ANNALS OF ECONOMICS AND FINANCE **12-1**, 89–107 (2011)

Efficient Estimation of Moment Condition Models with Heterogenous Populations

Zhiguo Xiao*

School of Management, Fudan University, Shanghai 200433, China
E-mail: zhiguo_xiao@fudan.edu.cn

A wide range of econometric and statistical models are specified through moment conditions. Efficient estimation of such models essentially employs two distinct ideas: optimally combining estimation equations (e.g., the optimal estimating equations of Godambe (1976), the generalized method of moments of Hansen (1982) and the empirical likelihood of Qin and Lawless (1994)), and optimally combining estimators (e.g., the weighted method of moments of Xiao (2010)). This paper extends these methods to moment condition models with heterogeneous populations. Comparison of the finite sample performance of the proposed estimators is conducted through Monte Carlo simulations.

Key Words: Moment condition models; Heterogenous populations; Optimal estimating equations; Generalized method of moments; Weighted method of moments; Empirical likelihood.
JEL Classification Numbers: C13, C14.

1. INTRODUCTION

A wide range of econometric and statistical models are specified through over-identified moment restrictions in the name of *moment condition models* (also known as the *estimating equation models*). Efficient estimation of such models essentially employs two types of ideas: optimally combining estimation equations (e.g., the optimal estimating equations (OEE) of Godambe (1976), the generalized method of moments (GMM) of Hansen (1982), the generalized estimating equations (GEE) of Liang and Zeger (1986) and the empirical likelihood (EL) of Owen (1988) and Qin and

* I thank the managing editor and a referee for helpful comments. Part of the paper was written when the author was a visiting scholar at the Sloan School of Management of MIT. The research is supported by the National Science Foundation of China (Grant No. 11001059).

1529-7373/2011

Lawless (1994)), and optimally combining estimators (e.g., the weighted method of moments (WMM) of Xiao (2010)). Those methods were primarily developed for data from a homogeneous population, i.e., all samples are generated by a common data generating process.

For the majority of applications, assuming all data observations come from a single population can serve the purpose of statistical modeling satisfactorily; however, in practice, empirical researchers often face situations where data were collected from multiple sources (e.g., data obtained from cluster sampling or from multiple- center/region/country studies), or data have different structure or dimensions (e.g., incomplete longitudinal/panel data due to design or random missing), and under those circumstances, the homogenous-population assumption may seem untenable (see e.g., Jöreskog (1977), Muthén (1989), Muthén et al. (1997), Girma (2000), and Shao et al. (2011)).

There are some sporadic studies that generalized to some extent the aforementioned methods to heterogenous populations. Under the structural equation modeling framework, Wansbeek and Meijer (2000, pp.217) suggested that one can minimize a weighted sum of subgroup-based GMM objective functions to extend GMM to data with multiple subpopulations; however, asymptotic properties (e.g., asymptotic normality and asymptotic efficiency) were not established therein. Similar to Wansbeek and Meijer (2000), Shao et al. (2011) developed a multiple-population based GMM estimator and established its asymptotic properties for linear unbalanced panel data with errors-in-variables. They also discussed the extension of the OEE idea to multiple populations. However, their results are specific to a class of linear moment condition models and need to be generalized. For EL methodology, Owen (2001, pp. 223-226) discussed the hypothesis testing with multiple samples and just-identified moment conditions. To the best of our knowledge, general theories about the OEE, the GMM, the EL and the WMM methods in multiple populations are not available.

The purpose of this paper is to extend these four estimation methods to general nonlinear moment condition models with heterogeneous populations. In Section 2 we outline the model setup and some regularity conditions. In Section 3 we discuss in detail how to extend the ideas of OEE, GMM, EL and WMM to multiple samples. We prove the asymptotic properties of the resulting estimators, and establish their asymptotic equivalence. Section 4 discusses the relationship among the four estimators, and compares them from the perspective of finite sample bias. Section 5 presents Monte Carlo simulations to compare the finite sample properties of the proposed estimators, and Section 6 concludes. All proofs are relegated to the Appendix. Throughout the paper, $\stackrel{p}{\to}$ denotes convergence in probability, and $\stackrel{d}{\to}$ denotes convergence in distribution.

2. MODEL SETUP

We have independent observations $X_{k,1}, ..., X_{k,n_k} \in R^{q_k} \sim F_{k,0}$ ($k = 1, ..., K$), where $q_1, ..., q_K, K$ are fixed positive integers, and $F_{1,0}, ..., F_{K,0}$ are unknown distributions with supports on $\mathcal{X}_1, ..., \mathcal{X}_K$–subsets of $R^{q_1}, ..., R^{q_K}$, respectively. The parameter of interest is θ_0, an interior point of Θ, which is assumed to be a compact subset of R^p. Assume that θ_0 is the unique solution to the following moment condition

$$E[\psi_k(X_{k,i}, \theta)] = 0, k = 1, ..., K, \tag{1}$$

where $\psi_k : R^{p+q_k} \mapsto R^{m_k}, k = 1, ..., K$, are known vector-valued functions.[1] Assume that $m_k \geq p$, i.e., the set of moment conditions for every population is identified. The total sample size is $n = \sum_{k=1}^{K} n_k$. In the following, $\|\cdot\|$ denotes the Euclidean norm.

Assumption 1. *For any $k = 1, \ldots, K$, and for any $x_k \in \mathcal{X}_k$, $\psi_k(x_k, \theta)$ is a continuously differentiable function of θ on Θ.*

Assumption 2. *For any $k = 1, \ldots, K$, $E[\sup_{\theta \in \Theta} \|\psi_k(X_{k,i}, \theta)\|] < \infty$.*

Assumption 3. *There exist a neighborhood $U(\theta_0)$ of θ_0 and integrable functions $\Psi_1, \ldots, \Psi_K$ such that for any $k = 1, \ldots, K$,*

$$\sup_{\theta \in U(\theta_0)} \|\psi_k(x_k, \theta)\|^3 \leq \Psi_k(x_k), \text{ for any } x_k \in \mathcal{X}_k.$$

Assumption 4. *For any $k = 1, \ldots, K$, $G_k = E[\frac{\partial \psi_k(X_{k,i}, \theta_0)}{\partial \theta}]$ exists and $rank(G_k) = p$.*

Assumption 5. *There exist a neighborhood $V_1(\theta_0)$ of θ_0 and integrable functions $\Phi_1, \ldots, \Phi_K$ such that for any $k = 1, \ldots, K$,*

$$\sup_{\theta \in V_1(\theta_0)} \left\| \frac{\partial \psi_k(x_k, \theta)}{\partial \theta} \right\| \leq \Phi_k(x_k), \text{ for any } x_k \in \mathcal{X}_k.$$

[1] A special case of model (1)–often adopted by applied researchers–is that $\theta = (\theta_1, ..., \theta_K)$, and $\psi_k(X_{k,i}, \theta)$ depends on θ only through θ_k.

Assumption 6. *There exist a neighborhood $V_2(\theta_0)$ of θ_0 and integrable functions $\Upsilon_1, \ldots, \Upsilon_K$ such that for any $k = 1, \ldots, K$, $\frac{\partial^2 \psi_k(x_k,\theta)}{\partial\theta\partial\theta'}$ is continuous in θ on $V_2(\theta_0)$, and*

$$\sup_{\theta \in V_2(\theta_0)} \left\| \frac{\partial^2 \psi_k(x_k, \theta)}{\partial\theta\partial\theta'} \right\|^2 \leq \Upsilon_k(x_k), \text{ for any } x_k \in \mathcal{X}_k.$$

Assumption 7. *For any $k = 1, \ldots, K$, $\Sigma_k = E[\psi_k(X_{k,i}, \theta_0)\psi_k(X_{k,i}, \theta_0)']$ is positive definite.*

Assumption 8. *For $k = 1, \ldots, K$, there exists a constant α_k such that $\frac{n_k}{n} \to \alpha_k \in (0, 1)$, as $n \to \infty$.*

3. OEE, GMM, EL AND WMM FOR MULTIPLE POPULATIONS

In this section we extend the OEE method, the GMM method, the EL method and the WMM method to moment condition models with multiple populations.

3.1. OEE for multiple populations

The idea of OEE is to linearly combine all moment conditions into a set of just-identified moment conditions such that the resulting estimator has the smallest asymptotic variance. Extending the OEE method for multiple populations is simple. For a set of K linear combination coefficient matrices $\Pi_1, ..., \Pi_K$, with dimensions $p \times m_1, ..., p \times m_K$ respectively, we solve for an estimator $\hat{\theta}_{EE}$ from the following estimating equation

$$\sum_{k=1}^{K} \Pi_k \sum_{i=1}^{n_k} \psi_k(X_{k,i}, \theta) = 0. \tag{2}$$

The following theorem summarizes the asymptotic properties of $\hat{\theta}_{EE}$.

Theorem 1. *If Assumptions 1,2,4,7 and 8 hold, then as $n \to \infty$, $\sqrt{n}(\hat{\theta}_{EE} - \theta_0) \xrightarrow{d} N(0, \Omega_1(\Pi))$, and for any $\Pi = (\Pi_1, ..., \Pi_K)$, we have that*

$$\Omega_1(\Pi) \geq \Omega^* = \left(\sum_{k=1}^{K} \alpha_k G_k' \Sigma_k^{-1} G_k \right)^{-1},$$

in the sense of being nonnegative definite. The equality is achieved when $\Pi_k = G_k' \Sigma_k^{-1}$ *for* $k = 1, ..., K$*. If practice, if* $\tilde{\theta}$ *is a first step consistent estimator of* θ_0*, then* Π_k *can be consistently estimated by* $\hat{\Pi}_k = \hat{G}_k' \hat{\Sigma}_k^{-1}$*, with*

$$\hat{G}_k(\tilde{\theta}) = \frac{1}{n_k} \sum_{i=1}^{n_k} \left[\frac{\partial \psi_k(X_{k,i}, \tilde{\theta})}{\partial \theta} \right] \tag{3}$$

and

$$\hat{\Sigma}_k(\tilde{\theta}) = \frac{1}{n_k} \sum_{i=1}^{n_k} \left[\psi_k(X_{k,i}, \tilde{\theta}) \psi_k(X_{k,i}, \tilde{\theta})' \right]. \tag{4}$$

The solution $\hat{\theta}_{OEE}$ to (2), with Π_k replaced by $\hat{\Pi}_k = \hat{G}_k' \hat{\Sigma}_k^{-1}$, is call the OEE estimator.

3.2. GMM for multiple populations

Define $\overline{\psi}_k(\theta) = \frac{1}{n_k} \sum_{i=1}^{n_k} \psi_k(X_{k,i}, \theta), k = 1, ..., K$. Let $\hat{W}_k$ be an $m_k \times m_k$ positive semidefinite matrix and that $\hat{W}_k \xrightarrow{p} W_k$ as $n_k \to \infty$, where W_k is also a positive semidefinite matrix. Using $\hat{W}_k$ as weighting matrix, the GMM objective function for population k is

$$J_k(\theta) = n_k \overline{\psi}_k(\theta)' \hat{W}_k \overline{\psi}_k(\theta). \tag{5}$$

To obtain an GMM estimator using all data information, it is natural to consider minimizing the following objective function, as suggested by Wansbeek and Meijer's (2000):

$$J(\theta) = \sum_{k=1}^{K} J_k(\theta) = \sum_{k=1}^{K} n_k \overline{\psi}_k(\theta)' \hat{W}_k \overline{\psi}_k(\theta).$$

Let $\hat{\theta}_{GMM}$ be the minimizer of $J(\theta)$. We establish the following result about the asymptotic distribution of $\hat{\theta}_{GMM}$.

THEOREM 2. *If Assumptions 1,2,4,7 and 8 hold, then as* $n \to \infty$, $\sqrt{n}(\hat{\theta}_{GMM} - \theta_0) \xrightarrow{d} N(0, \Omega_2(W))$*, with*

$$\Omega_2(W) = \left(\sum_{k=1}^{K} \alpha_k G_k' W_k G_k \right)^{-1} \left(\sum_{k=1}^{K} \alpha_k G_k' W_k \Sigma_k W_k G_k \right) \left(\sum_{k=1}^{K} \alpha_k G_k' W_k G_k \right)^{-1}.$$

For any $W = (W_1, ..., W_K)$*, we have that* $\Omega_2(W) \geq \Omega^*$*. The equality is achieved when* $W_k = \Sigma_k^{-1}$ *for every* k*.*

Let $\hat{\Sigma}_k$ be a consistent estimators for Σ_k, $k = 1, ..., K$. For example, $\hat{\Sigma}_k$ can be specified by (4). Then the GMM estimator

$$\hat{\theta}_{eGMM} = \arg\min_{\theta \in \Theta} \sum_{k=1}^{K} n_k \overline{\psi}_k(\theta)' \hat{\Sigma}_k^{-1} \overline{\psi}_k(\theta)$$

has an asymptotic variance that achieves the lower bound Ω^*. It is called an efficient GMM estimator of θ_0. Setting all $\hat{W}_k$ as identity matrices, we obtain a consistent GMM estimator, which can serve as the first step consistent estimator $\tilde{\theta}$ of θ_0.

3.3. EL for multiple populations

We consider generalizing EL in the following manner. We first solve the following restricted optimization problem:

$$\max_{p_{k,i}: i=1,...,n_k, k=1,...,K} \prod_{k=1}^{K} \prod_{i=1}^{n_k} n_k p_{k,i},$$

subject to the constraints that for any i and k,

$$p_{k,i} \geq 0, \ \sum_{i=1}^{n_k} p_{k,i} = 1, \ \sum_{i=1}^{n_k} p_{k,i} \psi_k(X_{k,i}, \theta) = 0.$$

For any given θ, we solve the maximization problem by Lagrange multiplier. Define

$$\mathcal{L} = \sum_{k=1}^{K} \sum_{i=1}^{n_k} \log p_{k,i} + \sum_{k=1}^{K} \gamma_k \left(1 - \sum_{i=1}^{n_k} p_{k,i}\right) - \sum_{k=1}^{K} n_k \lambda_k' \left(\sum_{i=1}^{n_k} p_{k,i} \psi_k(X_{k,i}, \theta)\right),$$

where γ_k and $\lambda_k = (\lambda_{k,1}, ..., \lambda_{k,m_k})'$, $k = 1, ..., K$, are Lagrange multipliers. From the first order conditions we have

$$p_{k,i} = \frac{1}{n_k} \frac{1}{1 + \lambda_k' \psi_k(X_{k,i}, \theta)}, \tag{6}$$

and

$$\sum_{i=1}^{n_k} \frac{1}{1 + \lambda_k' \psi_k(X_{k,i}, \theta)} \psi_k(X_{k,i}, \theta) = 0. \tag{7}$$

For given $\mathbf{X}_k = (X_{k,1}, ..., X_{k,n_k})$, eqn. (7) defines an implicit functional relationship between λ_k and θ. Solving for λ_k in terms of $\mathbf{X}_k$ and θ, we have $\lambda_k = \lambda_k(\mathbf{X}_k, \theta)$. For notational convenience let's write $\lambda_k(\mathbf{X}_k, \theta)$ as $\lambda_k(\theta)$. Substituting (6) into $\mathcal{L}$ and denoting the resulting objective function by $R(\theta, \lambda)$, we have

$$R(\theta, \lambda) = -\sum_{k=1}^{K} n_k \log n_k - \sum_{k=1}^{K}\sum_{i=1}^{n_k} \log(1 + \lambda_k' \psi_k(X_{k,i}, \theta)).$$

Note that $\lambda(\theta) = (\lambda_1(\theta)', ..., \lambda_K(\theta)')'$ is the minimizer of $R(\theta, \lambda)$, i.e.,

$$\lambda(\theta) = \arg\min_{\lambda} R(\theta, \lambda).$$

The profile empirical log-likelihood for θ is

$$R(\theta) = R(\theta, \lambda(\theta)) = -\sum_{k=1}^{K} n_k \log n_k - \sum_{k=1}^{K}\sum_{i=1}^{n_k} \log(1 + \lambda_k(\theta)' \psi_k(X_{k,i}, \theta)).$$

Maximizing $R(\theta)$ is equivalent to minimizing the empirical log-likelihood ratio

$$l_E(\theta) = \sum_{k=1}^{K}\sum_{i=1}^{n_k} \log(1 + \lambda_k(\theta)' \psi_k(X_{k,i}, \theta)).$$

The minimizer of $l_E(\theta)$, denoted by $\hat{\theta}_{EL}$, is called the EL estimator of θ_0. The following result summarizes the asymptotic properties of $\hat{\theta}_{EL}$.

THEOREM 3. *Suppose Assumptions 1,3,5,7 and 8 hold. (i) As $n \to \infty$, with probability one $l_E(\theta)$ attains its minimum at some point $\hat{\theta}_{EL}$ in the interior of ball $B_{\theta_0} = \{\theta : \|\theta - \theta\| \le n^{-\frac{1}{3}}\}$, and $\hat{\theta}_{EL}$ and $\hat{\lambda}_k = \lambda_k(\hat{\theta}_{EL}), k = 1, ..., K$ satisfy $Q_1(\hat{\theta}_{EL}, \hat{\lambda}) = 0,\ Q_2(\hat{\theta}_{EL}, \hat{\lambda}) = 0$, where*

$$Q_1(\theta, \lambda) = -\frac{1}{n}\frac{\partial R(\theta, \lambda)}{\partial \lambda} = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n_1} \left[\frac{1}{1+\lambda_1'\psi_1(X_{1,i},\theta)} \psi_1(X_{1,i}, \theta)\right] \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n_K} \left[\frac{1}{1+\lambda_K'\psi_K(X_{K,i},\theta)} \psi_K(X_{K,i}, \theta)\right] \end{bmatrix},$$

and

$$Q_2(\theta, \lambda) = -\frac{1}{n}\frac{\partial R(\theta, \lambda)}{\partial \theta} = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n_1} \left[\frac{1}{1+\lambda_1'\psi_1(X_{1,i},\theta)} (\frac{\partial \psi_1(X_{1,i},\theta)}{\partial \theta})' \lambda_1\right] \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n_K} \left[\frac{1}{1+\lambda_K'\psi_K(X_{K,i},\theta)} (\frac{\partial \psi_K(X_{K,i},\theta)}{\partial \theta})' \lambda_K\right] \end{bmatrix}.$$

(ii) Suppose further that Assumption 6 holds. Then as $n \to \infty$, $\sqrt{n}(\hat{\theta}_{EL} - \theta_0) \xrightarrow{d} N(0, \Omega^*)$.

3.4. WMM for multiple populations

As will be shown in Section 4, OEE, GMM and EL are all methods to optimally combining moment conditions. Xiao (2010) proposed the WMM method from a new perspective–optimally combining method of moment estimators. We now discuss the extension of WMM to multiple populations.

For each k, we have m_k moment conditions. We can select p conditions out of those m_k conditions to form a just-identified model and solve for a method of moment estimator. There are $M_k = C_{m_k}^p$ ways to select moments, thus producing M_k inefficient method of moments estimators $\hat{\theta}_{WMM,l}^{(k)}, l = 1, ..., M_k$. Repeating the procedure for all populations, we obtain $M = \sum_{k=1}^{K} M_k$ method of moments estimators. The idea of WMM is to construct an efficient estimator of θ_0 by linearly combining the M method of moments estimators. Specifically, the WMM estimator is defined as

$$\hat{\theta}_{WMM} = \sum_{k=1}^{K}\sum_{l=1}^{M_k} \Lambda_{k,l} \hat{\theta}_{WMM,l}^{(k)}, \tag{8}$$

where $\Lambda_{k,l}$ $(k = 1, ..., K; l = 1, ..., M_k)$ are $p \times p$ linear combination coefficient matrices such that $\sum_{k=1}^{K}\sum_{l=1}^{M_k} \Lambda_{k,l} = I_p$, with I_p the identity matrix of order p.

THEOREM 4. *If Assumptions 1,2,4,7 and 8 hold, then as* $n \to \infty$, $\sqrt{n}(\hat{\theta}_{WMM} - \theta_0) \xrightarrow{d} N(0, \Omega_3(\Lambda))$, *and for any* $\Lambda = \{\Lambda_{k,l}, k = 1, ..., K; l = 1, ..., M_k\}$, *we have that* $\Omega_3(\Lambda) \geq \Omega^*$. *The equality is achieved when*

$$\Lambda_{k,l} = \frac{\alpha_k}{C_{m_k-1}^{p-1}} \Omega^* G_k' \Sigma_k^{-1} P_{S_l^k}' P_{S_l^k} G_k \tag{9}$$

for $k = 1, ..., K; l = 1, ..., M_k$, *where* $P_{S_l^k}$ *is the* l^{th} m_k*-select-p matrix. If practice, if* $\tilde{\theta}$ *is a first step consistent estimator of* θ_0, *then* $\Lambda_{k,l}$ *can be consistently estimated by*

$$\hat{\Lambda}_{k,l} = \frac{n_k}{n C_{m_k-1}^{p-1}} \hat{\Omega}^*(\tilde{\theta}) [\hat{G}_k(\tilde{\theta})]' [\hat{\Sigma}_k(\tilde{\theta})]^{-1} P_{S_l^k}' P_{S_l^k} \hat{G}_k(\tilde{\theta}), \tag{10}$$

where $\hat{G}_k(\tilde{\theta})$ and $\hat{\Sigma}_k(\tilde{\theta})$ are defined in Theorem 1, and

$$\hat{\Omega}^*(\tilde{\theta}) = \left[\sum_{k=1}^{K} \frac{n_k}{n} [\hat{G}_k(\tilde{\theta})]' [\hat{\Sigma}_k(\tilde{\theta})]^{-1} \hat{G}_k(\tilde{\theta}) \right]^{-1}.$$

The estimator

$$\hat{\theta}_{eWMM} = \sum_{k=1}^{K} \sum_{l=1}^{M_k} \hat{\Lambda}_{k,l} \hat{\theta}^{(k)}_{WMM,l}$$

with $\hat{\Lambda}_{k,l}$ specified by (10) is called the efficient WMM estimator.

4. CONNECTIONS OF OEE, GMM, EL AND WMM

By definition, the OEE estimator is efficient, in the sense that it has the smallest variance among all estimators obtained by linearly combining all the moment conditions in (1). Since the GMM, the EL and the WMM estimators have the same asymptotic distribution as the OEE estimator, they are also efficient. The four estimators were derived based on different principles, therefore, in general they should be different in finite samples. The following result states that for linear models, OEE, GMM and WMM produce identical estimators.

Theorem 5. *For linear moment condition models, i.e., models where $\psi_k(X_{k,i}, \theta), k = 1, ..., K$ are linear functions of θ, $\hat{\theta}_{OEE} = \hat{\theta}_{eGMM} = \hat{\theta}_{eWMM}$.*

The proof of Theorem 5 is by some matrix algebra and is omitted. We now follow Newey and Smith (2004) to discuss the finite sample biases of the four estimators. It is easy to see that $\hat{\theta}_{OEE}$ satisfies

$$\sum_{k=1}^{K} n_k \sum_{i=1}^{n_k} [\hat{G}_k(\tilde{\theta})]' [\hat{\Sigma}_k(\tilde{\theta})]^{-1} \psi_k(X_{k,i}, \hat{\theta}_{OEE}) = 0. \tag{11}$$

From the first order condition of GMM, $\hat{\theta}_{eGMM}$ satisfies

$$\sum_{k=1}^{K} n_k \sum_{i=1}^{n_k} [\hat{G}_k(\hat{\theta}_{eGMM})]' [\hat{\Sigma}_k(\tilde{\theta})]^{-1} \psi_k(X_{k,i}, \hat{\theta}_{eGMM}) = 0. \tag{12}$$

A direct extension of Theorem 2.3 of Newey and Smith (2004) implies that $\hat{\theta}_{EL}$ satisfies the following first order condition:

$$\sum_{k=1}^{K} n_k \sum_{i=1}^{n_k} [\tilde{G}_k(\hat{\theta}_{EL})]'[\tilde{\Sigma}_k(\hat{\theta}_{EL})]^{-1}\psi_k(X_{k,i},\hat{\theta}_{EL}) = 0, \tag{13}$$

where $\tilde{G}_k(\theta) = \frac{1}{n_k}\sum_{i=1}^{n_k}\left[\hat{p}_{k,i}\frac{\partial\psi_k(X_{k,i},\theta)}{\partial\theta}\right]$, $\tilde{\Sigma}_k(\theta) = \frac{1}{n_k}\sum_{i=1}^{n_k}\left[\hat{p}_{k,i}\psi_k(X_{k,i},\theta)\psi_k(X_{k,i},\theta)\right]'$, and $\hat{p}_{k,i}$ $(i = 1, ..., n_k, k = 1, ..., K)$ are empirical probabilities.

Comparing (11), (12) and (13), we can see that OEE, GMM and EL share a common structure, all being estimators derived by linearly combining moment conditions. The ideal (yet infeasible) way to linearly combine moment conditions is

$$\sum_{k=1}^{K} n_k \sum_{i=1}^{n_k} G_k' \Sigma_k^{-1}\psi_k(X_{k,i},\hat{\theta}_{ideal}) = 0. \tag{14}$$

According to the stochastic higher order expansion theory of Newey and Smith (2004), EL should have smaller finite sample bias than OEE and G-MM, since $\tilde{G}_k(\hat{\theta}_{EL}), \tilde{\Sigma}_k(\hat{\theta}_{EL})$ are efficient estimators of G_k, Σ_k respectively, while $\hat{G}_k(\tilde{\theta}), \hat{\Sigma}_k(\tilde{\theta})$ and $\hat{G}_k(\hat{\theta}_{eGMM}), \hat{\Sigma}_k(\tilde{\theta})$ are not. WMM uses the same first step information as OEE, we expect that its finite sample performance should be similar to that of OEE.

5. MONTE CARLO STUDIES

In this section we inspect the finite sample performance of OEE, GM-M, EL and WMM using simulations.[2] Consider $K = 2$, i..e., we have two independent populations. Suppose the moment conditions for the first population is $E[g(X_i, \theta_0)] = 0$, with $g(x, \theta) = (x - \theta, x^2 - \theta^2 - 2\theta)'$, and $X_1, \cdots, X_{n_1}$ are i.i.d. observations from chi-square distribution with degree of freedom θ_0. The moment conditions for the second population is $E[h(Y_i, \theta_0)] = 0$, with $h(y, \theta) = (y - \frac{\theta}{2}, y^2 - \frac{3}{4}\theta^2)'$, and $Y_1, \cdots, Y_{n_2}$ are i.i.d. observations from gamma distribution with scale parameter θ_0 and shape parameter 1/2.

We experiment with $\theta_0 = 2$ and total sample sizes $n = 150, 600, 1500$. Denote by α the fraction of observations from population 1. We consider

[2]In econometric literature, it is well known that GMM has finite sample bias and EL can improve upon GMM. See, e.g., Altonji and Segal (1996), Imbens (2002), Newey and Smith (2004) and Kitamura (2006).

two scenarios: $\alpha = 1/3$ and $\alpha = 2/3$. We run 10,000 replications each time. Estimation results are reported in Table 1.[3]

TABLE 1.

Performance of the multiple-population OEE, GMM, EL and WMM estimators

	150 Observations							
	$\alpha = 1/3$				$\alpha = 2/3$			
	GMM	WMM	OEE	EL	GMM	WMM	OEE	EL
Estimate	1.8780	1.8608	1.8599	1.9103	1.8929	1.8659	1.8750	1.9137
Std. Err.	0.2126	0.2320	0.2308	0.1829	0.1877	0.2201	0.2095	0.1803
	600 Observations							
	$\alpha = 1/3$				$\alpha = 2/3$			
	GMM	WMM	OEE	EL	GMM	WMM	OEE	EL
Estimate	1.9712	1.9725	1.9716	1.9693	1.9720	1.9711	1.9722	1.9744
Std. Err.	0.0978	0.0993	0.0991	0.1133	0.0881	0.0925	0.0906	0.1116
	1500 Obervations							
	$\alpha = 1/3$				$\alpha = 2/3$			
	GMM	WMM	OEE	EL	GMM	WMM	OEE	EL
Estimate	1.9894	1.9908	1.9902	1.9762	1.9903	1.9909	1.9910	1.9846
Std. Err.	0.0612	0.0613	0.0613	0.1004	0.0527	0.0532	0.0529	0.1028

We observe from Table 1 that when sample size is small, all four estimators have nonnegligible biases. Among them, EL has the smallest bias and smallest variance. WMM and OEE are very similar to each other, and their biases and variances are bigger than those of GMM. As sample size increases, the biases of all estimators decrease. In terms of the rate of bias diminishing, WMM and OEE are the fastest, and EL is the slowest. In large samples, WMM and OEE have apparent advantage (smaller bias and smaller variance) over EL. Change in α does not change the above pattern.

6. CONCLUDING REMARKS

In this article, we discussed the extensions of four types of semiparametrically efficient estimators (OEE, GMM, EL and WMM) to moment condition models with multiple independent populations. The paper didn't address the extension of the GEE of Liang and Zeger (1986). However, since the idea of GEE is in line with the idea of OEE, the extension is immediate. Although our focus is parameter estimation, asymptotic hypothesis testing results can be easily obtained. Results in the paper were

[3]Computation of the EL estimator is based on a modified algorithm of Hansen (2003). For each simulation run, the starting value of the EL algorithm is a randomly generated number from the uniform distribution on $[1.5, 2.5]$, the center of which is θ_0.

proved using standard techniques combined with matrix algebra. The assumptions we adopted are standard regularity assumptions, but can be relaxed to some extent without jeopardizing validity of the results. For example, Assumption 8 can be dropped, and all the asymptotic results still hold, with the asymptotic covariance matrices appropriately redefined, as $\min\{n_k|k=1,...,K\} \to \infty$. The methods proposed in this paper are applicable in a number of situations, such as cluster sampling and unbalanced panel data. Our simulation results confirmed the small sample bias advantage of EL as established by Newey and Smith (2004). For future research, it is of interest to investigate the cause of the persistent bias of EL even for large samples.

APPENDIX: PROOFS OF RESULTS

We first establish the following lemma, which is an extension of the matrix version of the Cauchy-Schwartz inequality. A special case of this results is given in Wansbeek and Meijer (2000, pp.359-360).

LEMMA 1. *Let* C_k *and* $\Gamma_k, k=1,...,K$ *be* $p_k \times q_k$ *and* $p_k \times p_k$ *matrices, respectively,* Γ_k *is positive definite for any* k*, and that* $(\sum_{k=1}^{K} C_k' \Gamma_k^{-1} C_k)^{-1}$ *exists. Let* $\Pi_k, k=1...,K$*, of order* $q_k \times p_k$ *respectively, be a collection of matrices such that* $\left(\sum_{k=1}^{K} \Pi_k C_k\right)^{-1}$ *exists. Then*

$$\left(\sum_{k=1}^{K}\Pi_k C_k\right)^{-1}\left(\sum_{k=1}^{K}\Pi_k\Gamma_k\Pi_k'\right)\left(\sum_{k=1}^{K}C_k'\Pi_k'\right)^{-1} \geq \left(\sum_{k=1}^{K}C_k'\Gamma_k^{-1}C_k\right)^{-1}. \tag{A.1}$$

The equality holds when $\Pi_k = C_k'\Gamma_k^{-1}$ *for every* k.

Proof (**Proof of Lemma 1**). Define $\Pi = \begin{bmatrix} \Pi_1 & & \\ & \ddots & \\ & & \Pi_K \end{bmatrix}, C = \begin{bmatrix} C_1 & & \\ & \ddots & \\ & & C_K \end{bmatrix}$, and $\Pi = \begin{bmatrix} \Gamma_1 & & \\ & \ddots & \\ & & \Gamma_K \end{bmatrix}$. Then (A.1) can be expressed as

$$(\Pi C)^{-1}(\Pi\Gamma\Pi')(C'\Pi')^{-1} \geq (C'\Gamma^{-1}C)^{-1}.$$

Let $M = (\Pi C)^{-1}\Pi$, $U = \Gamma^{\frac{1}{2}}M'$ and $V = \Gamma^{-\frac{1}{2}}C$. Then $U'V = I$, hence

$$U'U - (V'V)^{-1} = [U - V(V'V)^{-1}V'U]'[U - V(V'V)^{-1}V'U] \geq 0.$$

Therefore $(\Pi C)^{-1}(\Pi\Gamma\Pi')(C'\Pi')^{-1} = M\Gamma M' = U'U \geq (V'V)^{-1} = (C'\Gamma^{-1}C)^{-1}$. ▮

Proof (**Proof of Theorem 1**). Let $H(\theta) = \sum_{k=1}^{K}\Pi_k\sum_{i=1}^{n_k}\psi_k(X_{k,i},\theta)$.Take a first order Taylor expansion of $H(\theta)$ around θ_0 we have

$$H(\theta) = H(\theta_0) + \left(\sum_{k=1}^{K} n_k\Pi_k\frac{\partial\overline{\psi}_k(\tilde{\theta})}{\partial\theta'}\right)(\theta-\theta_0).$$

Solving $H(\hat{\theta}_{EE}) = 0$ we obtain

$$\hat{\theta}_{EE} - \theta_0 = -\left(\sum_{k=1}^{K} n_k\Pi_k\frac{\partial\overline{\psi}_k(\tilde{\theta})}{\partial\theta'}\right)H(\theta_0).$$

Hence

$$\begin{aligned}\sqrt{n}(\hat{\theta}_{EE}-\theta_0) &= -\left(\sum_{k=1}^{K}\frac{n_k}{n}\Pi_k\frac{\partial\overline{\psi}_k(\tilde{\theta})}{\partial\theta'}\right)\frac{1}{\sqrt{n}}H(\theta_0)\\ &= -\left(\sum_{k=1}^{K}\frac{n_k}{n}\Pi_k\frac{\partial\overline{\psi}_k(\tilde{\theta})}{\partial\theta'}\right)\left(\sum_{k=1}^{K}\sqrt{\frac{n_k}{n}}\Pi_k\sqrt{n_k}\overline{\psi}_k(\theta_0)\right)\\ &\xrightarrow{d} N(0,\Omega_1(\Pi)),\end{aligned}$$

with

$$\Omega_1(\Pi) = \left(\sum_{k=1}^{K}\alpha_k\Pi_kG_k\right)\left(\sum_{k=1}^{K}\alpha_k\Pi_k\Sigma_k\Pi_k'\right)\left(\sum_{k=1}^{K}\alpha_k\Pi_kG_k\right)'.$$

By Lemma 1, $\Omega_1(\Pi) \geq \left(\sum_{k=1}^{K}\alpha_kG_k'\Sigma_k^{-1}G_k\right)^{-1}$ for any Π, with the equality hold if $\Pi_k = G_k'\Sigma_k^{-1}$ for any k. ▮

Proof (**Proof of Theorem 2**). The first order condition for (5) is

$$2H(\hat{\theta}) = \frac{\partial J(\hat{\theta})}{\partial\theta} = \sum_{k=1}^{K}2n_k\left[\frac{\partial\overline{\psi}_k(\hat{\theta})}{\partial\theta}\right]'\hat{W}_k\overline{\psi}_k(\hat{\theta}) = 0.$$

Taking a Taylor expansion of $H(\theta)$ around θ_0 we have

$$0 - H(\theta_0) = H(\hat{\theta}) - H(\theta_0) = \frac{\partial H(\overline{\theta})}{\partial\theta'}(\hat{\theta}-\theta_0),$$

hence

$$\begin{aligned}\hat{\theta}-\theta_0 &= -\left[\frac{\partial H(\overline{\theta})}{\partial\theta'}\right]^{-1} H(\theta_0) = -\left[\frac{1}{n}\frac{\partial H(\overline{\theta})}{\partial\theta'}\right]^{-1}\frac{1}{n}H(\theta_0)\\ \sqrt{n}(\hat{\theta}-\theta_0) &= -\left[\frac{1}{n}\frac{\partial H(\overline{\theta})}{\partial\theta'}\right]^{-1}\frac{1}{\sqrt{n}}H(\theta_0).\end{aligned}$$

While

$$\begin{aligned}\frac{1}{\sqrt{n}}H(\theta_0) &= \frac{1}{\sqrt{n}}\sum_{k=1}^{K} n_k\left[\frac{\partial\overline{\psi}_k(\theta_0)}{\partial\theta}\right]'\hat{W}_k\overline{\psi}_k(\theta_0)\\ &= \sum_{k=1}^{K}\sqrt{\frac{n_k}{n}}\left[\frac{\partial\overline{\psi}_k(\theta_0)}{\partial\theta}\right]'\hat{W}_k\sqrt{n_k}\overline{\psi}_k(\theta_0)\\ &\xrightarrow{d} N(0,V),\end{aligned}$$

with $V=\sum_{k=1}^{K}(\alpha_k G_k' W_k\Sigma_k W_k G_k)$, where the last convergence is due to the Law of Large Numbers, the Central Limit Theorem , the Slutsky Theorem, and the independence of populations. Now

$$\begin{aligned}\frac{1}{n}\frac{\partial H(\overline{\theta})}{\partial\theta'} &= \frac{1}{n}\frac{\partial}{\partial\theta'}\sum_{k=1}^{K} n_k\left[\frac{\partial\overline{\psi}_k(\overline{\theta})}{\partial\theta}\right]'\hat{W}_k\overline{\psi}_k(\overline{\theta})\\ &= \sum_{k=1}^{K}\frac{n_k}{n}\left[\frac{\partial\overline{\psi}_k(\overline{\theta})}{\partial\theta}\right]'\hat{W}_k\frac{\partial\overline{\psi}_k(\overline{\theta})}{\partial\theta}+\sum_{k=1}^{K}\frac{n_k}{n}[\overline{\psi}_k(\overline{\theta})]'\hat{W}_k\frac{\partial}{\partial\theta'}(\frac{\partial\overline{\psi}_k(\overline{\theta})}{\partial\theta})\\ &\xrightarrow{p}\sum_{k=1}^{K}\alpha_k G_k' W_k G_k,\end{aligned}$$

since $\overline{\theta}\xrightarrow{p}\theta_0$, $\frac{\partial\overline{\psi}_k(\overline{\theta})}{\partial\theta}\xrightarrow{p}G_k, \overline{\psi}_k(\overline{\theta})\xrightarrow{p}0$, and that $\frac{\partial}{\partial\theta'}(\frac{\partial\overline{\psi}_k(\overline{\theta})}{\partial\theta})=O_p(1)$. Therefore by Slutsky Theorem again, we have

$$\sqrt{n}(\hat{\theta}-\theta_0)\xrightarrow{d}\left(\sum_{k=1}^{K}\alpha_k G_k' W_k G_k\right)^{-1} N(0,V)\stackrel{d}{=}N(0,\Omega_2(W)),$$

where

$$\Omega_2(W)=\left(\sum_{k=1}^{K}\alpha_k G_k' W_k G_k\right)^{-1}\sum_{k=1}^{K}(\alpha_k G_k' W_k\Sigma_k W_k G_k)\left(\sum_{k=1}^{K}\alpha_k G_k' W_k G_k\right)^{-1}.$$

By Lemma 1, we have that $\Omega_2(W) \geq \left(\sum_{k=1}^{K} \alpha_k G_k' W_k G_k\right)^{-1}$ for any W. ▮

Proof (**Proof of Theorem 3**). The proof of Theorem 3 is a direct generalization of the proofs of Lemma 1 and Theorem 1 in Qin and Lawless (1994). (i) Since $\lambda_k(\theta)$ is defined by (7), similar to the proof of Owen (1990), we have that uniformly for $\theta \in \{\theta : \|\theta - \theta_0\| \leq n^{-\frac{1}{3}}\}$, $\lambda_k(\theta) = O(n^{-\frac{1}{3}})$, for any k. For any given $\theta \in \{\theta : \|\theta - \theta_0\| = n^{-\frac{1}{3}}\}$, we can express θ as $\theta = \theta_0 + un^{-\frac{1}{3}}$, where $\|u\| = 1$. By a Taylor expansion, we have that

$$
\begin{aligned}
l_E(\theta) &= \sum_{k=1}^{K}\sum_{i=1}^{n_k} \lambda_k' \psi_k(X_{k,i},\theta) - \frac{1}{2}\sum_{k=1}^{K}\sum_{i=1}^{n_k}[\lambda_k'\psi_k(X_{k,i},\theta)]^2 + o(n^{\frac{1}{3}}) \\
&= \sum_{k=1}^{K}\frac{n_k}{2}\left[\frac{1}{n_k}\sum_{i=1}^{n_k}\psi_k(X_{k,i},\theta)\right]'\left[\frac{1}{n_k}\sum_{i=1}^{n_k}\psi_k(X_{k,i},\theta)\psi_k(X_{k,i},\theta)'\right]^{-1} \\
&\quad \times \left[\frac{1}{n_k}\sum_{i=1}^{n_k}\psi_k(X_{k,i},\theta)\right] + o(n^{\frac{1}{3}}) \\
&= \sum_{k=1}^{K}\frac{n_k}{2}\left[\frac{1}{n_k}\sum_{i=1}^{n_k}\psi_k(X_{k,i},\theta_0) + \frac{1}{n_k}\sum_{i=1}^{n_k}\frac{\partial\psi_k(X_{k,i},\theta_0)}{\partial\theta}un^{-\frac{1}{3}}\right]' \\
&\quad \times \left[\frac{1}{n_k}\sum_{i=1}^{n_k}\psi_k(X_{k,i},\theta)\psi_k(X_{k,i},\theta)'\right]^{-1} \\
&\quad \times \left[\frac{1}{n_k}\sum_{i=1}^{n_k}\psi_k(X_{k,i},\theta_0) + \frac{1}{n_k}\sum_{i=1}^{n_k}\frac{\partial\psi_k(X_{k,i},\theta_0)}{\partial\theta}un^{-\frac{1}{3}}\right] + o(n^{\frac{1}{3}}) \\
&= \sum_{k=1}^{K}\frac{n_k}{2}\left[O(n^{-1}\log\log n)^{\frac{1}{2}} + E\left(\frac{\partial\psi_k(X_{k,i},\theta_0)}{\partial\theta}\right)un^{-\frac{1}{3}}\right]' \\
&\quad \times E[\psi_k(X_{k,i},\theta_0)\psi_k(X_{k,i},\theta_0)'] \\
&\quad \times \left[O(n^{-1}\log\log n)^{\frac{1}{2}} + E[\frac{\partial\psi_k(X_{k,i},\theta_0)}{\partial\theta}]un^{-\frac{1}{3}}\right] + o(n^{\frac{1}{3}}) \\
&\geq \sum_{k=1}^{K} c_k n_k^{\frac{1}{3}} \\
&= Cn^{\frac{1}{3}},
\end{aligned}
$$

where C is a positive constant. By a similar argument, we have

$$\begin{aligned} l_E(\theta_0) &= \sum_{k=1}^{K}\frac{n_k}{2}\left[\frac{1}{n_k}\sum_{i=1}^{n_k}\psi_k(X_{k,i},\theta_0)\right]'\left[\frac{1}{n_k}\sum_{i=1}^{n_k}\psi_k(X_{k,i},\theta_0)\psi_k(X_{k,i},\theta_0)'\right]^{-1} \\ &\quad \left[\frac{1}{n_k}\sum_{i=1}^{n_k}\psi_k(X_{k,i},\theta_0)\right] + o(1) \\ &= O(\log\log n). \end{aligned}$$

Thus $\frac{l_E(\theta_0)}{l_E(\theta)} = o(1)$ for any θ on the surface of the ball $\{\theta : \|\theta-\theta_0\| \le n^{-\frac{1}{3}}\}$. Therefore the minimum of $l_E(\theta)$ must be attained in the interior of $\{\theta : \|\theta-\theta_0\| \le n^{-\frac{1}{3}}\}$. By definition, $(\hat{\theta}_{EL},\hat{\lambda})$ minimizes $R(\theta,\lambda)$, thus it satisfies

$$\begin{aligned} R_\theta(\hat{\theta}_{EL},\hat{\lambda}) &= \frac{\partial R(\hat{\theta}_{EL},\hat{\lambda})}{\partial\theta} = 0, \\ R_\lambda(\hat{\theta}_{EL},\hat{\lambda}) &= \frac{\partial R(\hat{\theta}_{EL},\hat{\lambda})}{\partial\lambda} = 0. \end{aligned}$$

(ii) Let $Q(\theta,\lambda) = [R_\theta(\theta,\lambda)', R_\lambda(\theta,\lambda)']'$. Then $(\hat{\theta}_{EL},\hat{\lambda})$ is the solution to the equation $Q(\hat{\theta}_{EL},\hat{\lambda}) = 0$. Taking a Taylor expansion of $Q(\theta,\lambda)$ around $(\theta_0, 0)$ we have

$$Q(\theta,\lambda) = Q(\theta_0,0) + \left[\frac{\partial Q(\theta_0,0)}{\partial\theta'}, \frac{\partial Q(\theta_0,0)}{\partial\lambda'}\right]\begin{bmatrix}\theta-\theta_0\\ \lambda-0\end{bmatrix} + o(\delta),$$

with $\delta = \|\theta-\theta_0\|+\|\lambda\|$. Hence $0 = Q(\theta_0,0)+\left[\frac{\partial Q(\theta_0,0)}{\partial\theta'}, \frac{\partial Q(\theta_0,0)}{\partial\lambda'}\right]\begin{bmatrix}\hat{\theta}_{EL}-\theta_0\\ \hat{\lambda}-0\end{bmatrix}+ o(\hat{\delta})$, with $\hat{\delta} = \|\hat{\theta}_{EL}-\theta_0\| + \|\hat{\lambda}\|$. Therefore

$$\begin{aligned} \begin{bmatrix}\hat{\theta}_{EL}-\theta_0\\ \hat{\lambda}\end{bmatrix} &= -\left[\frac{\partial Q(\theta_0,0)}{\partial\theta'}, \frac{\partial Q(\theta_0,0)}{\partial\lambda'}\right]^{-1}[Q(\theta_0,0)+o(\hat{\delta})] \\ &= -\left[\frac{1}{n}\frac{\partial Q(\theta_0,0)}{\partial\theta'}, \frac{1}{n}\frac{\partial Q(\theta_0,0)}{\partial\lambda'}\right]^{-1}\left[\frac{1}{n}Q(\theta_0,0)+o(\frac{\hat{\delta}}{n})\right] \\ &= -S_n^{-1}\begin{bmatrix}\frac{1}{n}R_\theta(\theta_0,0)+o(\frac{\hat{\delta}}{n})\\ \frac{1}{n}R_\lambda(\theta_0,0)+o(\frac{\hat{\delta}}{n})\end{bmatrix}, \end{aligned}$$

where $S_n = \begin{bmatrix}S_{n,11} & S_{n,12}\\ S_{n,21} & S_{n,22}\end{bmatrix} = \begin{bmatrix}\frac{1}{n}R_{\theta\theta}(\theta_0,0) & \frac{1}{n}R_{\theta\lambda}(\theta_0,0)\\ \frac{1}{n}R_{\lambda\theta}(\theta_0,0) & \frac{1}{n}R_{\lambda\lambda}(\theta_0,0)\end{bmatrix}$. By calculus we have that $S_{n,11} = 0$, $S_{n,12} = \left(\frac{1}{n}\sum_{i=1}^{n_1}\frac{\partial\psi_1(X_{1,i},\theta_0)}{\partial\theta'}', ..., \frac{1}{n}\sum_{i=1}^{n_K}\frac{\partial\psi_K(X_{K,i},\theta_0)}{\partial\theta'}'\right)'$, and $S_{n,22}$ is a block-diagonal matrix whose k^{th} diagonal element is

$$\frac{1}{n}\sum_{i=1}^{n_k}\left[\psi_k(X_{k,i},\theta_0)\psi_k(X_{k,i},\theta_0)'\right],$$

for $k = 1,\ldots,K$. Hence $S_n \xrightarrow{p} S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22}\end{bmatrix}$, where $S_{11} = 0, S_{12} = S_{21}' = [\alpha_1 G_1',\ldots,\alpha_K G_K']$, and S_{22} is a block-diagonal matrix whose k^{th} diagonal element is $\alpha_k\Sigma_k$, for $k = 1,\ldots,K$. Therefore

$$\begin{aligned}\sqrt{n}(\hat\theta_{EL}-\theta_0) &= [S_{12}S_{22}^{-1}S_{21}]^{-1}S_{12}^{-1}S_{22}^{-1}\frac{1}{\sqrt{n}}R_\lambda(\theta_0,0)+o(1)\\ &\xrightarrow{d} N\left(0,[S_{12}S_{22}^{-1}S_{21}]^{-1}\right)\\ &= N\left(0,\left(\sum_{k=1}^{K}\alpha_k G_k'\Sigma_k^{-1}G_k\right)^{-1}\right).\end{aligned}$$

∎

Proof (**Proof of Theorem 4**). Let $P_{S_l^k}$ be the l^{th} m_k-select-p matrix. For population k, we can use the l^{th} set of (just-identified) moment conditions $P_{S_l^k}E[\psi_k(X_{k,i},\theta)] = 0$ to construct a method of moments estimator $\hat\theta^k_{WMM,l}$, which is the solution of $P_{S_l^k}\overline{\psi}_k(\theta) = 0$. Applying a Taylor expansion around θ_0 we have

$$0 - P_{S_l^k}\overline{\psi}_k(\theta_0) = P_{S_l^k}\frac{\partial\overline{\psi}_k(\tilde\theta_{k,l})}{\partial\theta'}(\hat\theta^{(k)}_{WMM,l}-\theta_0),$$

hence

$$\hat\theta^{(k)}_{WMM,l}-\theta_0 = -\left[P_{S_l^k}\frac{\partial\overline{\psi}_k(\tilde\theta_{k,l})}{\partial\theta'}\right]^{-1}P_{S_l^k}\overline{\psi}_k(\theta_0).$$

Now

$$\begin{aligned}\sqrt{n}(\hat\theta_{WMM}-\theta_0) &= \sum_{k=1}^{K}\sum_{l=1}^{M_k}\Lambda_{k,l}\sqrt{n}(\hat\theta^{(k)}_{WMM,l}-\theta_0)\\ &= -\sum_{k=1}^{K}\sum_{l=1}^{M_k}\Lambda_{k,l}\frac{n_k}{\sqrt{n}}\left[\frac{1}{n}P_{S_l^k}\frac{\partial\overline{\psi}_k(\tilde\theta_{k,l})}{\partial\theta'}\right]^{-1}P_{S_l^k}\overline{\psi}_k(\theta_0)\\ &= -\sum_{k=1}^{K}\sum_{l=1}^{M_k}\sqrt{\frac{n_k}{n}}\Lambda_{k,l}\left[\frac{1}{n}P_{S_l^k}\frac{\partial\overline{\psi}_k(\tilde\theta_{k,l})}{\partial\theta'}\right]^{-1}P_{S_l^k}\sqrt{n_k}\overline{\psi}_k(\theta_0)\\ &\xrightarrow{d} N(0,\Omega_3(\Lambda)),\end{aligned}$$

with

$$\Omega_3(\Lambda) = \sum_{k=1}^{K}\sum_{l=1}^{M_k}\alpha_k \Lambda_l^k [P_{S_l^k} G_k]^{-1} P_{S_l^k} \Sigma_k P'_{S_l^k} [G'_k P'_{S_l^k}]^{-1} (\Lambda_{k,l})'.$$

By Lemma 1, $\Omega_3(\Lambda) \geq \left(\sum_{k=1}^{K} \alpha_k G'_k \Sigma_k^{-1} G_k\right)^{-1}$ for any Λ, with the equality hold if

$$\Lambda_{k,l} = \frac{\alpha_k}{C_{m_k-1}^{p-1}} \left(\sum_{k=1}^{K} \alpha_k G'_k \Sigma_k^{-1} G_k\right)^{-1} G'_k \Sigma_k^{-1} P'_{S_l^k} P_{S_l^k} G_k.$$

∎

REFERENCES

Altonji, Joseph G. and Lewis M. Segal, 1996. Small-sample bias in GMM estimation of covariance structures. *Journal of Business and Economic Statistics* **14**, 353-366.

Girma, Sourafel, 2000. A quasi-differencing approach to dynamic modelling from a time series of independent cross-sections. *Journal of Econometrics* **98**, 365-383.

Godambe, Vidyadhar P., and Christopher C. Heyde, 1987. Quasi-likelihood and optimal estimation. *International Statistical Review* **55**, 231-244.

Hansen, Bruce E., 2003. The *elike* MATLAB program (http://www.ssc.wisc.edu/∼bhansen/). University of Wisconsin.

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

Imbens, Guido W., 2002. Generalized method of moments and empirical likelihood. *Journal of Business and Economic Statistics* **20**, 493-506.

Jöreskog, Karl G., 1971. Simultaneous factor analysis in several populations. *Psychometrika* **36**, 409-426.

Kitamura, Yuichi, 2006. Empirical likelihood methods in econometrics: Theory and practice. Cowles Foundation Discussion Paper No. 1569.

Muthén, Bengt O., 1989. Latent variable modeling in heterogenous populations. *Psychometrika* **54**, 557-585.

Muthén, Bengt O., Siek-Toon Khoo. and Jan-Eric Gustafsson, 1997, Multilevel Latent Variable Modeling in Multiple Populations. Unpublished Manuscript.

Liang, Kung-Yee,and Scott L. Zeger, 1986. Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13-22.

Newey, Whitney K. and Richard J. Smith, 2004. Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* **72**, 219-255.

Owen, Art B., 1988. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237-249.

Owen, Art B., 1990. Empirical likelihood confidence regions. *The Annals of Statistics* **18**, 90-120.

Owen, Art B., 2001. *Empirical Likelihood*. New York: Chapman and Hall.

Qin, Jin and Jerry Lawless, 1994. Empirical likelihood and general estimating equations. *The Annals of Statistics* **22**, 300-325.

Shao, Jun, Zhiguo Xiao, and Ruifeng Xu, 2011. Estimation with unbalanced panel data having covariate measurement error. *Journal of Statistical Planning and Inference* **141**, 800-808.

Wansbeek, Tom and Erik Meijer, 2000. *Measurement Error and Latent Variables in Econometrics*. Amsterdam: North-Holland.

Xiao, Zhiguo, 2010. The weighted method of moments for moment condition models. *Economics Letters* **107**, 183-186.