Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators

Whitney K. Newey
Richard J. Smith

THE INSTITUTE FOR FISCAL STUDIES
DEPARTMENT OF ECONOMICS, UCL
cemmap working paper CWP04/03

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 Warwick

First Draft: January 2000
Final Version, June 2003

Abstract

In an effort to improve the small sample properties of generalized method of moments (GMM) estimators, a number of alternative estimators have been suggested. These include 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 GEL has no asymptotic bias due to correlation of the moment functions with their Jacobian, eliminating an important source of bias for GMM in models with endogeneity. We also find that EL has no asymptotic bias from estimating the optimal weight matrix, eliminating a further important source of bias for GMM in panel data models. We give bias corrected GMM and GEL estimators. We also show that bias corrected EL inherits the higher order property of maximum likelihood, that it is higher order 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 a fellow at the Center for Advanced Study in the Behavioral Sciences. The NSF under Grant No. 0136869 and the ESRC under Grant No. R000237386 provided partial financial support for Newey and Smith respectively. We thank J. Horowitz, G. Imbens, L. Hansen, T. Rothenberg, P. Guggenberger and three referees for helpful comments. I. Fernandez provided capable research assistance. Aspects of this research have been presented at the 2002 Winter Econometric Society Meetings, the 2002 Econometric Society European Meetings, the 2000 EC^2 Conference on "Likelihood in Econometrics", an ESRC Econometric Study Group meeting, the 2001 NSF-Berkeley Symposium on "Identification and Inference for Econometric Models", the 2001 Canadian Econometrics Study Group Conference, University of Waterloo, and seminars at Amsterdam, Bristol, Copenhagen, Chicago, Harvard-MIT, Iowa, LSE, Northwestern, Rome, Tilburg and Warwick.

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 as are estimators from the Cressie and Read (1984) power divergence family of discrepancies. All of these estimators and GMM have the same asymptotic distribution but different higher order asymptotic properties. We use the GEL structure, which helps simplify calculations and comparisons, 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 GMM often does. Consequently, with many moment conditions the bias of EL will be less than the bias of GMM. 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. 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 order efficient relative to other bias corrected estimators. This property has a simple explanation. When the data are discrete, having finite support, the nonparametric (one probability per observation, unknown cells) EL estimator is asymptotically equal to the parametric (one probability per cell) 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 the MLE (e.g. see Rao, 1963 and Pfanzagl and Wefelmeyer, 1978). Then, since nothing in the higher order variance depends on discreteness, this result extends to any distribution. More precisely, by Lemma 3 of Chamberlain (1987), we can find a discrete distribution that matches all the moments that make up the higher order variances of any two estimators, so the efficiency of bias corrected EL for the discrete distribution implies efficiency for the true one.

Although the small bias property of EL is nice, there are methods of removing all of the asymptotic bias. These include the bootstrap, as in Horowitz (1998) for GMM, the jackknife, 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.

It is also of interest to compare higher order efficiency when the full bias correction is not used, so that EL need not be higher order efficient. We do this for estimators that improve asymptotic efficiency (relative to least squares) under unknown heteroskedasticity, as considered in Amemiya (1983), Chamberlain (1982), and Cragg (1983). We impose auxiliary assumptions that give zero bias for GMM and GEL, even though the estimated bias corrections are not zero (so that EL is not higher order efficient), and compare higher order variances. We find that with a Gaussian disturbance GEL and GMM have the same higher order variances with many moments, but that with conditional kurtosis, GMM is efficient relative to EL with thick tailed errors whereas EL is better with thin tailed errors. This provides an example where there is no bias concern, the only issue being efficiency, and where EL may not be best in terms of MSE.

Some previous work on higher order properties of these estimators has been done. Koenker et al. (1994) and Rilstone, Srivastava, 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 (1996) showed that for a single equation of a Gaussian, homoskedastic linear simultaneous equations model the asymptotic bias of EL is the same as the limited information MLE and that a bias corrected EL is higher order efficient relative to a bias corrected GMM estimator. Imbens and Spady (2001) showed in a special model that the higher order MSE for GMM grows faster with the number of moment conditions than for EL. 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 β, 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

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

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, 2001). 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 $\hat{\Lambda}_n(\beta) = \{\lambda : \lambda' g_i(\beta) \in \mathcal{V}, i = 1, ..., n\}$. The estimator is the solution to a saddle point problem

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

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 (ET) 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 does not affect the estimator of β. It is satisfied by the $\rho(v)$ given above for EL and ET.

The continuous updating estimator (CUE) of Hansen, Heaton, and Yaron (1996) is also a GEL estimator as we now show. 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 $AA^{-}A = A$.[1] The following 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

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

where $\hat{\lambda} = \arg\max_{\lambda\in\hat{\Lambda}_n(\hat{\beta})}\sum_{i=1}^{n}\rho(\lambda'\hat{g}_i)/n$. The empirical probabilities $\hat{\pi}_i$, $(i = 1, ..., n)$, 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 conditions 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); see also Smith (1997). 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 helps explain the form of the probabilities in equation (2.4) and connects our results with the existing literature. Let $h(\pi)$ be a convex function of a scalar π, and consider the estimator

$$\bar{\beta} = \arg\min_{\beta\in\mathcal{B},\pi_1,...,\pi_n}\sum_{i=1}^{n} h(\pi_i), s.t. \sum_{i=1}^{n}\pi_i g_i(\beta) = 0, \sum_{i=1}^{n}\pi_i = 1. \tag{2.5}$$

[1]The CUE of Hansen, Heaton, and Yaron (1996) actually minimizes $\tilde{Q}(\beta) = \hat{g}(\beta)'[\hat{\Omega}(\beta) - \hat{g}(\beta)\hat{g}(\beta)']^{-1}\hat{g}(\beta)$ rather than $\hat{Q}(\beta) = \hat{g}(\beta)'\hat{\Omega}(\beta)^{-1}\hat{g}(\beta)$, but equality of the two estimators follows by $\tilde{Q}(\beta) = \hat{Q}(\beta)/[1 - \hat{Q}(\beta)]$.

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 in which $h(\pi) = [\gamma(\gamma+1)]^{-1}[(n\pi)^{\gamma+1} - 1]/n$. To describe this result, note that the Lagrangian for MD is

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

where α is an $m \times 1$ vector of Lagrangian 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 $\mathcal{B}$, $\hat{\lambda}$ 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 $\hat{\lambda} = \bar{\alpha}$ for $\gamma = 0$.

The duality between MD and GEL estimators is known for EL ($\gamma = -1$, Qin and Lawless, 1994) and for ET ($\gamma = 0$, Kitamura and Stutzer, 1997), but is new for the CUE ($\gamma = 1$) as well as for all the other members of the Cressie and Read (1984) family. Duality is useful because it shows how the computationally less complex GEL estimators are related to MD estimators of the Cressie-Read family, which has become a common standard for comparison in the empirical likelihood literature (e.g. see Owen, 2001). Also, duality justifies the $\hat{\pi}_i$ in equation (2.4) as MD estimates, which aids the interpretation of the first order conditions of the estimators.

2.2 The First Order Conditions

Some interpretations of the first order conditions are useful for understanding our asymptotic bias results. Let $G_i(\beta) \equiv \partial g_i(\beta)/\partial\beta$. The GMM first order conditions imply

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

We can also obtain an analogous expression for any GEL estimator $\hat{\beta}$.[2] Let $k(v) = [\rho_1(v)+1]/v, v \neq 0$ and $k(0) = -1$. Also, let $\hat{v}_i = \hat{\lambda}'\hat{g}_i$, $\hat{k}_i = k(\hat{v}_i)/\sum_{j=1}^n k(\hat{v}_j)$, and $\hat{\pi}_i$ be as given in equation (2.4).

Theorem 2.3: *The GEL first order conditions imply*

$$[\sum_{i=1} \hat{\pi}_i G_i(\hat{\beta})]'[\sum_{i=1} \hat{k}_i g_i(\hat{\beta}) g_i(\hat{\beta})']^{-1}\hat{g}(\hat{\beta}) = 0,$$

where $\hat{k}_i = \hat{\pi}_i$ *for EL and* $\hat{k}_i = 1/n$ *for CUE.*

In comparing the GMM and GEL 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 while GEL uses an efficient estimator of the Jacobian term. Also EL uses an efficient estimator of the second moments, CUE uses the sample average, and other GEL estimators use other weighted averages. An important property of efficient moment estimators is that they are asymptotically uncorrelated with $\hat{g}(\hat{\beta})$, 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 GEL there will be no asymptotic bias from estimation of the Jacobian and, furthermore, for EL there will also be no asymptotic bias from estimating the second moments.[3] We will also see that the absence of second moment bias holds for any GEL estimator with $\rho_3 = -2$, which can be explained by the fact that $k(v) = \rho_1(v) + o(v)$ in this case, and hence $\hat{k}_i$ is approximately equal to $\hat{\pi}_i$.

[2] Bonnal and Renault (2001) independently obtained a similar result for the CUE.

[3] Donald and Newey (2000) previously discussed the absence of Jacobian bias for the CUE.

3 Stochastic Expansion

We find the asymptotic bias and higher order variance using 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 those 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 moment tests, as in Imbens, Spady, and Johnson (1998); see also Smith (1997, 2001).

Consistency and asymptotic normality are important prerequisites for stochastic expansions, so we first briefly consider these properties for any GEL estimator $\hat{\beta}$. We make use of the following identification and regularity condition. 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 the existence of slightly higher moments than consistency for two step efficient GMM, as in Hansen (1982), but otherwise is the same.

Theorem 3.1: *If Assumption 1 is satisfied then* $\hat{\beta} \xrightarrow{p} \beta_0$, $\hat{g}(\hat{\beta}) = O_p(n^{-1/2})$, $\hat{\lambda} = \arg\max_{\lambda \in \hat{\Lambda}_n(\hat{\beta})} \sum_{i=1}^n \rho(\lambda' g_i(\hat{\beta}))/n$ *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 the 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 $\Sigma = (G'\Omega^{-1}G)^{-1}$, $H = \Sigma G'\Omega^{-1}$, 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 GEL estimators, and that, properly normalized, the saddle point objective function has a limiting chi square 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 $0 \le j \le 4$ and all z, $\nabla^j g(z,\beta)$ 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 $\beta \in \mathcal{N}$, $\|\nabla^4 g(z,\beta) - \nabla^4 g(z,\beta_0)\| \le b(z)\|\beta - \beta_0\|$, $\rho(v)$ is four times continuously differentiable with Lipschitz fourth derivative in a neighborhood of zero.

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

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

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

$$-[\sum_{i=1}^{n} G_i(\hat{\beta}_{GMM})/n]'\hat{\lambda}_{GMM} = 0, -\hat{g}(\hat{\beta}_{GMM}) - \hat{\Omega}(\tilde{\beta})\hat{\lambda}_{GMM} = 0. \tag{3.2}$$

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.*

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

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

Expressions for each of the terms in the expansions of Theorems 3.3 and 3.4 are given in the respective proofs of these results because they are quite complicated. Implicit in this result for GMM is that the expansion depends on the preliminary estimator $\tilde{\beta}$ only through the limit W and influence function $\xi(z_i)$. For example, all efficient GMM estimators that have been iterated at least twice, so that $\hat{W} = \hat{\Omega}(\bar{\beta})$ and $\bar{\beta}$ is itself an efficient GMM estimator, have the same expansion. Also, similarly to Pfanzagl and Wefelmeyer (1978), Rothenberg (1984), and Robinson (1988), after three iterations that start at an initial $\sqrt{n}$-consistent estimator, numerical procedures for solving the GEL first order conditions will produce an estimator with the same leading three terms in the expansion of equation (3.1).

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 $O(n^{-2})$. To describe the precise form of the bias we need some additional notation. Let $H_W = (G'W^{-1}G)^{-1}G'W^{-1}$, $\bar{\Omega}_{\beta_j} = E[\partial\{g_i(\beta_0)g_i(\beta_0)'\}/\partial\beta_j]$, a be an $m \times 1$ 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 $g_i(\beta)$, and e_j the jth unit vector. For GMM we have the following result:

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

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

Each of the terms has an interesting interpretation. 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 as in minimum distance estimation, see Section 4.2, $B_G = 0$ but B_G is generally nonzero whenever 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 this case. The term B_W arises from the choice of preliminary estimator. It is zero if W is a scalar multiple 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.

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 preliminary estimator.

The absence of any bias from the preliminary estimator is to be expected from the one step nature of the GEL 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 Jacobian term is an important source of bias in IV estimators. Because the GMM bias includes such effects but GEL does not, we expect that GEL will have relatively small bias in such 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 optimal minimum distance. Given that the EL bias does not include this effect we expect that it will have relatively small bias for minimum distance. We can verify this intuition in some specific models.

4.1 Conditional Moment Restrictions

One important model, that is useful for considering IV estimation, involves a conditional moment restriction. To describe this model let $u(z, \beta)$ be a scalar residual satisfying

$$E[u(z_i, \beta_0)|x_i] = 0. \tag{4.6}$$

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

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

Theorem 4.5: *If Assumptions 1 - 4 and equation* (4.6) *are satisfied and* $\hat{\beta}_{GMM}$ *a GMM estimator with* $W = \Omega$,

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

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* $q(x)$

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

Here $\inf\{e_j'\delta_i\} = \sup\{C : \Pr(e_j'\delta_i \geq C) = 1\}$. In the general heteroskedastic case, we find that the asymptotic bias of GMM grows linearly with the number of overidentifying restrictions when $\inf\{e_j'\delta_i\} = \sup\{C : \Pr(e_j'\delta_i \geq C) = 1\} > 0$, 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 $\sup\{e_j'\delta_i\} < 0$. Donald, Imbens, and Newey (2002) show that these comparisons between asymptotic biases are also correct when m is allowed to grow with the sample size. If the preliminary estimator $\tilde{\beta}$ is inefficient, the additional term $B_W = 2\Sigma E[\bar{d}_i q_i'(H_W - H)'\kappa_i]/n$ should be included in $Bias(\hat{\beta}_{GMM})$.

An important special case is a homoskedastic linear model, where $u_{\beta\beta i} = 0$ and $\kappa_i = \kappa$, $\sigma_i^2 = \sigma^2$, $\mu_{3i} = \mu_3$ are constants. Here $Bias(\hat{\beta}_{EL}) = -\Sigma\kappa/\sigma^2$, which is the same as the bias of the Gaussian limited information MLE, as shown by Rothenberg (1996) for Gaussian disturbances. Also, when $\mu_3 = 0$, $Bias(\hat{\beta}_{GEL}) = Bias(\hat{\beta}_{EL})$ and $Bias(\hat{\beta}_{GMM}) = (m - p - 1)\Sigma\kappa/\sigma^2$, which is the Nagar (1959) bias of two stage least squares. When $\mu_3 \neq 0$ all the estimators, except for EL and GEL with $\rho_3 = -2$, have an additional bias term from estimating the weight matrix.

4.2 Minimum Distance Estimation

The second example is classical 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 previous model, when $h(\beta)$ can be interpreted 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 $\mathcal{N}$ 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 4.6: *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 or from the preliminary estimator $\tilde{\beta}$ as $\bar{\Omega}_{\beta_j} = 0$, $(j = 1, ..., p)$. 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 $h(\beta) = \beta\iota$, where ι is an $m \times 1$ vector of units. 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 $G = -\iota$, so that $\Sigma = \sigma^2/m$ and $P = (I_m - \iota\iota'/m)/\sigma^2$. It follows that

$$\begin{aligned} Bias(\hat{\beta}_{EL}) &= 0, Bias(\hat{\beta}_{GMM}) = Bias(\hat{\beta}_{CUE}) = \left(\frac{m-1}{m}\right)\left(\frac{\mu_3}{\sigma^2}\right)/n, \\ \frac{Bias(\hat{\beta}_{GMM})}{\sqrt{\Sigma}} &= \sqrt{m}\left(\frac{m-1}{m}\right)\left(\frac{\mu_3}{\sigma^3}\right)/n. \end{aligned}$$

Here the bias of GMM relative to its asymptotic standard error grows with the square root of the number of overidentifying restrictions. Dividing by the standard error is an appropriate normalization, since it goes to zero as m grows.

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, jackknife, and analytical methods. Here we use an analytical approach, bias correcting GMM and GEL using the asymptotic bias formulae we have derived. These bias corrections are much simpler computationally than the bootstrap or jackknife methods, particularly in nonlinear models. They can be constructed using the same ingredients as the estimator of Σ along with the second derivatives of the moment indicators.

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 estimator, nor the higher order asymptotic variance. It can be shown that the effect of using the GEL probabilities, rather than empirical distribution, enters only through the appearance in Q_2 of a linear combination of $\sqrt{n}\hat{g}(\hat{\beta})$, and that $\sqrt{n}\hat{g}(\hat{\beta})$ is asymptotically uncorrelated with $\tilde{\psi}$. Consequently, since Q_2 enters the higher order variance only through its asymptotic correlation with $\tilde{\psi}$, see Section 6, using a GEL estimator of the distribution has no effect on the higher order variance of $\hat{\beta}$ (although it will on $\hat{\lambda}$).

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

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

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

$$\begin{aligned}
\widehat{Bias}(\hat{\beta}_{GMM}) &= [-\hat{H}(\hat{a} + \sum_{i=1} \hat{G}_i\hat{\psi}_i^\beta/n) \\
&\quad -\hat{\Sigma}\sum_{i=1} \hat{G}_i'\hat{P}\hat{g}_i/n - \sum_{i=1} \hat{\psi}_i^\beta \hat{g}_i'\hat{P}\hat{g}_i/n - \hat{H}\sum_{j=1} \hat{\Omega}_{\beta_j}(\hat{H}_W - \hat{H})'e_j]/n.
\end{aligned}$$

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 formulae, 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 empirical probabilities, and

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

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

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

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

The precision of different estimators can be compared based on their higher order MSE, given by

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

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 $o(n^{-1})$ are dropped. Here the term Ξ is the additional, n^{-1} variance term for $\sqrt{n}(\hat{\theta} - \theta_0)$. One estimator is higher order efficient relative to another if its MSE matrix is smaller than that of the other, in the positive semidefinite sense. This property is often referred to in the literaure as third order efficiency, motivated by the presence of three terms in the expansion of equation (3.1) (see e.g. Pfanzagl and Wefelmeyer, 1978). In general, although they may be derived relatively straightforwardly from the Appendix, the expressions for Ξ for GMM and GEL are extremely complicated, and so are not given here, although some comparisons can be made.

It turns out that bias corrected EL is third or higher order efficient relative to other bias corrected GMM or GEL estimators, in the sense that Ξ is smaller for EL. An explanation of this result was given in the introduction. Here we give a rigorous proof. Let Ξ_{EL} denote the higher order variance of bias corrected EL.

Theorem 6.1: *If Assumptions 1-4 are satisfied and Ξ is the higher order variance of any bias corrected GEL or GMM estimator with $\hat{W} = \hat{\Omega}(\tilde{\beta})$ and $\tilde{\beta}$ an efficient GMM estimator, then $\Xi - \Xi_{EL}$ is positive semi-definite.*

The third order efficiency of EL will be shared by any GEL estimator for which $\rho_3 = -2$ and $\rho_4 = -6$, because they all have the same expansion (3.1) as EL. Rothenberg (1996) showed the third order efficiency of a bias corrected EL versus a bias corrected GMM in the linear case of equation (4.5) with Gaussian disturbances. These higher order variance comparisons correspond to a quadratic loss function. As shown by Pfanzagl and Wefelmeyer (1978), the MLE for discrete data is also third order efficient for a wide class of quasi-convex loss functions satisfying the smoothness condition of their Theorem $1'$. Consequently, it can also be shown that EL is higher order efficient for any such loss function.

The higher order efficiency of EL only holds among bias corrected estimators. If the bias corrections are dropped, then EL may not have the smallest MSE. Intuitively, the estimated bias corrections from Section 5 are asymptotically correlated with $\tilde{\psi}$, so dropping them may change the higher order MSE ranking of Theorem 6.1. Estimators of parametric models are known to behave analogously. For instance, Amemiya (1980) showed that in logit models the higher order MSE of a minimum chi square estimator is smaller than that of maximum likelihood for a wide range of parameter values.

As an example of MSE comparisons of estimators without bias correction we consider heteroskedastic linear regression, a special case of that in Section 4.1. The model is

$$y_i = x_i'\beta_0 + u_i, E[u_i|x_i] = 0. \tag{6.1}$$

Amemiya (1983), Chamberlain (1982), and Cragg (1983) considered GMM estimators that are more efficient than least squares, based on moment indicators $g(z, \beta) = q(x)(y - x'\beta)$, where $q(x)$ includes x . For these moment indicators we compare the higher order variance of GMM with GEL without bias correction. We also assume that $E[u_i^3|x_i] = 0$, which implies (together with $E[u_i|x_i] = 0$) that GMM and GEL have no asymptotic bias. However, EL need not be higher order efficient, because omitting the estimated bias corrections affects the ranking of Theorem 6.1. Intuitively, $E[u_i^3|x_i] = 0$ generally does not hold for either the empirical distribution or the empirical likelihood $\hat{\pi}_i$ distribution, so that estimated bias corrections are non-zero. Dropping them therefore will also change the higher order variances.

Let $\sigma_i^2 = E[u_i^2|x_i]$, $\mu_{4i} = E[u_i^4|x_i]$, $\bar{x}_i = -G'\Omega^{-1}q_i\sigma_i^2 = E[\sigma_i^2(x_i/\sigma_i^2)q_i']\{E[\sigma_i^2 q_i q_i']\}^{-1}q_i\sigma_i^2$ and $K_i = q_i'Pq_i$.

Theorem 6.2: *If Assumptions 1-4 are satisfied,* $\hat{W} = \hat{\Omega}(\tilde{\beta})$, *and* $\tilde{\beta}$ *is an optimal GMM estimator then*

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

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 $m \to \infty$*,*

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

This result gives an explicit formula for the difference of higher order variances 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, which by Newey (1997) will hold for cubic splines if the density of x is bounded away from zero, the support of X is a rectangle, knots are evenly spaced, and σ_i^2 is twice differentiable in x_i. Koenker et al. (1994) also calculate the higher order variance of GMM, for a different choice of $\tilde{\beta}$ in $\hat{W} = \hat{\Omega}(\tilde{\beta})$. Our contribution is to compare GMM with GEL.

The limit result 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 limit higher order variance as GMM. For EL and ET, $\rho_3 < 0$ so that they have smaller limit higher order variance than GMM 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 $\mu_{4i} > 3\sigma_i^4$. In the latter case, GEL estimators with $\rho_3 > 0$ have smaller limit higher order variance than GMM.

Recently, Donald, Imbens, and Newey (2002) have carried out an analogous comparison when there is endogeneity, but still with zero conditional skewness given the instruments. They find that when m is allowed to increase with the sample size, the MSE of GMM generally exceeds that of GEL for large enough sample size. This occurs because the squared bias from Section 4 grows with m^2, whereas the variance grows only with m. They also find that the CUE has smaller higher order variance than a bias corrected GMM which only corrects for B_G. Furthermore, the higher order efficiency ranking among GEL estimators is similar to that from Theorem 6.2, with EL being higher order less efficient for thick tailed disturbances.

7 Conclusion

The usefulness of higher order bias and variance results depends on how well they help to explain finite sample properties of estimators. There are now several Monte Carlo experiments that are consistent with our results. For conditional moment restriction models, Hansen, Heaton, and Yaron (1996) found that the CUE had smaller bias than GMM, and that iterating on the preliminary estimator $\tilde{\beta}$ used to form the weighting matrix had little effect on bias. For IV estimation of a Gaussian linear equation, Ramalho (2001) and Judge and Mittlehammer (2001) found that, with several instruments, EL and ET have similar, lower bias than GMM. These findings are consistent with Theorem 4.5, which shows lower asymptotic bias for GEL when there are several instruments and zero third moments.

For minimum distance estimation in panel data models Imbens (1997) found that EL had smaller bias than GMM. Newey, Ramalho and Smith (2001) obtained similar bias results for EL and GMM and also reported that the bias of the CUE differed little from that of GMM. Moreover, the bias of ET although an improvement over GMM exceeded that of EL. These findings are consistent with the relatively small bias of EL found in Theorem 4.6. Newey, Ramalho, and Smith (2001) also found that for large enough sample size EL generally had smaller variance than a bias corrected GMM, consistent with the higher order efficiency of EL found in Theorem 6.1.

Overall, the theory in this paper, when coupled with existing Monte Carlo results, suggests some prescriptions for applied work. For IV estimation with many instruments of a single equation where bias from the estimating the weighting matrix is not important, GEL estimators should all have smaller bias than GMM. As yet, the Monte Carlo evidence provides little guidance on which GEL estimator to use, although the recent theoretical work of Donald, Imbens, and Newey (2002) for IV estimation shows that the CUE has smaller higher order variance than bias corrected GMM while EL and ET may not. In minimum distance estimation of panel data models, where bias from estimation of the weighting matrix can be a serious problem, the EL estimator has especially good properties. It eliminates the bias from estimation of the weighting matrix, and after correcting for bias arising from nonlinearity, is higher order efficient relative to bias corrected GMM. Thus, for both IV and minimum distance estimation, the theoretical and Monte Carlo work to date suggest that GEL estimation should be considered as an alternative to GMM in applied work.

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-Schwarz, 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 Lindeberg-Lévy 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 $n \times 1$ 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 choice of generalized inverse as is the CUE objective function $\iota' A(A'A)^{-}A'\iota/n$. Also, $A'A(A'A)^{-}A' = A'$. By $\rho(v)$ quadratic, a second order Taylor expansion is exact, giving

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

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

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

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 $\bar{\pi}_i = [\gamma(\bar{\mu} + \bar{\alpha}' g_i(\bar{\beta}))]^{1/\gamma}/n$. The other MD first order conditions are $\sum_{i=1}^{n} \bar{\pi}_i = 1$ and, for $G_i(\beta) = \partial g_i(\beta)/\partial\beta$,

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

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 theorem 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

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

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 $\sum_{i=1}^{n} \bar{\pi}_i = 1$,

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

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 Lagrangian, the first order conditions for $\bar{\pi}_i$ are $1 + \ln(\bar{\pi}_i) = \bar{\mu} + \bar{\alpha}' g_i(\bar{\beta})$. Solving,

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

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

Proof of Theorem 2.3: Let $\hat{G}_i = G_i(\hat{\beta})$ and $\hat{g}_i = g_i(\hat{\beta})$. By eq. (A.4) and the definition of $k(v)$,

$$0 = \sum_{i=1} \rho_1(\hat{v}_i)\hat{g}_i = \sum_{i=1} [\rho_1(\hat{v}_i) + 1]\hat{g}_i - n\hat{g}(\hat{\beta}) = \sum_{i=1} k(\hat{v}_i)\hat{g}_i\hat{g}_i'\hat{\lambda} - n\hat{g}(\hat{\beta}).$$

Solving for $\hat{\lambda}$, plugging into the first part of eq. (A.4), and multiplying by $\sum_{j=1}^{n} k(\hat{v}_j)/n$ gives the first result. Note that for EL $k(v) = [-(1 - v)^{-1} + 1]/v = -(1 - v)^{-1} = \rho_1(v)$ and for CUE $k(v) = [-(1 + v) + 1]/v = -1$ is constant. Q.E.D.

Let $b_i = \sup_{\beta \in \mathcal{B}} \|g_i(\beta)\|$.

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

Proof: By Assumption 1 it follows by M that $\max_{1\le i\le n} b_i = O_p(n^{1/\alpha})$. Then by CS,

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

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 $\bar\lambda = \arg\max_{\lambda\in\hat\Lambda_n(\bar\beta)} \hat P(\bar\beta,\lambda)$ exists w.p.a.1, $\bar\lambda = O_p(n^{-1/2})$, and $sup_{\lambda\in\hat\Lambda_n(\bar\beta)}\hat P(\bar\beta,\lambda) \le \rho_0 + O_p(n^{-1})$.*

Proof: By UWL $\bar\Omega \overset{def}{=} \hat\Omega(\bar\beta) \xrightarrow{p} \Omega$. 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 A1. By Lemma A1 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 $\tilde\lambda = \arg\max_{\lambda\in\Lambda_n} \hat P(\bar\beta,\lambda)$ 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 A1 and $\rho_2 = -1$, $\max_{1\le i\le n} \rho_2(\dot\lambda'\bar g_i) < -1/2$ 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 $\bar g \overset{def}{=} \hat g(\bar\beta)$,

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

Subtracting $\rho_0 - C\|\tilde\lambda\|^2$ from both sides and dividing by $\|\tilde\lambda\|$ we find that $C\|\tilde\lambda\| \le \|\bar g\|$, 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 $\partial\hat P(\bar\beta,\tilde\lambda)/\partial\lambda = 0$, the first order conditions for an interior maximum. By Lemma A1, w.p.a.1 $\tilde\lambda \in \hat\Lambda_n(\bar\beta)$, so by concavity of $\hat P(\bar\beta,\lambda)$ and convexity of $\hat\Lambda_n(\bar\beta)$ it follows that $\hat P(\bar\beta,\tilde\lambda) = \sup_{\lambda\in\hat\Lambda_n(\bar\beta)} \hat P(\bar\beta,\lambda)$, giving the first and second conclusions with $\bar\lambda = \tilde\lambda$. Then by the last inequality of above equation, $\|\bar g\| = O_p(n^{-1/2})$, and $\|\bar\lambda\| = O_p(n^{-1/2})$, we obtain $\hat P(\bar\beta,\bar\lambda) \le \rho_0 + \|\bar\lambda\|\|\bar g\| - C\|\bar\lambda\|^2 = \rho_0 + O_p(n^{-1})$. Q.E.D.

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

Proof: Let $\hat g_i = g_i(\hat\beta)$, $\hat g = \hat g(\hat\beta)$, and for ζ in Lemma A1, $\tilde\lambda = -n^{-\zeta}\hat g/\|\hat g\|$. By Lemma A1, $\max_{i\le n} |\tilde\lambda'\hat g_i| \xrightarrow{p} 0$ and $\tilde\lambda \in \hat\Lambda_n(\hat\beta)$ 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 UWL, $\sum_i \hat{g}_i\hat{g}_i'/n \leq (\sum_i b_i^2/n)I \xrightarrow{p} CI$, 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 A2 are satisfied by $\bar{\beta} = \beta_0$. By $\hat{\beta}$ and $\hat{\lambda}$ being a saddle point, this equation and Lemma A2 give

$$\rho_0 + n^{-\zeta}\|\hat{g}\| - Cn^{-2\zeta} \leq \hat{P}(\hat{\beta}, \tilde{\lambda}) \leq \hat{P}(\hat{\beta}, \hat{\lambda}) \leq \sup_{\lambda \in \hat{\Lambda}_n(\beta_0)} \hat{P}(\beta_0, \lambda) \leq \rho_0 + O_p(n^{-1}). \quad \text{(A.5)}$$

Also, by $\zeta < 1/2$, $\zeta - 1 < -1/2 < -\zeta$. Solving eq. (A.5) for $\|\hat{g}\|$ then gives

$$\|\hat{g}\| \leq O_p(n^{\zeta-1}) + Cn^{-\zeta} = O_p(n^{-\zeta}). \quad \text{(A.6)}$$

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.6), so that $\bar{\lambda} \in \Lambda_n$ w.p.a.1. Then, as in eq. (A.5),

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

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 UWL, $\sup_{\beta \in \mathcal{B}} \|\hat{g}(\beta) - g(\beta)\| \xrightarrow{p} 0$ and $g(\beta)$ is continuous. By T $g(\hat{\beta}) \xrightarrow{p} 0$. Since $g(\beta) = 0$ has a unique zero at 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. The 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 3.1 and Lemma A1, $\max_{i\le n}|\tilde{\lambda}'\hat{g}_i| \xrightarrow{p} 0$. 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 $\theta_0 = (\beta_0', 0')'$,

$$0 = \begin{pmatrix} 0 \\ -\hat{g}(\beta_0) \end{pmatrix} + \bar{M}(\hat{\theta} - \theta_0), \tag{A.7}$$

$$\bar{M} = \begin{pmatrix} 0 & \sum_{i=1}^n \rho_1(\hat{\lambda}'\hat{g}_i)G_i(\hat{\beta})'/n \\ \sum_{i=1}^n \rho_1(\bar{\lambda}'\hat{g}_i)G_i(\bar{\beta})/n & \sum_{i=1}^n \rho_2(\bar{\lambda}'\hat{g}_i)g_i(\bar{\beta})\hat{g}_i'/n \end{pmatrix},$$

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 $\max_{i\le n}|\bar{\lambda}'\hat{g}_i| \xrightarrow{p} 0$. Therefore, $\max_{i\le n}|\rho_1(\bar{\lambda}'\hat{g}_i)+1| \xrightarrow{p} 0$ and $\max_{i\le n}|\rho_2(\bar{\lambda}'\hat{g}_i)+1| \xrightarrow{p} 0$. It then follows from UWL 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.7) then gives

$$\begin{aligned}\sqrt{n}(\hat{\theta} - \theta_0) &= -\bar{M}^{-1}(0, -\sqrt{n}\hat{g}(\beta_0)')' = -M^{-1}(0, -\sqrt{n}\hat{g}(\beta_0)')' + o_p(1) \qquad (A.8)\\ &= -(H', P)'\sqrt{n}\hat{g}(\beta_0) + o_p(1).\end{aligned}$$

The first conclusion follows from this equation and the CLT. For the second conclusion, note that an expansion and eq. (A.8) give

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

Also,

$$\begin{aligned}\hat{P}(\hat{\beta}, \hat{\lambda}) &= \rho_0 - \hat{\lambda}'\hat{g}(\hat{\beta}) + \hat{\lambda}'[\sum_{i=1} \rho_2(\bar{\lambda}'\hat{g}_i)\hat{g}_i\hat{g}_i'/n]\hat{\lambda}/2 \qquad (A.9)\\ &= \rho_0 - \hat{\lambda}'\hat{g}(\hat{\beta}) - \hat{\lambda}'\Omega\hat{\lambda}/2 + o_p(n^{-1}) = \rho_0 + \hat{g}(\hat{\beta})'\Omega^{-1}\hat{g}(\hat{\beta})/2 + o_p(n^{-1}).\end{aligned}$$

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 second conclusion follows from eq. (A.9). 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 $m(z,\theta)$ satisfy a) $\hat{\theta} = \theta_0 + O_p(n^{-1/2})$; b) $\hat{m}(\hat{\theta}) = \sum_{i=1}^n m(z_i,\hat{\theta})/n = 0$ w.p.a.1; c) For some $\zeta > 2$, $d(z)$ with $E[d(z)] < \infty$, and $T_n = \{\theta : \|\theta - \theta_0\| \le n^{-1/\zeta}\}$, w.p.a.1 for $i = 1, ..., n$, $m(z_i,\theta)$ is three times continuously differentiable on T_n and for $\theta \in T_n$,*

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

d) $E[\|m(z,\theta_0)\|^6]$, $E[\|\partial m(z,\theta_0)/\partial\theta\|^6]$, $E[\|\partial^2 m(z,\theta_0)/\partial\theta_j\partial\theta\|^6]$, and $E[\|\partial^3 m(z,\theta_0)/\partial\theta_j\partial\theta_k\partial\theta\|^2]$, $(j,k = 1,...,q)$, are finite; e) $E[m(z,\theta_0)] = 0$ and $M = E[\partial m(z,\theta_0)/\partial\theta]$ exists and is nonsingular. Let

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

Then eq. (3.1) is satisfied for $\sqrt{n}(\hat{\theta} - \theta_0)$ with

$$\begin{aligned}
Q_1(\tilde{\psi},\tilde{a}) &= -M^{-1}[\tilde{A}\tilde{\psi} + \sum_{j=1}\tilde{\psi}_j M_j\tilde{\psi}/2], Q_2(\tilde{\psi},\tilde{a},\tilde{b}) = -M^{-1}[\tilde{A}Q_1(\tilde{\psi},\tilde{a}) \qquad (A.10)\\
&\quad + \sum_{j=1}\{\tilde{\psi}_j M_j Q_1(\tilde{\psi},\tilde{a}) + Q_{1j}(\tilde{\psi},\tilde{a})M_j\tilde{\psi} + \tilde{\psi}_j\tilde{B}_j\tilde{\psi}\}/2 + \sum_{j,k=1}\tilde{\psi}_j\tilde{\psi}_k M_{jk}\tilde{\psi}/6].
\end{aligned}$$

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,

$$\begin{aligned}
0 &= \hat{m}(\theta_0) + \hat{M}(\theta_0)(\hat{\theta} - \theta_0) + \sum_{j=1}(\hat{\theta}_j - \theta_{j0})[\partial\hat{M}(\theta_0)/\partial\theta_j](\hat{\theta} - \theta_0)/2 \qquad (A.11)\\
&\quad + \sum_{j,k=1}(\hat{\theta}_j - \theta_{j0})(\hat{\theta}_k - \theta_{k0})[\partial^2\hat{M}(\bar{\theta})/\partial\theta_k\partial\theta_j](\hat{\theta} - \theta_0)/6.
\end{aligned}$$

By T, M, the CLT, and the Lipschitz hypothesis,

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

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

$$\begin{aligned}\hat{\theta}-\theta_0 &= \tilde{\psi}/\sqrt{n} - M^{-1}[\tilde{A}(\hat{\theta}-\theta_0)/\sqrt{n} + \sum_{j=1}(\hat{\theta}_j-\theta_{j0})M_j(\hat{\theta}-\theta_0)/2 \qquad \text{(A.12)} \\ &+ \sum_{j=1}(\hat{\theta}_j-\theta_{j0})(\tilde{B}_j/\sqrt{n})(\hat{\theta}-\theta_0)/2 \\ &+ \sum_{j,k=1}(\hat{\theta}_j-\theta_{j0})(\hat{\theta}_k-\theta_{k0})M_{jk}(\hat{\theta}-\theta_0)/6] + O_p(n^{-2}).\end{aligned}$$

As all the terms except $\tilde{\psi}/\sqrt{n}$ are $O_p(n^{-1})$, it follows that

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

Next, the last three terms (including the remainder) in eq. (A.12) are $O_p(n^{-3/2})$, and replacing $\hat{\theta}-\theta_0$ by $\tilde{\psi}/\sqrt{n}$ in the second and third terms also generates an error that is $O_p(n^{-3/2})$, we obtain

$$\begin{aligned}\hat{\theta}-\theta_0 &= \tilde{\psi}/\sqrt{n} - M^{-1}[\tilde{A}\tilde{\psi} + \sum_{j=1}\tilde{\psi}_j M_j\tilde{\psi}/2]/n + O_p(n^{-3/2}) \qquad \text{(A.13)} \\ &= \tilde{\psi}/\sqrt{n} + Q_1(\tilde{\psi},\tilde{a})/n + O_p(n^{-3/2}).\end{aligned}$$

Finally, replacing $\hat{\theta}-\theta_0$ in the second and third terms of eq. (A.12) by $\tilde{\psi}/\sqrt{n}+Q_1(\tilde{\psi},\tilde{a})/n$ and in the fourth and fifth terms by $\tilde{\psi}/\sqrt{n}$ gives the conclusion. Q.E.D.

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

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

Then for $\tilde{\lambda} = -\hat{W}^{-1}\hat{g}(\tilde{\beta})$, $\hat{\theta} = (\tilde{\beta}', \tilde{\lambda}')'$, *and for* $\tilde{\psi}$, $\tilde{a}$, *and* $Q_1(\cdot,\cdot)$ *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 $\theta = (\beta', \lambda')'$, $\lambda_0 = 0$, and $m(z,\theta) = -(\lambda'\partial g(z,\beta)/\partial\beta, g(z,\theta)' + \lambda'[W + \xi(z)])'$. It follows from Theorem 3.4 of Newey and McFadden (1994) that $\hat{\theta} = \theta_0 + O_p(n^{-1/2})$. Note that for this choice of $m(z,\theta)$ we have $M_i = \partial m(z_i, \theta_0)/\partial\theta$ and M_j in Lemma A4 as in the statement of Lemma A5. 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

$$0 = \hat{m}(\hat{\theta}) + [0, -\tilde{\lambda}'(\hat{W} - W - \sum_i \xi(z_i)/n)]' = \hat{m}(\hat{\theta}) + O_p(n^{-3/2}). \tag{A.14}$$

Then expanding as in eq. (A.13) gives the result. Q.E.D.

Lemma A6: *Suppose that Assumptions 1-4 are satisfied and let* $\Omega_{i\beta_j} = \partial[g_i(\beta_0)g_i(\beta_0)']/\partial\beta_j$, $\bar{\Omega}_{\beta_j} = E[\Omega_{i\beta_j}]$, $\tilde{\Omega}_{\beta_j} = \sum_i(\Omega_{i\beta_j} - \bar{\Omega}_{\beta_j})/\sqrt{n}$, $\bar{\Omega}_{\beta_j\beta_k} = E[\partial^2\{g_i(\beta_0)g_i(\beta_0)'\}/\partial\beta_k\partial\beta_j]$, *and let a superscript* W *denote objects from the conclusion of Lemma A5, that is let* ψ_i^W, M_i^W, M_j^W, a_i^W, *and* $Q_1^W(\cdot,\cdot)$ *be as there without the superscript* W. *Also, let* $\xi_i^\Omega = g_i g_i' - \Omega + \sum_{j=1}^p \bar{\Omega}_{\beta_j} e_j' \psi_i^W$ *and*

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

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

Proof: Similarly to the proof of Lemma A4, expanding gives

$$\begin{aligned}\hat{\Omega}(\tilde{\beta}) &= \hat{\Omega}(\beta_0) + \sum_{j=1} \bar{\Omega}_{\beta_j}(\tilde{\beta}_j - \beta_{j0}) + \sum_{j=1} (\tilde{\Omega}_{\beta_j}/\sqrt{n})(\tilde{\beta}_j - \beta_{j0}) \\ &\quad + \sum_{j,k=1} \bar{\Omega}_{\beta_k\beta_j}(\tilde{\beta}_j - \beta_{j0})(\tilde{\beta}_k - \beta_{k0})/2 + O_p(n^{-3/2}).\end{aligned} \tag{A.15}$$

By Lemma A5, $\tilde{\beta}_j - \beta_{j0} = e_j'\tilde{\psi}^W/\sqrt{n} + O_p(n^{-1}) = e_j'\tilde{\psi}^W/\sqrt{n} + e_j'Q_1^W(\tilde{\psi}^W, \tilde{a}^W)/n + O_p(n^{-3/2})$. The conclusion follows by substituting the first equality for the last two terms in eq. (A.15) and by substituting the second equality for the second term. Q.E.D.

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

$$
\begin{aligned}
B_j^1(z_i) &= -\begin{pmatrix} 0 & G_i^{j\prime} - E[G_i^j]' \\ G_i^j - E[G_i^j] & 0 \end{pmatrix}, (j \leq p), \\
B_{p+j}^1(z_i) &= -\begin{pmatrix} \partial^2 g_{ij}(\beta_0)/\partial\beta\partial\beta' - E[\partial^2 g_{ij}(\beta_0)/\partial\beta\partial\beta'] & 0 \\ 0 & 0 \end{pmatrix}, (j \leq m).
\end{aligned}
$$

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

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

Let $Q_1(\cdot,\cdot)$ and $Q_{21}(\cdot,\cdot,\cdot)$ be equal to Q_1 and Q_2 as given in the conclusion of Lemma A4, with $\tilde{\psi}$, M, M_j, M_{jk}, $A(z)$, as specified here (and as in Lemma A5 with $W = \Omega$). Also, let $b^1(z)$ be the vector elements of every $B_j(z)$ and $T = \theta_0 + \tilde{\psi}/\sqrt{n} + Q_1(\tilde{\psi}, \tilde{a})/n + Q_{21}(\tilde{\psi}, \tilde{a}, \tilde{b}^1)/n^{3/2}$. Then as $-\hat{\lambda}'\tilde{Q}_1^\Omega/n = O_p(n^{-3/2})$ we can solve for $\hat{\theta} - \theta_0$ as in the conclusion of Lemma A4 to obtain

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

Then by $\hat{\lambda} = [0, I_m]\tilde{\psi}/\sqrt{n} + O_p(n^{-1})$ we can substitute for $\hat{\lambda}$ to obtain

$$\hat{\theta} = T + M^{-1} diag[0, \tilde{Q}_1^\Omega]\tilde{\psi}/n^{3/2} + O_p(n^{-2}).$$

The conclusion then follows by including in $b(z)$ all the components of $b^1(z)$ as well as those of every variable that appears as $\sqrt{n}$ times a sample average in $\tilde{Q}_1^\Omega$. Then we find the second order term in the expansion for GMM to be

$$Q_2(\tilde{\psi}, \tilde{a}, \tilde{b}) = Q_{21}(\tilde{\psi}, \tilde{a}, \tilde{b}^1) + M^{-1} diag[0, \tilde{Q}_1^\Omega]\tilde{\psi},$$

giving the conclusion. 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

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

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 1(d). Then by Lemma A1, Assumption 3, and $\rho_1(v)$ three times continuously differentiable on a neighborhood of 0, $m(z_i,\theta)$ is three times continuously differentiable on T_n of Lemma A4, $i=1,...,n$, to which we henceforth restrict attention. Let $m_i(\theta)=m(z_i,\theta)$, $v_i(\theta)=\lambda' g_i(\theta)$, $h_i(\theta)=\partial v_i(\theta)/\partial\theta=(\lambda' G_i(\beta),g_i(\beta)')'$, and $h_i(\theta)_j$ denote the j^{th} element of $h_i(\theta)$. 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 $b_i=b(z_i)$,

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

Also, by Assumption 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 $J\le 4$. 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 $\rho_1=\rho_2=-1$,

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

and 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

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

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 $j \le p$ and $k \le p$,

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

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

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

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

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

Then by the conclusion of Lemma A4, eq. (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 $\xi_i^\Omega = g_i g_i' - \Omega - \sum_{j=1}^p \bar{\Omega}_{\beta_j} e_j' H_W g_i$. 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 $\sum_{j=1}^p e_j' S_k \Sigma e_j/2 = \sum_{j=1}^p tr(\Sigma e_j e_j' S_k)/2 = a_k$. Then for $\hat{\lambda} = -\hat{\Omega}(\tilde{\beta})^{-1}\hat{g}(\hat{\beta})$ 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

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

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

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

Proof of Theorem 4.5: 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

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

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

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

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

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

By T and CS we then have

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

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 $\eta_i = e'_j\Sigma(\kappa_i + \bar{d}_i\mu_{3i})/\sigma_i^2 = e'_j\delta_i$,

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

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

Proof of Theorem 4.6: 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

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

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 $\widehat{Bias}(\hat{\beta}) = \tau(\sum_i d_i(\hat{\beta})/n)/n$, 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. For GEL $\widehat{Bias}(\hat{\beta}) = \tau(\sum_i \hat{\pi}_i d_i(\hat{\beta}))/n$. The conclusion follows similarly for GEL, with τ and $d(z, \beta)$ corresponding to the bias formula for GEL, and

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

Before proving Theorem 6.1, we will prove the following intermediate result:

Lemma A7: *If the hypotheses of Theorem 6.1 are satisfied and z_i has finite support $\{Z_1, ..., Z_J\}$ then for the bias corrected estimators, $\Xi - \Xi_{EL}$ is positive semi-definite.*

Proof: Let $n_j = \sum_{i=1}^{n} 1(z_i = Z_j)$ and consider the multinomial, moment restricted MLE given by

$$\hat{\beta}_{ML} = \arg \max_{\beta \in \mathcal{B}, \Pi_1, ..., \Pi_J} \sum_{j=1} n_j \ln \Pi_j, s.t. \sum_{J=1} \Pi_j g(Z_j, \beta) = 0, \sum_{j=1} \Pi_j = 1.$$

By standard theory for MLE $\hat{\beta}_{ML}$ is consistent and there is neighborhood $\mathcal{N}$ of β_0 such that w.p.a.1 $\hat{\beta}_{ML}$ is the unique β in $\mathcal{N}$ solving the first order conditions. Also, $\sum_{j=1}^{J} n_j \ln \Pi_j$ is a monotonic increasing transformation of $\sum_{j=1}^{J} n_j \ln(\Pi_j / n_j)$. Let $I_j = \{i : z_i = Z_j\}$. Note that holding $\Pi_j > 0$ fixed, by strict concavity the maximum of $\sum_{i \in I_j} \ln(\pi_i)$ subject to $\Pi_j = \sum_{i \in I_j} \pi_i$ is $n_j \ln(\Pi_j / n_j)$. Then, similarly to Section 2.3 of Owen (2001),

$$\begin{aligned} \hat{\beta}_{ML} &= \arg \max_{\beta \in \mathcal{B}, \pi_1, ..., \pi_n} \sum_{j=1} \sum_{i \in I_j} \ln(\pi_i) \\ &: s.t. \sum_{j=1} \Pi_j g(Z_j, \beta) = 0, \sum_{j=1} \Pi_j = 1, \Pi_j = \sum_{i \in I_j} \pi_i. \qquad \text{(A.23)} \\ &= \arg \max_{\beta \in \mathcal{B}, \pi_1, ..., \pi_n} \sum_{i=1} \ln(\pi_i), s.t. \sum_{i=1} \pi_i g(z_i, \beta) = 0, \sum_{i=1} \pi_i = 1. \end{aligned}$$

Therefore, w.p.a.1 $\hat{\beta}_{ML} = \hat{\beta}_{MD}$ for $h(\pi) = -\ln(\pi)$. Now consider $\hat{\beta}_{EL}$ defined as the solution to eq. (2.2). By Theorem 3.1, $\hat{\beta}_{EL}$ is consistent, so that $\hat{\beta}_{EL} \in int(B)$ and $\hat{\lambda}$ exists w.p.a.1. Also, similarly to the proof of Theorem 3.1 it follows that $\sum_{i=1}^{n} \rho_2(\hat{\lambda}' \hat{g}_i) \hat{g}_i \hat{g}_i' / n$ is nonsingular so that all the hypotheses of Theorem 2.2 are satisfied, w.p.a.1. Then by consistency, $\hat{\beta}_{EL} \in \mathcal{N}$ and by Theorem 2.2 has the same first order conditions as $\hat{\beta}_{ML} = \hat{\beta}_{MD}$, so $\hat{\beta}_{ML} = \hat{\beta}_{EL}$, w.p.a.1. Furthermore, from Corollary 4.3 we know that there are known functions $\tau(d)$ and $d(z, \beta)$ with

$$\begin{aligned} Bias(\hat{\beta}_{EL}) &= \tau(E[d(z, \beta_0)])/n = B(\Pi_{10}, ..., \Pi_{J0}, \beta_0)/n, \\ B(\Pi_1, ..., \Pi_J, \beta) &= \tau(\sum_{j=1} \Pi_j d(Z_j, \beta)). \end{aligned}$$

Then by $\hat{\pi}_i = \hat{\Pi}_j / n_j$ and $\hat{\beta}_{EL} = \hat{\beta}_{ML} = \hat{\beta}$,

$$\widehat{Bias}(\hat{\beta}) = \tau(\sum_{i=1} \hat{\pi}_i d(z_i, \hat{\beta}))/n = \tau(\sum_{j=1} \hat{\Pi}_j d(Z_j, \hat{\beta}))/n = B(\hat{\Pi}_1, ..., \hat{\Pi}_J, \hat{\beta})/n.$$

Thus, the EL bias estimate $\widehat{Bias}(\hat{\beta})$ equals the MLE bias estimate obtained by plugging the MLE into the bias formula. Since EL equals MLE, and the EL bias correction equals the MLE bias correction, the bias corrected EL estimator is equal to the bias corrected MLE.

Next we show that the Pfanzagl and Wefelmeyer (1978) (PW henceforth) conditions for third order efficiency of MLE relative to the other estimator are satisfied. We consider a reparameterization as in Lemma 1 of Chamberlain (1987), where it is shown that there exists a $J - (m - p) - 1$ subvector γ of $\Pi = (\Pi_1, ..., \Pi_J)'$ such that for $\theta = (\beta', \gamma')'$, there is $\Pi(\theta) = (\Pi_1(\theta), ..., \Pi_J(\theta))'$ and an open set Θ containing θ_0 with

$$\begin{aligned} \sum_{j=1} \Pi_j(\theta) g(Z_j, \beta) &= 0, \Pi_j(\theta) \geq C > 0, \\ \sum_{j=1} \Pi_j(\theta) \partial^2 \ln \Pi_j(\theta)/\partial\theta\partial\theta' &: \text{ is nonsingular.} \end{aligned}$$

Consider the multinomial log likelihood $\ell(z, \theta) = \sum_{j=1}^{J} 1(z = Z_j) \ln \Pi_j(\theta)$. In the notation of PW, the score vector is $l(z, \theta) = \sum_{j=1}^{J} 1(z = Z_j) \Pi_j(\theta)^{-1} \partial \Pi_j(\theta)/\partial\theta$. Then it follows from the implicit function theorem similarly to Lemma 1 of Chamberlain (1987) that $\Pi(\theta)$ is four times continuously differentiable with Lipschitz fourth derivative, giving L_4 and M_4 of PW. Conditions i), ii), iii), and I_3 of PW follow similarly, so that $l(\cdot, \cdot)$ satisfies all the conditions of Theorem 1' of PW. Furthermore, it follows by $\hat{\beta}$ being equal to the MLE, as shown above, and by invariance of the MLE to reparameterization, that there is $\hat{\gamma}$ such that w.p.a.1, $\hat{\theta} = (\hat{\beta}', \hat{\gamma}')$ satisfies $\sum_{i=1}^{n} l(z_i, \hat{\theta}) = 0$. Therefore, all the conditions of Theorem 1' of PW for the MLE and the likelihood are satisfied.

Next, consider the other GMM or GEL estimator $\hat{\beta}$. Let $\tilde{\theta} = (\hat{\beta}', \hat{\gamma}'_{MLE})$. It follows by Theorem 3.3 or 3.4 and the previous paragraph that the estimator has a stochastic expansion as in eq. (3.1). Then eq. (3.1) of PW is satisfied, without the remainder condition (which we will not need). By Lemma A5 all the terms in the expansion are polynomials in means of random variables, with coefficients that are Lipschitz in θ, so that the Condition B requirements on p.5 of PW are satisfied. Furthermore, the normalizations required by PW for the random variables in the expansion can be satisfied by adding and subtracting appropriate terms (including the mean square projections given there). Then the expansion for $\tilde{\theta}$ satisfies all the conditions of PW.

Finally, we show how the results of PW can be adapted to show that the higher order variance of the bias corrected MLE is less than or equal to that of any one of the other estimators. To do this, let $\tilde{Q}_1 = Q_1(\tilde{\psi}, \tilde{a}, F_0)$, $\tilde{Q}_2 = Q_2(\tilde{\psi}, \tilde{a}, \tilde{b}, F_0)$ be the terms in the stochastic expansion of any of the estimators and

$$\tilde{Y} = \tilde{\psi} + \tilde{Q}_1/\sqrt{n} + \tilde{Q}_2/n$$

be the expansion without the remainder. Also, for any positive definite matrix A let $L(u) = u'Au$. Then, as noted in Remark 16 of PW (see also Rothenberg, 1984, p.904), for the polynomial (quadratic) loss function $L(u)$ and the polynomial (in $\tilde{\psi}, \tilde{a}, \tilde{b}$) stochastic expansion $\tilde{Y}$, the expected loss computed from a formal Edgeworth expansion equals $E[L(\tilde{Y})] + o(n^{-1})$. Then, as in the square brackets on the top of p. 25 of PW,

$$E[L(\tilde{Y})] = \int \bar{\chi}_n(u, v, w) L(u) du dv dw + o(n^{-1}),$$

where $\bar{\chi}_n(u, v, w)$ is given in eq. (6.15) of PW. It then follows as in the remainder of the argument on pp.25-26 of PW that $\int \bar{\chi}_n(u, v, w) L(u) du dv dw + o(n^{-1})$ is minimized at the bias corrected MLE. It is also the case that by the expression for the higher order variance Ξ in Section 6,

$$\begin{aligned} E[\tilde{\psi}' A \tilde{\psi}] &= tr(A\Sigma), \\ E[\tilde{Q}_1' A \tilde{Q}_1]/n + 2E[\tilde{\psi}' A \tilde{Q}_1]/\sqrt{n} + 2E[\tilde{\psi}' A \tilde{Q}_2]/n &= tr(A\Xi)/n + o(n^{-1}), \\ E[\tilde{Q}_2' A \tilde{Q}_2]/n^2 + 2E[\tilde{Q}_1' A \tilde{Q}_2]/n^{3/2} &= o(n^{-1}), \end{aligned}$$

so that

$$E[L(\tilde{Y})] = tr(A\Sigma) + tr(A\Xi)/n + o(n^{-1}).$$

Subtracting, we obtain

$$\int \bar{\chi}_n(u,v,w)L(u)dudvdw = tr(A\Sigma) + \ tr(A\Xi)/n + o(n^{-1}).$$

Therefore, $tr(A\Sigma)+ \ tr(A\Xi)/n$ is minimized at the bias corrected MLE. Since Σ is the same for each estimator, it follows that $tr(A\Xi)/n$ is minimized at the bias corrected MLE, and thus

$$0 \leq tr(A\Xi) - tr(A\Xi_{EL}) = tr(A\Delta), \Delta = \Xi - \Xi_{EL}.$$

Since this inequality holds for any positive definite matrix A it follows that Δ is positive semi-definite. (For $\Delta = B\Lambda B'$ with $B'B = I$ and Λ a diagonal matrix of eigenvalues of Δ let $A = B(e_j e_j' + \varepsilon \sum_{k \neq j} e_k e_k')B'$ for any $\varepsilon > 0$, so that $tr(A\Delta) = \Lambda_{jj} + \varepsilon \sum_{k \neq j} \Lambda_{kk} \geq 0$ for any $\varepsilon > 0$ implies $\Lambda_{jj} \geq 0$.) Q.E.D.

Proof of Theorem 6.1: By Lemma A7 it suffices to show that there is a distribution with finite support $\{Z_1, ..., Z_J\}$ such that Assumptions 1-4 are satisfied and both Ξ_{EL} and Ξ have the same values as under the true distribution. To do this, we show that there is a vector of known functions $V(z,\beta)$ and known functions $\tau_{EL}(\cdot)$ and $\tau(\cdot)$ such that

$$\Xi_{EL} = \tau_{EL}(E[V(z,\beta_0)]), \Xi = \tau(E[V(z,\beta_0)]). \tag{A.24}$$

For GEL, it follows as in the proof of Theorem 3.4 that eq. (A.12) is satisfied with $\theta = (\beta', \lambda')'$ and $m(z_i,\theta)$ as given in the proof of Theorem 3.4. Then, from the higher order variance formula given in Rilstone et. al. (1996), it follows that the higher order variance is a known function of expectations of first, second, and third derivatives of $m(z_i,\theta)$ with respect to θ, evaluated at the truth (forming the constant coefficients in the expansion), the covariance of $m(z_i,\theta_0)$ with itself and with its derivative with respect to θ, (forming $\lim_{n\to\infty} Var(\tilde{Q}_1)$), third moments of $m(z_i,\theta_0)$ and the third cross moment of derivatives of $m(z_i,\theta_0)$ with products of $m(z_i,\theta_0)$ (forming $\lim_{n\to\infty} E[\sqrt{n}\tilde{Q}_1\tilde{\psi}']$), and covariance of $m(z_i,\theta_0)$ with itself, its derivatives, and its third derivatives (forming $\lim_{n\to\infty} E[\tilde{Q}_2\tilde{\psi}']$, including the bias correction term in $\tilde{Q}_2$), all of which moments exist by Assumptions 1-4. Let $V(z,\theta)$ be any finite vector including all of these functions. Thus, eq. (A.24) is satisfied for Ξ_{EL} and for any Ξ corresponding to a GEL estimator. For GMM, it follows by the use of at least two iterations that the estimator has the same asymptotic expansion as an estimator solving eq. (3.2) with $\tilde{\beta} = \hat{\beta}_{GMM}$, i.e. that is fully iterated. This then is an M-estimator with $m(z, \theta) = (\lambda' \partial g(z, \beta)/\partial \beta, g(z, \beta)'(1 + \lambda' g(z, \beta))'$. It follows similarly to the proof of Theorem 3.4 that eq. (A.12) is satisfied, so that eq. (A.24) is satisfied for Ξ corresponding to a GMM estimator.

Next, by Lemma 3 of Chamberlain (1987) there is a distribution with support $\{Z_1, ..., Z_J\}$ and probabilities Π_{j0} of each Z_j such that for the expectation $\bar{E}[a(z)] = \sum_{j=1}^{J} \Pi_{j0} a(Z_j)$,

$$\begin{aligned}
\bar{E}[g(z, \beta_0)] &= E[g(z, \beta_0)] = 0, \\
\bar{E}[\partial g(z, \beta_0)/\partial \beta] &= E[\partial g(z, \beta_0)/\partial \beta], \bar{E}[g(z, \beta_0) g(z, \beta_0)'] = E[g(z, \beta_0) g(z, \beta_0)'], \\
\bar{E}[V(z, \beta_0)] &= E[V(z, \beta_0)], \bar{E}[b(z)] = E[b(z)].
\end{aligned}$$

Consider now the case where $z_i, (i = 1, ..., n)$, are i.i.d. with distribution $\bar{F}$. By construction this discrete distribution has the same Ξ_{EL} and Ξ as the true distribution. By $\bar{E}[b(z)] = E[b(z)] < \infty$ it follows that Assumptions 1-4 are satisfied for this distribution. Then by Lemma A7 it follows that $\Xi_{EL} \le \Xi$. Q.E.D.

Proof of Theorem 6.2: Let $g_i(\beta) = q_i(y_i - x_i'\beta)$. 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, $\bar{\Omega}_{\beta_j} = E[-2 q_i q_i' x_{ij} u_i] = 0$, 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. For $\tilde{b}^{GMM}$ and $\tilde{b}^{GEL}$, let $\tilde{Q}_2^{GMM} = Q_2(\tilde{\psi}, \tilde{a}, \tilde{b}^{GMM})$ and $\tilde{Q}_2^{GEL} = Q_2(\tilde{\psi}, \tilde{a}, \tilde{b}^{GEL})$ denote the second order terms for GMM and GEL respectively. From the form of Ξ given in Section 6 we see that the difference in higher order variances for GMM and GEL estimators of β reduces to

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

Thus, it suffices just to 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.10) 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 $\sum_{j=1}^{q} \tilde{\psi}_j \tilde{B}_j \tilde{\psi}/2$ and $\sum_{j,k=1}^{q} \tilde{\psi}_j \tilde{\psi}_k M_{jk} \tilde{\psi}/6$. For GMM, $B_j^1(z) = 0$ and $M_{jk} = 0$ (by linearity of $m_i(\theta)$) from the proof of Theorem 3.3. In addition for GMM, by efficiency of $\tilde{\beta}$, $\tilde{\psi}^W$ from Lemma A6 is equal to $\tilde{\psi}$, so that $\tilde{Q}_1^{\Omega} = \sum_{j=1}^{p} \tilde{\Omega}_{\beta_j} e_j' \tilde{\psi} + \sum_{j,k=1}^{p} \bar{\Omega}_{\beta_k \beta_j} e_j' \tilde{\psi}_j e_k' \tilde{\psi}_k / 2$. Let $\tilde{\psi}^{\beta} = [I_p, 0]\tilde{\psi}$ and $\tilde{\psi}^{\lambda} = [0, I_m]\tilde{\psi}$. We have

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

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

$$\begin{aligned} D &= D_1 + D_2, D_1 = \lim_{n\to\infty} E[\tilde{D}_1], D_2 = \lim_{n\to\infty} E[\tilde{D}_2], \\ \tilde{D}_1 &= [\Sigma, -H]\{\tilde{T}_{GMM} + \sum_{j=1} \tilde{\psi}_j \tilde{B}_j \tilde{\psi}/2\}\tilde{\psi}^{\beta\prime}, \tilde{T}_{GMM} = diag[0, \sum_{j=1} \tilde{\Omega}_{\beta_j} \tilde{\psi}_j^{\beta}]\tilde{\psi} \\ \tilde{D}_2 &= [\Sigma, -H]\{\sum_{j,k=1} \tilde{\psi}_j \tilde{\psi}_k M_{jk} \tilde{\psi}/6\}\tilde{\psi}^{\beta\prime}. \end{aligned}$$

Consider $\tilde{D}_1$. 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

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

Note also that for $j > 0$,

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

Then,

$$\begin{aligned} \tilde{T}_{GMM} + \sum_{j=1}^{m+p} \tilde{\psi}_j \tilde{B}_j \tilde{\psi}/2 &= \tilde{T}_{GMM}/2 + \sum_{j=1}^{m} \tilde{\psi}_j^{\lambda} \tilde{B}_{p+j} \tilde{\psi}/2 \\ &= \sum_{j=1}^{m} \tilde{\psi}_j^{\lambda} \tilde{B}_{p+j} \begin{pmatrix} 0 \\ I_m \end{pmatrix} \tilde{\psi}^{\lambda}/2 = -\sum_{j=1}^{m} \tilde{\psi}_j^{\lambda} \begin{pmatrix} 2\sum_i g_{ij} G_i'/\sqrt{n} \\ -\rho_3 \sum_i g_{ij} g_i g_i'/\sqrt{n} \end{pmatrix} \tilde{\psi}^{\lambda}/2. \end{aligned}$$

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

$$
\begin{aligned}
D_1 &= [\Sigma, -H] \lim_{n\to\infty} E[\sum_{j=1} \tilde{\psi}_j^\lambda \begin{pmatrix} 2\sum_i u_i q_{ij} x_i q_i'/\sqrt{n} \\ \rho_3 \sum_i u_i^3 q_{ij} q_i q_i'/\sqrt{n} \end{pmatrix} \tilde{\psi}^\lambda \tilde{\psi}^{\beta\prime}]/2 \\
&= \Sigma \sum_{j=1} \{2E[(u_i^2/\sigma_i^2) q_{ij} x_i q_i' P e_j \bar{x}_i'] + \rho_3 E[(\varepsilon_i^4/\sigma_i^4) q_{ij} \bar{x}_i q_i' P e_j \bar{x}_i']\}\Sigma/2 \\
&= \Sigma\{E[K_i x_i \bar{x}_i'] + (\rho_3/2) E[(\mu_{4i}/\sigma_i^3) K_i \bar{x}_i \bar{x}_i']\}\Sigma.
\end{aligned}
$$

Next, consider $\tilde{D}_2$. Let P_{jk} denote the $(j,k)^{\text{th}}$ element of P. We have by linearity of $g_i(\beta)$ in β, eq. (A.22),

$$
\begin{aligned}
\sum_{j,k=1} P_{jk} M_{p+j,p+k}[\Sigma, 0]' &= \sum_{j,k=1} P_{jk} E[\begin{pmatrix} -G_i' e_j e_k' G_i - G_i' e_k e_j' G_i \\ \rho_3 g_i (g_{ij} e_k' G_i + g_{ik} e_j' G_i + g_{ij} g_{ik} G_i) \end{pmatrix}]\Sigma \\
&= -\sum_{j,k=1} P_{jk} \begin{pmatrix} 2E[x_i x_i' q_{ij} q_{ik}] \\ 3\rho_3 E[\sigma_i^2 q_i x_i q_{ij} q_{ik}] \end{pmatrix} \Sigma \\
&= -\begin{pmatrix} 2E[K_i x_i x_i'] \\ 3\rho_3 E[\sigma_i^2 K_i q_i x_i'] \end{pmatrix} \Sigma.
\end{aligned}
$$

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

$$
\begin{aligned}
\sum_{j=1} \sum_{k=1} M_{p+j,k} \begin{pmatrix} 0 \\ P \end{pmatrix} e_j e_k' \Sigma &= -\sum_{j=1} \sum_{k=1} E[\begin{pmatrix} G_i' e_j g_i^{k\prime} + + G_{ijk} G_i' \\ -\rho_3 [G_{ijk} g_i g_i' + g_{ij}(g_i^k g_i + g_i g_i^{k\prime})] \end{pmatrix}] P e_j e_k' \Sigma \\
&= -\sum_{j=1} \sum_{k=1} \begin{pmatrix} 2E[x_i q_i' q_{ij} x_{ik}] \\ 3\rho_3 E[\sigma_i^2 q_i q_i' q_{ij} x_{ik}] \end{pmatrix} P e_j e_k' \Sigma \\
&= -\begin{pmatrix} 2E[K_i x_i x_i'] \\ 3\rho_3 E[\varepsilon_i^2 K_i q_i x_i'] \end{pmatrix} \Sigma.
\end{aligned}
$$

Note that for $j, k \le p$, for GEL in a linear model the 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, by standard V-statistic calculations and $M_{jk} = M_{kj}$,

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

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

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

It follows similarly that $E[K_i\bar{x}_i(\bar{x}_i - x_i)'] \to 0$, 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. (1980): "The n^{-2}-order Mean Square Errors of the Maximum Likelihood and Minimum Logit Chi-Square Estimators," *Annals of Statistics* 8, 488-505.

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.

Bonnal, H., and E. Renault (2001): "Minimum Chi-Square Estimation with Conditional Moment Restrictions," working paper, C.R.D.E., Université de Montréal. Paper presented at the 2001 Canadian Econometric Study Group Conference, University of Waterloo.

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 Jackknife Interpretation of the Continuous Updating Estimator," *Economics Letters* 67, 239-244.

Donald, S.G., G.W. Imbens, and W.K. Newey (2002): "Choosing the Number of Instruments for GMM and GEL Estimators," working paper, Department of Economics, M.I.T..

Hahn, J., J.A. Hausman, and G. Kuersteiner (2001): "Bias Corrected Instrumental Variables Estimators for Panel Data," working paper, Department of Economics, M.I.T..

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. and R.H. Spady (2001): ”The Performance of Empirical Likelihood and Its Generalizations,” paper presented at 2001 NSF-Berkeley Econometrics Symposium on “Identification and Inference for Econometric Models”.

Judge, G. and R. Mittelhammer (2001): ”Empirical Evidence Concerning the Finite Sample Performance of EL-Type Structural Equation Estimators,” paper presented at NSF-Berkeley Symposium on “Identification and Inference for Econometric Models”.

Kezdi, G., J. Hahn, and G. Solon (2001): ”Jackknife Minimum Distance Estimation,” mimeo, Department of Economics, Brown University.

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 (1994): ”Momentary Lapses: Moment Expansions and the Robustness of Minimum Distance Estimators,” *Econometric Theory* 10, 172-197.

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.

Newey, W.K., J.J.S. Ramalho, and R.J. Smith (2001): "Asymptotic Bias for GMM and GEL Estimators with Estimated Nuisance Parameters," Festschrift in honor of Tom Rothenberg, forthcoming.

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

Owen, A. (2001): *Empirical Likelihood*, New York: Chapman and Hall.

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.

Ramalho, J.J.S. (2001): *Alternative Estimation Methods and Specification Tests for Moment Condition Models*, unpublished Ph.D. thesis, Department of Economics, University of Bristol.

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.

Robinson, P.M. (1988): "The Stochastic Difference Between Econometric Statistics," *Econometrica* 56, 531-548.

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. (1996): "Empirical Likelihood Parameter Estimation Under Moment Restrictions," seminar notes, Harvard/M.I.T. and Bristol.

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

Smith, R. J. (2001): "GEL Methods for Moment Condition Models", working paper, Department of Economics, University of Bristol.