Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators*

Whitney K. Newey
Department of Economics
M.I.T.

Richard J. Smith
Department of Economics
University of Bristol

First Draft: January 2000
Revised, June 2001

Abstract

In an effort to improve the small sample properties of GMM, a number of alternative estimators have been suggested. These include the empirical likelihood (EL), continuous updating, and exponential tilting estimators. We show that these estimators share a common structure, being members of a class of Generalized Empirical Likelihood (GEL) estimators. We use this structure to compare their higher-order asymptotic properties.We find that the asymptotic bias of EL often does not grow with the number of moment restrictions, while that of GMM and other GEL estimators grows without bound. We also use the formulae to derive bias corrected GMM and GEL estimators. We find that bias corrected EL inherits the higher-order property of maximum likelihood, that it is asymptotically efficient relative to the other bias corrected estimators.

JEL Classification: C13, C30

Keywords: GMM, Empirical Likelihood, Bias, Higher-Order Efficiency, Stochastic Expansions.

*This research was partially completed while Newey was Benjamin Meaker Visiting Professor at the University of Bristol and while he was a fellow at the Center for Advanced Study in the Behavioral Sciences. The NSF provided partial financial support for Newey. We thank J. Horowitz, G. Imbens, L. Hansen, T. Rothenberg, three referees, and seminar participants at Bristol.and Iowa for helpful comments.

1 Introduction

In an effort to improve the small sample properties of GMM, a number of alternative estimators have been suggested. These include the empirical likelihood (EL) estimator of Owen (1988), Qin and Lawless (1994), and Imbens (1997), the continuous updating estimator (CUE) of Hansen, Heaton, and Yaron (1996), and the exponential tilting (ET) estimator of Kitamura and Stutzer (1997) and Imbens, Spady and Johnson (1998). As shown by Smith (1997), EL and ET share a common structure, being members of a class of generalized empirical likelihood (GEL) estimators. We show that the CUE is also a member of this class. All of these estimators and GMM have the same asymptotic distribution but different higher-order asymptotic properites. We use the GEL structure, which helps simplify calculations and comparsions, to analyze higher order properties like those of Nagar (1959). We derive and compare the (higher order) asymptotic bias for all of these estimators. We also derive bias-corrected GMM and GEL estimators and consider their higher-order efficiency.

We find that EL has two theoretical advantages. First, its asymptotic bias does not grow with the number of moment restrictions, while the bias of the others often does. Consequently, for large numbers of moment conditions the bias of EL will be less than the bias of the other estimators. This property is important in econometrics, where many moment conditions are often used. For example, Hansen and Singleton (1982), Holtz-Eakin, Newey, and Rosen (1988), and Abowd and Card (1989), all use quite large numbers of moment conditions in their empirical work. The relatively low asymptotic bias of EL indicates that it is an important alternative to GMM in such applications. Furthermore, we show that under a symmetry condition, which may be satisfied in some instrumental variable settings, all the GEL estimators inherit the small bias property of EL. This result may help explain Monte Carlo findings of Hansen, Heaton, and Yaron (1996) and Imbens, Spady, and Johnson (1998).

We provide intuition for the bias results by interpreting EL as a GMM estimator where the linear combination coefficients are efficiently estimated. Because of their efficiency these coefficients are asymptotically uncorrelated with the moment conditions, removing the primary source of asymptotic bias.

The second theoretical advantage of EL is that after it is bias corrected, using probabilities obtained from EL, it is higher efficient relative to the other estimators. This property has a simple explanation. When the data are discrete, having finite support, EL is equal to the maximum likelihood estimator (MLE). Furthermore, the bias correction based on EL probabilities is identical to the discrete data bias correction for the MLE. Consequently, for discrete data EL inherits the well known higher order efficiency of MLE (e.g. see Rao, 1963 and Pfanzagl and Wefelmeyer, 1978). Then, because discrete distributions can be used to approximate moments of a continuous distribution the efficiency of EL for the discrete case leads to efficiency in general. This explanation is a higher-order version of Chamberlain's (1987) result on first-order efficiency of GMM.

Although the small bias property of EL is nice, there are methods of removing all of the asymptotic bias from GMM estimators. These include the bootstrap, as in Horowitz (1998) for GMM, the jacknife, as in Kezdi, Hahn, and Solon (2001) for minimum distance, and analytical methods, as in Hahn, Hausman, and Kuersteiner (2001) for dynamic panel data. Here we give general analytical bias corrected versions of GMM and GEL. The higher-order efficiency of bias corrected EL gives it a theoretical advantage over all the other bias corrected estimators.

There are some important econometric models where GMM has no asymptotic bias. One of these is heteroskedasticity improved estimation as considered in Amemiya (1983), Chamberlain (1982), and Cragg (1982), with symmetric disturbances. Here it is interesting to compare the estimators without any bias corrections. We carry out a higher-order variance comparision for GMM and EL and find that with large numbers of moments the estimator that is best depends on conditional kurtosis. In the Gaussian case they have the same higher-order variance, EL is better with thin-tailed errors, and GMM with thick tailed errors.

Some previous work on higher order properties of these estimators has been done. Koenker et. al. (1992) and Rilstone, Srivastavaa, and Ullah (1996) give some higher-order variance and bias calculations for special cases of GMM. Corcoran (1998) showed that in a class of minimum discrepancy estimators, EL has the only objective function that is Bartlett correctable. Rothenberg (1999) showed that for a single equation of a homoskedastic linear simultaneous equations model the asymptotic bias of EL is the same as the limited information maximum likelihood estimator. Imbens (2000) showed that a canonical example of GMM has mean-square error that grows at the same rate as the square of the number of moment restrictions. We obtain bias formulae and corrections for fully general GMM and GEL estimators and show EL has relatively small bias and is higher-order efficient after bias correction.

The outline of the paper is as follows. In Section 2 the model and estimators are described, and new interpretations of some of the estimators are given. Section 3 gives the asymptotic expansions on which the results are based, including a new consistency result for GEL. Section 4 presents the results on asymptotic bias. Bias corrected versions of GMM and GEL are given in Section 5. Section 6 presents the results on higher-order efficiency. Section 7 concludes. Proofs are given in the Appendix.

2 The Model and Estimators

The model we consider is one with a finite number of moment restrictions. To describe it, let z_i, $(i = 1, ..., n)$, be i.i.d. observations on a data vector z. Also, let β be a $p \times 1$ parameter vector and $g(z, \beta)$ be an $m \times 1$ vector of functions of the data observation z and the parameter, where $m \geq p$. The model has a true parameter β_0 satisfying the moment condition

$$E[g(z, \beta_0)] = 0,$$

where $E[.]$ denotes expectation taken with respect to the distribution of z_i.

An important estimator of β is the two-step GMM estimator of Hansen (1982). To describe it, let $g_i(\beta) \equiv g(z_i, \beta)$, $\hat{g}(\beta) \equiv n^{-1} \sum_{i=1}^{n} g_i(\beta)$, and $\hat{\Omega}(\beta) \equiv n^{-1} \sum_{i=1}^{n} g_i(\beta) g_i(\beta)'$. Also, let $\tilde{\beta}$ be some preliminary estimator, given by $\tilde{\beta} = \arg\min_{\beta \in \mathcal{B}} \hat{g}(\beta)' \hat{W}^{-1} \hat{g}(\beta)$ where $\mathcal{B}$ denotes the parameter space, and $\hat{W}$ is a random matrix with properties to be specified below. The GMM estimator we consider is

$$\hat{\beta}_{GMM} = \arg\min_{\beta\in\mathcal{B}} \hat{g}(\beta)'\hat{\Omega}(\tilde{\beta})^{-1}\hat{g}(\beta), \tag{2.1}$$

where $\mathcal{B}$ denotes the parameter space. We will compare the properties of this estimator to a class of alternative estimators.

The alternatives to GMM we consider are generalized empirical likelihood (GEL) estimators, as in Smith (1997). To describe GEL let $\rho(v)$ be a function of a scalar v that is concave on its domain, an open interval $\mathcal{V}$ containing zero. Let $\hat{\Lambda}(\beta) = \{\lambda : \lambda' g_i(\beta) \in \mathcal{V}, i = 1, ..., n\}$. The estimator is the solution to a saddle point problem

$$\hat{\beta}_{GEL} = \arg\min_{\beta\in\mathcal{B}} \sup_{\lambda\in\hat{\Lambda}(\beta)} \sum_{i=1}^{n} \rho(\lambda' g_i(\beta)). \tag{2.2}$$

The empirical likelihood (EL) estimator is a special case with $\rho(v) = \ln(1 - v)$ and $\mathcal{V} = (-\infty, 1)$, as shown by Qin and Lawless (1994) and Smith (1997). The exponential tilting estimator is a special case with $\rho(v) = -e^v$, as shown by Kitamura and Stutzer (1997) and Smith (1997).

It will be convenient to impose a normalization on $\rho(v)$. Let $\rho_j(v) = \partial^j\rho(v)/\partial v^j$ and $\rho_j = \rho_j(0)$, $(j = 0, 1, 2, ...)$. We normalize so that $\rho_1 = \rho_2 = -1$. As long as $\rho_1 \neq 0$ and $\rho_2 < 0$, which we will assume to be true, this normalization can always be imposed by replacing $\rho(v)$ by $[-\rho_2/\rho_1^2]\rho([\rho_1/\rho_2]v)$, which replacement does not affect the estimator of β. It is satisfied by the $\rho(v)$ we have given for EL and ET.

We can show that that the continuous updating estimator (CUE) of Hansen, Heaton, and Yaron (1996) is also a GEL estimator. The CUE is analogous to GMM except that the objective function is simultaneously minimized over β in $\hat{\Omega}(\beta)^{-1}$. It is given by

$$\hat{\beta}_{CUE} = \arg\min_{\beta\varepsilon\mathcal{B}} \hat{g}(\beta)'\hat{\Omega}(\beta)^{-}\hat{g}(\beta), \tag{2.3}$$

where A^- denotes any generalized inverse of a matrix A, satisfying $AA^-A = A$. The followng result shows that this estimator is a GEL estimator for quadratic $\rho(v)$.

Theorem 2.1: *If $\rho(v)$ is quadratic then $\hat{\beta}_{GEL} = \hat{\beta}_{CUE}$.*

Associated with each GEL estimator are empirical probabilities for the observations. Because these probabilities are important for our analysis we give a brief description. For a given function $\rho(v)$, an associated GEL estimator $\hat{\beta}$, and $\hat{g}_i = g_i(\hat{\beta})$ they are

$$\hat{\pi}_i = \rho_1(\hat{\lambda}'\hat{g}_i)/\sum_{j=1}^{n}\rho_1(\hat{\lambda}'\hat{g}_j), (i = 1, ..., n). \tag{2.4}$$

These $\hat{\pi}_i$ sum to one by construction, satisfy the sample moment condition $\sum_{i=1}^{n}\hat{\pi}_i\hat{g}_i = 0$ when the first-order condtions for $\hat{\lambda}$ hold, and are positive when $\hat{\lambda}'\hat{g}_i$ is small uniformly in i. For EL they were given by Owen (1988), for ET by Kitamura and Stutzer (1997), for quadratic $\rho(v)$ by Back and Brown (1993), and for the general case by Brown and Newey (1992). For any function $a(z, \beta)$ and GEL estimator $\hat{\beta}$ these can be used to form an efficient estimator $\sum_{i=1}^{n}\hat{\pi}_i a(z_i, \hat{\beta})$ of $E[a(z, \beta_0)]$, as in Brown and Newey (1998).

2.1 Duality for GEL

Comparing GEL with another type of estimator provides useful insights. Let $h(\pi)$ be a convex function of a scalar π, and consider the estimator

$$\bar{\beta} = \arg\min_{\beta\in\mathcal{B}}\sum_{i=1}^{n}h(\pi_i), s.t. \sum_{i=1}^{n}\pi_i g_i(\beta) = 0, \sum_{i=1}^{n}\pi_i = 1. \tag{2.5}$$

This general class of minimum discrepancy (MD) estimators was formulated by Corcoran (1998). Like GEL, this class also includes as special cases EL and ET, where $h(\pi)$ is $-\ln(\pi)$ and $\pi\ln(\pi)$ respectively.

For each MD estimator there is a dual GEL estimator when $h(\pi)$ is a member of the Cressie and Read (1984) family of discrepancies where $h(\pi) = [\gamma(\gamma+1)]^{-1}[(n\pi)^{\gamma+1} - 1]/n$. To describe this result, note that the Lagrangean for MD is

$$L = \frac{1}{\gamma(\gamma+1)}\sum_{i=1}^{n}[(\frac{1/n}{\pi_i})^{-(\gamma+1)} - 1]/n - \alpha'\sum_{i=1}^{n}\pi_i g_i(\beta) + \mu(1 - \sum_{i=1}^{n}\pi_i).$$

where α is an m-vector of Lagrange multipliers associated with the first constraint and μ a scalar multiplier for the second constraint. Let $\bar{\pi}_i$, $\bar{\alpha}$, and $\bar{\mu}$ denote the solutions to the MD optimization problem, along with $\bar{\beta}$. We interpret expressions as limits for $\gamma = 0$ or $\gamma = -1$.

Theorem 2.2: *If $g(z,\beta)$ is continuously differentiable in β, for some scalar γ*

$$\rho(v) = -(1+\gamma v)^{(\gamma+1)/\gamma}/(\gamma+1), \tag{2.6}$$

the solutions to equation (2.5) and (2.2) occur in the interior of B, $\hat{\lambda}$ exists, and $\sum_{i=1}^{n} \rho_2(\hat{\lambda}'\hat{g}_i)\hat{g}_i\hat{g}_i'$ is nonsingular, then the first-order conditions for GEL and MD coincide for $\hat{\beta} = \bar{\beta}$, $\hat{\pi}_i = \bar{\pi}_i, (i = 1, ..., n)$, and $\hat{\lambda} = \bar{\alpha}/(\gamma\bar{\mu})$ for $\gamma \neq 0$ and $\hat{\lambda} = \hat{\alpha}$ for $\gamma = 0$.

The following Table summarizes the relationships between MD and GEL for EL, ET, and CUE, with $h(\pi)$ corresponding to a linear transformation of that given above. Let $\hat{\beta}$ denote the estimator corresponding to each row of the table and $\hat{g}_i = g_i(\hat{\beta})$.

Table 1

Estimator	γ	$\rho(v)$	$h(\pi)$	$\hat{\pi}_i$	$\hat{\lambda}$
EL	-1	$\ln(1-v)$	$-\ln(\pi)$	$(1-\hat{\lambda}'\hat{g}_i)^{-1}/\sum_{j=1}^{n}(1-\hat{\lambda}'\hat{g}_j)^{-1}$	$-\bar{\alpha}/\bar{\mu}$
ET	0	$-e^v$	$\pi\ln(\pi)$	$\exp(\hat{\lambda}'\hat{g}_i)/\sum_{j=1}^{n}\exp(\hat{\lambda}'\hat{g}_j)$	$\bar{\alpha}$
CUE	1	$-(1+v)^2/2$	π^2	$(1+\hat{\lambda}'\hat{g}_i)/\sum_{j=1}^{n}(1+\hat{\lambda}'\hat{g}_j)$	$\bar{\alpha}/\bar{\mu}$

The duality between MD and GEL estimators is known for EL (Qin and Lawless, 1994) and for ET (Kitamura and Stutzer, 1997), but is new for the CUE as well as for all the other members of the Cressie and Read (1984) family. Duality is useful because it shows that the computationally complicated MD maximization can be replaced by a simpler GEL one. Also, duality justifies the $\hat{\pi}_i$ in equation (2.4) as MD estimates. We conjecture that for $h(\pi)$ outside the Cressie and Read (1984) family, or $\rho(v)$ outside the family of Theorem 2.2, an explicit dual relationship between MD and GEL does not exist, due to non-homogeneity of $h(\pi)$ and $\rho(v)$.

2.2 The First-Order Conditions

Some interpretations of the first-order conditions are useful for understanding our asymptotic bias results. The GMM first order conditions imply

$$[\sum_{i=1}^{n} G_i(\hat{\beta}_{GMM})/n]'\hat{\Omega}(\tilde{\beta})^{-1}\hat{g}(\hat{\beta}_{GMM}) = 0. \tag{2.7}$$

Also, for $\hat{\pi}_i^{CUE}$ equal the CUE $\hat{\pi}_i$ given in Table 1, as shown in Donald and Newey (2000), the first order conditions for CUE imply

$$[\sum_{i=1}^{n} \hat{\pi}_i^{CUE} G_i(\hat{\beta}_{CUE})]' \hat{\Omega}(\hat{\beta}_{CUE})^{-1} \hat{g}(\hat{\beta}_{CUE}) = 0.$$

Indeed, we show in the proof of Theorem 2.2 in the appendix that each GEL estimator involves an analogous estimator of the Jacobian term, using corresponding probabilities $\hat{\pi}_i$.

The EL estimator first-order conditions have a special interpretation:

Theorem 2.3: *For $\hat{\pi}_i^{EL} = (1-\hat{\lambda}'\hat{g}_i)^{-1} / \sum_{j=1}^{n}(1-\hat{\lambda}'\hat{g}_j)^{-1}$ the EL first-order conditions imply*

$$[\sum_{i=1}^{n} \hat{\pi}_i^{EL} G_i(\hat{\beta}_{EL})]' [\sum_{i=1}^{n} \hat{\pi}_i^{EL} g_i(\hat{\beta}_{EL}) g_i(\hat{\beta}_{EL})']^{-1} \hat{g}(\hat{\beta}_{EL}) = 0.$$

In comparing the GMM, CUE, and EL first order conditions, we see that each can be viewed as setting a linear combination of $\hat{g}(\beta)$ equal to zero, but the linear combination coefficients are estimated in different ways. GMM uses sample averages, CUE (and other GEL estimators) use an efficient esimator of the Jacobian term, and EL uses an efficient estimator of both the Jacobian and second moment terms. An important property of efficient moment estimators is that they are asymptotically uncorrelated with the $\hat{g}(\beta_0)$, eliminating correlations between corresponding terms in the first-order conditions which are an important source of nonzero expectation for the first-order conditions, and hence of bias. Consequently, as we will show, for the CUE there will be no asymptotic bias from estimation of the Jacobian and for EL there will be no asymptotic bias from estimating either the Jacobian or the second moments.

3 Stochastic Expansion

We find the asymptotic bias and higher-order variance using a stochastic expansions for each estimator. Let F denote the distribution of z, $\psi(z, F)$ a function of z and F with $E[\psi(z, F_0)] = 0$, and $\tilde{\psi} = \sum_{i=1}^{n} \psi(z_i, F_0)/\sqrt{n}$. Also define $a(z, F)$, $\tilde{a}$, $b(z, F)$, and $\tilde{b}$

analogously. For each estimator we derive an expansion

$$\sqrt{n}(\hat{\beta}-\beta_0)=\tilde{\psi}+Q_1(\tilde{\psi},\tilde{a},F_0)/\sqrt{n}+Q_2(\tilde{\psi},\tilde{a},\tilde{b},F_0)/n+R_n, \tag{3.1}$$

where Q_1 is quadratic in its first two arguments, Q_2 is cubic in its first three arguments, and $R_n = O_p(n^{-3/2})$. As discussed in Rothenberg (1984), valid higher-order bias and variance calculations can be based on the expectation and variance of the sum of the first three terms in this expansion. Under certain regularity conditions, including continuous distributions, this bias and variance will coincide with that of an Edgeworth approximation to the distribution. Furthermore, even when the data are discrete, so that an Edgeworth approximation is not valid, these calculations can be used for higher-order efficiency comparisons, as in Pfanzagl and Wefelmeyer (1978). We also note that in the Appendix we give a corresponding expansion for $\hat{\lambda}$, which may be of interest for the analysis of overidentifying tests these tests, as in Imbens, Spady, and Johnson (1998).

Consistency and asymptotic normality are important prerequisites for stochastic expansions, so we first briefly consider these properties. We make use of the following identification and regularity conditon. Let $\Omega = E[g_i(\beta_0)g_i(\beta_0)']$.

Assumption 1: **(a)** $\beta_0 \in \mathcal{B}$ is the unique solution to $E[g(z,\beta)] = 0$; **(b)** $\mathcal{B}$ is compact; **(c)** $g(z,\beta)$ is continuous at each $\beta \in \mathcal{B}$ with probability one; **(d)** $E\left[\sup_{\beta\in\mathcal{B}} \|g(z,\beta)\|^{\alpha}\right] < \infty$ for some $\alpha > 2$; **(e)** Ω is nonsingular; **(f)** $\rho(v)$ is twice continuously differentiable in a neighborhood of zero.

This assumption requires existence of a slightly higher moments than consistency for two-step efficient GMM, in Hansen (1982), but otherwise is the same.

Theorem 3.1: *If Assumption 1 is satisfied then* $\hat{\beta} \xrightarrow{p} \beta_0$, $\hat{g}(\hat{\beta}) = O_p(n^{-1/2})$, $\hat{\lambda} = \arg\max_{\lambda\in\hat{\Lambda}(\hat{\beta})} \sum_{i=1}^{n} \rho(\lambda' g_i(\hat{\beta}))/n$ *exists with probability approaching one, and* $\hat{\lambda} = O_p(n^{-1/2})$.

This result is new in making no auxiliary assumption about $\hat{\beta}$ or $\hat{\lambda}$. Also, the proof is based directly on the global concavity of $\rho(v)$ and saddle point form of GEL. Additional conditions are needed for asymptotic normality. Let $G = E[\partial g_i(\beta_0)/\partial\beta]$.

Assumption 2: **(a)** $\beta_0 \in int(\mathcal{B})$; **(b)** $g(z,\beta)$ is continuously differentiable in a neighborhood $\mathcal{N}$ of β_0 and $E[\sup_{\beta\in\mathcal{N}} \|\partial g_i(\beta)/\partial\beta'\|] < \infty$; **(c)** $rank(G) = p$.

Let $\Sigma = (G'\Omega^{-1}G)^{-1}$ and $P = \Omega^{-1} - \Omega^{-1}G\Sigma G'\Omega^{-1}$.

Theorem 3.2: *If Assumptions 1 and 2 are satisfied then*

$$\sqrt{n}\begin{pmatrix} \hat{\beta}-\beta_0 \\ \hat{\lambda} \end{pmatrix} \xrightarrow{d} N(0, diag(\Sigma, P)), 2n[\sum_{i=1}^{n}\rho(\hat{\lambda}'g_i(\hat{\beta}))/n - \rho(0)] \xrightarrow{d} \chi^2(m-p).$$

This result shows asymptotic normality of the GEL estimators, and that, properly normalized, the saddle-point objective function has a limiting chi-squared distribution. This is an overidentification test statistic that was formulated by Smith (1997). It is included here because we thought that this test statistic might have independent interest.

Additional smoothness and moment conditions are needed for the stochastic expansion. Let ∇^j denote a vector of all distinct partial derivatives with respect to β of order j.

Assumption 3: There is $b(z)$ with $E[b(z_i)^6] < \infty$ such that for $0 \le j \le 4$, $\nabla^j g(z,\beta)$ exists on a neighborhood $\mathcal{N}$ of β_0, $\sup_{\beta\in\mathcal{N}} \|\nabla^j g(z,\beta)\| \le b(z)$, and for each $\beta \in \mathcal{N}$, $\|\nabla^4 g(z,\beta) - \nabla^4 g(z,\beta_0)\| \le b(z)\|\beta-\beta_0\|$.

Also, for the GMM estimator we need to specify conditions concerning the initial weighting matrix $\hat{W}$.

Assumption 4: There is W and $\psi^w(z_i)$ such that $\hat{W} = W + \sum_{i=1}^{n}\psi^w(z_i)/n + O_p(n^{-1})$, W is positive definite $E[\psi^w(z_i)] = 0$, and $E[\|\psi^w(z_i)\|^6] < \infty$.

We derive the stochastic expansion for GMM using an auxiliary parameter $\hat{\lambda}_{GMM}$ that is analogous to that for GEL. Specifically, we consider GMM first order conditions of the form

$$-[\sum_{i=1}^{n} G_i(\hat{\beta}_{GMM})/n]'\hat{\lambda}_{GMM} = 0, -\hat{g}(\hat{\beta}_{GMM}) - \hat{\Omega}(\tilde{\beta})\hat{\lambda}_{GMM} = 0.$$

This formulation simplifies calculations, because it removes the inverse matrix from the first-order conditions. A different way to do this was proposed by Rilstone et. al. (1996). The next result shows that GMM has a stochastic expansion.

Theorem 3.3: *If Assumptions 1 - 4 are satisfied then equation (3.1) is satisfied for the GMM estimator.*

Expressions for each of the terms in the expansion are given in the proof of this result because they are quite complicated. Implicit in this result is that the expansion for GMM depends only on the limit W and influence function $\psi^w(z_i)$. For example, this means that iterating the GMM estimator two or more times results in the same expansion, where $\hat{W} = \hat{\Omega}(\bar{\beta})$ and $\bar{\beta}$ is itself an efficient GMM estimator.

The final result of this Section is the stochastic expansion for GEL.

Theorem 3.4: *If Assumptions 1 - 3 are satisfied then for the GEL estimator equation (3.1) is satisfied.*

4 Asymptotic Bias

The asymptotic (higher-order) bias formula is given by

$$Bias(\hat{\beta}) = E[Q_1(\psi_i, a_i, F_0)]/n, \tag{4.1}$$

with other terms in the expansion being $O_p(n^{-2})$. To describe the precise form of the bias we need some additional notation. Let $H = \Sigma G'\Omega^{-1}$, $H_W = (G'W^{-1}G)^{-1}G'W^{-1}$, $\Omega_{\beta_j} = E[\partial\{g_i(\beta_0)g_i(\beta_0)'\}/\partial\beta]$, and a be an m-vector such that

$$a_j \equiv tr(\Sigma E[\partial^2 g_{ij}(\beta_0)/\partial\beta\partial\beta'])/2, (j = 1, ..., m), \tag{4.2}$$

where $g_{ij}(\beta)$ denotes the jth element of $g_i(\beta)$. For GMM we have the following result:

Theorem 4.1: *If Assumptions 1 - 4 are satisfied then*

$$\begin{aligned} Bias(\hat{\beta}_{GMM}) &= B_I + B_G + B_\Omega + B_W, B_I = H(-a + E[G_i H g_i])/n, B_G = -\Sigma E[G_i' P g_i]/n, \\ B_\Omega &= H E[g_i g_i' P g_i]/n, B_W = H \sum_{j=1}^{p} \Omega_{\beta_j}(H_W - H)' e_j. \end{aligned} \tag{4.3}$$

Each of the terms has an interesting intrepretation. The first term B_I is precisely the asymptotic bias for a GMM estimator with the optimal (asymptotic variance minimizing, Hansen, 1982) linear combination $G'\Omega^{-1}g(z,\beta)$. The term B_G arises from estimation of G. If G_i is constant $B_G = 0$, but B_G is generally nonzero when there is endogeneity. Similarly the term B_Ω arises from estimation of the second moment matrix Ω. It is zero if third moments are zero, but is generally nonzero. Both B_G and B_Ω will be zero with exact identification, where $m = p$, because P is zero in that case. The term B_W arises from the choice of first step estimator. It is zero if W is a scalar muptiple of Ω. This result is consistent with the Monte Carlo example of Hansen, Heaton, and Yaron (1996), where multiple iterations on $\tilde{\beta}$ had little effect on bias.

We now turn to the bias formula for GEL. Let $\rho_3 = \partial^3\rho(0)/\partial v^3$.

Theorem 4.2: *If Assumptions 1 - 3 are satisfied then*

$$Bias(\hat{\beta}_{GEL}) = B_I + (1 + \frac{\rho_3}{2})B_\Omega$$

In comparison with the GMM bias, we find that B_G and B_W drop out, i.e. there is no asymptotic bias from estimation of the Jacobian or from the first step estimator. The absence of any bias from the first step is to be expected from the one step nature of the estimator. Also, as noted in Section 2, the absence of bias from Jacobian estimation can be explained by the presence of an efficient estimator of the Jacobian in the first-order conditions. In addition, as noted in Section 2, EL uses an efficient second moment estimator, leading to the following result.

Corollary 4.3: *If Assumptions 1 - 3 are satisfied then*

$$Bias(\hat{\beta}_{EL}) = B_I. \tag{4.4}$$

Thus, for EL the bias is exactly the same as for an estimator with moment functions $G'\Omega^{-1}g(z,\beta)$. This same property would be shared by any GEL estimator with $\rho_3 = -2$. It will also be shared by any GEL estimator when third moments are zero.

Corollary 4.4: *If Assumptions 1 - 3 are satisfied and* $E[g_i g_i' g_{ij}] = 0, (j = 1, ..., m)$, *then*

$$Bias(\hat{\beta}_{GEL}) = Bias(\hat{\beta}_{EL}) = B_I. \tag{4.5}$$

This third moment condition will hold in an IV setting, when disturbances are symmetrically distributed. When it does hold one can actually show something slightly stronger, that $\hat{\beta}_{GEL} - \hat{\beta}_{EL} = O_p(n^{-3/2})$.

It is well known that, in overidentified linear models, estimation of the optimal linear combination coefficients is an important source of bias in IV estimators. Because the GMM bias includes such effects but EL does not, we expect that EL will have relatively small bias in IV settings. Also, from Altonji and Segal (1996) we know that, in covariance parameter models, estimation of Ω can be an important source of bias in GMM. Given that the EL bias does not include this effect we expect that it will also have relatively small bias for minimum distance. We can verify this intution in some specific models.

4.1 Homoskedastic Linear Instrumental Variables

The first model we consider is a homoskedastic linear model. Let y denote a scalar endogenous variable, w a $p \times 1$ vector of variables that may be endogenous, and x an $m \times 1$ vector of instrumental variables. The model is

$$y = w'\beta_0 + \varepsilon, E[\varepsilon|x] = 0, var(\varepsilon|x) = \sigma_\varepsilon^2, E[\varepsilon w|x] = \sigma_{\varepsilon w}, E[\varepsilon^3|x] = \mu_3. \tag{4.6}$$

Consider moment conditions where $g(z, \beta) = x(y - w'\beta)$. Here the GMM and GEL estimators have the same asymptotic distribution as two-stage least squares. The following result gives the asymptotic bias formulae. Let $\bar{\delta} = \Sigma\sigma_{\varepsilon w}/\sigma_\varepsilon^2$ where Σ is the asymptotic variance:

Theorem 4.5: *If Assumptions 1 - 4 and equation (4.6) are satisfied then*

$$\begin{aligned} Bias(\hat{\beta}_{EL}) &= -\bar{\delta}/n. \\ Bias(\hat{\beta}_{GEL}) &= -\bar{\delta}/n + (1 + \frac{\rho_3}{2})B_\Omega, B_\Omega = \mu_3 \Sigma G' \Omega^{-1} E[x_i x_i' P x_i]/n, \\ Bias(\hat{\beta}_{GMM}) &= (m - p - 1)\bar{\delta}/n + B_\Omega. \end{aligned}$$

When $\mu_3 = 0$ the asymptotic bias of all the GEL estimators is the same, and is equal to that of limited information maximum likelihood. For EL this result was shown by Rothenberg (1999). Also, the asymptotic bias of GMM is the same as two stage least squares, as given in Nagar (1959), and so grows linearly with the number of overidentifying restrictions. Thus, for 2 or more overidentifying restrictions, the bias of GMM exceeds that of EL in magnitude. For $\mu_3 \neq 0$, except for EL there is an additional bias term from estimation of Ω.

4.2 Conditional Moment Restrictions

The second example is a well known generalization of the linear model that allows for nonlinearity and/or heteroskedasticty. Let $u(z,\beta)$ be a scalar residual satisfying the conditional moment restriction

$$E[u(z_i,\beta_0)|x_i] = 0. \tag{4.7}$$

Consider moment conditions where $g(z,\beta) = q(x)u(z,\beta)$ and $q(x)$ is a vector of $m \times 1$ instrumental variables. To derive the bias, let $u_i = u(z_i,\beta_0)$, $u_{\beta i} = \partial u(z_i,\beta_0)/\partial\beta$, $\sigma_i^2 = E[u_i^2|x_i]$, and $q_i = q(x_i)$. Also, for $\sigma_i^2 > 0$, let $d_i = E[u_{\beta i}|x_i]/\sigma_i^2$, $\kappa_i = -E[u_{\beta i}u_i|x_i]$,

$$\bar{d}_i = G'\Omega^{-1}q_i, H_i = E[\partial^2 u(z_i,\beta_0)/\partial\beta\partial\beta'|x_i], \mu_{3i} = E[u_i^3|x_i], \delta_i = \Sigma(\kappa_i + \bar{d}_i\mu_{3i})/\sigma_i^2.$$

Theorem 4.6: *If Assumptions 1 - 4 and equation* (4.7) *are satisfied,*

$$\begin{aligned}
Bias(\hat{\beta}_{EL}) &= \Sigma(-E[\bar{d}_i tr(\Sigma H_i)]/2 + E[\bar{d}_i\bar{d}_i'\Sigma\kappa_i])/n, \\
Bias(\hat{\beta}_{GEL}) &= Bias(\hat{\beta}_{EL}) + (1 + \frac{\rho_3}{2})B_\Omega, B_\Omega = \Sigma E[\bar{d}_i\mu_{3i}q_i'Pq_i]/n, \\
Bias(\hat{\beta}_{GMM}) &= Bias(\hat{\beta}_{EL}) + \Sigma E[\kappa_i q_i'Pq_i]/n + B_\Omega,
\end{aligned}$$

Also, if $E[\|H_i\|^2/\sigma_i^2] < \infty$, $E[\sigma_i^2\|d_i\|^2] < \infty$, and κ_i/σ_i^2 is bounded, there are constants C_1 and C_2 such that for all $q(x)$

$$\|Bias(\hat{\beta}_{EL})\| \leq C_1\|\Sigma\|^2/n, e_j'Bias(\hat{\beta}_{GMM}) - e_j'Bias(\hat{\beta}_{EL}) \geq C_2(m-p)\inf\{e_j'\delta_i\}/n.$$

Here $\inf\{e_j'\delta_i\} = \sup\{C : \Pr(e_j'\delta_i \geq C) = 1\}$. Similarly to the previous model, we find that the asymptotic bias of GMM grows linearly with the number of overidentifying restrictions when $\inf\{e_j'\delta_i\} > 0$, while the bias of EL is bounded. In this case the bias of GMM will exceed the bias of EL in magnitude when the number of overidentifying restrictions is large enough. We can also show this result when $\sup\{e_j'\delta_i\} < 0$. This condition should be interpreted as a sign restriction on a conditional version of the IV bias. For example, note that if $\mu_{3i} = 0$ then $\delta_i = -\Sigma E[u_{\beta i}u_i|x_i]/\sigma_i^2$, which is analogous to the term $\bar{\delta}$ that enters the bias for IV.

4.3 Minimum Distance Estimation

The third model is one that leads to optimal minimum distance estimation. Consider moment conditions where $g(z,\beta) = r(z) - h(\beta)$, for $r(z)$ a vector of functions of the data and $h(\beta)$ a vector of functions of the unknown parameters. Here $G = -\partial h(\beta_0)/\partial\beta$, $\Omega = Var(r(z_i))$, and $a_j = -tr(\Sigma\partial^2 h_j(\beta_0)/\partial\beta\partial\beta')/2$. We can derive a bound on the bias of β that only depends on Σ, analogous to that for the second example, when $h(\beta)$ can be thought of as the expectation with respect to the pdf for some model. The following assumption imposes this condition along with some smoothness.

Assumption 5: *There is a family of densities $f(z|\beta)$ such that for any $r(z)$, $h(\beta) = \int r(z)f(z|\beta)dz$. Also, $f(z|\beta)$ is twice continuously differentiable in a neighborhood N of β_0, $\int(1+\|r(z)\|)\sup_{\beta\in\mathcal{N}}\|\partial f(z|\beta)/\partial\beta\|dz < \infty$, $\int(1+\|r(z)\|)\sup_{\beta\in\mathcal{N}}\|\partial^2 f(z|\beta)/\partial\beta\partial\beta'\|dz < \infty$, and for $s_i = \partial\ln f(z_i|\beta_0)/\partial\beta$ and $F_i = \partial^2\ln f(z_i|\beta_0)/\partial\beta\partial\beta' + s_i s_i'$, we have $E[\|s_i\|^2] < \infty$, and $E[\|F_i\|^2] < \infty$.*

Theorem 4.7: *If Assumptions 1 - 4 are satisfied and $g(z,\beta) = r(z) - h(\beta)$ then*

$$\begin{aligned}
Bias(\hat{\beta}_{EL}) &= -\Sigma G'\Omega^{-1}a/n, \\
Bias(\hat{\beta}_{GEL}) &= Bias(\hat{\beta}_{EL}) + (1+\frac{\rho_3}{2})\Sigma G'\Omega^{-1}E[g_i g_i' P g_i]/n, \\
Bias(\hat{\beta}_{GMM}) &= Bias(\hat{\beta}_{CUE}) = Bias(\hat{\beta}_{EL}) + \Sigma G'\Omega^{-1}E[g_i g_i' P g_i]/n.
\end{aligned}$$

Also, if $h(\beta)$ is linear in β then $Bias(\hat{\beta}_{EL}) = 0$. Furthermore, if Assumption 5 is also satisfied then

$$\|Bias(\hat{\beta}_{EL})\| \leq p\|\Sigma\|^2\sqrt{E[\|s_i\|^2]E[\|F_i\|^2]}/2n.$$

Here the bias for GMM is identical to that for CUE, which occurs because there is no asymptotic bias from estimation of the Jacobian. Also, we find that the asymptotic bias of EL is zero in the special case of a linear $h(\beta)$ function, and that it does not grow with the number of overidentifying restrictions.

For optimal minimum distance it seems difficult to give a general result showing how the bias of GMM grows with the number of moment restrictions, but an example provides some insight. Suppose that β is a scalar, $r(z) = (z_1, ..., z_m)'$, and $h(\beta) = \beta e$, where e is an $m \times 1$ vector of ones. Also, suppose that the components of z are mutually independent and identically distributed. Let $\sigma^2 = Var(z_{ji})$ and $\mu_3 = E[(z_{ji} - \beta_0)^3]$. Then $\Omega = \sigma^2 I_m$ and $G = e$, so that $\Sigma = \sigma^2/m$ and $P = (I_m - ee'/m)/\sigma^2$. It follows that

$$\begin{aligned} Bias(\hat{\beta}_{EL}) &= 0, Bias(\hat{\beta}_{GMM}) = Bias(\hat{\beta}_{CUE}) = \left(\frac{m-1}{m}\right)\left(\frac{\mu_3}{\sigma^4}\right), \\ \frac{Bias(\hat{\beta}_{GMM})}{\sqrt{\Sigma}} &= \sqrt{m}\left(\frac{m-1}{m}\right)\left(\frac{\mu_3}{\sigma^3}\right)/n. \end{aligned}$$

Here the bias of GMM relative to its asymptotic standard error grows with the square root of the number of overidentifying restrictions. Dividing by $\sqrt{\Sigma}$ is an appropriate normalization, since the asymptotic variance is going to zero with $1/m$.

5 Bias Corrected GMM and GEL

Although we have established that EL has smaller asymptotic bias than GMM in several important cases, it is also possible to remove all the asymptotic bias. As mentioned in the introduction, there are several approaches to bias correction, including the bootstrap, jacknife, and analytical methods. Here we use an analytical approach, bias correcting GMM and GEL using the asymptotic bias formulas we have derived. This bias correction is much simpler computationally than the bootstrap or jacknife methods, particularly in nonlinear models.

The basic idea of analytical bias corrections is simple and well known, and consists of estimating the asymptotic bias and subtracting from $\hat{\beta}$. Here we use the general formula of equation (4.1) to construct the bias estimate. For an estimator $\hat{F}$ of the distribution of a single observation, the bias corrected estimator is

$$\hat{\beta}^c = \hat{\beta} - \widehat{Bias}(\hat{\beta}), \widehat{Bias}(\hat{\beta}) = \int Q_1(\psi(z,\hat{F}), a(z,\hat{F}), \hat{F})\hat{F}(dz)/n. \tag{5.1}$$

The distribution estimator $\hat{F}$ can be chosen to be the empirical distribution or a distribution based on the GEL probabilities in equation (2.4). This choice does not affect the asymptotic bias of the variance but will affect its higher-order variance. We use the empirical distribution for the GMM bias correction and for GEL we use the corresponding probabilities. As we discuss in the next Section, the higher order efficiency of bias corrected EL depends on the use of the EL probabilities in the corresponding bias correction.

The describe the specific form of the bias correction for GMM and we need to introduce some notation. Let $\hat{\beta}_{GMM}$ denote the GMM estimator and

$$\begin{aligned}
\hat{g}_i &= g_i(\hat{\beta}_{GMM}), \hat{G}_i = G_i(\hat{\beta}_{GMM}), \hat{G} = \sum_{i=1}^{n} \hat{G}_i/n, \hat{\Omega} = \hat{\Omega}(\hat{\beta}_{GMM}), \\
\hat{\Sigma} &= (\hat{G}'\hat{\Omega}\hat{G})^{-1}, \hat{H} = \hat{\Sigma}\hat{G}'\hat{\Omega}^{-1}, \hat{\psi}_i = -\hat{H}\hat{g}_i, \hat{P} = \hat{\Omega}^{-1} - \hat{\Omega}^{-1}\hat{G}\hat{\Sigma}\hat{G}'\hat{\Omega}^{-1}, \\
\hat{a}_j &\equiv tr(\hat{\Sigma}\sum_{i=1}^{n} \partial^2 g_{ij}(\hat{\beta}_{GMM})/\partial\beta\partial\beta'/n)/2, (j = 1, ..., m),
\end{aligned}$$

Then for the bias formula given in Theorem 4.1, and using the empirical distribution $\hat{F}$ to estimate the expectations in this formula, the estimator of the bias term is

$$\widehat{Bias}(\hat{\beta}_{GMM}) = [\hat{H}(-\hat{a} + \sum_{i=1}^{n} \hat{G}_i\hat{\psi}_i/n) - \hat{\Sigma}\sum_{i=1}^{n} \hat{G}_i'\hat{P}\hat{g}_i/n + \sum_{i=1}^{n} \hat{\psi}_i\hat{g}_i'\hat{P}\hat{g}_i/n]/n$$

The bias corrected GMM estimator is then $\hat{\beta}^c_{GMM} = \hat{\beta}_{GMM} - \widehat{Bias}(\hat{\beta}_{GMM})$.

To form a bias corrected GEL estimator we use analogous formulas, replacing the empirical distribution $\hat{F}$ by one based on the GEL probabilities of equation (2.4). Let $\hat{\beta}_{GEL}$ denote the estimator, $\hat{\pi}_i, (i = 1, ..., n)$ the associated probabilities, and

$$\tilde{g}_i = g_i(\hat{\beta}_{GEL}), \tilde{G}_i = G_i(\hat{\beta}_{GEL}), \tilde{G} = \sum_{i=1}^{n} \hat{\pi}_i\tilde{G}_i/n, \tilde{\Omega} = \sum_{i=1}^{n} \hat{\pi}_i\tilde{g}_i\tilde{g}_i',$$

$$\begin{aligned}
\tilde{\Sigma} &= (\tilde{G}'\hat{\Omega}\tilde{G})^{-1}, \tilde{H} = \tilde{\Sigma}\tilde{G}'\tilde{\Omega}^{-1}, \tilde{\psi}_i = -\tilde{H}\tilde{g}_i, \tilde{P} = \tilde{\Omega}^{-1} - \tilde{\Omega}^{-1}\tilde{G}\tilde{\Sigma}\tilde{G}'\tilde{\Omega}^{-1}, \\
\hat{a}_j &\equiv tr(\tilde{\Sigma}\sum_{i=1}^{n}\hat{\pi}_i\partial^2 g_{ij}(\hat{\beta}_{GEL})/\partial\beta\partial\beta')/2, (j = 1, ..., m),
\end{aligned}$$

Then for the bias formula in Theorem 4.2, the estimator of the GEL asymptotic bias is

$$\widehat{Bias}(\hat{\beta}_{GEL}) = -\tilde{H}(\tilde{a} + \sum_{i=1}^{n}\hat{\pi}_i\hat{G}_i\tilde{\psi}_i) - (1 + \frac{\rho_3}{2})\sum_{i=1}^{n}\hat{\pi}_i\hat{\psi}_i\hat{g}_i'\hat{P}\hat{g}_i.$$

The bias corrected GEL estimator is then $\hat{\beta}^c_{GEL} = \hat{\beta}_{GEL} - \widehat{Bias}(\hat{\beta}_{GEL})$.

We can show under the conditions already given that these bias corrected estimators have expansions with zero asymptotic bias.

Theorem 5.1: *If Assumptions 1 - 4 are satisfied then $\hat{\beta}^c_{GEL}$ and $\hat{\beta}^c_{GMM}$ satisfy equation (3.1) with $Bias(\hat{\beta}^c_{GEL}) = Bias(\hat{\beta}^c_{GMM}) = 0$.*

6 Higher Order Efficiency of Empirical Likelihood

Different asymptotically unbiased estimators, such as the bias-corrected GMM and GEL, can be compared on the basis of their higher-order variance. The higher-order variance is given by

$$\begin{aligned}
Var(\sqrt{n}(\hat{\theta} - \theta_0)) &= \Sigma + \Xi/n, \\
\Xi &= \lim_{n\to\infty}\{Var(\tilde{Q}_1) + E[\tilde{\psi}(\sqrt{n}\tilde{Q}_1 + \tilde{Q}_2)'] + E[(\sqrt{n}\tilde{Q}_1 + \tilde{Q}_2)\tilde{\psi}']\},
\end{aligned}$$

where $\tilde{Q}_1 = Q_1(\tilde{\psi}, \tilde{a}, F_0)$, $\tilde{Q}_2 = Q_2(\tilde{\psi}, \tilde{a}, \tilde{b}, F_0)$, and terms that are $o_p(n^{-1})$ are dropped. Here the term Ξ is the additional, n^{-1} variance term for $\hat{\theta}$. Different bias corrected estimators can be compared on the basis of the corresponding Ξ terms, one being higher order efficient relative to another if its Ξ matrix is no larger, in the positive semidefinite sense.

It is straightforward to show that bias corrected EL is higher-order efficient relative to other bias corrected GMM or GEL estimators. This is a consequence of the fact that bias-corrected EL is equal to the bias corrected maximum likelihood estimator (MLE) for discrete data, as shown by the following result.

Theorem 6.1: *If Assumptions 1-3 are satisfied and z_i has finite support then with probability approaching one, $\hat{\beta}$ exists and is equal to the MLE, and the bias corrected EL estimator is equal to the bias corrected MLE.*

When combined with the higher-order efficiency of MLE, as shown by Pfanzagl and Wefelmeyer (1978) in general, an immediate consequence of Theorem 6.1 is that, when z_i has finite support, EL is higher-order efficient relative to bias corrected GMM or GEL. Then, since none of the properties of the estimators are sensitive to the discreteness of the support, EL will be higher-order efficient in general. More precisely, as shown by Chamberlain (1987), one can find a discrete distribution that replicates any set of expectations for the true distribution, including those that make up Σ and Ξ, so that higher-order efficiency of EL in the discrete case extends immediately to the general case.

Theorem 6.2: *If Assumptions 1-4 are satisfied then EL is higher-order efficient relative to GMM and GEL.*

The ranking of estimators based on Ξ corresponds to a quadratic loss function. As shown by Pfanzagl and Wefelmeyer (1978), the higher-order efficiency of MLE also holds for a wide class of quasi-convex loss functions satisfying certain properties. Consequently, it can also be shown that EL is higher-order efficient for any such loss function.

The use of the EL probabilities in forming the bias correction is apparently critical for EL higher-order efficiency. When the data are discrete these probabilities are asymptotically equal to the MLE estimates of the probabilities of the outcomes, and hence the bias correction depends only on the MLE estimator. If sample averages were used instead in the bias correction then additional terms would be introduced into the higher-order variance, affecting the efficiency of EL.

Comparisons of estimators that have no asymptotic bias under some auxiliary assumptions, in addition to $E[g(z, \beta_0)] = 0$, may also be of interest. There are important examples of GMM estimators that have no asymptotic bias under a few additional conditions, where it would be interesting to know if EL still improves higher order efficiency.

Because the higher-order variances are complicated, it is difficult to do many such comparisons. We do consider one important example, improved estimators in a heteroskedastic linear regression model.

Consider the model

$$y_i = x_i'\beta_0 + \varepsilon_i, E[\varepsilon_i|x_i] = 0.$$

It is well known that least squares may be inefficient due to heteroskedasticity in this model. Amemiya (1983), Chamberlain (1982), and Cragg (1982) proposed estimators with improved asymptotic efficiency. Consider forming moment conditions $g(z, \beta) = q(x)(y - x'\beta)$, of the form considered above for conditional moment restrictions. Then if x is included among the elements of $q(x)$, the GMM estimator based on $g(z, \beta)$ will be asymptotically no less efficient than least squares, because it uses more moments.

Using the above expansions we can compare the higher-order efficiency of the GMM and EL estimators for this model. We do this under the conditional moment condition above and the conditional symmetry condition $E[\varepsilon_i^3|x_i] = 0$, where the estimators have no asymptotic bias. For this result let $\sigma_i^2 = E[\varepsilon_i^2|x_i]$, $\mu_{4i} = E[\varepsilon_i^4|x_i]$, and $\bar{x}_i = -G'\Omega^{-1}q_i\sigma_i^2 = E[\sigma_i^2(x_i/\sigma_i^2)q_i']\{E[\sigma_i^2 q_i q_i']\}^{-1}q_i\sigma_i^2$.

Theorem 6.3: *If Assumptions 1-4 are satisfied, the first step estimator for GMM is optimal, then for the estimators of* β,

$$\begin{aligned}\Xi_{GMM} - \Xi_{GEL} &= D + D', D = \Sigma\{(\rho_3/2)E[(\mu_{4i}/\sigma_i^4 - 3)K_i x_i \bar{x}_i'] \\ &\quad + E[K_i(\bar{x}_i - x_i)'] + (3\rho_3/2)E[K_i\bar{x}_i(\bar{x}_i - x_i)']\}\Sigma\end{aligned}$$

Furthermore, if σ_i^2 *is bounded and bounded away from zero,* μ_{4i} *is bounded,* $E[q_i q_i']$ *is nonsingular for each* m, *and there exists* γ_m *such that for the support* X *of* x, *as* $m \to \infty$,

$$\sup_{x \in X}\{q(x)'(E[q_i q_i'])^{-1}q(x)\}^2 E[\|x_i/\sigma_i^2 - \gamma_m q_i\|^2] \to 0.$$

then as $m \to \infty$,

$$\Xi_{GMM} - \Xi_{GEL} - \rho_3\Sigma E[(\mu_{4i}/\sigma_i^4 - 3)K_i x_i x_i']\Sigma \to 0.$$

This result gives an explicit formula for the difference of the higher-order variance terms as well as a limit result as the number of moments gets large. The hypothesis for the limit result combines an approximation property for $q(x)$ with a bound on its size. For example, it follows form Newey (1997) that this condition will hold with splines if the density is bounded away from zero, the support of X is a rectangle, knots are evenly spaced, and $\sigma(x)$ is twice differentiable.

The limit form of the difference has a nice interpretation. If the disturbances are conditionally normal, so that $\mu_{4i} = 3\sigma_i^4$. Then in the limit the higher-order variances are equal. Also $\rho_3 = 0$ for CUE, so that it has the same higher-order variance as GMM in the limit. For EL and ET, $\rho_3 < 0$ so that they have smaller higher-order variance than GMM in the limit when the disturbances are thinner tailed than normal, in the sense that $\mu_{4i} < 3\sigma_i^4$, and higher when they are thick tailed, in the sense that $\mu_{4i} > 3\sigma_i^4$.

What allows EL to be less higher-order efficient than GMM here in some cases is that the bias corrections have not been applied. The use of the extra information that $E[\varepsilon_i|x_i] = 0$ and $E[\varepsilon_i^3|x_i] = 0$ to avoid bias corrections changes the efficiency properties of EL relative to GMM. If this extra information were ignored and the bias corrections from Section 5 were applied then Theorem 6.2 would lead higher-order efficiency of the bias-corrected EL estimator relative to the bias corrected GMM estimator.

7 Conclusion

The bias results we have derived are consistent with much of the existing Monte Carlo results for GMM and GEL estimators. Hansen, Heaton, and Yaron (1996) showed that the CUE has smaller bias for IV estimators of asset pricing models with several overidentifying restrictions, which is consistent with the asymptotic bias of GMM growing with the moment conditions but that of the CUE not, as is the case under symmetry. Also, Imbens (1997) gave a Monte Carlo example where the bias of EL is less than that of GMM, again consistent with our results for large numbers of moment conditions.

One potentially useful implication of our results is that EL may eliminate the bias for minimum distance estimation in panel data. As documented by Altonji and Segal (1996), the estimation of the optimal weighting matrix can be a large source of bias. We have shown that to first-order at least, EL eliminates this bias. Thus, EL may prove a useful alternative to the bootstrap bias correction of Horowitz (1998) or the analytical bias correction for GMM we have derived.

Little is known about the variability of different bias corrected estimators. It would be interesting to explore this topic in the future. However, given the large biases that have been found for GMM, as in other settings (see Rothenberg, 1983, p. 909), the corrections for bias implicit in EL and the explicit corrections for GMM we have derived may be more important than second-order variance comparisons for choosing among estimators.

Appendix: Proofs

Throughout the Appendix, C will denote a generic positive constant that may be different in different uses, and CS, M, and T the Cauchy-Schwartz, Markov, and triangle inequalities respectively. Also, with probability approaching one will be abbreviated as w.p.a.1, positive semi-definite as p.s.d., UWL will denote a uniform weak law of large numbers such as Lemma 2.4 of Newey and McFadden (1994), and CLT will refer to the Lindbergh-Levy central limit theorem. We let $\hat{P}(\beta,\lambda) = \sum_{i=1}^{n} \rho(\lambda' g_i(\beta))/n$.

Proof of Theorem 2.1: Let $A = [g_1(\beta), ..., g_n(\beta)]'/\sqrt{n}$ and $\iota = (1, ..., 1)'$ be an n-vector of units. Thus, $\hat{g}(\beta) = A'\iota/\sqrt{n}$ and $\hat{\Omega}(\beta) = A'A$. By Rao (1973, 1b.5(vi),(viii)), $A(A'A)^{-}A'$ is invariant to the generalized inverse (ginv) and $A'A(A'A)^{-}A' = A'$ for any ginv.Then the CUE objective function $\iota' A(A'A)^{-}A'\iota/n$ is invariant to ginv. By $\rho(v)$ quadratic, a second-order Taylor expansion is exact, giving

$$\hat{P}(\beta,\lambda) = \rho_0 - \hat{g}(\beta)'\lambda - \frac{1}{2}\lambda'\hat{\Omega}(\beta)\lambda. \qquad \text{(A.1)}$$

By concavity of $\hat{P}(\beta,\lambda)$ in λ, any solution $\hat{\lambda}(\beta)$ to the first-order conditions

$$0 = \hat{g}(\beta) + \hat{\Omega}(\beta)\lambda$$

will maximize $\hat{P}(\beta,\lambda)$ with respect to λ holding β fixed. Then, $\hat{\Omega}(\beta)\hat{\Omega}(\beta)^{-}\hat{g}(\beta) = A'A(A'A)^{-}A'\iota/\sqrt{n} = \hat{g}(\beta)$, so that $\hat{\lambda}(\beta) = -\hat{\Omega}(\beta)^{-}\hat{g}(\beta)$ solves the first-order conditions. Since

$$\hat{P}(\beta,\hat{\lambda}(\beta)) = \rho_0 + \hat{g}(\beta)'\hat{\Omega}(\beta)^{-}\hat{g}(\beta)/2. \tag{A.2}$$

the GEL objective function $\hat{P}(\beta,\hat{\lambda}(\beta))$ is a monotonic increasing transformation of the CUE objective function, so that the set of GEL estimators coincides with the set of CUE estimators. Q.E.D.

Proof of Theorem 2.2: We first consider the case where $\gamma \neq 0$. The first-order conditions for $\bar{\pi}_i$ are $(n\bar{\pi}_i)^{\gamma}/\gamma - \bar{\alpha}'g_i(\bar{\beta}) - \bar{\mu} = 0$. Solving gives $\bar{\pi}_i = [\gamma(\bar{\mu}+\bar{\alpha}'g_i)]^{1/\gamma}/n$. The other MD first-order conditions are $\sum_{i=1}^{n}\bar{\pi}_i = 1$ and, for $G_i(\beta) = \partial g_i(\beta)/\partial\beta$,

$$\sum_{i=1}^{n}\bar{\pi}_i G_i(\bar{\beta})'\bar{\alpha} = 0, \sum_{i=1}^{n}\bar{\pi}_i g_i(\bar{\beta}) = 0. \tag{A.3}$$

The first-order conditions for $\hat{\lambda}$ are $\sum_{i=1}^{n}\rho_1(\hat{\lambda}'g_i(\hat{\beta}))g_i(\hat{\beta}) = 0$. By the implicit function there is a neighborhood of $\hat{\beta}$ where the solution $\lambda(\beta)$ to $\sum_{i=1}^{n}\rho_1(\lambda'g_i(\beta))g_i(\beta) = 0$ exists and is continuously differentiable. Then by the envelope theorem the first-order conditions for GEL are

$$\sum_{i=1}^{n}\hat{\pi}_i G_i(\hat{\beta})'\hat{\lambda} = 0, \sum_{i=1}^{n}\hat{\pi}_i g_i(\hat{\beta}) = 0, \tag{A.4}$$

where $\hat{\pi}_i = \rho_1(\hat{\lambda}'g_i(\hat{\beta}))/\sum_{j=1}^{n}\rho_1(\hat{\lambda}'g_j(\hat{\beta}))$. Then for $\bar{\lambda} = \bar{\alpha}/(\gamma\bar{\mu})$, by $\sum_{i=1}^{n}\bar{\pi}_i = 1$,

$$\bar{\pi}_i = [(\gamma\bar{\mu})^{1/\gamma}/n](1+\gamma\bar{\lambda}'g_i(\bar{\beta}))^{1/\gamma} = (1+\gamma\bar{\lambda}'g_i(\bar{\beta}))^{1/\gamma}/\sum_{j=1}^{n}(1+\gamma\bar{\lambda}'g_j(\bar{\beta}))^{1/\gamma}$$

Noting that $\rho_1(v) = -(1+\gamma v)^{1/\gamma}$, we see from the respective first order conditions that the conclusion holds for $\hat{\pi}_i = \bar{\pi}_i$ and $\hat{\lambda} = \bar{\lambda}$.

For the $\gamma = 0$ case, we note that $\rho(v) = -e^{v}$ and that under the constraint $\sum_{i=1}^{n}\pi_i = 1$, $\sum_{i=1}^{n}h(\pi_i) = \sum_{i=1}^{n}\ln(n\pi_i)\pi_i = \sum_{i=1}^{n}\ln(\pi_i)\pi_i + \ln(n)$. Then using this objective function in the Lagrangean, the first-order conditions for $\bar{\pi}_i$ are $1+\ln(\bar{\pi}_i) = \bar{\mu}+\bar{\alpha}'g_i(\bar{\beta})$. Solving,

$$\bar{\pi}_i = \exp(\bar{\mu}-1+\bar{\alpha}'g_i(\bar{\beta})) = \exp(\bar{\lambda}'g_i(\bar{\beta}))/\sum_{j=1}^{n}\exp(\bar{\lambda}'g_j(\bar{\beta})),$$

where $\bar{\lambda} = \bar{\alpha}$. The conclusion then follows as before. Q.E.D.

Proof of Theorem 2.3: For notational simplicity drop the EL superscript and subscript, let $\hat{G}_i = G_i(\hat{\beta})$ and $\hat{g}_i = g_i(\hat{\beta})$, and let $Y = 1/\sum_{i=1}^{n} \rho_1(\hat{\lambda}'\hat{g}_i)$. Note that for EL, $\hat{\pi}_i = \hat{\pi}_i \hat{g}_i' \hat{\lambda} - Y$, so by the second part of eq. (A.4),

$$0 = \sum_{i=1}^{n} \hat{\pi}_i \hat{g}_i = \sum_{i=1}^{n} \hat{\pi}_i \hat{g}_i \hat{g}_i' \hat{\lambda} - nY\hat{g}(\hat{\beta}).$$

Solving for $\hat{\lambda}$, plugging into the first part of eq. (A.4), and dividing through by nY gives the result. Q.E.D.

Lemma A1: *If Assumption 1 is satisfied then for any ζ with $1/\alpha < \zeta < 1/2$ and $\Lambda_n = \{\lambda : \|\lambda\| \le n^{\zeta}\}$, $\sup_{\beta \in \mathcal{B}, \lambda \in \Lambda_n, 1 \le i \le n} |\lambda' g_i(\beta)| \xrightarrow{p} 0$ and w.p.a.1, $\Lambda_n \subseteq \hat{\Lambda}(\beta)$ for all $\beta \in B$.*

Proof: For $b_i = \sup_{\beta \in \mathcal{B}} \|g_i(\beta)\|$, by Assumption 4.1 it follows by M that $\max_{1 \le i \le n} b_i = O_p(n^{1/\alpha})$.Then by CS,

$$\sup_{\beta \in \mathcal{B}, \lambda \in \Lambda_n, 1 \le i \le n} |\lambda' g_i(\beta)| \le n^{-\gamma} \max_{1 \le i \le n} b_i = O_p(n^{-\gamma + 1/\alpha}) \xrightarrow{p} 0,$$

giving the first conclusion, so w.p.a.1 $\lambda' g_i(\beta) \in \mathcal{V}$ for all $\beta \in \mathcal{B}$ and $\|\lambda\| \le n^{-\zeta}$. Q.E.D.

Lemma A2: *If Assumption 1 is satisfied, $\bar{\beta} \in \mathcal{B}$, $\bar{\beta} \xrightarrow{p} \beta_0$, and $\hat{g}(\bar{\beta}) = O_p(n^{-1/2})$, then $\bar{\lambda} = \arg\max_{\lambda \in \hat{\Lambda}(\bar{\beta})} \hat{P}(\bar{\beta}, \lambda)$ exists w.p.a.1, $\bar{\lambda} = O_p(n^{-1/2})$, and $sup_{\lambda \in \hat{\Lambda}(\bar{\beta})} \hat{P}(\bar{\beta}, \lambda) \le \rho_0 + O_p(n^{-1})$.*

Proof: By LLN $\bar{\Omega} \stackrel{def}{=} \hat{\Omega}(\bar{\beta}) \xrightarrow{p} \Omega$. Then by nonsingularity of Ω the smallest eigenvalue of $\bar{\Omega}$ is bounded away from zero w.p.a.1. Let Λ_n be as defined in Lemma A2. By Lemma A2 and twice continuous differentiability of $\rho(v)$ in a neighborhood of zero, $\hat{P}(\bar{\beta}, \lambda)$ is twice continuously differentiable on Λ_n w.p.a.1. Then $\tilde{\lambda} = \arg\max_{\lambda \in \Lambda_n} \hat{P}(\beta_0, \lambda)$ exists w.p.a.1. Furthermore, for $\bar{g}_i = g_i(\bar{\beta})$ and any $\dot{\lambda}$ on the line joining $\tilde{\lambda}$ and 0, by Lemma A2 and $\rho_2 = -1$, $\max_{1 \le i \le n} \rho_2(\lambda' \bar{g}_i) < -1/2$ w.p.a.1. Then by a Taylor expansion around

$\lambda = 0$ with Lagrange remainder, there is $\dot{\lambda}$ on the line joining $\tilde{\lambda}$ and 0 such that for $\bar{g} \stackrel{def}{=} \hat{g}(\bar{\beta})$,

$$\begin{aligned}
\rho_0 &= \hat{P}(\bar{\beta}, 0) \leq \hat{P}(\bar{\beta}, \tilde{\lambda}) = \rho_0 - \tilde{\lambda}'\bar{g} + (1/2)\tilde{\lambda}'[\sum_{i=1}^{n} \rho_2(\dot{\lambda}'g_i)g_i g_i'/n]\tilde{\lambda} \qquad \text{(A.5)} \\
&\leq \rho_0 - \tilde{\lambda}'\bar{g} - (1/4)\tilde{\lambda}'\bar{\Omega}\tilde{\lambda} \leq \rho_0 + \|\tilde{\lambda}\|\|\bar{g}\| - C\|\tilde{\lambda}\|^2,
\end{aligned}$$

Subtracting $\rho_0 - C\|\tilde{\lambda}\|^2$ from both sides and dividing by $\|\tilde{\lambda}\|$ we find that $C\|\tilde{\lambda}\| \leq \|\bar{g}\|$, w.p.a.1. By assumption, $\bar{g} = O_p(n^{-1/2})$, and hence $\|\tilde{\lambda}\| = O_p(n^{-1/2}) = o_p(n^{-\zeta})$. Therefore, w.p.a.1 $\tilde{\lambda} \in int(\Lambda_n)$ and hence $\partial\hat{P}(\beta_0, \tilde{\lambda})/\partial\lambda = 0$, the first-order conditions for an interior maximum. By Lemma A2, w.p.a.1 $\tilde{\lambda} \in \hat{\Lambda}(\beta_0)$, so by concavity of $\hat{P}(\beta_0, \lambda)$ and convexity of $\hat{\Lambda}(\beta_0)$ it follows that $\hat{P}(\bar{\beta}, \tilde{\lambda}) = \sup_{\lambda \in \hat{\Lambda}(\bar{\beta})} \hat{P}(\bar{\beta}, \lambda)$, giving the first and second conclusions with $\bar{\lambda} = \tilde{\lambda}$. Then by the last inequality of equation (A.5), $\|\bar{g}\| = O_p(n^{-1/2})$, and $\|\tilde{\lambda}\| = O_p(n^{-1/2})$, we obtain $\hat{P}(\bar{\beta}, \bar{\lambda}) \leq \rho_0 + \|\bar{\lambda}\|\|\bar{g}\| - C\|\bar{\lambda}\|^2 = \rho_0 + O_p(n^{-1})$. Q.E.D.

Lemma A3: *If Assumption 1 is satisfied, then* $\|\hat{g}(\hat{\beta})\| = O_p(n^{-1/2})$.

Proof: Let $\hat{g}_i = g_i(\hat{\beta})$, $\hat{g} = \hat{g}(\beta_0)$, and for ζ in Lemma A2, $\tilde{\lambda} = -n^{-\zeta}\hat{g}/\|\hat{g}\|$. By Lemma A2, $\max_{i \leq n} |\tilde{\lambda}'\hat{g}_i| \xrightarrow{p} 0$ and $\tilde{\lambda} \in \hat{\Lambda}(\hat{\beta})$ w.p.a.1. Thus, for any $\dot{\lambda}$ on the line joining $\tilde{\lambda}$ and 0, w.p.a.1 $\rho_2(\dot{\lambda}'\hat{g}_i) \geq -C$, $(i = 1, ..., n)$. Also, by CS and LLN, $\sum_i \hat{g}_i\hat{g}_i'/n \leq (\sum_i \delta_i^2/n)I \xrightarrow{p} CI$, so that the largest eigenvalue of $\sum_i \hat{g}_i\hat{g}_i'/n$ is bounded above w.p.a.1. An expansion then gives

$$\begin{aligned}
\hat{P}(\hat{\beta}, \tilde{\lambda}) &= \rho_0 - \tilde{\lambda}'\hat{g} + (1/2)\tilde{\lambda}'[\sum_i \rho_2(\dot{\lambda}'\hat{g}_i)\hat{g}_i\hat{g}_i'/n]\tilde{\lambda} \\
&\geq \rho_0 + n^{-\zeta}\|\hat{g}\| - C(1/2)\tilde{\lambda}'[\sum_i \hat{g}_i\hat{g}_i'/n]\tilde{\lambda} \geq \rho_0 + n^{-\zeta}\|\hat{g}\| - Cn^{-2\zeta}.
\end{aligned}$$

w.p.a.1. By the CLT the hypotheses of Lemma A3 are satisfied by $\bar{\beta} = \beta_0$. Noting that by $\hat{\beta}$ and $\hat{\lambda}$ being a saddle point, this equation and Lemma A3 give

$$\rho_0 + n^{-\zeta}\|\hat{g}\| - Cn^{-2\zeta} \leq \hat{P}(\hat{\beta}, \tilde{\lambda}) \leq \hat{P}(\hat{\beta}, \hat{\lambda}) \leq \sup_{\lambda \in \hat{\Lambda}(\beta_0)} \hat{P}(\beta_0, \lambda) \leq \rho_0 + O_p(n^{-1}). \qquad \text{(A.6)}$$

Also, by $\zeta < 1/2$, $\zeta - 1 < -1/2 < -\zeta$. Solving for $\|\hat{g}\|$ then gives, by

$$\|\hat{g}\| \leq O_p(n^{-1+\zeta}) + Cn^{-\zeta} = O_p(n^{-\zeta}). \qquad \text{(A.7)}$$

Now, consider any $\varepsilon_n \to 0$. Let $\bar{\lambda} = -\varepsilon_n \hat{g}$. Note that $\bar{\lambda} = o_p(n^{-\zeta})$ by eq. (A.7), so that $\bar{\lambda} \in \Lambda_n$ w.p.a.1. Then, as in equation (A.6),

$$\rho_0 + \varepsilon_n \|\hat{g}\|^2 - C\varepsilon_n^2 \|\hat{g}\|^2 = \rho_0 - \bar{\lambda}'\hat{g} - C\|\bar{\lambda}\|^2 \leq \rho_0 + O_p(n^{-1}).$$

Since, for all n large enough, $1 - \varepsilon_n C$ is bounded away from zero, it follows that $\varepsilon_n \|\hat{g}\|^2 = O_p(n^{-1})$. The conclusion then follows by a standard result from probability theory, that if $\varepsilon_n Y_n = O_p(n^{-1})$ for all $\varepsilon_n \to 0$, then $Y_n = O_p(n^{-1})$. Q.E.D.

Proof of Theorem 3.1: Let $g(\beta) = E[g(z,\beta)]$. By Lemma A3, $\hat{g}(\hat{\beta}) \xrightarrow{p} 0$, and by LLN, $\sup_{\beta \in \mathcal{B}} \|\hat{g}(\beta) - g(\beta)\| \xrightarrow{p} 0$ and $g(\beta)$ is continuous. The triangle inequality then gives $g(\hat{\beta}) \xrightarrow{p} 0$. Since $g(\beta) = 0$ has a unique zero β_0, $\|g(\beta)\|$ must be bounded away from zero outside any neighborhood of β_0. Therefore, $\hat{\beta}$ must be inside any neighborhood of β_0 w.p.a.1, i.e. $\hat{\beta} \xrightarrow{p} \beta_0$, giving the first conclusion. Then second conclusion follows by Lemma A3. Also, note by the first two conclusions the hypotheses of Lemma A2 are satisfied for $\bar{\beta} = \hat{\beta}$, so that the last conclusion follows from Lemma A2. Q.E.D.

Proof of Theorem 3.2: For $\hat{g}_i = g_i(\hat{\beta})$, by Theorem 4.1 and Lemma A1, $\max_{i \leq n} |\hat{\lambda}'\hat{g}_i| \xrightarrow{p} 0$. Therefore, the first-order conditions $\sum_{i=1}^n \rho_1(\hat{\lambda}'\hat{g}_i)\hat{g}_i = 0$ are satisfied w.p.a.1. Also, $\tilde{\Omega} = \sum_{i=1}^n \rho_2(\hat{\lambda}'\hat{g}_i)\hat{g}_i\hat{g}_i'/n \xrightarrow{p} \rho_2\Omega$ so that $\tilde{\Omega}$ is nonsingular w.p.a.1. Then as in the proof of Theorem 2.2, the first-order conditions of eq. (A.4) are satisfied w.p.a.1. Then by a mean-value expansion of the second part of these first order conditions we have, for $\hat{\theta} = (\hat{\beta}', \hat{\lambda}')'$ and $\theta_0 = (\beta_0', 0')'$,

$$\begin{aligned} 0 &= \begin{pmatrix} 0 \\ -\hat{g}(\beta_0) \end{pmatrix} + \bar{M}(\hat{\theta} - \theta_0), \qquad \text{(A.8)} \\ \bar{M} &= \begin{pmatrix} 0 & \sum_{i=1}^n \rho_1(\hat{\lambda}'\hat{g}_i)G_i(\hat{\beta})'/n \\ \sum_{i=1}^n \rho_1(\bar{\lambda}'\hat{g}_i)G_i(\bar{\beta})/n & \sum_{i=1}^n \rho_2(\bar{\lambda}'\hat{g}_i)g_i(\bar{\beta})\hat{g}_i'/n \end{pmatrix}, \end{aligned}$$

where $\bar{\beta}$ and $\bar{\lambda}$ are mean-values that actually differ from row to row of the matrix $\bar{M}$. By $\bar{\lambda} = O_p(n^{-1/2})$, it follows as in Lemma A1 that $\max_{i \leq n} |\bar{\lambda}'\hat{g}_i| \xrightarrow{p} 0$. Therefore, $\max_{i \leq n} |\rho_1(\tilde{\lambda}'\hat{g}_i) + 1| \xrightarrow{p} 0$ and $\max_{i \leq n} |\rho_2(\bar{\lambda}'\hat{g}_i) + 1| \xrightarrow{p} 0$. It then follows from LLN that

$\bar{M} \xrightarrow{p} M$, where

$$M = -\begin{pmatrix} 0 & G' \\ G & \Omega \end{pmatrix}, M^{-1} = -\begin{pmatrix} -\Sigma & H \\ H' & P \end{pmatrix}.$$

Inverting and solving in eq. (A.8) then gives

$$\begin{aligned} \sqrt{n}(\hat{\theta} - \theta_0) &= -\bar{M}^{-1}(0, -\sqrt{n}\hat{g}(\beta_0)')' = -M^{-1}(0, -\sqrt{n}\hat{g}(\beta_0)')' + o_p(1) \qquad (A.9) \\ &= -(H', P)'\sqrt{n}\hat{g}(\beta_0) + o_p(1). \end{aligned}$$

The first conclusion follows from this equation and the CLT. The second conclusion follows similarly. For the third conclusion, note that an expansion and eq. (A.9) give

$$\hat{g}(\hat{\beta}) = \hat{g}(\beta_0) - GH\hat{g}(\beta_0) + o_p(n^{-1/2}) = -\Omega\hat{\lambda} + o_p(n^{-1/2}).$$

Expanding,

$$\begin{aligned} \hat{P}(\hat{\beta}, \hat{\lambda}) &= \rho_0 - \hat{\lambda}'\hat{g}(\hat{\beta}) + \hat{\lambda}'[\sum_{i=1}^{n} \rho_2(\bar{\lambda}'\hat{g}_i)\hat{g}_i\hat{g}_i'/n]\hat{\lambda}/2 \qquad (A.10) \\ &= \rho_0 - \hat{\lambda}'\hat{g}(\hat{\beta}) - \hat{\lambda}'\Omega\hat{\lambda}/2 + o_p(n^{-1}) = \rho_0 + \hat{g}(\hat{\beta})'\Omega^{-1}\hat{g}(\hat{\beta})/2 + o_p(n^{-1}). \end{aligned}$$

It follows as in Hansen (1982) that $n\hat{g}(\hat{\beta})'\Omega^{-1}\hat{g}(\hat{\beta}) \xrightarrow{d} \chi^2(m-p)$, so the conclusion follows from eq. (A.10). Q.E.D.

We now give some Lemmas that are used to derive asymptotic expansions. The next one is like Lemma 3.3 of Rilstone et. al. (1996), except that we expand in a shrinking neighborhood to allow for $\hat{\lambda}$ in GEL. For notational simplicity we will suppress the F argument.

Lemma A4: *Suppose that the estimator $\hat{\theta}$ and vector of functions $m(z, \theta)$ satisfies a) $\hat{\theta} = \theta_0 + O_p(n^{-1/2})$; b) $\sum_{i=1}^{n} m(z_i, \hat{\theta}) = 0$ w.p.a.1; c) For some $\zeta > 2$, $d(z)$ with $E[d(z)] < \infty$, and $T_n = \{\theta : \|\theta - \theta_0\| \leq n^{-1/\zeta}\}$, w.p.a.1 for $i = 1, ..., n$, $m(z_i, \theta)$ is three times continuously differentiable on T_n and for $\theta \in T_n$,*

$$\|\partial^3 m(z_i, \theta)/\partial\theta_j\partial\theta_k\partial\theta_\ell - \partial^3 m(z_i, \theta_0)/\partial\theta_j\partial\theta_k\partial\theta_\ell\| \leq d(z_i)\|\theta - \theta_0\|;$$

d) $E[m(z,\theta_0)] = 0$ *and* $M = E[\partial m(z,\theta_0)/\partial\theta]$ *exists and is nonsingular. Let*

$$\begin{aligned}
M_j &= E[\partial^2 m(z,\theta_0)/\partial\theta_j\partial\theta], M_{jk} = E[\partial^3 m(z,\theta_0)/\partial\theta_k\partial\theta_j\partial\theta], \\
A(z) &= \partial m(z,\theta_0)/\partial\theta - M, B_j(z) = \partial^2 m(z,\theta_0)/\partial\theta_j\partial\theta - M_j, \\
\psi(z) &= -M^{-1}m(z,\theta_0), a(z) = vecA(z), b(z) = vec[B_1(z),...,B_q(z)].
\end{aligned}$$

Suppose that $E[\|\psi(z)\|^6]$, $E[\|A(z)\|^6]$, *and* $E[\|B_j(z)\|^6]$ *are finite. Then eq. (3.1) is satisfied for*

$$\begin{aligned}
Q_1(\tilde\psi,\tilde a) &= -M^{-1}[\tilde A\tilde\psi + \sum_{j=1}^q \tilde\psi_j M_j\tilde\psi/2], Q_2(\tilde\psi,\tilde a,\tilde b) = -M^{-1}[\tilde A Q_1(\tilde\psi,\tilde a) \qquad \text{(A.11)} \\
&+ \sum_{j=1}^q\{\tilde\psi_j M_j Q_1(\tilde\psi,\tilde a) + Q_{1j}(\tilde\psi,\tilde a)M_j\tilde\psi + \tilde\psi_j\hat B_j\tilde\psi\}/2 + \sum_{j,k=1}^q \tilde\psi_j\tilde\psi_k M_{jk}\tilde\psi/6].
\end{aligned}$$

Proof: Let $\hat M(\theta) = n^{-1}\sum_{i=1}^n \partial m(z_i,\theta)/\partial\theta$. A Taylor expansion with Lagrange remainder gives,

$$\begin{aligned}
0 &= \hat m(\theta_0) + \hat M(\theta_0)(\hat\theta-\theta_0) + \sum_{j=1}^q(\hat\theta_j-\theta_{j0})[\partial\hat M(\theta_0)/\partial\theta_j](\hat\theta-\theta_0)/2 \qquad \text{(A.12)} \\
&+. \sum_{j,k=1}^q(\hat\theta_j-\theta_{j0})(\hat\theta_k-\theta_{k0})[\partial^2\hat M(\bar\theta)/\partial\theta_k\partial\theta_j](\hat\theta-\theta_0)/6.
\end{aligned}$$

By M, the CLT, and the Lipschitz hypothesis,

$$\begin{aligned}
\|\partial^2\hat M(\bar\theta)/\partial\theta_k\partial\theta_j - M_{jk}\| &\le \|\partial^2\hat M(\bar\theta)/\partial\theta_k\partial\theta_j - \partial^2\hat M(\theta_0)/\partial\theta_k\partial\theta_j\| + \|\partial^2\hat M(\theta_0)/\partial\theta_k\partial\theta_j - M_{jk}\| \\
&\le [\sum_{i=1}^n d(z_i)/n]\|\hat\theta-\theta_0\| + O_p(n^{-1/2}) = O_p(n^{-1/2}).
\end{aligned}$$

It follows then for $\hat M = \hat M(\theta_0)$ that by adding, subtracting, and solving gives

$$\begin{aligned}
\hat\theta-\theta_0 &= \tilde\psi/\sqrt n - M^{-1}[\tilde A(\hat\theta-\theta_0)/\sqrt n + \sum_{j=1}^q(\hat\theta_j-\theta_{j0})M_j(\hat\theta-\theta_0)/2 \qquad \text{(A.13)} \\
&+ \sum_{j=1}^q(\hat\theta_j-\theta_{j0})(\tilde B_j/\sqrt n)(\hat\theta-\theta_0)/2 \\
&+ \sum_{j,k=1}^q(\hat\theta_j-\theta_{j0})(\hat\theta_k-\theta_{k0})M_{jk}(\hat\theta-\theta_0)/6] + O_p(n^{-2}) \\
&= \tilde\psi/\sqrt n - M^{-1}[\tilde A(\hat\theta-\theta_0)/\sqrt n + \sum_{j=1}^q(\hat\theta_j-\theta_{j0})M_j(\hat\theta-\theta_0)/2 \\
&+ \sum_{j=1}^q\tilde\psi_j\hat B_j\tilde\psi/2n^{3/2} + \sum_{j,k=1}^q\tilde\psi_j\tilde\psi_k M_{jk}\tilde\psi/6n^{3/2}] + O_p(n^{-2}).
\end{aligned}$$

Noting that the last five terms are $O_p(n^{-1})$. Then solving this equation for $\tilde{\theta}$ gives

$$\hat{\theta} - \theta_0 = \tilde{\psi}/\sqrt{n} + O_p(n^{-1}).$$

Next, the last three terms in equation (A.13) are $O_p(n^{-3/2})$, and replacing $\hat{\theta}-\theta_0$ by $\tilde{\psi}/\sqrt{n}$ in the last third and fourth terms also generates an error that is $O_p(n^{-3/2})$, we obtain

$$\begin{aligned} \hat{\theta} - \theta_0 &= \tilde{\psi}/\sqrt{n} - M^{-1}[(\hat{M} - M)\tilde{\psi}/\sqrt{n} + \sum_{j=1}^{q} \tilde{\psi}_j M_j \tilde{\psi}/2n] + O_p(n^{-3/2}) \quad \text{(A.14)} \\ &= \tilde{\psi}/\sqrt{n} + Q_1(\tilde{\psi}, \tilde{a})/n + O_p(n^{-3/2}). \end{aligned}$$

Finally, replacing $\hat{\theta} - \theta_0$ in eq. by the expression following the last equality gives the conclusion. Q.E.D.

Lemma A5: *Suppose that Assumptions 1-4 are satisfied and let* $\Sigma_W = (G'W^{-1}G)^{-1}$, $H_W = \Sigma_W G'W^{-1}$, $P_W = W^{-1} - W^{-1}G\Sigma_W$, $\psi_i = -[H_W', P_W]'g_i$, $G_i^j = E[\partial G_i(\beta_0)/\partial\beta_j]$,

$$\begin{aligned} M_i &= -\begin{pmatrix} 0 & G_i' \\ G_i & W + \psi_i^W \end{pmatrix}, M = -\begin{pmatrix} 0 & G' \\ G & W \end{pmatrix}, M^{-1} = -\begin{pmatrix} -\Sigma_W & H_W \\ H_W' & P_W \end{pmatrix}, \\ M_j &= -\begin{pmatrix} 0 & E[G_i^j]' \\ E[G_i^j] & 0 \end{pmatrix}, (j \le p), M_{p+j} = -\begin{pmatrix} E[\partial^2 g_{ij}(\beta_0)/\partial\beta\partial\beta'] & 0 \\ 0 & 0 \end{pmatrix}, (j \le m). \end{aligned}$$

Then for $\tilde{\lambda} = \hat{W}^{-1}\hat{g}(\tilde{\beta})$, $\hat{\theta} = (\tilde{\beta}', \tilde{\lambda}')'$, $\tilde{\psi} = \sum_i \psi_i/\sqrt{n}$, $\tilde{A} = \sum_i (M_i - M)/\sqrt{n}$, *and* $Q_1(\tilde{\psi}, \tilde{a})$ *as in Lemma A4 we have,*

$$\hat{\theta} = \theta_0 + \tilde{\psi}/\sqrt{n} + Q_1(\tilde{\psi}, \tilde{a})/n + O_p(n^{-3/2}).$$

Proof: Let $\theta = (\beta', \lambda')'$, $\lambda_0 = 0$, $m(z, \theta) = -(\lambda'\partial g(z, \beta)/\partial\beta, g(z, \theta)' + \lambda'[W + \psi^W(z)])'$. It follows from Theorem 3.4 of Newey and McFadden (1994) that $\hat{\theta} = \theta_0 + O_p(n^{-1/2})$. Then for $\hat{m}(\theta) = \sum_i m(z_i, \theta)/n$, by the first order conditions for $\tilde{\beta}$, the definition of $\tilde{\lambda}$, and Assumption 4 we have

$$0 = \hat{m}(\hat{\theta}) + [0, -\tilde{\lambda}'(\hat{W} - W - \sum_i \psi^W(z_i)/n)]' = \hat{m}(\hat{\theta}) + O_p(n^{-3/2}). \quad \text{(A.15)}$$

Then expanding as in eq. (A.12) and solving as in eq. (A.14) gives the result. Q.E.D.

Lemma A6: *Suppose that Assumptions 1-4 are satisfied and let* $\Omega_{i\beta_j} = \partial[g_i(\beta_0)g_i(\beta_0)']/\partial\beta_j$, $\bar{\Omega}_{\beta_j} = E[\Omega_{i\beta_j}]$, $\tilde{\Omega}_{\beta_j} = \sum_i(\Omega_{i\beta_j} - \bar{\Omega}_{\beta_j})/\sqrt{n}$, $\bar{\Omega}_{\beta_j\beta_k} = E[\partial^2\{g_i(\beta_0)g_i(\beta_0)'\}/\partial\beta_k\partial\beta_j]$, $\psi_i^\beta =$

$-H_W g_i$, $\tilde{\psi}^\beta = \sum_i \psi_i^\beta/\sqrt{n}$, $\psi_i^\Omega = g_i g_i' - \Omega + \sum_{j=1}^p \bar{\Omega}_{\beta_j} e_j' \psi_i^\beta$, *and for* $Q_1(\tilde{\psi}, \tilde{a})$ *from the conclusion of Lemma A5 let*

$$\tilde{Q}_1^\Omega = \sum_{j=1}^p \tilde{\Omega}_{\beta_j} e_j' \tilde{\psi}^\beta + \sum_{j=1}^p \bar{\Omega}_{\beta_j} e_j' Q_1(\tilde{\psi}, \tilde{a}) + \sum_{j,k=1}^p \bar{\Omega}_{\beta_k \beta_j} e_j' \tilde{\psi}^\beta e_k' \tilde{\psi}^\beta/2.$$

Then $\hat{\Omega}(\tilde{\beta}) = \Omega + \tilde{\psi}^\Omega/\sqrt{n} + \tilde{Q}_1^\Omega/n + O_p(n^{-3/2})$.

Proof: Expanding gives

$$\begin{aligned}\hat{\Omega}(\tilde{\beta}) &= \Omega + \tilde{\psi}^\Omega/\sqrt{n} + \sum_{j=1}^p (\tilde{\Omega}_{\beta_j}/\sqrt{n})(\tilde{\beta}_j - \beta_{j0}) + \sum_{j=1}^p \bar{\Omega}_{\beta_j}(\tilde{\beta}_j - \beta_{j0}) \\ &\quad + \sum_{j,k=1}^p \bar{\Omega}_{\beta_k \beta_j}(\tilde{\beta}_j - \beta_{j0})(\tilde{\beta}_k - \beta_{k0})/2 + O_p(n^{-3/2}),\end{aligned}$$

so the conclusion follows by substituting for $\tilde{\beta}$ from Lemma A5. Q.E.D.

Proof of Theorem 3.3: Let $m_i(\theta) = -(\lambda' G_i(\beta), g_i(\beta)' + \lambda'(\Omega + \psi_i^\Omega))'$, $A(z_i)$, M, M_j be as in Lemma A5 with $W = \Omega$. Also, let $\psi_i = -[H', P']' g_i$ and $\tilde{\psi} = \sum_i \psi_i/\sqrt{n}$.

$$\begin{aligned} B_j(z_i) &= -\begin{pmatrix} 0 & G_i^{j\prime} - E[G_i^j]' \\ G_i^j - E[G_i^j] & 0 \end{pmatrix}, (j \le p), \\ B_{p+j}(z_i) &= -\begin{pmatrix} \partial^2 g_{ij}(\beta_0)/\partial\beta\partial\beta' - E[\partial^2 g_{ij}(\beta_0)/\partial\beta\partial\beta'] & 0 \\ 0 & 0 \end{pmatrix}, (j \le m). \end{aligned}$$

Let $\hat{\lambda} = -\hat{\Omega}(\tilde{\beta})^{-1}\hat{g}(\hat{\beta})$. Then $\hat{\lambda} = O_p(n^{-1/2})$, e.g. as shown in Newey and McFadden (1994). Then the first-order conditions for GMM and Lemma A6 imply

$$0 = \hat{m}(\hat{\theta}) + [0, -\hat{\lambda}'(\tilde{Q}_1^\Omega/n + O_p(n^{-3/2}))]' = \hat{m}(\hat{\theta}) + [0, -\hat{\lambda}'\tilde{Q}_1^\Omega]' + O_p(n^{-2}). \tag{A.16}$$

Then for $Q_1(\tilde{\psi}, \tilde{a})$ and $Q_2(\tilde{\psi}, \tilde{a}, \tilde{b})$ as given in the conclusion of Lemma A4, with $\tilde{\psi}$, M, M_j, M_{jk}, $A(z)$, and $B(z)$ as specified here (and in Lemma A5 with $W = \Omega$), and for $T = \theta_0 + \tilde{\psi}/\sqrt{n} + Q_1(\tilde{\psi}, \tilde{a})/n + Q_2(\tilde{\psi}, \tilde{a}, \tilde{b})/n^{3/2}$, solving for $\hat{\theta} - \theta_0$ as in the conclusion of Lemma A4 gives

$$\hat{\theta} = T - M^{-1}[0, -\hat{\lambda}'\tilde{Q}_1^\Omega]' + O_p(n^{-2}) = T + M^{-1}[0, \hat{\lambda}'\tilde{Q}_1^\Omega]' + O_p(n^{-2}).$$

This equation implies that $\hat{\lambda} = [0, I_m]\tilde{\psi} + O_p(n^{-2})$. Substituting we obtain

$$\hat{\theta} = T + M^{-1} diag[0, \tilde{Q}_1^\Omega]\tilde{\psi} + O_p(n^{-2}).$$

The conclusion follows by adding $M^{-1}diag[0, \tilde{Q}_1^{\Omega}]\tilde{\psi}$ to $Q_2(\tilde{\psi}, \tilde{a}, \tilde{b})$ (from Lemma A4) to obtain the second-order term in the stochastic expansion for GMM, with the first-order term being $Q_1(\tilde{\psi}, \tilde{a})$ (from Lemma A4). Q.E.D.

Proof of Theorem 3.4: We apply Lemma A4. Let $\theta = (\beta', \lambda')'$, $\theta_0 = (\beta_0', 0')'$, $\hat{\theta}$ be the GEL estimator, $G_i(\beta) = \partial g_i(\beta)/\partial \beta$, and

$$m(z_i, \theta) = \rho_1(\lambda' g_i(\beta)) \begin{pmatrix} G_i(\beta)'\lambda \\ g_i(\beta) \end{pmatrix}, \theta = (\beta', \lambda')'.$$

By Theorem 3.2, $\hat{\theta} = \theta_0 + O_p(n^{-1/2})$. Also, as shown in the proof of Theorem 3.2, $\sum_i m(z_i, \hat{\theta}) = 0$ w.p.a.1. Let $2 < \zeta < \alpha$ for α in Assumption 3.3. Then by Lemma A1, Assumption 3.3, and $\rho_1(v)$ three times continuously differentiable on a neighborhood of 0, $m(z_i, \theta)$ is three times continuously differentiable on the T_n from Lemma A3, $i = 1, ..., n$, to which we henceforth restrict attention. Let $m_i(\theta) = m(z_i, \theta)$, $v_i(\theta) = \lambda' g_i(\theta)$, and $h_i(\theta) = \partial v_i(\theta)/\partial \theta = (\lambda' G_i(\beta), g_i(\beta)')'$. Then

$$\begin{aligned}
\partial m_i(\theta)/\partial \theta &= \rho_2(v_i(\theta))h_i(\theta)h_i(\theta)' + \rho_1(v_i(\theta))\partial h_i(\theta)/\partial \theta, \qquad \text{(A.17)} \\
\partial^2 m_i(\theta)/\partial \theta_j \partial \theta &= \rho_3(v_i(\theta))h_i(\theta)_j h_i(\theta)h_i(\theta)' + \rho_2(v_i(\theta))\partial[h_i(\theta)h_i(\theta)']/\partial \theta_j \\
&\quad + \rho_2(v_i(\theta))h_i(\theta)_j \partial h_i(\theta)/\partial \theta + \rho_1(v_i(\theta))\partial^2 h_i(\theta)/\partial \theta_j \partial \theta. \\
\partial^3 m_i(\theta)/\partial \theta_k \partial \theta_j \partial \theta &= \rho_4(v_i(\theta))h_i(\theta)_k h_i(\theta)_j h_i(\theta)h_i(\theta)' + \rho_3(v_i(\theta))\partial[h_i(\theta)_j h_i(\theta)h_i(\theta)']/\partial \theta_k \\
&\quad + \rho_3(v_i(\theta))h_i(\theta)_k \partial[h_i(\theta)h_i(\theta)']/\partial \theta_j + \rho_2(v_i(\theta))\partial^2[h_i(\theta)h_i(\theta)']/\partial \theta_k \partial \theta_j \\
&\quad + \rho_3(v_i(\theta))h_i(\theta)_k h_i(\theta)_j \partial h_i(\theta)/\partial \theta + \rho_2(v_i(\theta))\partial[h_i(\theta)_j \partial h_i(\theta)/\partial \theta]/\partial \theta_k \\
&\quad + \rho_2(v_i(\theta))h_i(\theta)_k \partial^2 h_i(\theta)/\partial \theta_j \partial \theta + \rho_1(v_i(\theta))\partial^3 h_i(\theta)/\partial \theta_k \partial \theta_j \partial \theta.
\end{aligned}$$

By hypothesis $\rho_j(v)$ is Lipschitz in a neighborhood of zero so that for $b_i = b(z_i)$,

$$|\rho_j(v_i(\theta)) - \rho_j| \leq C|v_i(\theta)| \leq C\|\lambda\|\|g_i(\beta)\| \leq Cb_i\|\theta - \theta_0\|.$$

Also, by Assumption 3.3, all of the terms involving $h_i(\theta)$ and its derivatives in the third derivative for $m_i(\theta)$ are bounded above by Cb_i^4 on T_n. Then the norm of the difference of $\partial^3 m_i(\theta)/\partial \theta_k \partial \theta_j \partial \theta$ and the same expression with $v_i(\theta)$ replaced by $v_i(\theta_0) = 0$ is bounded above by $Cb_i^5\|\theta-\theta_0\|$. Also, it follows by similar reasoning that the difference of each expression involving $h_i(\theta)$ and its value at θ_0 is bounded by $Cb_i^J\|\theta-\theta_0\|$ for some integer $J \leq 4$. Thus, the Lipschitz hypothesis of Lemma A4 holds by $E[b_i^5]<\infty$.

Next, let $g_i = g_i(\beta_0)$ and $G_i = G_i(\beta_0)$. Note that $h_i(\theta_0) = (0', g_i')'$, so that by $\rho_1 = \rho_2 = -1$,

$$\partial m_i(\theta_0)/\partial\theta = -\begin{pmatrix} 0 & G_i' \\ G_i & g_i g_i' \end{pmatrix}, M = -\begin{pmatrix} 0 & G' \\ G & \Omega \end{pmatrix}. \tag{A.18}$$

M is nonsingular, as shown in the proof of Theorem 3.2. Now let $G_i^j = \partial^2 g_i(\beta_0)/\partial\beta_j\partial\beta$, $g_i^j = \partial g_i(\beta_0)/\partial\beta_j$, $t = j-p$ for $j>p$, let e_t denote the t^{th} unit vector, and a t subscript denote the t^{th} element of a vector. Then evaluate at $\theta=\theta_0$ to obtain

$$\begin{aligned}\partial^2 m_i(\theta_0)/\partial\theta_j\partial\theta &= -\begin{pmatrix} 0 & G_i^{j\prime} \\ G_i^j & g_i^j g_i' + g_i g_i^{j\prime} \end{pmatrix}, (j\leq p), \\ &= -\begin{pmatrix} \partial^2[e_t' g_i(\beta_0)]/\partial\beta\partial\beta' & G_i' e_t g_i' + g_{it} G_i' \\ g_i e_t' G_i + g_{it} G_i & -\rho_3 g_{it} g_i g_i' \end{pmatrix}, (j>p).\end{aligned} \tag{A.19}$$

Next, let $G_i^{jk} = \partial^3 g_i(\beta_0)/\partial\beta_k\partial\beta_j\partial\beta$ and $g_i^{jk} = \partial^2 g_i(\beta_0)/\partial\beta_k\partial\beta_j$. Then for the second derivatives corresponding to β, with $j\leq p$ and $k \leq p$,

$$\partial^3 m_i(\theta_0)/\partial\theta_k\partial\theta_j\partial\theta = -\begin{pmatrix} 0 & G_i^{jk\prime} \\ G_i^{jk} & g_i^{jk} g_i' + g_i^j g_i^{k\prime} + g_i^k g_i^{j\prime} + g_i g_i^{jk\prime} \end{pmatrix}. \tag{A.20}$$

For the cross-partial between λ_t and β_j, with $j\leq p$, $k>p$, and $t=k-p$,

$$\begin{aligned}&\partial^3 m_i(\theta_0)/\partial\theta_k\partial\theta_j\partial\theta \\ = \; -&\begin{pmatrix} \partial^3 g_{it}(\beta_0)/\partial\beta_j\partial\beta\partial\beta' & G_i' e_t g_i^{j\prime} + G_i^{j\prime} e_t g_i' + G_{itj} G_i' + g_{it} G_i^{j\prime} \\ g_i^j e_g G_i + g_i e_t G_i^j + G_{itj} G_i + g_{it} G_i^j & -\rho_3[G_{itj} g_i g_i' + g_{it}(g_i^j g_i + g_i g_i^{j\prime})] \end{pmatrix}.\end{aligned} \tag{A.21}$$

For the second partial derivatives between λ_t and λ_u, with $j>p$, $k>p$, $t=j-p$, and $u=k-p$,

$$\begin{aligned}\partial^3 m_i(\theta_0)/\partial\theta_k\partial\theta_j\partial\theta &= \begin{pmatrix} -G_i' e_t e_u' G_i - G_i' e_u e_t' G_i & \rho_3(g_{it} G_i' e_u + g_{iu} G_i' e_t) g_i' \\ \rho_3 g_i (g_{it} e_u' G_i + g_{iu} e_t' G_i) & \rho_4 g_{it} g_{iu} g_i g_i' \end{pmatrix} \\ &- \begin{pmatrix} g_{it}\partial^2 g_{tu}(\beta_0)/\partial\beta\partial\beta' + g_{it}\partial^2 g_{tu}(\beta_0)/\partial\beta\partial\beta' & -\rho_3 g_{it} g_{iu} G_i' \\ -\rho_3 g_{it} g_{iu} G_i & 0 \end{pmatrix}.\end{aligned} \tag{A.22}$$

Then by the conclusion of Lemma A4, equation (3.1) is satisfied, for Q_1, Q_2, $a(z)$, and $b(z)$ as given in the statement of Lemma A4, and $m_i(\theta)$ and its derivatives as given in this proof. Q.E.D.

Proof of Theorem 4.1: By Lemma A6 it follows that Assumption 4 is satisfied for $W = \Omega$ and $\psi_i^W = g_i g_i' - \Omega - \sum_{j=1}^p \bar{\Omega}_{\beta_j} e_j' H_W g_i$. Note that $E[e_j' H_W g_i P g_i] = PE[g_i g_i'] H_W' e_j = (H_W - H)' e_j$. Also, for $S_k = E[\partial^2 g_{ik}(\beta_0)/\partial\beta\partial\beta']$, the k^{th} element of $\sum_{j=1}^p E[G_i^j]\Sigma e_j/2$ is $\sum_{j=1}^p e_j' S_k \Sigma e_j/2 = \sum_{j=1}^p tr(\Sigma e_j e_j' S_k)/2 = a_k$. Also, Then for $\hat{\lambda} = -\hat{\Omega}(\tilde{\beta})^{-1}\hat{g}(\beta_0)$ the bias of $\hat{\theta} = (\hat{\beta}', \hat{\lambda}')'$ can be obtained as the expectation of the term from Lemma A5 with $W = \Omega$, giving

$$\begin{aligned}
Bias(\hat{\theta}) &= E[Q_1(\psi_i, a_i)]/n = -M^{-1}\{E[-\begin{pmatrix} 0 & G_i' \\ G_i & \psi_i^\Omega \end{pmatrix}\begin{pmatrix} -Hg_i \\ -Pg_i \end{pmatrix}] \\
&\quad -\sum_{j=1}^p \begin{pmatrix} 0 & E[G_i^j]' \\ E[G_i^j] & 0 \end{pmatrix}\begin{pmatrix} \Sigma \\ 0 \end{pmatrix} e_j/2 - \sum_{j=1}^m \begin{pmatrix} S_j & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 \\ P \end{pmatrix} e_j/2\}/n \\
&= -M^{-1}\begin{pmatrix} E[G_i' P g_i] \\ E[G_i H g_i] - a + E[g_i g_i' P g_i] - \sum_{j=1}^p \bar{\Omega}_{\beta_j}(H_W - H)' e_j \end{pmatrix}/n.
\end{aligned}$$

Then $[I_p, 0]M^{-1} = [\Sigma, -H]$ and the previous equation gives the result. Q.E.D.

Proof of Theorem 4.2: By the proof of Theorem 3.4 $\hat{\theta} = (\hat{\beta}', \hat{\lambda}')'$ satisfies eq. (3.1) with Q_1 as in the statement of Lemma A4 with $\psi(z_i) = -[H', P]g_i$, for $H = \Sigma G' \Omega^{-1}$, $A(z_i) = \partial m_i(\theta_0)/\partial\theta - E[\partial m_i(\theta_0)/\partial\theta]$ for $\partial m_i(\theta_0)/\partial\theta$ from eq. (A.18), and $M_j = E[\partial^2 m_i(\theta_0)/\partial\theta_j\partial\theta]$ for $\partial^2 m_i(\theta_0)/\partial\theta_j\partial\theta$ from eq. (A.19). Note that $E[\psi_i\psi_i'] = diag[\Sigma, P]$ and

$$E[A(z_i)\psi_i] = \begin{pmatrix} E[G_i' P g_i] \\ E[G_i H g_i + g_i g_i' P g_i] \end{pmatrix}.$$

Also, $\sum_{j=1}^m P e_j g_{ij} = \sum_{j=1}^m P e_j e_j' g_i = P g_i$, and by symmetry of P, $\sum_{j=1}^m G_i' e_j g_i' P e_j = \sum_{j=1}^m G_i' e_j e_j' P g_i = G_i' P g_i$. Then

$$\begin{aligned}
&\sum_{j=1}^q M_j E[\psi_i\psi_i'] e_j/2 = \sum_{j=1}^p M_j[\Sigma, 0]' e_j/2 + \sum_{j=1}^m M_{j+p}[0, P]' e_j/2\} \\
&= -\sum_{j=1}^p \begin{pmatrix} 0 \\ E[G_i^j]\Sigma e_j/2 \end{pmatrix} - \sum_{j=1}^m \begin{pmatrix} E[G_i' e_j g_i' + g_{ij} G_i'] P e_j/2 \\ -\rho_3 E[g_{ij} g_i g_i'] P e_j/2 \end{pmatrix} = \begin{pmatrix} -E[G_i' P g_i] \\ -a + \rho_3 E[g_i g_i' P g_i]/2 \end{pmatrix}.
\end{aligned}$$

Then by Lemma A4, $Bias(\hat{\theta})$ is the first p elements of

$$\begin{aligned}
E[Q_1(\psi_i, a_i, F_0)]/n &= -M\{E[A(z_i)\psi_i] + \sum_{j=1}^q M_j E[\psi_i\psi_i'] e_j/2\}/n \\
&= -M\{\begin{pmatrix} 0 \\ -a + E[G_i H g_i] + (1 + \rho_3/2) E[g_i g_i' P g_i] \end{pmatrix}\}/n. \text{ Q.E.D.}
\end{aligned}$$

Proof of Theorem 4.5: For $Q = E[x_i x_i']$ we have $\Omega = \sigma_\varepsilon^2 Q$ and $G = -E[x_i w_i']$. Also, note that $E[x_i' P x_i] = \sigma_\varepsilon^{-2} E[g_i' P g_i] = \sigma_\varepsilon^{-2}(m-p)$. It then follows that

$$E[G_i' P g_i] = -E[\varepsilon_i w_i x_i' P x_i] = -\sigma_{\varepsilon w} E[x_i' P x_i] = -(\sigma_{\varepsilon w}/\sigma_\varepsilon^2)(m-p).$$

Also, since $w_i' \Sigma G' \Omega^{-1} x_i$ is a scalaar,

$$\begin{aligned} E[G_i \Sigma G' \Omega^{-1} g_i] &= -E[x_i w_i' \Sigma G' \Omega^{-1} x_i \varepsilon_i] = -E[x_i x_i' \Omega^{-1} G \Sigma w_i \varepsilon_i] \\ &= -E[x_i x_i'] \Omega^{-1} G \Sigma \sigma_{\varepsilon w} = -G\bar{\delta}. \end{aligned}$$

It then follows that

$$\begin{aligned} B_I &= \Sigma G' \Omega^{-1} E[G_i \Sigma G' \Omega^{-1} g_i] = \Sigma G' \Omega^{-1}(-G\bar{\delta})/n = -\bar{\delta}/n, \\ B_G &= -\Sigma E[G_i' P g_i] = (m-p)\bar{\delta}/n, B_\Omega = \mu_3 \Sigma G' \Omega^{-1} E[x_i x_i' P x_i]. \end{aligned}$$

The conclusion then follows by Theorem 4.1 and 4.2. Q.E.D.

Proof of Theorem 4.6: Note that $tr(\Sigma \partial^2 g_{ij}(\beta_0)/\partial\beta\partial\beta') = q_j(x_i) tr(\Sigma u_{\beta\beta i})$, so that $a_j = E[q_j(x_i) tr(\Sigma u_{\beta\beta i})]/2 = E[q_j(x_i) tr(\Sigma H_i)]/2$. Also, note that $G_i = q_i u_{\beta i}'$, so that $G' \Omega^{-1} G_i = \bar{d}_i u_{\beta i}'$. Then we have

$$\begin{aligned} \Sigma G' \Omega^{-1} a &= \Sigma G' \Omega^{-1} E[q_i tr(\Sigma H_i)] = \Sigma E[\bar{d}_i tr(\Sigma H_i)]/2, \\ \Sigma G' \Omega^{-1} E[G_i \Sigma G' \Omega^{-1} g_i] &= \Sigma E[\bar{d}_i u_{\beta i}' \Sigma \bar{d}_i u_i] = \Sigma E[\bar{d}_i \bar{d}_i' \Sigma u_{\beta i} u_i] = \Sigma E[\bar{d}_i \bar{d}_i' \Sigma \delta_i], \\ B_G &= \Sigma E[u_{\beta i} q_i' P q_i u_i] = E[\delta_i q_i' P q_i], \\ B_\Omega &= \Sigma E[\bar{d}_i u_i^3 q_i' P q_i] = \Sigma E[\bar{d}_i \mu_{3i} q_i' P q_i]. \end{aligned}$$

Note next that $\bar{d}_i$ is the mean-square projection of d_i on q_i for the expectation operator $\bar{E}$ given by $\bar{E}[a(x_i)] = E[\sigma_i^2 a(x_i)]/E[\sigma_i^2]$. Therefore, it follows that $E[\sigma_i^2 \|\bar{d}_i\|^2] \le E[\sigma_i^2 \|d_i\|^2]$. By standard results for matrix norms, $|tr(\Sigma H_i)| \le p\|\Sigma H_i\| \le p\|\Sigma\|\|H_i\|$. Then by CS

$$\begin{aligned} \|E[\bar{d}_i tr(\Sigma H_i)]/2\| &\le p\|\Sigma\| E[\sigma_i \|\bar{d}_i\| \|H_i\|/\sigma_i]/2 \le p\|\Sigma\| \sqrt{E[\sigma_i \|\bar{d}_i\|^2]} \sqrt{E[\|H_i\|^2/\sigma_i^2]}/2 \\ &\le p\|\Sigma\| \sqrt{E[\sigma_i \|d_i\|^2]} \sqrt{E[\|H_i\|^2/\sigma_i^2]}/2. \end{aligned}$$

Also, we have for $\Delta = \sup_x \|\delta(x)/\sigma^2(x)\|$,

$$\|E[\bar{d}_i \bar{d}_i' \Sigma \delta_i]\| \le \|\Sigma\| E[\|\bar{d}_i\|^2 \|\delta_i\|] \le \|\Sigma\| E[\sigma_i^2 \|\bar{d}_i\|^2] \Delta \le \|\Sigma\| E[\sigma_i^2 \|d_i\|^2] \Delta.$$

By the T and CS we then have

$$\|Bias(\hat{\beta}_{EL})\| \leq \|\Sigma\|^2 (p\sqrt{E[\sigma_i \|d_i\|^2]}\sqrt{E[\|H_i\|^2/\sigma_i^2]}/2 + E[\sigma_i^2 \|d_i\|^2]\Delta)/n,$$

giving the first conclusion. For the second conclusion, note that $E[\sigma_i^2 q_i' P q_i] = E[g_i' P g_i] = m - p$, so that for $\eta_i = e_j' \Sigma(\delta_i + \bar{d}_i \mu_{3i})/\sigma_i^2$,

$$e_j'(Bias(\hat{\beta}_{GMM}) - Bias(\hat{\beta}_{EL})) = e_j' \Sigma E[(\delta_i + \bar{d}_i \mu_{3i}) q_i' P q_i]/n = E[\eta_i \sigma_i^2 q_i P q_i'].$$

The second conclusion then follows from $\sigma_i^2 q_i' P q_i \geq 0$, so that when $\eta_i \geq C_2$, $E[\eta_i \sigma_i^2 q_i P q_i'] \geq C_2 E[\sigma_i^2 q_i' P q_i] = C_2(m - p)$. Q.E.D.

Proof of Theorem 4.7: The bias formulae follow immediately from Theorems 4.1 and 4.2, since by $G_i = G$,

$$E[G_i' P g_i] = E[G' P g_i] = G' P E[g_i] = 0, E[G_i H g_i] = E[G H g_i] = 0.$$

To obtain the bound, note that differentiating the equality $h(\beta) = \int r(z) f(z|\beta) dz$ under the integral is allowed by the conditions, as is differentiating the identity $1 = \int f(z|\beta) dz$. Twice differentiating the second gives $E[s_i] = 0$ and $E[F_i] = 0$. Twice differentiating the first gives each gives

$$\begin{aligned} G &= -\int r(z)[\partial f(z|\beta_0)/\partial \beta] dz = -E[r(z_i) s_i] = -E[g_i s_i], \\ a_j &= -tr(\Sigma \int r_j(z)[\partial^2 f(z|\beta_0)/\partial \beta \partial \beta'] dz)/2 = -E[r_j(z_i) tr(\Sigma F_i)]/2 = -E[g_{ij} tr(\Sigma F_i)]/2. \end{aligned}$$

Stacking the formulae for a_j we find that for $\tau_i = tr(\Sigma F_i)$, $a = -E[g_i \tau_i]/2$, so that

$$Bias(\hat{\beta}_{EL}) = -\Sigma E[s_i g_i'](E[g_i g_i'])^{-1} E[g_i \tau_i]/2n.$$

Note that $\tau_i^2 \leq p^2 \|\Sigma\|^2 \|F_i\|^2$, so that by CS,

$$\|Bias(\hat{\beta}_{EL})\| \leq \|\Sigma\| \sqrt{E[\|s_i\|^2] E[\tau_i^2]}/2n \leq p\|\Sigma\|^2 \sqrt{E[\|s_i\|^2] E[\|F_i\|^2]}/2n. Q.E.D.$$

Proof of Theorem 5.1: In the case of GMM, the bias correction takes the form $Bias(\hat{\beta}) = \tau(\sum_i d_i(\hat{\beta})/n)/n$, where $d_i(\beta) = d(z_i, \beta)$ is a vector of products of $g(z, \beta)$ and its derivatives to second order and τ is a function that is twice continuously differentiable in a neighborhood of $d_0 = E[d_i(\beta_0)]$. Then by Assumption 3 and a standard expansion,

$$\widehat{Bias}(\hat{\beta}) = \tau(d_0)/n + \tau_d(d_0)\sum_i \psi_i^\tau/n^2 + O_p(n^{-2}), \psi_i^\tau = d_i(\beta_0) - d_0 - E[\partial d_i(\beta_0)/\partial\beta]Hg_i.$$

Then for $\tilde{\psi}$, Q_1, and Q_2 from Theorem 3.3,

$$\sqrt{n}(\hat{\beta}_c - \beta_0) = \tilde{\psi} + [Q_1(\tilde{\psi}, \tilde{a}) - \tau(d_0)]/\sqrt{n} + [Q_2(\tilde{\psi}, \tilde{a}, \tilde{b}, F_0) + \tau_d(d_0)\tilde{\psi}^\tau]/n + O_p(n^{-3/2}),$$

giving the conclusion for GMM. The conclusion follows similarly for GEL, with τ and $d(z, \beta)$ corresponding to the bias formula for EL, and

$$\psi_i^\tau = d_i(\beta_0) - d_0 - E[d_i(\beta_0)g_i']\Omega^{-1}g_i - E[\partial d_i(\beta_0)/\partial\beta]Hg_i.Q.E.D.$$

Proof of Theorem 6.1: By Theorem 3.1 and Lemma A1, $\max_{i\le n}|\hat{\lambda}'g_i(\hat{\beta})| \xrightarrow{p} 0$, so that for $\hat{\pi}_i$ as in Table 1, $\hat{\pi}_i > 0$, $(i = 1, ..., n)$. Then the first-order conditions for maximization over λ imply that $\sum_{i=1}^n \hat{\pi}_i g_i(\hat{\beta}) = 0$. Also, for $\bar{\mu} = -\sum_{j=1}^n (1 - \hat{\lambda}'\hat{g}_j)^{-1}$ and $\bar{\alpha} = -\bar{\mu}\hat{\lambda}$, the formula implies that $\hat{\pi}_i = -1/[\bar{\mu} + \bar{\alpha}'g_i(\hat{\beta})]$. Therefore, the first-order conditions for the Lagrangean are satisfied, and hence by global concavity of the objective function, $\hat{\pi}_i, \hat{\beta}$ are a solution to the constrained maximization problem in eq. ?? with $h(\pi) = -\ln\pi$. Suppose that the support of z_i *is* $\{z^1, ..., z^J\}$, and let $I_j = \{i : z_i = z^j\}$. Then $\hat{\pi}_i$ maximizes the objective functions

$$\sum_{j=1}^J \sum_{i\in I_j} \ln(\pi_i) s.t \sum_{j=1}^J (\sum_{i\in I_j} \pi_i) g(z^j, \beta) = 0, \sum_{j=1}^J (\sum_{i\in I_j} \pi_i) = 1.$$

For given $\pi^j = \sum_{i\in I_j} \pi_i$, by strict concavity of $\ln\pi$, this is maximized at $\pi_i = \pi^j/n_j, i \in I_j$, where n_j is equal to the number of elements of I_j. Then $\hat{\pi}^j$ and $\hat{\beta}$ solve the concentrated maximization problem,

$$\sum_{j=1}^J n_j \ln(\pi^j) - \sum_{j=1}^J n_j \ln(n_j)\ s.t \sum_{j=1}^J \pi^j g(z^j, \beta) = 0, \sum_{j=1}^J \pi^j = 1.$$

which is precisely the maximum likelihood objective function, showing the first conclusion.

To show the second, note that $E[A_n] = \tau(E[m(z,\beta_0)])$ for vectors $\tau(m)$ and $m(z,\beta)$, not necessarily of the same dimension, which in the discrete case is $\tau(\sum_{j=1}^J \pi_0^j m(z^j,\beta_0))$. Also,

$$\sum_{i=1}^n \hat{\pi}_i m(z_i,\hat{\beta}) = \sum_{j=1}^J (\sum_{i\in I_j} \hat{\pi}_i) m(z^j,\hat{\beta}) = \sum_{j=1}^J (n_j \hat{\pi}^j / n_j) m(z^j,\hat{\beta}) = \sum_{j=1}^J \hat{\pi}^j m(z^j,\hat{\beta}),$$

so that the EL bias correction $\tau(\sum_{i=1}^n \hat{\pi}_i m(z_i,\hat{\beta}))/n$ is equal to the discrete maximum likelihood bias correction $\tau(\sum_{j=1}^J \hat{\pi}^j m(z^j,\hat{\beta}))$. Q.E.D.

Proof of Theorem 6.2: Consider the case where z_i has finite support. It follows from the smoothness and nonsingularity conditions of Assumptions 1 - 3, similarly to the proof of Lemma 1 of Chamberlain (1987), that the multinomial likelihood for the discrete data with fixed support satisfies all the regularity conditions of Pfanzagl and Wefelmeyer (1978) (PF henceforth). The conclusion for discrete data then follows from Theorem 5.1 and Corollary 1' of PF, for the case of Remark 3.16 of PF.

To show that the result holds in general, we note that by standard V-statistic results and the form of Q_1 and Q_2 derived in the proofs, it follows that for each estimator Ξ consists of a function of expectations, i.e. it takes the form $\Xi = \tau(E[d(z,\beta_0)])$ for some function τ that does not depend on the true distribuiton. By Lemma 3 of Chamberlain (1987), for EL and the bias-corrected or GMM pair of estimators there exists a distribution with finite support satisfying the moment restrictions and Assumptions 1-3 such that the corresponding expectation vectors $E[d(z,\beta_0)]$ under the discrete distribution are equal to those under the true distribution. Then by the efficiency result for discrete distributions, which implies that Ξ is smaller for EL than for the other estimator, we also have the same ranking for the true distribution. Q.E.D.

Proof of Theorem 6.3: Let $g_i(\beta) = q_i(y - x_i'\beta)$. Note that, by comparing the proof of Theorems 3.3 and 3.4, the M and ψ_i for GMM and GEL are identical. Also, in Lemma A6, $\bar{\Omega}_{\beta_j} = E[-2q_i q_i x_i \varepsilon_i] = 0$, so that $\psi_i^\Omega = g_i g_i' - \Omega$. It then follows that the $A(z)$ in the statement of Lemma A4 for GMM and GEL are identical to one another. Furthermore, it is straightforward to show that $M_j = 0$ for both GMM and GEL. Therefore, $Q_1(\tilde{\psi},\tilde{a})$

coincides for the two estimators. Let $\tilde{b}^{GMM}$, $\tilde{b}^{GEL}$, $\tilde{Q}_2^{GMM} = Q_2^{GMM}(\tilde{\psi}, \tilde{a}, \tilde{b}^{GMM})$, $\tilde{Q}_2^{GEL} = Q_2^{GEL}(\tilde{\psi}, \tilde{a}, \tilde{b}^{GEL})$ denote the second order terms for GMM and GEL respectively. From the form of Ξ given in Section 5 we see that the difference in higher-order variances for GMM and GEL estimators of β is

$$\Xi_{GMM} - \Xi_{GEL} = D + D', D = [I_p, 0] \lim_{n\to\infty} E[(\tilde{Q}_2^{GMM} - \tilde{Q}_2^{GEL})\tilde{\psi}'][I_p, 0]'.$$

Thus, it suffices to just calculate the difference of second-order terms. Furthermore, by $\tilde{A}$ and $\tilde{Q}_1$ identical for GMM and GEL, the first term in the formula for $\tilde{Q}_2$ in eq. (A.11) is identical for GMM and GEL. Also, for GEL $M_j = 0$ for all j so that we only have to calculate the last two terms in $\tilde{Q}_2$ for GEL, namely $\sum_{j=1}^{q} \tilde{\psi}_j \hat{B}_j \tilde{\psi}/2$ and $\sum_{j,k=1}^{q} \tilde{\psi}_j \tilde{\psi}_k M_{jk} \tilde{\psi}/6$. For GMM, and $B_j(z) = 0$ and $M_{jk} = 0$ (by linearity of $m_i(\theta)$ from the proof of Theorem 3.3). In addition for GMM, by efficiency of $\tilde{\beta}$, $\tilde{\psi}^{\beta}$ from Lemma A6 is equal to $[I_p, 0]\tilde{\psi}$, so that defining and although $\tilde{Q}_1^{\Omega} = \sum_{j=1}^{p} \tilde{\Omega}_{\beta_j} e_j' \tilde{\psi}^{\beta} + \sum_{j,k=1}^{p} \bar{\Omega}_{\beta_k\beta_j} \tilde{\psi}_j^{\beta} \tilde{\psi}_k^{\beta}/2$, we have

$$E[M^{-1} diag[0, \sum_{j,k=1}^{p} \bar{\Omega}_{\beta_k\beta_j} \tilde{\psi}_j^{\beta} \tilde{\psi}_k^{\beta}/2] \tilde{\psi} \tilde{\psi}_{\beta}'] = -H \sum_{j,k=1}^{p} \bar{\Omega}_{\beta_k\beta_j} E[\tilde{\psi}_j^{\beta} \tilde{\psi}_k^{\beta} \tilde{\psi}_{\lambda} \tilde{\psi}_{\beta}'] = O(n^{-2}),$$

where the last equality follows by existence of fourth moments of g_i and by $\hat{\beta}$ and $\hat{\lambda}$ having zero asymptotic covariance. Therefore, for M_{jk} and $\tilde{B}_j$ from GEL, we have, by $[I_p, 0]M^{-1} = [\Sigma, -H]$,

$$\begin{aligned} D &= D_1 + D_2, D_1 = \lim_{n\to\infty} E[\tilde{D}_1], D_2 = \lim_{n\to\infty} E[\tilde{D}_2], \\ \tilde{D}_1 &= [\Sigma, -H]\{\tilde{T}_{GMM} + \sum_{j=1}^{m+p} \tilde{\psi}_j \hat{B}_j \tilde{\psi}/2\}\tilde{\psi}_{\beta}', \tilde{T}_{GMM} = diag[0, \sum_{j=1}^{p} \tilde{\Omega}_{\beta_j} \tilde{\psi}_j^{\beta}]\tilde{\psi} \\ \tilde{D}_2 &= [\Sigma, -H]\{\sum_{j,k=1}^{m+p} \tilde{\psi}_j \tilde{\psi}_j M_{jk} \tilde{\psi}/6\}\tilde{\psi}_{\beta}'. \end{aligned}$$

Consider next $\tilde{D}_1$. Note that for $j \le p$ and $\Omega_{i\beta_j} = \partial[g_i(\beta_0)g_i(\beta_0)']/\partial\beta_j$ as defined above, from eq. (A.19), $\tilde{B}_j = -diag[0, \tilde{\Omega}_{\beta j}]$, so that

$$\sum_{j=1}^{p} \tilde{\psi}_j \tilde{B}_j \tilde{\psi}/2 = -diag[0, \sum_{j=1}^{p} \tilde{\Omega}_{\beta j} \tilde{\psi}_j^{\beta}]\tilde{\psi}/2 = -\tilde{T}_{GMM}/2.$$

Note also that for $j > 0$,

$$\sum_{j=1}^{m}\tilde{\psi}_j^{\lambda}\tilde{B}_{j+p}\left(\begin{array}{c} I_p \\ 0 \end{array}\right)\tilde{\psi}_{\beta}/2 = -\sum_{j=1}^{m}\tilde{\psi}_j^{\lambda}\left(\begin{array}{c} 0 \\ [\tilde{\Omega}_{\beta 1}e_j, ..., \tilde{\Omega}_{\beta p}e_j] \end{array}\right)\tilde{\psi}^{\beta} = -\tilde{T}_{GMM}/2.$$

Then for $\tilde{\psi}^{\lambda} = [0, I_m]\tilde{\psi}$,

$$\begin{aligned}
\tilde{T}_{GMM} + \sum_{j=1}^{m+p}\tilde{\psi}_j\hat{B}_j\tilde{\psi}/2 &= \tilde{T}_{GMM}/2 + \sum_{j=1}^{m}\tilde{\psi}_j^{\lambda}\tilde{B}_{j+p}\tilde{\psi}/2 \\
&= \sum_{j=1}^{m}\tilde{\psi}_j^{\lambda}\tilde{B}_{j+p}\left(\begin{array}{c} 0 \\ I_m \end{array}\right)\tilde{\psi}^{\lambda}/2 = -\sum_{j=1}^{m}\tilde{\psi}_j^{\lambda}\left(\begin{array}{c} 2\sum_i g_{ij}G_i'/\sqrt{n} \\ -\rho_3\sum_i g_{ij}g_ig_i'/\sqrt{n} \end{array}\right)\tilde{\psi}^{\lambda}/2.
\end{aligned}$$

Then using $g_i = q_i\varepsilon_i$, $G_i = -q_ix_i'$, and letting $\bar{x}_i = -G'\Omega^{-1}q_i\sigma_i^2$, and $K_i = q_i'Pq_i$, so that $\psi_i^{\beta} = \Sigma\bar{x}_i\varepsilon_i/\sigma_i^2$, it follows by $E[\tilde{\psi}^{\lambda}\tilde{\psi}^{\beta\prime}] = 0$ and fourth moments bounded that

$$\begin{aligned}
D_1 &= [\Sigma, -H]\lim_{n\to\infty}E[\sum_{j=1}^{m}\tilde{\psi}_j^{\lambda}\left(\begin{array}{c} 2\sum_i\varepsilon_iq_{ij}x_iq_i'/\sqrt{n} \\ \rho_3\sum_i\varepsilon_i^3q_{ij}q_iq_i'/\sqrt{n} \end{array}\right)\tilde{\psi}^{\lambda}\tilde{\psi}^{\beta\prime}]/2 \\
&= \Sigma\sum_{j=1}^{m}\{2E[\varepsilon_i^2q_{ij}x_iq_i'Pe_j\psi_i^{\beta\prime}] + \rho_3E[(\varepsilon_i^4/\sigma_i^4)q_{ij}\bar{x}_iq_i'Pe_j\bar{x}_i']\}\Sigma/2 \\
&= \Sigma\{E[K_ix_i\bar{x}_i'] + (\rho_3/2)E[(\mu_{4i}/\sigma_i^3)K_i\bar{x}_i\bar{x}_i']\}\Sigma.
\end{aligned}$$

Next, we have by linearity of $g_i(\beta)$ in β, eq. (A.22),

$$\begin{aligned}
\sum_{j,k=1}^{m}P_{jk}M_{p+j,p+k}[\Sigma, 0]' &= \sum_{j,k=1}^{m}P_{jk}E[\left(\begin{array}{c} -G_i'e_te_u'G_i - G_i'e_ue_t'G_i \\ \rho_3g_i(g_{it}e_u'G_i + g_{iu}e_t'G_i + g_{it}g_{iu}G_i) \end{array}\right)]\Sigma \\
&= \sum_{j,k=1}^{m}P_{jk}E[\left(\begin{array}{c} -2E[x_ix_i'q_{ij}q_{ik}] \\ -3\rho_3E[\sigma_i^2q_ix_iq_{ij}q_{ik}] \end{array}\right)]\Sigma \\
&= E[\left(\begin{array}{c} -2E[K_ix_ix_i'] \\ -3\rho_3E[\sigma_i^2K_iq_ix_i'] \end{array}\right)]\Sigma
\end{aligned}$$

Also, by eq. (A.21), $\sum_{j=1}^{m}q_{ij}Pe_j = Pq_i$, and $\sum_{k=1}^{p}x_{ik}e_k'\Sigma = x_i'\Sigma$,

$$\begin{aligned}
\sum_{j=1}^{m}\sum_{k=1}^{p}M_{p+j,k}\left(\begin{array}{c} 0 \\ P \end{array}\right)e_je_k'\Sigma &= -\sum_{j=1}^{m}\sum_{k=1}^{p}E[\left(\begin{array}{c} G_i'e_tg_i^{j\prime} + +G_{itj}G_i' \\ -\rho_3[G_{itj}g_ig_i' + g_{it}(g_i^jg_i + g_ig_i^{j\prime})] \end{array}\right)]Pe_je_k'\Sigma \\
&= \sum_{j=1}^{m}\sum_{k=1}^{p}E[\left(\begin{array}{c} -2E[x_iq_i'q_{ij}x_{ik}] \\ -3\rho_3E[\sigma_i^2q_iq_i'q_{ij}x_{ik}] \end{array}\right)]Pe_je_k'\Sigma \\
&= E[\left(\begin{array}{c} -2E[K_ix_ix_i'] \\ -3\rho_3E[\varepsilon_i^2K_iq_ix_i'] \end{array}\right)]\Sigma
\end{aligned}$$

Note that for $j,k \leq p$, for GEL in a linear model the upper left block of M_{jk} is zero, so that $M_{jk}[\Sigma, 0]' = 0$ and hence $M_{jk}E[\psi_i \psi_{ij}] = 0$. Also, For GEL we also have by standard V-statistic calculations and $M_{jk} = M_{kj}$

$$\begin{aligned} D_2 &= [\Sigma, -H] \lim_{n\to\infty} E[\sum_{j,k=1}^{m+p} \tilde{\psi}_j \tilde{\psi}_k M_{jk} \tilde{\psi}\tilde{\psi}^{\beta\prime}]/6 = [\Sigma, -H]\{\sum_{j,k=1}^{m+p} E[\psi_{ij}\psi_{ik}]M_{jk} \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} \\ &\quad +2\sum_{j,k=1}^{m+p} M_{jk}E[\psi_i \psi_{ij}]E[\psi_{ik}\psi_i^{\beta\prime}]\}/6 \\ &= [\Sigma, -H]\{\sum_{j,k=1}^{m} P_{jk} M_{j+p,k+p} \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} + 2\sum_{j=1}^{m}\sum_{k=1}^{p} M_{j+p,k} \begin{pmatrix} 0 \\ P \end{pmatrix} e_j e_k' \Sigma\}/6 \\ &= \Sigma\{-E[K_i x_i x_i'] - (3\rho_3/2)E[K_i \bar{x}_i x_i']\}\Sigma. \end{aligned}$$

Then summing D_1 and D_2 gives the first conclusion. For the second conclusion, note that $K_i \leq q_i'\Omega^{-1}q_i \leq Cq_i'(E[q_i q_i'])^{-1}q_i \leq C\zeta(m)^{1/2}$ for $\zeta(m) = \{\sup_{x\in X} q(x)'(E[q_i q_i'])^{-1}q(x)\}^2$. Then since $G'\Omega^{-1}$ are the population least squares coefficients from a regression of $x_i \backslash \sigma_i^2$ on q_i,

$$\begin{aligned} \|E[K_i(\bar{x}_i - x_i)']\|^2 &\leq E[K_i^2]E[\|\bar{x}_i - x_i\|^2] \leq C\zeta(m)E[\sigma_i^4\|G'\Omega^{-1}q_i - x_i/\sigma_i^2\|^2] \\ &\leq C\zeta(m)E[\sigma_i^2\|G'\Omega^{-1}q_i - x_i/\sigma_i^2\|^2] \leq C\zeta(m)E[\sigma_i^2\|\gamma_m q_i - x_i/\sigma_i^2\|^2] \to 0. \end{aligned}$$

It follows similarly that $E[K_i\bar{x}_i(\bar{x}_i - x_i)'] \to 0$ and $E[(\mu_{4i}/\sigma_i^4 - 3)K_i x_i(\bar{x}_i - x_i)] \to 0$, giving the second conclusion. Q.E.D..

References

Abowd, J. and D. Card (1989), "On the Covariance Structure of Earnings and Hours Changes," *Econometrica* 57, 411-445.

Altonji, J. and L.M. Segal (1996): "Small Sample Bias in GMM Estimation of Covariance Structures," *Journal of Economic and Business Statistics* 14, 353-366.

Amemiya, T. (1983), "Partially Generalized Least Squares and Two-Stage Least Squares Estimators," *Journal of Econometrics* 23, 275-283.

Back, K., and D. Brown (1993): "Implied Probabilities in GMM Estimators," *Econometrica* 61, 971-976.

Brown, B.W. and W.K. Newey (1992): "Bootstrapping for GMM", *mimeo*, M.I.T.

Brown, B.W. and W.K. Newey (1998): "Efficient Semiparametric Estimation of Expectations," *Econometrica* 66, 453-464.

Chamberlain, G. (1982): "Multivariate Regression Models for Panel Data," *Journal of Econometrics 18, 5-42.*

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

Corcoran, S.A. (1998): "Bartlett Adjustment of Empirical Discrepancy Statistics", *Biometrika* 85, 967-972.

Cragg, J.G. (1983): "More Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form," *Econometrica* 51, 751-763.

Cressie, N., and T. Read (1984): "Multinomial Goodness-of-Fit Tests", *Journal of the Royal Statistical Society*, *Series B*, 46, 440-464.

Donald, S.G. and W.K. Newey (2000): "A Jacknife Interpretation of the Continuous Updating Estimator," *Economics Letters*, forthcoming.

Hahn, J., J.A. Hausman, and G. Kuersteiner (2001): "Bias Corrected Instrumental Variables Estimators for Panel Data," working paper, MIT.

Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica* 50, 1029-1054.

Hansen, L.P. and K.J. Singleton. (1982): "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica* 50, 1269-1286.

Hansen, L.P., J. Heaton and A. Yaron (1996): “Finite-Sample Properties of Some Alternative GMM Estimators”, *Journal of Business and Economic Statistics* 14, 262-280.

Holtz-Eakin, D., W.K. Newey, and H.S. Rosen (1988), ”Estimating Vector Autoregressions With Panel Data,” *Econometrica* 56, 1371-1396.

Horowitz, J.L. (1998): ”Bootstrap Methods for Covariance Structures,” *Journal of Human Resources* 33, 38-61.

Imbens, G.W. (1997): ”One-Step Estimators for Over-Identified Generalized Method of Moments Models,” *Review of Economic Studies* 64, 359-383.

Imbens, G.W., R.H. Spady and P. Johnson (1998): ”Information Theoretic Approaches to Inference in Moment Condition Models,” *Econometrica* 66, 333-357.

Imbens, G.W. (2000): ”GMM Estimation with Many Moments,” mimeo.

Kezdi, G., J. Hahn, and G. Solon (2001): ”Jacknife Minimum Distance Estimation,” mimeo.

Kitamura, Y., and M. Stutzer (1997): “An Information-Theoretic Alternative to Generalized Method of Moments Estimation”, *Econometrica* 65, 861-874.

Koenker, R., J.A.F. Machado, C.L. Skeels, and A.H.I. Welsh (1992): ”Momentary Lapses: Moment Expansions and the Robustness of Minimum Distance Estimators,” mimeo.

Nagar, A.L. (1959): ”The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations”, *Econometrica* 27, 573-595.

Newey, W.K. (1997): ”Convergence Rates and Asymptotic Normality for Series Estimators,” *Journal of Econometrics* 79, 147-168.

Newey, W.K. and D. McFadden (1994): ”Large Sample Estimation and Hypothesis Testing,” in Engle, R. and D. McFadden, eds., *Handbook of Econometrics, Vol. 4,* New York: North Holland.

Owen, A. (1988): "Empirical Likelihood Ratio Confidence Intervals for a Single Functional," *Biometrika* 75, 237-249.

Pfanzagl, J. and W. Wefelmeyer (1978): "A Third-Order Optimum Property of the Maximum Likelihood Estimator," *Journal of Multivariate Analysis 8, 1-29,*

Qin, J. and Lawless, J. (1994): "Empirical Likelihood and General Estimating Equations", *Annals of Statistics* 22, 300-325.

Rao, C.R. (1963): "Criteria of Estimation in Large Samples," *Sankhya, Series A* 25, 189-206.

Rao, C.R. (1973): *Linear Statistical Inference and Its Applications*, New York: Wiley.

Rilstone, P., V.K. Srivastava, and A. Ullah (1996): "The Second-order Bias and Mean-Squared Error of Nonlinear Estimators," *Journal of Econometrics* 75, 369-395.

Rothenberg, T.J (1984): "Approximating the Distributions of Econometric Estimators and Test Statistics," in Griliches, Z. and M.D. Intriligator, eds., *Handbook of Econometrics, Vol. 2*, New York: North-Holland.

Rothenberg, T.J. (1999): "Higher Order Properties of Empirical Likelihood for Simultaneous Equations," seminar notes, Harvard-MIT econometrics seminar.

Smith, R. J. (1997): "Alternative Semi-Parametric Likelihood Approaches to Generalized Method of Moments Estimation", *Economic Journal* 107, 503-519.

Smith, R. J. (1999): "Generalized Empirical Likelihood Methods for Moment Condition Models", *mimeo*, University of Bristol. Revision of paper presented at Cowles Foundation Econometrics Conference on "New Developments in Time Series Econometrics", Yale University, 1999.

Stock, J.H., and J. Wright (2000): "GMM with Weak Identification", *Econometrica*, forthcoming.