$, MILE again performs better than BCOLS for small values of T.

7 Appendix of Proofs

7.1 Proofs of Results Stated in Section 3

Proof of Lemma 1. Parts (a) and (b) follow from Newey and McFadden (1994) or Potscher and Prucha (1997). Part (c) follows from Theorem 12.2.3 of Lehmann and Romano (2005) and Lemma 8.14 of van der Vaart (1998).

Proof of Proposition 1. For part (a), we need to show that $M(y_{i\cdot}) = M(\widetilde{y}_{i\cdot})$ if and only if $\widetilde{y}_{i\cdot} = y_{i\cdot} + \widetilde{g} \cdot 1_T$ for some $\widetilde{g}$. Clearly, $M(y_{i\cdot})$ is an invariant statistic:

$$M(y_{i\cdot} + g \cdot 1_T) = D(y_{i\cdot} + g \cdot 1_T) = Dy_{i\cdot} + g \cdot D1_T = Dy_{i\cdot} = M(y_{i\cdot}).$$

Now, suppose that $M(y_{i\cdot}) = M(\widetilde{y}_{i\cdot})$. This implies that $Dz_i = 0$ for $z_i = \widetilde{y}_{i\cdot} - y_{i\cdot}$, which means that z_i belongs to the space orthogonal to the row space of D. Because $rank(D) = T - 1$, the orthogonal space has dimension one. As this space contains the vector 1_T, it must be the case that $z_i = \widetilde{g} \cdot 1_T$ for some scalar $\widetilde{g}$. Therefore, $\widetilde{y}_{i\cdot} = y_{i\cdot} + \widetilde{g} \cdot 1_T$.

Part (b) follows from the fact that the group of transformations acts transitively on η_i. Part (c) follows from the formula of the density of a normal distribution.

Proof of Proposition 2. For part (a), let M_{it} be the rank of y_{it} in the collection $y_{i1}, ..., y_{iT}$. Formally, we can define M_{it} through $y_{it} = y_{i(M_{it})}$. We shall abbreviate the notation, e.g., $(g(y_{i1}), g(y_{i2}), ..., g(y_{iT}))$ as $g(y_{i\cdot})$. The maximal invariant is $M_i = (M_{i1}, ...M_{iT}) = M(y_{i\cdot})$. We need to show that $M(y_{i\cdot}) = M(\widetilde{y}_{i\cdot})$ if and only if $\widetilde{y}_{i\cdot} = \widetilde{g}(y_{i\cdot})$. Consider the case that if $t \neq \widetilde{t}$, then $y_{it} \neq y_{i\widetilde{t}}$ (this set has probability measure equal to one). Clearly, M_i is an invariant statistic. Now, suppose that $M(y_{i\cdot}) = M(\widetilde{y}_{i\cdot})$. This implies that $M_{i1} = \widetilde{M}_{i1}, ..., M_{iT} = \widetilde{M}_{iT}$. Therefore, $y_{i1} < ... < y_{iT}$ and $\widetilde{y}_{i1} < ... < \widetilde{y}_{iT}$. There is a continuous, strictly increasing transformation $\widetilde{g}$ such that $\widetilde{y}_{it} = \widetilde{g}(y_{it})$, $t = 1, ..., T$.

Part (b) follow from the fact that the group of transformations acts transitively on η_i.

For part (c), we note that because η_i is an increasing transformation, M_{it} is also the rank in the collection $y^*_{i1}, ..., y^*_{iT}$, where $y^*_{it} = x'_{it}\beta + u_{it}$. We note that $y^*_{i1}, ..., y^*_{iT}$ are jointly independent with marginal densities

$$f_{it}(z_{it}; \beta) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(z_{it} - x'_{it}\beta)^2\right\}.$$

Now, we note that

$$P(M_{i1} = m_{i1}, ..., M_{iT} = m_{iT}) = \int ... \int f_{i1}(z_{i1}; \beta) ... f_{iT}(z_{iT}; \beta)\, dz_{i1} ... dz_{iT},$$

integrated over the set in which z_{it} is the m_{it}-th smallest element of $z_{i1}, ..., z_{iT}$. Transforming $w_{m_{it}} = z_{it}$, we obtain

$$P\left(M_{i1} = m_{i1}, ..., M_{iT} = m_{iT}\right) = \int_A \prod_{t=1}^{T} f_{it}\left(w_{m_{it}}; \beta\right) dw = \int_A \prod_{t=1}^{T} \frac{f_{it}\left(w_{m_{it}}; \beta\right)}{f\left(w_{m_{it}}\right)} f\left(w_{m_{it}}\right) dw,$$

where $f\left(w_t\right)$ is the density of a $N\left(0,1\right)$ distribution and $A = \left\{w \in \mathbb{R}^T; w_1 < ... < w_T\right\}$. Simple algebraic manipulations show that

$$P\left(M_i = m_i\right) = \int_A \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\left(w_{m_{it}} - x_{it}'\beta\right)^2 + \frac{1}{2}\sum_{t=1}^{T} w_{m_{it}}^2\right\} \prod_{t=1}^{T} f\left(w_{m_{it}}\right) dw$$

$$= \int_A \exp\left\{\sum_{t=1}^{T} w_{m_{it}} x_{it}'\beta - \frac{1}{2}\sum_{t=1}^{T}\left(x_{it}'\beta\right)^2\right\} \prod_{t=1}^{T} f\left(w_{m_{it}}\right) dw$$

$$= \frac{1}{T!}\int_A \exp\left\{\left(\sum_{t=1}^{T} w_{m_{it}} x_{it}'\right)\beta - \frac{1}{2}\beta'\left(\sum_{t=1}^{T} x_{it} x_{it}'\right)\beta\right\} T! \prod_{t=1}^{T} f\left(w_{m_{it}}\right) dw,$$

where $T! \prod_{t=1}^{T} f\left(w_t\right)$ for $w_1 < ... < w_T$ is the pdf of the order statistics $V_{(1)}, ..., V_{(T)}$.

7.2 Proofs of Results Stated in Section 4

Proof of Proposition 3. For part (a), we need to show that $M\left(R_1, R_2\right) = M(\widetilde{R}_1, \widetilde{R}_2)$ if and only if $(\widetilde{R}_1, \widetilde{R}_2) = (\widetilde{g} R_1, R_2)$ for some $\widetilde{g} \in O\left(K\right)$. Clearly, $M\left(y_{i\cdot}\right)$ is an invariant statistic:

$$M\left(gR_1, R_2\right) = \left(R_1' g' g R_1, R_2\right) = \left(R_1' R_1, R_2\right) = M\left(R_1, R_2\right).$$

Now, suppose that $M\left(R_1, R_2\right) = M(\widetilde{R}_1, \widetilde{R}_2)$. This is equivalent to $R_1' R_1 = \widetilde{R}_1' \widetilde{R}_1$ and $R_2 = \widetilde{R}_2$. But this implies that $\widetilde{R}_1 = \widetilde{g} R_1$ (and, of course, $R_2 = \widetilde{R}_2$).

Part (b) follows analogously.

Proof of Theorem 1. The matrix M has a noncentral Wishart distribution with K degrees of freedom, covariance matrix Σ, and noncentrality matrix $\Omega = \Sigma^{-1}\overline{M}'\overline{M}$ where $\overline{M} = (Z'Z)^{1/2}\pi a'$. We write that M is $W_2(K, \Sigma, \Omega)$. Following Anderson (1946), the density function of M at q is

$$f\left(q\right) = C_{1,K} \cdot \exp\left(-\frac{1}{2} tr(\Sigma^{-1}\overline{M}'\overline{M})\right) \left|\Sigma\right|^{-K/2} \left|q\right|^{\frac{K-3}{2}} \exp\left(-\frac{1}{2} tr(\Sigma^{-1} q)\right)$$

$$\times \left(\sqrt{tr(q\Sigma^{-1}\overline{M}'\overline{M}\Sigma^{-1})}\right)^{-\frac{K-2}{2}} I_{\frac{K-2}{2}}\left(\sqrt{tr(q\Sigma^{-1}\overline{M}'\overline{M}\Sigma^{-1})}\right).$$

Using the fact that $\Sigma^{-1}\overline{M}'\overline{M} = \Sigma^{-1}a\pi'Z'Z\pi a'$, we obtain

$$tr(\Sigma^{-1}\overline{M}'\overline{M}) = (N\lambda_N)a'\Sigma^{-1}a \text{ and } tr(q\Sigma^{-1}a\pi'Z'Z\pi a'\Sigma^{-1}) = (N\lambda_N)a'\Sigma^{-1}q\Sigma^{-1}a.$$

As a result, the density function of M at q simplifies to

$$f(q) = C_{1,K}\cdot\exp\left(-\frac{N\lambda_N}{2}a'\Sigma^{-1}a\right)|\Sigma|^{-K/2}|q|^{\frac{K-3}{2}}\exp\left(-\frac{1}{2}tr(\Sigma^{-1}q)\right)$$
$$\times\left(\sqrt{N\lambda_N\cdot a'\Sigma^{-1}q\Sigma^{-1}a}\right)^{-\frac{K-2}{2}} I_{\frac{K-2}{2}}\left(\sqrt{N\lambda_N\cdot a'\Sigma^{-1}q\Sigma^{-1}a}\right).$$

The density function of W_N is then

$$g(w;\beta,\lambda_N) = f(q(w))\cdot|q'(w)| = f(q(w))\,N^{\frac{2\cdot3}{2}},$$

which simplifies to (4).

Proof of Theorem 2. The log-likelihood function divided by N is

$$Q_N(\theta) = -\frac{1}{2}\lambda\cdot a'\Sigma^{-1}a + \frac{1}{N}\ln\left(Z_N^{-\frac{K-2}{2}} I_{\frac{K-2}{2}}\left(\frac{N}{2}Z_N\right)\right) \quad (18)$$
$$-\frac{K}{2N}\ln|\Sigma| - \frac{K-3}{2N}\ln|W_N| - \frac{1}{2}tr(\Sigma^{-1}W_N) + \frac{1}{N}\ln(2^{\frac{K-2}{2}}N^{\frac{3K+2}{4}}C_{1,K}),$$

where $Z_N = 2\sqrt{\lambda\cdot a'\Sigma^{-1}W_N\Sigma^{-1}a}$.

All terms in the second line converge under both SIV and MWIV asymptotics (the only exception is $\ln|W_N|$ under SIV asymptotics and under MWIV asymptotics with $\alpha = 0$). For example, the last term is

$$\frac{1}{N}\ln\left(2^{\frac{K-2}{2}}N^{\frac{K+2}{2}}C_{1,K}\right) = \frac{1}{N}\ln\left(\frac{2^{\frac{K-2}{2}}N^{\frac{K+2}{2}}}{2^{\frac{K+2}{2}}\pi^{\frac{1}{2}}\Gamma\left(\frac{K-1}{2}\right)}\right) = \frac{1}{N}\ln\left(\frac{N^{\frac{K+2}{2}}}{\Gamma\left(\frac{K-1}{2}\right)}\right) + o(1)$$

under both SIV and MWIV asymptotics. Under SIV asymptotics,

$$\frac{1}{N}\ln\left(\frac{N^{\frac{K+2}{2}}}{\Gamma\left(\frac{K-1}{2}\right)}\right) \to 0.$$

Under MWIV asymptotics, we can use Stirling's formula to obtain

$$\frac{1}{N}\ln\left(\frac{N^{\frac{K+2}{2}}}{\Gamma\left(\frac{K-1}{2}\right)}\right) = \frac{1}{N}\ln\left(\frac{N^{\frac{K+2}{2}}}{(2\pi)^{1/2}\left(\frac{K-1}{2}\right)^{\frac{K-2}{2}}\exp\left(-\frac{K-1}{2}\right)}\right) + o(1)$$
$$\to -\frac{\alpha}{2}\left\{1 + \ln\left(\frac{\alpha}{2}\right)\right\}.$$

However, the second line in (18) does not depend on θ. As a result, these terms can be ignored in finding the limiting behavior of $\widehat{\theta}_N$. Hence, define the objective function

$$\widehat{Q}_N(\theta) = -\frac{1}{2}\lambda \cdot a'\Sigma^{-1}a + \frac{1}{N}\ln\left(Z_N^{-\frac{K-2}{2}} I_{\frac{K-2}{2}}\left(\frac{N}{2}Z_N\right)\right).$$

The quantity Z_N depends on W_N. Following Muirhead (2005, Section10.2):

$$E(W_N) = \frac{K \cdot \Sigma + \overline{M}'\overline{M}}{N} = \frac{K \cdot \Sigma + \pi' Z' Z \pi \cdot a^* a^{*\prime}}{N} = \frac{K}{N}\Sigma + \lambda_N^* \cdot a^* a^{*\prime},$$

From here, we split the result into SIV or MWIV with $\alpha = 0$ asymptotics, and MWIV with $\alpha > 0$.

For part (a), define

$$W_N^* \equiv \lambda_N^* \cdot a^* a^{*\prime}.$$

Because $V(W_N) \to 0$, we have $W_N = W_N^* + o_p(1)$. Hence, $Z_N = Z_N^* + o_p(1)$, where

$$Z_N^* \equiv 2\sqrt{\lambda \cdot \lambda_N^* \left(a'\Sigma^{-1}a^*\right)^2}.$$

The same holds for nonnormal errors as long as $V(W_N) \to 0$.

Because K is fixed and $N \to \infty$, $\widehat{Q}_N(\theta) = \overline{Q}_N(\theta) + o_p(1)$ (uniformly in $\theta \in \Theta$ compact), where

$$\overline{Q}_N(\theta) = -\frac{1}{2}\lambda \cdot a'\Sigma^{-1}a + \lambda^{1/2}\lambda_N^{*1/2} a^{*\prime}\Sigma^{-1}a.$$

The first order condition (FOC) for $\overline{Q}_N(\theta)$ is given by

$$\begin{aligned}
\frac{\partial \overline{Q}_N(\theta)}{\partial \beta} &= -\lambda \cdot a'\Sigma^{-1}e_1 + \lambda^{1/2}\lambda_N^{*1/2} a^{*\prime}\Sigma^{-1}e_1 \\
\frac{\partial \overline{Q}_N(\theta)}{\partial \lambda} &= -\frac{1}{2}a'\Sigma^{-1}a + \frac{1}{2}\lambda^{-1/2}\lambda_N^{*1/2} a^{*\prime}\Sigma^{-1}a.
\end{aligned}$$

The value $\theta^* = (\beta^*, \lambda_N^*)$ minimizes $\overline{Q}_N(\theta)$, setting the FOC to zero.

For parts (a)(i),(ii), $\overline{Q}_N(\theta) \to_p \overline{Q}(\theta)$ given by

$$\overline{Q}(\theta) = -\frac{1}{2}\lambda \cdot a'\Sigma^{-1}a + \lambda^{1/2}\lambda^{*1/2} a^{*\prime}\Sigma^{-1}a.$$

Since $\theta \in \Theta$ compact and $\overline{Q}(\theta)$ is continuous, $\widehat{\theta}_N \to_p \theta$.

For part (a)(iii), we can define $\tau(\theta, \theta_N^*) \equiv \overline{Q}_N(\theta)$ which is continuous. For each point θ_N^*, the function $\tau(\theta, \theta_N^*)$ reaches the minimum at $\theta = \theta_N^*$. Because $\theta \in \Theta$ compact and $\tau(\cdot, \theta_N^*)$ is continuous,

$$\sup_{\theta\in\Theta;\left\|\theta-\theta_N^*\right\|\geq\epsilon} \overline{Q}_N(\theta) - \overline{Q}_N(\theta_N^*) = \max_{\theta\in\Theta;\left\|\theta-\theta_N^*\right\|\geq\epsilon} \overline{Q}_N(\theta) - \overline{Q}_N(\theta_N^*) \equiv \delta(\theta_N^*) < 0.$$

Because $0 < \liminf \lambda_N^*$ and $\limsup \lambda_N^* < \infty$, there exists a compact set Θ^* such that $0 \notin \Theta^*$ in which $\theta_N^* \in \Theta^*$ eventually. Using continuity of $\delta(\cdot)$,

$$\inf_{\theta_N^*} \delta(\theta_N^*) = \min_{\theta_N^*\in\Theta^*} \delta(\theta_N^*) = \delta < 0$$

for large enough N. This implies θ_N^* is an identifiably unique sequence of maximizers of $\overline{Q}_N(\theta)$:

$$\limsup \sup_{\theta\in\Theta;\left\|\theta-\theta_N^*\right\|\geq\epsilon} \overline{Q}_N(\theta) - \overline{Q}_N(\theta_N^*) < 0.$$

The result now follows from Potscher and Prucha (1997, Lemma 3.1).

For part (b), define

$$W_N^* = \alpha\Sigma + \lambda_N^* \cdot a^* a^{*\prime}.$$

Because $V(W_N)$ goes to zero under SIV and MWIV asymptotics, we have $W_N = W_N^* + o_p(1)$. Hence, $Z_N = Z_N^* + o_p(1)$, where Z_N^* is defined as

$$Z_N^* \equiv 2\sqrt{\lambda \cdot a'\Sigma^{-1}\left(\alpha\Sigma + \lambda_N^* \cdot a^* a^{*\prime}\right)\Sigma^{-1} a}.$$

The same holds for nonnormal errors as long as $V(W_N) \to 0$. Because $K/N \to \alpha > 0$, $\widehat{Q}_N(\theta) = \overline{Q}_N(\theta) + o_p(1)$ (uniformly in $\theta \in \Theta$ compact), where

$$\overline{Q}_N(\theta) = -\frac{1}{2}\lambda \cdot a'\Sigma^{-1}a + \frac{\alpha}{2}\left(1 + \frac{Z_N^{*2}}{\alpha^2}\right)^{1/2} - \frac{\alpha}{2}\ln\left(1 + \left(1 + \frac{Z_N^{*2}}{\alpha^2}\right)^{1/2}\right).$$

The first order condition (FOC) for $\overline{Q}_N(\theta)$ is given by

$$\begin{aligned}
\frac{\partial \overline{Q}_N(\theta)}{\partial \beta} &= -\lambda \cdot a'\Sigma^{-1}e_1 + \frac{2\lambda}{\alpha} \frac{\alpha \cdot a'\Sigma^{-1}e_1 + \lambda_N^* \cdot a^{*\prime}\Sigma^{-1}a \cdot a^{*\prime}\Sigma^{-1}e_1}{1 + \left(1 + \frac{Z_N^{*2}}{\alpha^2}\right)^{1/2}} \\
\frac{\partial \overline{Q}_N(\theta)}{\partial \lambda} &= -\frac{1}{2}a'\Sigma^{-1}a + \frac{a'\Sigma^{-1}a}{\alpha} \frac{\alpha + \lambda_N^* \cdot a'\Sigma^{-1}a}{1 + \left(1 + \frac{Z_N^{*2}}{\alpha^2}\right)^{1/2}}.
\end{aligned}$$

The value $\theta_N^* = (\beta^*, \lambda_N^*)$ minimizes $\overline{Q}_N(\theta)$, setting the FOC to zero.

For parts (b)(i),(ii), $\overline{Q}_N(\theta) \to_p \overline{Q}(\theta)$ given by

$$\overline{Q}(\theta) = -\frac{1}{2}\lambda \cdot a'\Sigma^{-1}a + \frac{\alpha}{2}\left(1 + \frac{Z_N^{*2}}{\alpha^2}\right)^{1/2} - \frac{\alpha}{2}\ln\left(1 + \left(1 + \frac{Z_N^{*2}}{\alpha^2}\right)^{1/2}\right),$$

where $Z^* \equiv 2\sqrt{\lambda \cdot a'\Sigma^{-1}\left(\alpha\Sigma + \lambda^* \cdot a^* a^{*\prime}\right)\Sigma^{-1}a}$. Since $\theta \in \Theta$ compact and $\overline{Q}(\theta)$ is continuous, $\widehat{\theta}_N \to_p \theta$.

Part (b)(iii) follows analogously to Part (a)(iii).

Proof of Proposition 4. It follows from Chamberlain (2007) that (in his notation) the Bayes estimator of ϕ (integrated over Haar measures for orthogonal groups of transformations) equals the MLE. The integrated likelihood equals the marginal likelihood of the maximal invariant and ϕ is a transformation of β. As a result, MILE is equivalent to LIMLK.

Proof of Theorem 3. For part (a), when K is fixed or $K/N \to 0$,

$$\widehat{Q}_N(\theta) = -\frac{1}{2}\lambda \cdot a'\Sigma^{-1}a + \lambda^{1/2}\left(a'\Sigma^{-1}W_N\Sigma^{-1}a\right)^{1/2} + o_p\left(N^{-1}\right). \tag{19}$$

All results below hold up to $o_p\left(N^{-1/2}\right)$ order.

The components of the score function $S_N(\theta)$ are

$$\begin{aligned}
\frac{\partial Q_N(\theta)}{\partial\beta} &= -\lambda \cdot a'\Sigma^{-1}e_1 + \lambda^{1/2}\frac{a'\Sigma^{-1}W_N\Sigma^{-1}e_1}{\left(a'\Sigma^{-1}W_N\Sigma^{-1}a\right)^{1/2}} \\
\frac{\partial Q_N(\theta)}{\partial\lambda} &= -\frac{a'\Sigma^{-1}a}{2} + \frac{\left(a'\Sigma^{-1}W_N\Sigma^{-1}a\right)^{1/2}}{2\lambda^{1/2}}.
\end{aligned}$$

The components of the Hessian matrix $H_N(\theta) \equiv H(W_N;\theta)$ are

$$\begin{aligned}
\frac{\partial^2 Q_N(\theta)}{\partial\beta^2} &= -\lambda \cdot e_1'\Sigma^{-1}e_1 + \lambda^{1/2}\frac{e_1'\Sigma^{-1}W_N\Sigma^{-1}e_1}{\left(a'\Sigma^{-1}W_N\Sigma^{-1}a\right)^{1/2}} - \lambda^{1/2}\frac{\left(a'\Sigma^{-1}W_N\Sigma^{-1}e_1\right)^2}{\left(a'\Sigma^{-1}W_N\Sigma^{-1}a\right)^{3/2}} \\
\frac{\partial^2 Q_N(\theta)}{\partial\beta\partial\lambda} &= -a'\Sigma^{-1}e_1 + \frac{a'\Sigma^{-1}W_N\Sigma^{-1}e_1}{2\lambda^{1/2}\left(a'\Sigma^{-1}W_N\Sigma^{-1}a\right)^{1/2}} \\
\frac{\partial^2 Q_N(\theta)}{\partial\lambda^2} &= -\frac{1}{4}\frac{\left(a'\Sigma^{-1}W_N\Sigma^{-1}e_1\right)^{1/2}}{\lambda^{3/2}}.
\end{aligned}$$

Because $W_N \to_p W^*$, $H_N(\theta) \to_p -I_0(\theta^*)$. Furthermore, $H_N(\theta) \to_p H(W_N^*;\theta)$ uniformly on $\theta = (\beta, \lambda)$ for a compact set containing θ^* as long as $\lambda > 0$. This completes part (a)(ii). To show part (a)(i), we write

$$\sqrt{N}S_N(\theta^*) \equiv \sqrt{N}S(W_N;\theta^*) \equiv \sqrt{N}\left[S(W_N;\theta^*) - S(W^*;\theta^*)\right].$$

Using $vec(W_N) = \mathcal{D}_T vech(W_N)$, where $\mathcal{D}_T$ is the duplication matrix (e.g. Magnus and Neudecker (1988)), we write

$$\sqrt{N} S_N(\theta^*) \equiv \sqrt{N}\left[L(vech(W_N);\theta^*) - L(vech(W^*);\theta^*)\right],$$

where $L : \mathbf{R}^3 \to \mathbf{R}^2$. Now, $\sqrt{N}(vech(W_N) - vech(W^*))$ converges to a normal distribution by a standard CLT. As a result, using the delta method and the information identity, $\sqrt{N} S_N(\theta^*)$ converges to a normal distribution with zero mean and variance $I_\alpha(\theta^*)$. Part (iii) follows from Newey and McFadden (1994).

For part (b), when $K/N \to \alpha > 0$,

$$\widehat{Q}_N(\theta) = -\frac{1}{2}\lambda \cdot a'\Sigma^{-1}a + \frac{\alpha}{2}\left(1 + \frac{Z_N^2}{\alpha^2}\right)^{1/2} - \frac{\alpha}{2}\ln\left(1 + \left(1 + \frac{Z_N^2}{\alpha^2}\right)^{1/2}\right) \tag{20}$$

up to an $o_p(N^{-1})$ term. All results below hold up to $o_p\left(N^{-1/2}\right)$ order.

The components of the score function $S_N(\theta)$ are

$$\begin{aligned}
\frac{\partial Q_N(\theta)}{\partial \beta} &= -\lambda \cdot a'\Sigma^{-1}e_1 + \frac{2\lambda}{\alpha}\frac{a'\Sigma^{-1}W_N\Sigma^{-1}e_1}{1 + \left(1 + \frac{Z_N^2}{\alpha^2}\right)^{1/2}} \\
\frac{\partial Q_N(\theta)}{\partial \lambda} &= -\frac{a'\Sigma^{-1}a}{2} + \frac{1}{\alpha}\frac{a'\Sigma^{-1}W_N\Sigma^{-1}a}{1 + \left(1 + \frac{Z_N^2}{\alpha^2}\right)^{1/2}}.
\end{aligned}$$

The components of the Hessian matrix $H_N(\theta)$ are

$$\begin{aligned}
\frac{\partial^2 Q_N(\theta)}{\partial \beta^2} &= -\lambda \cdot e_1'\Sigma^{-1}e_1 + \frac{2\lambda}{\alpha}\frac{e_1'\Sigma^{-1}W_N\Sigma^{-1}e_1}{1 + \left(1 + \frac{Z_N^2}{\alpha^2}\right)^{1/2}} - \frac{8\lambda^2}{\alpha^3\left(1 + \frac{Z_N^2}{\alpha^2}\right)^{1/2}}\frac{\left(a'\Sigma^{-1}W_N\Sigma^{-1}e_1\right)^2}{\left(1 + \left(1 + \frac{Z_N^2}{\alpha^2}\right)^{1/2}\right)^2} \\
\frac{\partial^2 Q_N(\theta)}{\partial \beta \partial \lambda} &= -a'\Sigma^{-1}e_1 + \frac{2}{\alpha}\frac{a'\Sigma^{-1}W_N\Sigma^{-1}e_1}{1 + \left(1 + \frac{Z_N^2}{\alpha^2}\right)^{1/2}} - \frac{4\lambda \cdot a'\Sigma^{-1}W_N\Sigma^{-1}e_1}{\alpha^3\left(1 + \frac{Z_N^2}{\alpha^2}\right)^{1/2}}\frac{a'\Sigma^{-1}W_N\Sigma^{-1}a}{\left(1 + \left(1 + \frac{Z_N^2}{\alpha^2}\right)^{1/2}\right)^2} \\
\frac{\partial^2 Q_N(\theta)}{\partial \lambda^2} &= -\frac{2}{\alpha^3\left(1 + \frac{Z_N^2}{\alpha^2}\right)^{1/2}}\frac{\left(a'\Sigma^{-1}W_N\Sigma^{-1}a\right)^2}{\left(1 + \left(1 + \frac{Z_N^2}{\alpha^2}\right)^{1/2}\right)^2}.
\end{aligned}$$

Parts (b)(i)-(iii) follow analogously to parts (a)(i)-(iii).

Proof of Corollary 1. The determinant of $I_\alpha(\theta^*)$ simplifies to

$$|I_\alpha(\theta^*)| = \frac{\lambda^{*2}\left(a^{*\prime}\Sigma^{-1}a^*\right)^2}{\alpha + 2\lambda^* \cdot a^{*\prime}\Sigma^{-1}a^*}\frac{a^{*\prime}\Sigma^{-1}a^* \cdot e_1'\Sigma^{-1}e_1 - \left(a^{*\prime}\Sigma^{-1}e_1\right)^2}{2\left(\alpha + \lambda^* \cdot a^{*\prime}\Sigma^{-1}a^*\right)}.$$

Hence, the entry $(1,1)$ of the inverse of $I_\alpha(\theta^*)$ equals

$$\begin{aligned}\left(I_\alpha(\theta^*)^{-1}\right)_{11} &= \frac{(a^{*\prime}\Sigma^{-1}a^*)^2}{2(\alpha+2\lambda^* a^{*\prime}\Sigma^{-1}a^*)}\left|I_\alpha(\theta^*)\right|^{-1} \\ &= \frac{\alpha+\lambda^*\cdot a^{*\prime}\Sigma^{-1}a^*}{\lambda^{*2}\cdot a^{*\prime}\Sigma^{-1}a^*}\frac{a^{*\prime}\Sigma^{-1}a^*}{a^{*\prime}\Sigma^{-1}a^*\cdot e_1'\Sigma^{-1}e_1'-(a^{*\prime}\Sigma^{-1}e_1)^2} \\ &= \frac{\sigma_u^2}{\lambda^{*2}}\left\{\lambda^*+\frac{\alpha}{a^{*\prime}\Sigma^{-1}a^*}\right\}.\end{aligned}$$

This expression coincides with the asymptotic variance of LIMLK as described in equation (4.7) of Bekker (1994):

$$\left(I_\alpha(\theta^*)^{-1}\right)_{11} = \frac{\sigma_u^2}{\lambda^{*2}}\left\{\lambda^*+\alpha\cdot e_2'\Sigma e_2-\alpha\frac{(b'\Sigma e_2)^2}{b'\Sigma b}\right\}.$$

Proof of Theorem 4. This result follows from standard limit of experiment arguments; see Chioda and Jansson (2007). Part (a) follows from expansions based on (19). Part (b) follows from expansions based on (20).

7.3 Proofs of Results Stated in Section 5

For the next proofs, define the following four quantities:

$$\begin{aligned}c_1 &= tr\left(DB^*B^{*\prime}D'\right)+\lambda_N^* 1_T' B^{*\prime}D'DB^*1_T \\ c_2 &= 1_T'DB^*B^{*\prime}D'1_T+\lambda_N^*\left(1_T'DB^*1_T\right)^2 \\ c_3 &= 1_T'F1_T+(\rho^*-\rho)\,1_T'F'F1_T+\lambda^*1_T'DB^*1_T\cdot 1_T'F1_T \\ c_4 &= (\rho^*-\rho)\,tr\left(F'F\right)+\lambda^*\left\{1_T'F1_T+(\rho^*-\rho)\,1_T'F'F1_T\right\}.\end{aligned}$$

Proof of Proposition 5. We omit the original proof here as it has been generalized by Chamberlain and Moreira (2006).

Proof of Theorem 5. The density function of M at q is

$$\begin{aligned}f(q) = C_{2,N}\cdot\exp\left(-\frac{1}{2}tr(\Sigma^{-1}\overline{M}'\overline{M})\right)|\Sigma|^{-N/2}|q|^{\frac{N-T-1}{2}}\exp\left(-\frac{1}{2}tr(\Sigma^{-1}q)\right) \\ \times\left(\sqrt{tr(q\Sigma^{-1}\overline{M}'\overline{M}\Sigma^{-1})}\right)^{-\frac{K-2}{2}} I_{\frac{K-2}{2}}\left(\sqrt{tr(q\Sigma^{-1}\overline{M}'\overline{M}\Sigma^{-1})}\right).\end{aligned}$$

We obtain

$$\Sigma^{-1}=\frac{D'D}{\sigma^2},\ tr(\Sigma^{-1}\overline{M}'\overline{M})=\frac{\eta'\eta}{\sigma^2}T,\text{ and } tr(q\Sigma^{-1}\overline{M}'\overline{M}\Sigma^{-1})=\frac{\eta'\eta}{(\sigma^2)^2}1_T'DqD'1_T$$

to simplify the density function of M to

$$f\left(q\right)=C_{2,N}\cdot\exp\left(-\frac{\eta'\eta}{2\sigma^{2}}T\right)\left(\sigma^{2}\right)^{-\frac{NT}{2}}\left|q\right|^{\frac{N-T-1}{2}}\exp\left(-\frac{1}{2\sigma^{2}}tr(DqD')\right)$$
$$\times\left(\sqrt{\frac{\eta'\eta}{\left(\sigma^{2}\right)^{2}}1_T'DqD'1_T}\right)^{-\frac{K-2}{2}}I_{\frac{K-2}{2}}\left(\sqrt{\frac{\eta'\eta}{\left(\sigma^{2}\right)^{2}}1_T'DqD'1_T}\right).$$

The density function of W_N is then

$$g\left(w;\beta,\lambda_N\right)=f\left(q\left(w\right)\right)\cdot\left|q'(w)\right|=f\left(q\left(w\right)\right)N^{\frac{T(T+1)}{2}},$$

which simplifies to (13).

Proof of Theorem 6. The log-likelihood divided by NT is

$$Q_N\left(\theta\right)=-\frac{1}{2}\ln\sigma^{2}-\frac{1}{2\sigma^{2}}\frac{tr(DW_ND')}{T}-\frac{1}{2}\lambda+\frac{1}{NT}\ln\left(Z_N^{-\frac{N-2}{2}}I_{\frac{N-2}{2}}\left(\frac{N}{2}Z_N\right)\right)$$
$$+\frac{N-T-1}{2NT}\ln\left|W_N\right|+\frac{1}{NT}\ln\left(2^{\frac{N-2}{2}}N^{\frac{NT}{2}-\frac{N-2}{2}}C_{2,N}\right),\tag{21}$$

where $Z_N=2\sqrt{\lambda\frac{1_T'DW_ND'1_T}{\sigma^{2}}}$.

The second line is well-behaved when $N\rightarrow\infty$ with T fixed. Using Stirling's formula,

$$\begin{aligned}\frac{1}{NT}\ln\left(2^{\frac{N-2}{2}}N^{\frac{NT}{2}-\frac{N-2}{2}}C_{2,N}\right)&=\frac{1}{T}\ln\left(\frac{N^{\frac{NT}{2}-\frac{N-2}{2}}2^{1/2}}{\prod_{t=1}^{T-1}\left(N-t\right)^{\frac{N-t-1}{2N}}\exp\left(-\frac{N-t}{2N}\right)}\right)+o(1)\\&=\frac{\ln\left(2\right)}{2T}-\frac{1}{T}\ln\left(\prod_{t=1}^{T-1}\left(1-\frac{t}{N}\right)^{1/2}\exp\left(-\frac{1}{2}\right)\right)+o\left(1\right)\\&=\frac{\ln\left(2\right)}{2T}+\frac{T-1}{2T}+o\left(1\right).\end{aligned}$$

In addition,

$$E\left(W_N\right)=\frac{N\cdot\Sigma+\overline{M}'\overline{M}}{N}=\sigma^{*2}B^{*}\left(I_T+\lambda_N^{*}1_T1_T'\right)B^{*\prime}\equiv W_N^{*},$$

Because $V\left(W_N\right)\rightarrow0$, we have $W_N=W_N^{*}+o_p\left(1\right)$. Now,

$$\left|W_N^{*}\right|=\left|B^{*}\right|\cdot\left|\sigma^{*2}\left(I_T+\lambda_N^{*}1_T1_T'\right)\right|\cdot\left|B^{*\prime}\right|=\left(\sigma^{*2}\right)^{T}\left|I_T+\lambda_N^{*}1_T1_T'\right|=\left(\sigma^{*2}\right)^{T}\left(1+\lambda_N^{*}T\right).$$

As a result, $\ln\left(W_N\right)=T\ln\left(\sigma^{*2}\right)+\ln\left(1+\lambda_N^{*}T\right)+o_p\left(1\right).$

It is unknown whether the second line in (21) is well-behaved with $T \to \infty$. However, since it does not depend on θ, it can be ignored when finding the limiting behavior of $\widehat{\theta}_N$. Hence, define the objective function

$$\widehat{Q}_N(\theta) = -\frac{1}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\frac{tr(DW_ND')}{T} - \frac{1}{2}\lambda + \frac{1}{NT}\ln\left(Z_N^{-\frac{N-2}{2}} I_{\frac{N-2}{2}}\left(\frac{N}{2}Z_N\right)\right).$$

From here, we split the result into fixed T and large T asymptotics.

For part (a), in which $N \to \infty$ with T fixed, $Z_N = Z_N^* + o_p(1)$, where

$$Z_N^* \equiv 2\sqrt{\lambda\frac{1_T'DW_N^*D'1_T}{\sigma^2}}.$$

Furthermore, $\widehat{Q}_N(\theta) = \overline{Q}_N(\theta) + o_p(1)$, where

$$\overline{Q}_N(\theta) = -\frac{1}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\frac{tr(DW_N^*D')}{T} - \frac{1}{2}\lambda + \frac{1}{2T}\left(1+Z_N^{*2}\right)^{1/2} - \frac{1}{2T}\ln\left(1+\left(1+Z_N^{*2}\right)^{1/2}\right).$$

The first order condition (FOC) for $\overline{Q}_N(\theta)$ is given by

$$\begin{aligned}
\frac{\partial\overline{Q}_N(\theta)}{\partial\rho} &= \frac{\sigma^{*2}}{\sigma^2}\frac{(\rho^*-\rho)\,tr(FF') + \lambda^*\left\{1_T'F1_T + (\rho^*-\rho)\,1_T'F'F1_T\right\}}{T} \\
&-\frac{\sigma^{*2}}{\sigma^2}\frac{\lambda^*}{1+(1+Z_N^{*2})^{1/2}}\frac{1_T'F1_T + (\rho^*-\rho)\,1_T'F'F1_T - \lambda^*\left(T + (\rho^*-\rho)\,1_T'F1_T\right)}{T} \\
\frac{\partial\overline{Q}_N(\theta)}{\partial\sigma^2} &= -\frac{1}{2\sigma^2} + \frac{\sigma^{*2}}{2(\sigma^2)^2}\frac{c_1}{T} - \frac{\sigma^{*2}}{(\sigma^2)^2}\frac{\lambda_N^*}{1+(1+Z_N^{*2})^{1/2}}\frac{c_2}{T} \\
\frac{\partial\overline{Q}_N(\theta)}{\partial\lambda} &= -\frac{1}{2} + \frac{\sigma^{*2}}{\sigma^2}\frac{1}{1+(1+Z_N^{*2})^{1/2}}\frac{c_2}{T}.
\end{aligned}$$

The value $\theta^* = (\rho^*, \sigma^{*2}, \lambda_N^*)$ minimizes $\overline{Q}_N(\theta)$, setting the FOC to zero.

For parts (a)(i),(ii), $\overline{Q}_N(\theta) \to_p \overline{Q}(\theta)$ (uniformly in Θ compact) given by

$$\overline{Q}_N(\theta) = -\frac{1}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\frac{tr(DW^*D')}{T} - \frac{1}{2}\lambda + \frac{1}{2T}\left(1+Z^{*2}\right)^{1/2} - \frac{1}{2T}\ln\left(1+\left(1+Z^{*2}\right)^{1/2}\right),$$

where W^* and Z^* are defined as

$$W^* = \sigma^{*2}B^*\left(I_T + \lambda^*1_T1_T'\right)B^{*\prime} \text{ and } Z^* = 2\sqrt{\lambda\frac{1_T'DW^*D'1_T}{\sigma^2}}. \qquad (22)$$

Since $\theta \in \Theta$ compact and $\overline{Q}(\theta)$ is continuous, $\widehat{\theta}_N \to_p \theta$.

Part (a)(iii) follows analogously to Theorem 2-(a)(iii).

For part (b), the dimension of W_N changes as $T \to \infty$. Yet, for $|\rho^*| < 1$,

$$\begin{aligned}
\frac{tr(DW_N D')}{T} &= \frac{tr(DW_N^* D')}{T} + o_p(1) \text{ and} \\
\frac{1_T' DW_N D' 1_T}{T^2} &= \frac{1_T' DW_N^* D' 1_T}{T^2} + o_p(1).
\end{aligned}$$

As a result, $Q_N(\theta) = \overline{Q}_N(\theta) + o_p(1)$, where

$$\overline{Q}_N(\theta) = -\frac{1}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\frac{tr(DW_N^* D')}{T} - \frac{1}{2}\lambda + \frac{1}{2}\frac{Z_N^*}{T}.$$

The first order condition (FOC) for $\overline{Q}_N(\theta)$ is given by

$$\begin{aligned}
\frac{\partial \overline{Q}_N(\theta)}{\partial \rho} &= \frac{\sigma^{*2}}{\sigma^2}\frac{(\rho^* - \rho)\, tr(FF') + \lambda^* \{1_T' F 1_T + (\rho^* - \rho)\, 1_T' F' F 1_T\}}{T} \\
&\quad - \frac{(\sigma^{*2})^{1/2}}{(\sigma^2)^{1/2}} \frac{\lambda^{*1/2}\lambda^{1/2}}{1 + (1 + Z_N^{*2})^{1/2}} \frac{1_T' F 1_T}{T} \\
\frac{\partial \overline{Q}_N(\theta)}{\partial \sigma^2} &= -\frac{1}{2\sigma^2} + \frac{\sigma^{*2}}{2(\sigma^2)^2}\frac{c_1}{T} - \frac{(\sigma^{*2})^{1/2}\lambda^{1/2}\lambda^{*1/2}}{2(\sigma^2)^{3/2}}\frac{1_T' DB^* 1_T}{T} \\
\frac{\partial \overline{Q}_N(\theta)}{\partial \lambda} &= -\frac{1}{2} + \frac{(\sigma^{*2})^{1/2}\lambda^{*1/2}}{2(\sigma^2)^{1/2}\lambda^{1/2}}\frac{1_T' DB^* 1_T}{T}.
\end{aligned}$$

The value $\theta^* = (\rho^*, \sigma^{*2}, \lambda_N^*)$ minimizes $\overline{Q}_N(\theta)$, setting the FOC to zero.

For parts (b)(i),(ii), $\overline{Q}_N(\theta) = \overline{Q}(\theta) + o_p(1)$ (uniformly in Θ compact), given by

$$\begin{aligned}
\overline{Q}(\theta) &= -\frac{1}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\lim_{T\to\infty}\frac{tr(DW^* D')}{T} - \frac{1}{2}\lambda \\
&\quad + \lim_{T\to\infty}\frac{1}{2T}\left(1 + Z^{*2}\right)^{1/2} - \lim_{T\to\infty}\frac{1}{2T}\ln\left(1 + \left(1 + Z^{*2}\right)^{1/2}\right).
\end{aligned}$$

where W^* and Z^* are defined in (22). Since $\theta \in \Theta$ compact and $\overline{Q}(\theta)$ is continuous, $\widehat{\theta}_N \to_p \theta$.

Part (b)(iii) follows analogously to Theorem 2-(a)(iii).

Proof of Theorem 7. First, we prove part (a). The objective function is

$$\widehat{Q}_N(\theta) = -\frac{\ln\sigma^2}{2} - \frac{tr(DW_N D')}{2\sigma^2 T} - \frac{\lambda}{2} + \frac{(1 + Z_N^2)^{1/2}}{2T} - \frac{\ln\left(1 + (1 + Z_N^2)^{1/2}\right)}{2T} \quad (23)$$

up to an $o_p(N^{-1})$ term. All results below hold up to $o_p\left(N^{-1/2}\right)$ order.

The components of the score function $S_N(\theta)$ are

$$\begin{aligned}
\frac{\partial Q_N(\theta)}{\partial \rho} &= \frac{1}{\sigma^2}\frac{tr(J_T W_N D')}{T} - \frac{2\lambda}{1+(1+Z_N^2)^{1/2}}\frac{1_T' J_T W_N D' 1_T}{T} \\
\frac{\partial Q_N(\theta)}{\partial \sigma^2} &= -\frac{1}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\frac{tr(DW_N D')}{T} - \frac{1}{(\sigma^2)^2}\frac{\lambda}{1+(1+Z_N^2)^{1/2}}\frac{1_T' D W_N D' 1_T}{T} \\
\frac{\partial Q_N(\theta)}{\partial \lambda} &= -\frac{1}{2} + \frac{1}{\sigma^2}\frac{1}{1+(1+Z_N^2)^{1/2}}\frac{1_T' D W_N D' 1_T}{T}.
\end{aligned}$$

The Hessian matrix $H_N(\theta) \to_p -I_T(\theta)$, whose components are

$$\begin{aligned}
\frac{\partial^2 \overline{Q}_N(\theta)}{\partial \rho^2} &= \frac{\sigma^{*2}}{\sigma^2}\frac{2\lambda}{1+(1+Z_N^{*2})^{1/2}}\frac{1_T' F' F 1_T + \lambda(1_T' F 1_T)^2}{T} - \frac{\sigma^{*2}}{\sigma^2}\frac{tr(F'F) + \lambda^* 1_T' F' F 1_T}{T} \\
&\quad - \left(\frac{\sigma^{*2}}{\sigma^2}\right)^2 \frac{8\lambda^2}{(1+(1+Z_N^{*2})^{1/2})^2}\frac{1}{(1+Z_N^{*2})^{1/2}}\frac{(c_3)^2}{T} \\
\frac{\partial_N^2 \overline{Q}(\theta)}{\partial \rho \partial \sigma^2} &= -\frac{\sigma^{*2}}{(\sigma^2)^2}\frac{c_4}{T} + \frac{\sigma^{*2}}{(\sigma^2)^2}\frac{2\lambda}{1+(1+Z_N^{*2})^{1/2}}\frac{c_3}{T} \\
&\quad \times \left\{1 - \frac{\sigma^{*2}}{\sigma^2}\frac{2\lambda c_2}{1+(1+Z_N^{*2})^{1/2}}\frac{1}{(1+Z_N^{*2})^{1/2}}\right\} \\
\frac{\partial_N^2 \overline{Q}(\theta)}{\partial \rho \partial \lambda} &= -\frac{\sigma^{*2}}{\sigma^2}\frac{2}{1+(1+Z_N^{*2})^{1/2}}\frac{c_3}{T}\left\{1 - \frac{\sigma^{*2}}{\sigma^2}\frac{2\lambda c_2}{1+(1+Z_N^{*2})^{1/2}}\frac{1}{(1+Z_N^{*2})^{1/2}}\right\} \\
\frac{\partial_N^2 \overline{Q}(\theta)}{\partial (\sigma^2)^2} &= -\frac{(\sigma^{*2})^2}{(\sigma^2)^4}\frac{2\lambda^2}{(1+(1+Z_N^{*2})^{1/2})^2}\frac{1}{(1+Z_N^{*2})^{1/2}}\frac{(c_2)^2}{T} \\
&\quad + \frac{1}{2(\sigma^2)^2} - \frac{\sigma^{*2}}{(\sigma^2)^3}\frac{c_1}{T} + \frac{\sigma^{*2}}{(\sigma^2)^3}\frac{2\lambda}{1+(1+Z_N^{*2})^{1/2}}\frac{c_2}{T} \\
\frac{\partial_N^2 \overline{Q}(\theta)}{\partial \sigma^2 \partial \lambda} &= -\frac{\sigma^{*2}}{(\sigma^2)^2}\frac{1}{1+(1+Z_N^{*2})^{1/2}}\frac{c_2}{T}\left\{1 - \frac{\sigma^{*2}}{\sigma^2}\frac{2\lambda c_2}{1+(1+Z_N^{*2})^{1/2}}\frac{1}{(1+Z_N^{*2})^{1/2}}\right\} \\
\frac{\partial_N^2 \overline{Q}(\theta)}{\partial \lambda^2} &= -\left(\frac{\sigma^{*2}}{\sigma^2}\right)^2 \frac{2}{(1+(1+Z_N^{*2})^{1/2})^2}\frac{1}{(1+Z_N^{*2})^{1/2}}\frac{(c_2)^2}{T}.
\end{aligned}$$

This convergence is uniform on $\theta = (\beta, \lambda)$ for a compact set containing θ^* as long as $\lambda > 0$. This completes part (a)(ii). To show part (a)(i), we write

$$\sqrt{NT} S_N(\theta^*) \equiv \sqrt{NT} S(W_N; \theta^*) \equiv \sqrt{NT}\left[S(W_N; \theta^*) - S(W^*; \theta^*)\right].$$

Using $vec(W_N) = \mathcal{D}_T vech(W_N)$, where $\mathcal{D}_T$ is the duplication matrix (e.g. Magnus and Neudecker (1988)), we write

$$\sqrt{NT} S_N(\theta^*) \equiv \sqrt{NT}\left[L(vech(W_N); \theta^*) - L(vech(W^*); \theta^*)\right],$$

where $L:\mathbf{R}^{\frac{T(T+1)}{2}}\rightarrow\mathbf{R}^3$. Now, $\sqrt{NT}\left(vech\left(W_N\right)-vech\left(W^*\right)\right)$ converges to a normal distribution by a standard CLT. As a result, using the delta method and the information identity, $\sqrt{NT}S_N\left(\theta^*\right)$ converges to a normal distribution with zero mean and variance $I_T\left(\theta\right)$. Part (iii) follows from Newey and McFadden (1994).

Part (b) follows from the asymptotic normality of the score (whose variance is given by the reciprocal of the inverse of the limit of the Hessian matrix). As the remainder terms from expansions based on (23) are asymptotically negligible, (14) holds true.

Proof of Proposition 6. First, we prove part (a). The first moment of W_N is

$$E_{\theta_N}\left[W_N\right]=\sigma^2B\left\{I_T+\lambda_N\cdot 1_T1_T'\right\}B. \tag{24}$$

The matrix $E_{\theta_N}\left[W_N\right]$ is symmetric and has $T\left(T+1\right)/2$ nonredundant elements.

For each observation i, $m\left(w_i;\rho,\sigma^2,\left(\eta_i/\sigma\right)^2\right)$ depends on a different parameter $\left(\eta_i/\sigma\right)^2$. By averaging out (15), we can identify the parameter $\theta_N^*=(\rho^*,\sigma^{*2},\lambda_N^*)$:

$$\begin{aligned}E_{\theta_N^*}\left[\overline{m}\left(W_N;\theta_N\right)\right] &= E_{\theta_N^*}\left[\frac{1}{N}\sum\nolimits_{i=1}^{N}m\left(w_i;\rho,\sigma^2,\left(\eta_i/\sigma\right)^2\right)\right]\\ &= E_{\theta_N^*}\left[vech\left(W_N-\sigma^2B\left\{I_T+\lambda_N\cdot 1_T1_T'\right\}B\right)\right]\\ &= 0 \text{ if and only if } \theta_N=\theta_N^*.\end{aligned}$$

Every GMM estimator of ρ that we are aware of is invariant to orthogonal transformations and implicitly uses a subset of the $T(T+1)/2$ moment conditions given by (16). This includes Arellano and Bond's (1991) and Ahn and Schmidt's (1995) estimators. Specifically, the existing GMM estimators use the moment given by

$$E_{\theta_N}\left[\overline{m}^0\left(W_N;\theta_N\right)\right]=0, \tag{25}$$

where $\overline{m}^0\left(W_N;\theta_N\right)=\delta^{0\prime}\overline{m}\left(W_N;\theta_N\right)$ for a suitably chosen matrix δ^0 with $T(T+1)/2$ columns. For example, Arellano and Bond's (1991) differentiates the data to construct a $(T-2)(T-1)/2$-dimensional function $\overline{m}^0\left(W_N;\theta_N\right)$ with entries

$$\frac{1}{N}\sum\nolimits_{i=1}^{N}y_{i,t'}\left(\triangle y_{i,t}-\rho\cdot\triangle y_{i,t-1}\right),\ t=3,\ldots,T,\ t'=1,\ldots,t-2.$$

This choice of $\overline{m}^0\left(w_i;\theta_N\right)$ yields the $(T-2)(T-1)/2\times T(T+1)/2$ block-diagonal matrix

$$\delta^0 = \begin{bmatrix} H_{T-2} & & & \\ & H_{T-3} & & \\ & & \ddots & \\ & & & H_1 \end{bmatrix}, \text{ where } H_t \text{ is a matrix with } t \text{ lines:}$$

$$H_t = \begin{bmatrix} \rho & -(1+\rho) & 1 & & & & \\ & \rho & -(1+\rho) & 1 & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & & \rho & -(1+\rho) & 1 \end{bmatrix}. \tag{26}$$

It is interesting to see Arellano and Bond's (1991) moment conditions using $\overline{m}^0(W_N;\theta_N) = \delta^{0\prime}\overline{m}(W_N;\theta_N)$. Their moment conditions arise exactly because

$$\begin{aligned} \delta^{0\prime}\overline{m}(W_N;\theta_N) &= \delta^{0\prime}\left[\left\{vech(W_N) - vech\left(\sigma^2 B\left\{I_T + \lambda_N \cdot 1_T 1_T'\right\} B'\right)\right\}\right] \\ &= \delta^{0\prime}vech(W_N) - \delta^{0\prime}\mathcal{D}_T^{+}(B\otimes B)vec(\sigma^2\left\{I_T + \lambda_N \cdot 1_T 1_T'\right\}) \\ &= \delta^{0\prime}vech(W_N), \end{aligned} \tag{27}$$

where $\mathcal{D}_T^{+}$ is the Moore-Penrose inverse of $\mathcal{D}_T$ (e.g. Magnus and Neudecker (1988)). Expressions (26) and (27) also illustrate how differentiating the data imposes particular structures on δ^0:

$$\begin{aligned} E_{\theta_N^*}\left[\overline{m}(W_N;\theta_N)\right] &= vech\left(\sigma^{*2}B^*\left\{I_T + \lambda_N^* \cdot 1_T 1_T'\right\}B^* - \sigma^2 B\left\{I_T + \lambda_N \cdot 1_T 1_T'\right\}B\right) \\ E_{\theta_N^*}\left[\overline{m}^0(W_N;\theta_N)\right] &= E_{\theta_N^*}\left[\delta^{0\prime}vech\left(\sigma^{*2}B^*\left\{I_T + \lambda_N^* \cdot 1_T 1_T'\right\}B^*\right)\right]. \end{aligned}$$

Part (b) follows from standard results, e.g., Theorems 2.1 and 3.2 of Newey and McFadden (1994).

For part (c), we assume that $(\eta_i^*/\sigma^*)^2$ is known to be fixed at λ^*, $i = 1, ..., N$:

$$m\left(w_i;\rho,\sigma^2,\lambda\right) = vech(w_i) - vech\left(\sigma^2 B\left\{I_T + \lambda \cdot 1_T 1_T'\right\}B\right).$$

Hence, we can apply Chamberlain's (1987) efficiency bound to moment conditions:

$$\left\{E\left(\frac{\partial m(w_i;\theta^*)}{\partial\theta}\right)'\left\{E\left(m(w_i;\theta^*)\,m(w_i;\theta^*)'\right)\right\}^{-1}E\left(\frac{\partial m(w_i;\theta^*)}{\partial\theta'}\right)\right\}^{-1} = (\zeta'\Xi\zeta)^{-1}.$$

Proof of Corollary 2. As a preliminary result, we need to find the limits of $T^{-1}tr(FF')$, $T^{-1}1_T'F1_T$, and $T^{-1}1_T'F'F1_T$, as $T\to\infty$. For the first term,

$$\frac{1}{T}tr(FF') = \frac{1}{T}\sum_{j=0}^{T-2}\sum_{i=0}^{j}\rho^{*2i} = \frac{T-1}{T}\sum_{i=0}^{T-1}\rho^{*2i} - \frac{1}{T}\sum_{i=0}^{T-1}i\rho^{*2i} \to \frac{1}{1-\rho^{*2}},$$

because $\sum_{i=0}^{T-1} i(\rho^{*2})^i$ is a convergent series. This is true because a sufficient condition for a series $\sum_{i=0}^{T} a_i$ to converge is that $\lim \sqrt[T]{|a_T|} < 1$ as $T \to \infty$. Taking $a_i = i(\rho^{*2})^i$, $\lim \sqrt[T]{|a_T|} = \lim \sqrt[T]{|T(\rho^{*2})^T|} = \rho^{*2} \lim \sqrt[T]{T} = \rho^{*2} < 1$. Analogously,

$$\frac{1}{T} 1_T' F 1_T = \frac{1}{T} \sum_{j=0}^{T-2} \sum_{i=0}^{j} \rho^{*i} = \frac{T-1}{T} \sum_{i=0}^{T-1} \rho^{*i} - \frac{1}{T} \sum_{i=0}^{T-1} i\rho^{*i} \to \frac{1}{1-\rho^*}.$$

because $\sum_{i=0}^{T-1} i\rho^{*i}$ also converges. Finally, by the Cauchy–Schwarz inequality,

$$\left(\frac{1}{T} 1_T' F 1_T\right)^2 \leq \frac{1}{T} 1_T' F' F 1_T = \frac{1}{T} \sum_{j=0}^{T-2} \left(\sum_{i=0}^{j} \rho^{*i}\right)^2 \leq \frac{T-1}{T} \left(\frac{1}{1-\rho^*}\right)^2.$$

Taking limits, we obtain

$$\frac{1}{(1-\rho^*)^2} \leq \liminf \frac{1}{T} 1_T' F' F 1_T \leq \limsup \frac{1}{T} 1_T' F' F 1_T \leq \frac{1}{(1-\rho^*)^2}.$$

Hence, the limit of $T^{-1} 1_T' F' F 1_T$ exists and equals $(1-\rho^*)^{-2}$. This result implies that

$$\frac{(T^{-1} 1_T' F 1_T)^2}{T^{-1} 1_T' F' F 1_T} = \frac{(1_T' F 1_T)^2}{(1_T' 1_T)(1_T' F' F 1_T)} = \frac{\langle x_T, y_T \rangle^2}{\langle x_T, x_T \rangle \langle y_T, y_T \rangle} \to 1 \text{ as } T \to \infty,$$

where x_T and y_T are sequences of elements in the Hilbert space (with $\langle x_T, y_T \rangle$ as the usual inner product) in which the first T entries equal 1_T and $F1_T$, respectively, and zero otherwise.

Therefore, the limiting information matrix $I_\infty(\theta^*)$ simplifies to

$$I_\infty(\theta^*) = \begin{bmatrix} \frac{1}{1-\rho^{*2}} + \frac{\lambda^*}{(1-\rho^*)^2} & \frac{\lambda^*}{2\sigma^{*2}(1-\rho^*)} & \frac{1}{2(1-\rho^*)} \\ \frac{\lambda^*}{2\sigma^{*2}(1-\rho^*)} & \frac{2+\lambda^*}{4(\sigma^{*2})^2} & \frac{1}{4\sigma^{*2}} \\ \frac{1}{2(1-\rho^*)} & \frac{1}{4\sigma^{*2}} & \frac{1}{4\lambda^*} \end{bmatrix}.$$

The entry $(1,1)$ of the inverse of $I_\infty(\theta^*)$ is

$$\left(I_\infty(\theta^*)^{-1}\right)_{11} = (A_{11} - A_{12} A_{22}^{-1} A_{21})^{-1},$$

where the matrices A_{jk} are partitions of $I_\infty(\theta^*)$:

$$A_{11} = \frac{1}{1-\rho^{*2}} + \frac{\lambda^*}{(1-\rho^*)^2}, \; A_{12} = \begin{bmatrix} \frac{\lambda^*}{2\sigma^{*2}(1-\rho^*)} & \frac{1}{2(1-\rho^*)} \end{bmatrix}$$
$$A_{21} = \begin{bmatrix} \frac{\lambda^*}{2\sigma^{*2}(1-\rho^*)} \\ \frac{1}{2(1-\rho^*)} \end{bmatrix}, \text{ and } A_{22} = \begin{bmatrix} \frac{2+\lambda^*}{4(\sigma^{*2})^2} & \frac{1}{4\sigma^{*2}} \\ \frac{1}{4\sigma^{*2}} & \frac{1}{4\lambda^*} \end{bmatrix}.$$

The following holds true:

$$
\begin{aligned}
A_{12}A_{22}^{-1}A_{21} &= \frac{1}{(1-\rho^*)^2}\left[\begin{array}{cc}\frac{\lambda^*}{\sigma^{*2}} & 1\end{array}\right]\left[\begin{array}{cc}\frac{2+\lambda^*}{(\sigma^{*2})^2} & \frac{1}{\sigma^{*2}} \\ \frac{1}{\sigma^{*2}} & \frac{1}{\lambda^*}\end{array}\right]^{-1}\left[\begin{array}{c}\frac{\lambda^*}{\sigma^{*2}} \\ 1\end{array}\right] \\
&= \frac{\lambda^*(\sigma^{*2})^2}{2(1-\rho^*)^2}\left[\begin{array}{cc}\frac{\lambda^*}{\sigma^{*2}} & 1\end{array}\right]\left[\begin{array}{cc}\frac{1}{\lambda^*} & \frac{-1}{\sigma^{*2}} \\ \frac{-1}{\sigma^{*2}} & \frac{2+\lambda^*}{(\sigma^{*2})^2}\end{array}\right]\left[\begin{array}{c}\frac{\lambda^*}{\sigma^{*2}} \\ 1\end{array}\right] \\
&= \frac{\lambda^*(\sigma^{*2})^2}{2(1-\rho^*)^2}\left\{\frac{\lambda^*}{(\sigma^{*2})^2}-\frac{2\lambda^*}{(\sigma^{*2})^2}+\frac{2}{(\sigma^{*2})^2}+\frac{\lambda^*}{(\sigma^{*2})^2}\right\} \\
&= \frac{\lambda^*}{(1-\rho^*)^2}.
\end{aligned}
$$

As a result, we obtain

$$
\left(I_\infty(\theta^*)^{-1}\right)_{11} = \left(\frac{1}{1-\rho^{*2}}+\frac{\lambda^*}{(1-\rho^*)^2}-\frac{\lambda^*}{(1-\rho^*)^2}\right)^{-1} = 1-\rho^{*2}.
$$

Proof of Theorem 8. When $T\to\infty$, the objective function is

$$
\widehat{Q}_N(\theta) = -\frac{1}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\frac{tr(DW_ND')}{T} - \frac{1}{2}\lambda - \frac{1}{2T}Z_N
$$

up to an $o_p(N^{-1})$ term. All results below hold up to $o_p\left(N^{-1/2}\right)$ order.

The components of the score function $S_N(\theta)$ are

$$
\begin{aligned}
\frac{\partial Q_N(\theta)}{\partial\rho} &= \frac{1}{\sigma^2}\frac{tr(J_TW_ND')}{T} - \frac{\lambda^{1/2}}{(\sigma^2)^{1/2}}\frac{1_T'J_TW_ND'1_T}{T(1_T'DW_ND'1_T)^{1/2}} \\
\frac{\partial Q_N(\theta)}{\partial\sigma^2} &= -\frac{1}{2\sigma^2}+\frac{1}{2(\sigma^2)^2}\frac{tr(DW_ND')}{T} - \frac{\lambda^{1/2}}{2(\sigma^2)^{3/2}}\frac{(1_T'DW_ND'1_T)^{1/2}}{T} \\
\frac{\partial Q_N(\theta)}{\partial\lambda} &= -\frac{1}{2}+\frac{1}{2(\sigma^2)^{1/2}\lambda^{1/2}}\frac{(1_T'DW_ND'1_T)^{1/2}}{T}.
\end{aligned}
$$

If $|\rho^*|$ is bounded away from one, as $T\to\infty$,

$$
\begin{aligned}
\frac{tr(J_TW_ND')}{T} &\to {}_p\lim\frac{tr(J_TW_N^*D')}{T},\ \frac{1_T'J_TW_ND'1_T}{T^2}\to_p\lim\frac{1_T'J_TW_N^*D'1_T}{T^2} \\
\frac{tr(DW_ND')}{T} &\to {}_p\lim\frac{tr(DW_N^*D')}{T},\ \text{and}\ \frac{1_T'DW_ND'1_T}{T^2}\to_p\lim\frac{1_T'DW_N^*D'1_T}{T^2}.
\end{aligned}
$$

As a result, the Hessian matrix $-H_N(\theta) \to_p I_\infty(\theta)$, whose components are limits of

$$
\begin{aligned}
-\frac{\partial^2 Q_N(\theta)}{\partial \rho^2} &= \frac{\sigma^{*2}}{\sigma^2} \frac{tr(F'F) + \lambda^* 1_T' F' F 1_T}{T} \\
-\frac{\partial^2 Q_N(\theta)}{\partial \rho \partial \sigma^2} &= \frac{\sigma^{*2}}{(\sigma^2)^2} \frac{c_4}{T} - \frac{\lambda^{1/2} \lambda^{*1/2} \left(\sigma^{*2}\right)^{1/2}}{2(\sigma^2)^{3/2}} \frac{1_T' F 1_T}{T} \\
-\frac{\partial^2 Q_N(\theta)}{\partial \rho \partial \lambda} &= \frac{\left(\sigma^{*2}\right)^{1/2} \lambda^{*1/2}}{2(\sigma^2)^{1/2} \lambda^{3/2}} \frac{1_T' F 1_T}{T} \\
-\frac{\partial^2 Q_N(\theta)}{\partial (\sigma^2)^2} &= \frac{\sigma^{*2}}{(\sigma^2)^3} \frac{c_1}{T} - \frac{3}{4} \frac{\left(\sigma^{*2}\right)^{1/2} \lambda^{1/2} \lambda^{*1/2}}{(\sigma^2)^{5/2}} \frac{1_T' D B^* 1_T}{T} \frac{1}{2(\sigma^2)^2} \\
-\frac{\partial^2 Q_N(\theta)}{\partial \sigma^2 \partial \lambda} &= \frac{\left(\sigma^{*2}\right)^{1/2} \lambda^{*1/2}}{4(\sigma^2)^{3/2} \lambda^{1/2}} \frac{1_T' D B^* 1_T}{T} \\
-\frac{\partial^2 Q_N(\theta)}{\partial \lambda^2} &= \frac{\left(\sigma^{*2}\right)^{1/2} \lambda^{*1/2}}{4(\sigma^2)^{1/2} \lambda^{3/2}} \frac{1_T' D B^* 1_T}{T}.
\end{aligned}
$$

This convergence is uniform on $\theta = (\beta, \lambda)$ for a compact set containing θ^* as long as $|\rho^*|$ is bounded away from one. This completes part (ii). To show part (i), define

$$
\begin{aligned}
\mathcal{W}_N &= \begin{pmatrix} \frac{tr(J_T W_N D^{*\prime})}{T} & \frac{1_T' J_T W_N D^{*\prime} 1_T}{T^2} & \frac{tr(D^* W_N' D^{*\prime})}{T} & \frac{1_T' D^* W_N' D^{*\prime} 1_T}{T^2} \end{pmatrix}' \text{ and} \\
\mathcal{W}_N^* &= \begin{pmatrix} \frac{tr(J_T W_N^* D^{*\prime})}{T} & \frac{1_T' J_T W_N^* D^{*\prime} 1_T}{T^2} & \frac{tr(D^* W_N^* D^{*\prime})}{T} & \frac{1_T' D^* W_N^* D^{*\prime} 1_T}{T^2} \end{pmatrix}',
\end{aligned}
$$

and write

$$
\sqrt{NT} S_N(\theta^*) \equiv \sqrt{NT} \left[L(\mathcal{W}_N; \theta^*) - L(\mathcal{W}_N^*; \theta^*) \right],
$$

where $L : \mathbf{R}^4 \to \mathbf{R}^3$. Now, $\sqrt{NT}(\mathcal{W}_N - \mathcal{W}_N^*)$ converges to a normal distribution by a standard CLT and the Cramér-Wold device. Using the delta method and the information identity, $\sqrt{NT} S_N(\theta^*)$ converges to a normal distribution with zero mean and variance $I_\infty(\theta^*)$ as long as $N \geq T$. Part (iii) follows from Newey and McFadden (1994).

References

Abrevaya, J. (2000): “Rank Estimation of a Generalized Fixed-Effects Regression Model,” *Journal of Econometrics*, 95, 1–23.

Ahn, S., and P. Schmidt (1995): “Efficient Estimation of Models for Dynamic Panel Data,” *Journal of Econometrics*, 68, 5–27.

Andersen, E. B. (1970): “Asymptotic Properties of Conditional Maximum Likelihood Estimators,” *Journal of the Royal Statistical Society, Series B*, 32, 283–301.

Anderson, T. W. (1946): “The Noncentral Wishart Distribution and Certain Problems of Multivariate Statistics,” *The Annals of Mathematical Statistics*, 17, 409–431.

Anderson, T. W., N. Kunitomo, and Y. Matsushita (2006): “A New Light from Old Wisdoms: Alternative Estimation Methods of Simultaneous Equations and Microeconometric Models,” *Unpublished Manuscript*, University of Tokyo.

Andrews, D. W. K., M. J. Moreira, and J. H. Stock (2006): “Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression,” *Econometrica*, 74, 715–752.

Arellano, M. (2003): “Panel Data Econometrics,” in *Advanced Texts in Econometrics*, ed. by M. Arellano, G. Imbens, G. Mizon, A. Pagan, and M. Watson. Oxford.

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, 277–297.

Arellano, M., and B. Honoré (1991): “Panel Data Models: Some Recent Developments,” in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 5, chap. 53, pp. 775–826. Elsevier Science, Amsterdam.

Bekker, P. A. (1994): “Alternative Approximations to the Distributions of Instrumental Variables Estimators,” *Econometrica*, 62, 657–681.

Bekker, P. A., and V. der Ploeg (2005): “Instrumental Variable Estimation Based on Grouped Data,” *Statistica Neerlandica*, 59, 239–267.

Chamberlain, G. (1987): "Asymptotic Efficiency in Estimation With Conditional Moment Restrictions," *Journal of Econometrics*, 34, 305–334.

——— (2007): "Decision Theory Applied to an Instrumental Variables Model," *Econometrica*, 75, 609–652.

Chamberlain, G., and M. J. Moreira (2006): "Decision Theory Applied to a Linear Panel Data Model," *Unpublished Manuscript*, Harvard University.

Chioda, L., and M. Jansson (2007): "Optimal Invariant Inference when the Number of Instruments is Large," *Unpublished Manuscript*, UC Berkeley.

Hahn, J., and G. Kuersteiner (2002): "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both N and T are Large," *Econometrica*, 70, 1639–1657.

Hall, W., R. Wijsman, and J. Ghosh (1965): "The Relationship Between Sufficiency and Invariance with Applications in Sequential Analysis," *Annals of Mathematical Statistics*, 36, 575–614.

Hansen, C., J. Hausman, and W. K. Newey (2006): "Estimation with Many Instrumental Variables," *Unpublished Manuscript*, University of Western Ontario.

Harville, D. (1974): "Bayesian Inference for Variance Components Using Only Error Contrasts," *Biometrika*, 61, 383–385.

Johnson, N. L., and S. Kotz (1970): *Distributions in Statistics: Continuous Multivariate Distributions*. New York: John Wiley and Sons.

Kunitomo, N. (1980): "Asymptotic Expansions of Distributions of Estimators in a Linear Functional Relationship and Simultaneous Equations," *Journal of the American Statistical Association*, 75, 693–700.

Lancaster, T. (2000): "The Incidental Parameter Problem Since 1948," *Journal of Econometrics*, 95, 391–413.

——— (2002): "Orthogonal Parameters and Panel Data," *Journal of Econometrics*, 69, 647–666.

Le Cam, L., and G. L. Yang (2000): *Asymptotics in Statistics: Some Basic Concepts*. 2nd edn., Springer-Verlag.

LEHMANN, E. L., AND J. P. ROMANO (2005): *Testing Statistical Hypotheses*. Third edn., Springer.

MAGNUS, J. R., AND H. NEUDECKER (1988): *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley, New York.

MOREIRA, M. J. (2001): "Tests with Correct Size When Instruments Can Be Arbitrarily Weak," *Center for Labor Economics Working Paper Series*, 37, UC Berkeley.

MORIMUNE, K. (1983): "Approximate Distributions of k-Class Estimators When the Degree of Overidentification is Large Compared With Sample Size," *Econometrica*, 51, 821–841.

MUIRHEAD, R. J. (2005): *Aspects of Multivariate Statistical Theory.*

NEWEY, W., AND D. L. MCFADDEN (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. 4, chap. 36, pp. 2111–2245. Elsevier Science, Amsterdam.

NEWEY, W. K. (2004): "Efficient Semiparametric Estimation Via Moment Restrictions," *Econometrica*, 72, 1877–1897.

NEYMAN, J., AND E. L. SCOTT (1948): "Consistent estimates based on partially consistent observations," *Econometrica*, 16, 1–32.

POTSCHER, B. M., AND I. R. PRUCHA (1997): *Dynamic Nonlinear Econometric Models*. Springer-Verlag.

SWEETING, T. J. (1989): "On Conditional Weak Convergence," *Journal of Theoretical Probability*, 2, 461–474.

VAN DER VAART, A. W. (1998): *Asymptotic Statistics*. Cambridge: Cambridge University Press.

VAN HASSELT, M. (2007): "Instrumental Variables Estimators, Many Instruments, and Nonnormality," *Unpublished Manuscript*, University of Western Ontario.

TABLE I

Performance of Estimators for the Autoregressive Parameter ρ

(random effects, normal errors, and $\rho = 0.50$)

		Mean				MSE			
T	N	MILE	BCOLS	AB	AS	MILE	BCOLS	AB	AS
2	5	0.4592	0.9651	*	*	0.1552	0.4602	*	*
2	10	0.4859	0.9500	*	*	0.0631	0.3109	*	*
2	25	0.4960	0.9523	*	*	0.0246	0.2394	*	*
2	100	0.4974	0.9474	*	*	0.0054	0.2083	*	*
3	5	0.4431	0.7695	-0.0578	0.8642	0.0631	0.1607	516.8489	0.3823
3	10	0.4789	0.7903	0.9766	0.8954	0.0280	0.1165	153.1105	0.2559
3	25	0.4908	0.8008	0.5705	0.9389	0.0115	0.1045	4.7087	0.2219
3	100	0.4979	0.8068	0.5372	0.9632	0.0024	0.0975	0.0724	0.2204
5	5	0.4626	0.6469	0.1980	0.6541	0.0231	0.0538	0.2323	0.0991
5	10	0.4802	0.6657	0.2386	0.7162	0.0116	0.0422	0.2145	0.0820
5	25	0.4935	0.6702	0.3768	0.7940	0.0044	0.0347	0.0869	0.1002
5	100	0.4991	0.6799	0.4650	0.8667	0.0010	0.0336	0.0136	0.1371
10	5	0.4731	0.5505	0.0385	0.3753	0.0122	0.0158	52.4500	0.0747
10	10	0.4861	0.5660	0.3249	0.4518	0.0049	0.0107	0.0489	0.0437
10	25	0.4937	0.5717	0.3977	0.5763	0.0021	0.0074	0.0211	0.0294
10	100	0.4993	0.5736	0.4625	0.7223	0.0005	0.0060	0.0058	0.0550
25	5	0.4871	0.5128	**	**	0.0048	0.0055	**	**
25	10	0.4930	0.5151	**	**	0.0025	0.0025	**	**
25	25	0.4966	0.5180	**	**	0.0010	0.0013	**	**
25	100	0.4997	0.5184	**	**	0.0002	0.0006	**	**
100	5	0.4941	0.5014	**	**	0.0014	0.0013	**	**
100	10	0.4978	0.5018	**	**	0.0007	0.0007	**	**
100	25	0.4990	0.5001	**	**	0.0003	0.0003	**	**
100	100	0.4997	0.5015	**	**	0.0001	0.0001	**	**

(*) The estimator is not available for T = 2.

(**) Computational cost is prohibitive for large T.

TABLE II

Performance of Estimators for the Autoregressive Parameter ρ

(nonconvergent effects, normal errors, and $\rho = 0.50$)

		Mean				MSE			
T	N	MILE	BCOLS	AB	AS	MILE	BCOLS	AB	AS
2	5	0.4770	1.0835	*	*	0.0818	0.5044	*	*
2	10	0.4911	1.1389	*	*	0.0196	0.4442	*	*
2	25	0.4989	1.1994	*	*	0.0037	0.4959	*	*
2	100	0.5000	1.2352	*	*	0.0002	0.5410	*	*
3	5	0.4773	0.8349	0.2500	0.9455	0.0346	0.1603	384.7828	0.3733
3	10	0.4908	0.9110	0.5705	0.9203	0.0087	0.1818	0.5864	0.2215
3	25	0.4981	0.9636	0.5160	0.8997	0.0013	0.2173	0.0173	0.1719
3	100	0.4992	0.9904	0.5013	0.8231	0.0001	0.2406	0.0009	0.1049
5	5	0.4727	0.6997	0.2452	0.7159	0.0165	0.0603	0.1766	0.0873
5	10	0.4918	0.7415	0.4475	0.7635	0.0043	0.0640	0.0339	0.0795
5	25	0.4991	0.7755	0.4912	0.7902	0.0007	0.0768	0.0046	0.0861
5	100	0.4997	0.7936	0.4988	0.7854	0.0000	0.0863	0.0002	0.0816
10	5	0.4789	0.5798	-0.9436	0.4278	0.0080	0.0151	1721.7952	0.0516
10	10	0.4908	0.6104	0.4005	0.5980	0.0024	0.0148	0.0197	0.0281
10	25	0.5027	0.6326	0.4806	0.7370	0.0014	0.0180	0.0022	0.0583
10	100	0.5000	0.6452	0.4988	0.7765	0.0000	0.0211	0.0001	0.0765
25	5	0.4884	0.5157	**	**	0.0040	0.0042	**	**
25	10	0.4949	0.5330	**	**	0.0014	0.0027	**	**
25	25	0.4995	0.5464	**	**	0.0003	0.0024	**	**
25	100	0.4999	0.5562	**	**	0.0000	0.0032	**	**
100	5	0.4964	0.4994	**	**	0.0013	0.0014	**	**
100	10	0.4987	0.5038	**	**	0.0006	0.0005	**	**
100	25	0.4994	0.5076	**	**	0.0002	0.0002	**	**
100	100	0.5001	0.5119	**	**	0.0000	0.0002	**	**

(*) The estimator is not available for T = 2.

(**) Computational cost is prohibitive for large T.

TABLE III

Performance of Estimators for the Autoregressive Parameter ρ

(random effects, nonnormal errors, and $\rho = 0.50$)

		Mean				MSE			
T	N	MILE	BCOLS	AB	AS	MILE	BCOLS	AB	AS
2	5	0.4520	0.9797	*	*	0.1430	0.5085	*	*
2	10	0.5024	0.9975	*	*	0.0869	0.3687	*	*
2	25	0.4993	0.9665	*	*	0.0414	0.2711	*	*
2	100	0.5042	0.9507	*	*	0.0105	0.2175	*	*
3	5	0.4666	0.7910	0.3562	0.8923	0.0687	0.1811	31.5729	0.4008
3	10	0.4803	0.8056	0.4189	0.9204	0.0343	0.1373	59.3092	0.2723
3	25	0.4951	0.8054	0.3363	0.9376	0.0143	0.1104	53.3848	0.2233
3	100	0.4992	0.8091	0.5244	0.9683	0.0030	0.0999	0.0839	0.2278
5	5	0.4712	0.6629	0.2628	0.6585	0.0268	0.0647	0.1905	0.1359
5	10	0.4821	0.6704	0.3211	0.6975	0.0150	0.0456	0.1282	0.0872
5	25	0.4928	0.6778	0.3899	0.7748	0.0045	0.0380	0.0810	0.0914
5	100	0.4967	0.6798	0.4717	0.8539	0.0011	0.0339	0.0128	0.1291
10	5	0.4722	0.5602	0.0781	0.3906	0.0110	0.0175	162.8453	0.0840
10	10	0.4893	0.5663	0.3471	0.4507	0.0047	0.0105	0.0405	0.0516
10	25	0.4946	0.5721	0.4084	0.5625	0.0020	0.0077	0.0178	0.0309
10	100	0.4984	0.5745	0.4740	0.7154	0.0005	0.0061	0.0035	0.0514
25	5	0.4819	0.5113	**	**	0.0052	0.0046	**	**
25	10	0.4890	0.5157	**	**	0.0024	0.0026	**	**
25	25	0.4974	0.5182	**	**	0.0010	0.0014	**	**
25	100	0.4990	0.5187	**	**	0.0003	0.0006	**	**
100	5	0.4949	0.4997	**	**	0.0015	0.0014	**	**
100	10	0.4972	0.5004	**	**	0.0007	0.0007	**	**
100	25	0.5000	0.5015	**	**	0.0003	0.0003	**	**
100	100	0.5000	0.5016	**	**	0.0001	0.0001	**	**

(*) The estimator is not available for T = 2.

(**) Computational cost is prohibitive for large T.

TABLE IV

Performance of Estimators for the Autoregressive Parameter ρ

(random effects, normal errors, and ρ = -0.50)

		Mean				MSE			
T	N	MILE	BCOLS	AB	AS	MILE	BCOLS	AB	AS
2	5	-0.5489	-0.5689	*	*	0.1706	0.2478	*	*
2	10	-0.5206	-0.5622	*	*	0.0694	0.1020	*	*
2	25	-0.5024	-0.5485	*	*	0.0269	0.0374	*	*
2	100	-0.5047	-0.5476	*	*	0.0058	0.0104	*	*
3	5	-0.4920	-0.4907	-0.0209	-0.3722	0.0801	0.0791	20.5152	0.3044
3	10	-0.5006	-0.4994	-0.4555	-0.4485	0.0326	0.0352	4.0370	0.1651
3	25	-0.5024	-0.5087	-0.4951	-0.4990	0.0117	0.0146	0.0409	0.0578
3	100	-0.5020	-0.5063	-0.4948	-0.5368	0.0031	0.0033	0.0080	0.0129
5	5	-0.4878	-0.4728	-0.5408	-0.3755	0.0339	0.0371	0.0549	0.1201
5	10	-0.4971	-0.4871	-0.5262	-0.4113	0.0156	0.0202	0.0326	0.0713
5	25	-0.5000	-0.5007	-0.5153	-0.4608	0.0069	0.0073	0.0136	0.0310
5	100	-0.4992	-0.5021	-0.5030	-0.4860	0.0017	0.0017	0.0033	0.0069
10	5	-0.4947	-0.4779	0.6536	-0.4602	0.0157	0.0181	3313.3070	0.0343
10	10	-0.4965	-0.4944	-0.5334	-0.4563	0.0083	0.0078	0.0098	0.0211
10	25	-0.4987	-0.4951	-0.5144	-0.4541	0.0031	0.0032	0.0046	0.0122
10	100	-0.4995	-0.4984	-0.5024	-0.4552	0.0008	0.0008	0.0014	0.0041
25	5	-0.4958	-0.4921	**	**	0.0061	0.0066	**	**
25	10	-0.4986	-0.4952	**	**	0.0033	0.0030	**	**
25	25	-0.4988	-0.4994	**	**	0.0013	0.0012	**	**
25	100	-0.4996	-0.4998	**	**	0.0003	0.0003	**	**
100	5	-0.4996	-0.4986	**	**	0.0016	0.0015	**	**
100	10	-0.5002	-0.4992	**	**	0.0008	0.0008	**	**
100	25	-0.4997	-0.4999	**	**	0.0003	0.0003	**	**
100	100	-0.5000	-0.4993	**	**	0.0001	0.0001	**	**

(*) The estimator is not available for T = 2.

(**) Computational cost is prohibitive for large T.

TABLE V

Performance of Estimators for the Autoregressive Parameter ρ

(random effects, normal errors, and $\rho = 1.00$)

		Mean				MSE			
T	N	MILE	BCOLS	AB	AS	MILE	BCOLS	AB	AS
2	5	0.9307	1.6990	*	*	0.1316	0.7595	*	*
2	10	0.9766	1.7115	*	*	0.0679	0.6034	*	*
2	25	1.0009	1.6943	*	*	0.0274	0.5166	*	*
2	100	0.9958	1.7047	*	*	0.0057	0.5048	*	*
3	5	0.9674	1.5029	1.0935	1.3267	0.0452	0.3211	36.9311	0.1953
3	10	1.0072	1.5032	1.0299	1.3320	0.0224	0.2776	5.5735	0.1386
3	25	0.9971	1.5156	1.0120	1.3469	0.0059	0.2733	0.0313	0.1318
3	100	0.9975	1.5216	0.9996	1.3624	0.0015	0.2740	0.0068	0.1345
5	5	0.9827	1.3241	0.9478	1.1497	0.0093	0.1190	0.0313	0.0363
5	10	0.9949	1.3341	0.9838	1.1531	0.0032	0.1165	0.0089	0.0289
5	25	0.9984	1.3403	0.9919	1.1659	0.0012	0.1174	0.0030	0.0294
5	100	0.9999	1.3442	0.9986	1.1760	0.0003	0.1189	0.0007	0.0315
10	5	0.9960	1.1774	1.2028	1.0534	0.0015	0.0330	55.2326	0.0065
10	10	0.9989	1.1838	0.9892	1.0621	0.0004	0.0343	0.0007	0.0053
10	25	0.9992	1.1839	0.9960	1.0680	0.0001	0.0340	0.0002	0.0051
10	100	1.0000	1.1854	0.9991	1.0687	0.0000	0.0344	0.0001	0.0048
25	5	0.9994	1.0765	**	**	0.0001	0.0059	**	**
25	10	1.0000	1.0767	**	**	0.0000	0.0059	**	**
25	25	0.9998	1.0776	**	**	0.0000	0.0060	**	**
25	100	1.0000	1.0776	**	**	0.0000	0.0060	**	**
100	5	1.0000	1.0197	**	**	0.0000	0.0004	**	**
100	10	0.9999	1.0198	**	**	0.0000	0.0004	**	**
100	25	1.0000	1.0198	**	**	0.0000	0.0004	**	**
100	100	1.0000	1.0198	**	**	0.0000	0.0004	**	**

(*) The estimator is not available for T = 2.

(**) Computational cost is prohibitive for large T.