UNIVERSIDADE DE ÉVORA

DEPARTAMENTO DE ECONOMIA

DOCUMENTO DE TRABALHO Nº **2003/05**

Asymptotic Bias for GMM and GEL Estimators with Estimated Nuisance Parameters*

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

Joaquim J.S. Ramalho
Universidade de Évora, Departamento de Economia

Richard J. Smith
Department of Economics, University of Warwick

*This paper has been prepared for the Festschrift in honour of Tom Rothenberg. We are grateful for helpfulcomments and criticisms by the Editor and two anonymous referees. The third author also appreciated the hospitality of the Cowles Foundation for Research in Economics, Yale University, and C.R.D.E, Departement des Sciences Economiques, University de Montreal, when the initial version was being prepared.

UNIVERSIDADE DE ÉVORA
DEPARTAMENTO DE ECONOMIA
Largo dos Colegiais, 2 – 7000-803 Évora – Portugal
Tel.: +351 266 740 894 Fax: +351 266 742 494
www.decon.uevora.pt wp.economia@uevora.pt

Abstract:

This papers studies and compares the asymptotic bias of GMM and generalized empirical likelihood (GEL) estimators in the presence of estimated nuisance parameters. We consider cases in which the nuisance parameter is estimated from independent and identical samples. A simulation experiment is conducted for covariance structure models. Empirical likelihood offers much reduced mean and median bias, root mean squared error and mean absolute error, as compared with two-step GMM and other GEL methods. Both analytical and bootstrap bias-adjusted two-step GMM estimators are compared. Analytical bias-adjustment appears to be a serious competitor to bootstrap methods in terms of finite sample bias, root mean squared error and mean absolute error. Finite sample variance seems to be little affected.

Keywords: GMM, Empirical Likelihood, Exponential Tilting, Continuous Updating, Bias, Stochastic Expansions.

JEL Classification: C13, C30

1 Introduction

It is now widely recognised that the most commonly used efficient two-step GMM (Hansen, 1982) estimator may have large biases for the sample sizes typically encountered in applications. See, for example, the Special Section, July 1996, of the *Journal of Business and Economic Statistics*. To improve the small sample properties of GMM, a number of alternative estimators have been suggested which include empirical likelihood (EL) [Owen (1988), Qin and Lawless (1994), and Imbens (1997)], continuous updating (CUE) [Hansen, Heaton, and Yaron (1996)] and exponential tilting (ET) [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. Newey and Smith (2002) show that CUE and members of the Cressie-Read (1984) power family are members of the GEL class; see also Smith (2001). All of these estimators and GMM have the same asymptotic distribution but different higher-order asymptotic properties. In a random sampling setting, Newey and Smith (2002) use the GEL structure, which helps simplify calculations and comparisons, to analyze higher-order properties using methods like those of Nagar (1959). Newey and Smith (2002) derive and compare the (higher-order) asymptotic bias for all of these estimators. They also derive bias-corrected GMM and GEL estimators and consider their higher-order efficiency.

Newey and Smith (2002) find that EL has two theoretical advantages. First, the asymptotic bias does not grow with the number of moment restrictions, while the bias of the others often does. Consequently, for large numbers of moment conditions the bias of EL will be less than the bias of the other estimators. The relatively low asymptotic bias of EL indicates that it is an important alternative to GMM in applications. Furthermore, under a symmetry condition, which may be satisfied in some instrumental variable settings, all the GEL estimators inherit the small bias property of EL. 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 the other estimators. This result generalizes the conclusions of Rothenberg (1996) who showed that for a single equation from a Gaussian, homoskedastic linear simultaneous equations model the asymptotic bias of EL is the same as the limited information maximum likelihood estimator and that bias-corrected EL is higher-order efficient relative to a bias-corrected GMM estimator.

This paper reconsiders Newey and Smith's (2002) results for scenarios in which GMM and GEL estimation criteria involve a preliminary nuisance parameter estimator. This type of situation arises in a number of familiar cases. Firstly, generated regressors employed in a regression model context require a preliminary estimator of a nuisance parameter; see Pagan (1984). Heckman's (1979) sample selectivity correction is a special case with the nuisance parameter estimator obtained from a selectivity equation. Secondly, covariance structure models typically require an initial estimator of the mean of the data which itself may not be of primary interest. Thirdly, but trivially, the use of a preliminary consistent GMM estimator to estimate the efficient GMM metric may be regarded as a nuisance parameter estimator and is thus a special case also. Consequently, the sample-splitting method for efficient two-step GMM metric estimation proposed to ameliorate the bias of efficient GMM estimators also falls within our analysis, the preliminary estimator being obtained from one sub-sample with the other sub-sample then used to implement efficient GMM. See *inter alia* Altonji and Segal (1996). The presence of the nuisance parameter estimator typically affects the first order asymptotic distribution of the estimator for the parameters of interest in the first and third examples, with sample-splitting inducing asymptotic inefficiency because of the reduction in sample size. There is no loss in efficiency in the second example because the Jacobian with respect to the nuisance parameter is null. However, the presence of the nuisance parameter estimator alters the higher-order asymptotic bias in all of these examples as compared to the nuisance parameter free situation.

To provide sufficient generality to deal with these various set-ups we define a sampling structure which permits the nuisance parameter estimator to be obtained from either an identical or independent sample. Sample selectivity and covariance structure models together with the standard method for estimation of the efficient GMM metric are examples of the first type whereas the sample-splitting example fits the latter category. We provide general stochastic expansions for GMM and GEL estimators. These expansions are then specialised for identical and independent samples and for the case when no nuisance parameters are present. The analytical expressions for asymptotic bias obtained from these expansions may be consistently estimated as in Newey and Smith (2002) to bias-correct GMM or GEL estimators. Some simulation experiments for covariance structure models show that these analytical methods for bias-adjustment of the efficient two-step GMM estimator may be efficacious as compared with bootstrap methods which are computationally more complex.

The outline of the paper is as follows. Section 2 describes the set-up and GMM and GEL estimators. Section 3 details the asymptotic biases for situations which involve either an independent or identical sample. A simulation experiment in section 4 for covariance structures with a single nuisance parameter estimated from the same sample considers the finite sample properties of GMM, CUE, ET and EL estimators and compares some bootstrap and analytical bias-adjusted versions of the efficient two-step GMM estimator. Appendix A contains general stochastic expansions for GMM and GEL estimators together with proofs of the results in the paper. For ease of reference, some notation used extensively in the paper is collected together in Appendix B.

2 The Estimators and Other Preliminaries

2.1 Moment Conditions

Consider the moment indicator $g^{\beta}(z, \alpha, \beta)$, an m_{β}-vector of functions of a data observation z and the p_{β}-vector β of unknown parameters which are the object of inferential interest, where $m_{\beta} \geq p_{\beta}$. The moment indicator $g^{\beta}(z, \alpha, \beta)$ also depends on α, a p_{α}-vector of nuisance parameters. It is assumed that the true parameter vector β_0 uniquely satisfies the moment condition

$$E[g^{\beta}(z, \alpha_0, \beta_0)] = 0,$$

where $E[.]$ denotes expectation.

Estimation of the nuisance parameter vector α_0 is based on the additional moment indicator $g^{\alpha}(x, \alpha)$, an m_{α}-vector of functions of a data observation x and α, where $m_{\alpha} \geq p_{\alpha}$. The true value α_0 of the nuisance parameter vector is assumed to satisfy uniquely the moment condition

$$E[g^{\alpha}(x, \alpha_0)] = 0.$$

2.2 Sample Structure

Let z_i, $(i = 1, ..., n_\beta)$, and x_j, $(j = 1, ..., n_\alpha)$, denote samples of i.i.d. observations on the data vectors z and x respectively. An additional i.i.d. sample of observations on z, z_k, $(k = 1, ..., n)$, is also assumed to be available. This second sample of observations on z is used to obtain the preliminary consistent estimator for β required to estimate the efficient GMM metric. We identify the indices i, j and k uniquely with these respective samples throughout the paper.

This sampling structure is sufficiently general to permit consideration of a number of scenarios of interest, including the various examples outlined in the introduction. Firstly, sample-splitting schemes are allowed by defining the samples z_i, $(i = 1, ..., n_\beta)$, and z_k, $(k = 1, ..., n)$, to be independent. Secondly, situations in which these samples are identical may be addressed by setting $k = i$, $(i = 1, ..., n_\beta)$, which allows generated regressors such as a sample selectivity correction to be considered in our analysis. Our framework also allows for the possibility that the nuisance parameter estimator for α is obtained from a sample which is either independent of or identical to the sample of observations z_i, $(i = 1, ..., n_\beta)$, the latter case obtained by setting $x = z$ and $j = i$, $(i = 1, ..., n_\beta)$.

2.3 GMM and GEL Estimation of α_0

Initially, we describe a two-step GMM estimator of the nuisance parameter α due to Hansen (1982). Let

$$g_j^\alpha(\alpha) \equiv g^\alpha(x_j, \alpha), \hat{g}^\alpha(\alpha) \equiv \sum_{j=1}^{n_\alpha} g_j^\alpha(\alpha)/n_\alpha.$$

A preliminary estimator for α_0 is given by $\tilde{\alpha} = \arg\min_{\alpha \in \mathcal{A}} \hat{g}^\alpha(\alpha)'(\hat{W}^{\alpha\alpha})^{-1}\hat{g}^\alpha(\alpha)$ where $\mathcal{A}$ denotes the parameter space, and $\hat{W}^{\alpha\alpha} = W^{\alpha\alpha} + \sum_{j=1}^{n_\alpha} \xi^\alpha(x_j)/n_\alpha + O_p(n_\alpha^{-1})$ with $W^{\alpha\alpha}$ positive definite and $E[\xi^\alpha(x)] = 0$. The two-step GMM estimator is one that satisfies

$$\hat{\alpha}_{2S} = \arg\min_{\alpha \in \mathcal{A}} \hat{g}^\alpha(\alpha)'[\hat{\Omega}^{\alpha\alpha}(\tilde{\alpha})]^{-1}\hat{g}^\alpha(\alpha), \quad (2.1)$$

where $\hat{\Omega}^{\alpha\alpha}(\alpha) \equiv \sum_{j=1}^{n_\alpha} g_j^\alpha(\alpha) g_j^\alpha(\alpha)'/n_\alpha$.

We also examine as alternatives to GMM generalized empirical likelihood (GEL) estimators, as in Smith (1997, 2001); see also Newey and Smith (2002). Let $\varphi = (\alpha', \mu')'$

where μ is a m_α-vector of auxiliary parameters, $\rho^\varphi(.)$ be a function that is concave on its domain, which is an open interval $\mathcal{V}_\alpha$ containing zero, and $\rho^\varphi_v(.)$, $\rho^\varphi_{vv}(.)$ and $\rho^\varphi_{vvv}(.)$ denote first, second and third derivatives of $\rho^\varphi(.)$ respectively. Without loss of generality we normalise the first and second order derivatives of $\rho^\varphi_v(.)$ at 0 as $\rho^\varphi_v(0) = \rho^\varphi_{vv}(0) = -1$. Let $\hat{\Lambda}^\alpha_{n_\alpha}(\alpha) = \{\mu : \mu' g^\alpha_j(\alpha) \in \mathcal{V}_\alpha, j = 1, ..., n_\alpha\}$.

The GEL estimation criterion is

$$\hat{P}^\varphi(\varphi) = \sum_{j=1}^{n_\alpha} \rho^\varphi(\mu' g^\alpha_j(\alpha))/n_\alpha. \tag{2.2}$$

Then a GEL estimator for α_0 is obtained as the solution to the saddle point problem

$$\hat{\alpha}_{GEL} = \arg\min_{\alpha \in \mathcal{A}} \sup_{\mu \in \hat{\Lambda}^\alpha_{n_\alpha}(\alpha)} \hat{P}^\varphi(\varphi). \tag{2.3}$$

The GEL criterion (2.2) admits a number of estimators as special cases: empirical likelihood (EL) with $\rho^\varphi(v) = \log(1 - v)$, [Imbens (1997) and Qin and Lawless (1994)], exponential tilting (ET) with $\rho^\varphi(v) = -\exp(v)$, [Imbens, Spady, and Johnson (1998) and Kitamura and Stutzer (1997)], continuous updating (CUE) with $\rho^\varphi(v)$ quadratic and $\rho^\varphi_v(0) \neq 0$ and $\rho^\varphi_{vv}(0) < 0$ [Hansen, Heaton, and Yaron (1996)] and the Cressie-Read (1984) power family $\rho^\varphi(v) = -(1 + \gamma v)^{(\gamma+1)/\gamma}/(\gamma + 1)$ for some scalar γ. See Newey and Smith (2001) for further discussion.

Let $\hat{\alpha}$ denote a consistent estimator for α_0 obtained as described above in (2.1) or (2.3).

2.4 GMM and GEL Estimation of β_0

Let

$$g^\beta_i(\alpha, \beta) \equiv g^\beta(z_i, \alpha, \beta),\ \hat{g}^\beta(\alpha, \beta) \equiv \sum_{i=1}^{n_\beta} g^\beta_i(\alpha, \beta)/n_\beta.$$

A two-step GMM estimator of β is obtained using $\hat{\alpha}$ as a plug-in estimator of α in $\hat{g}^\beta(\alpha, \beta)$. The second sample of observations on z, z_k, $(k = 1, ..., n)$, is used to obtain a preliminary consistent estimator $\tilde{\beta}$ for β_0 defined by $\tilde{\beta} = \arg\min_{\beta \in \mathcal{B}} \sum_{k=1}^n g^\beta_k(\hat{\alpha}, \beta)'(\hat{W}^{\beta\beta})^{-1} \sum_{k=1}^n g^\beta_k(\hat{\alpha}, \beta)$ where $\mathcal{B}$ denotes the parameter space, $g^\beta_k(\alpha, \beta) = g^\beta(z_k, \alpha, \beta)$, $(k = 1, ..., n)$. Similar to above, it is assumed that $\hat{W}^{\beta\beta} = W^{\beta\beta} + \sum_{i=1}^{n_\beta} \xi^\beta(z_i)/n_\beta + O_p(n_\beta^{-1})$ with $W^{\beta\beta}$ positive definite and

$E[\xi^{\beta}(z)] = 0$. This second sample is also used to estimate a GMM metric which has generic form

$$\hat{\Omega}^{\beta\beta}(\alpha, \beta) \equiv \sum_{k=1}^{n} g_k^{\beta}(\alpha, \beta) g_k^{\beta}(\alpha, \beta)'/n.$$

This structure for the GMM metric allows a number of important special cases. Sample-splitting schemes are included by specifying the samples z_i, $(i = 1, ..., n_\beta)$, and z_k, $(k = 1, ..., n)$, to be mutually independent. A set-up in which these samples are identical is permitted. Hence, generated regressors are a special case of our analysis. Our framework also allows the nuisance parameter estimator $\hat{\alpha}$ to be obtained from either an independent or the same sample of observations; in the latter case, we define $x = z$ and $k = i$, $(i = 1, ..., n_\beta)$. See section 3 for further details of these particular specialisations.

The two-step GMM estimator for β_0 is one that satisfies

$$\hat{\beta}_{2S} = \arg\min_{\beta \in \mathcal{B}} \hat{g}^{\beta}(\hat{\alpha}, \beta)'[\hat{\Omega}^{\beta\beta}(\hat{\alpha}, \tilde{\beta})]^{-1} \hat{g}^{\beta}(\hat{\alpha}, \beta). \tag{2.4}$$

For GEL estimators of β_0, let $\theta = (\beta', \lambda')'$ where λ is a m_β-vector of auxiliary parameters, $\rho^{\theta}(.)$ be a function that is concave on its domain, which is an open interval $\mathcal{V}_\beta$ containing zero, and $\rho_v^{\theta}(.)$, $\rho_{vv}^{\theta}(.)$ and $\rho_{vvv}^{\theta}(.)$ denote first, second and third derivatives of $\rho^{\theta}(.)$ respectively. As above we normalise $\rho_v^{\theta}(0) = \rho_{vv}^{\theta}(0) = -1$. Let $\hat{\Lambda}_{n_\beta}^{\beta}(\beta) = \{\lambda : \lambda' g_i^{\beta}(\hat{\alpha}, \beta) \in \mathcal{V}_\beta, i = 1, ..., n_\beta\}$.

When the samples z_i, $(i = 1, ..., n_\beta)$, and z_k, $(k = 1, ..., n)$, are mutually independent we assume that they are pooled for GEL estimation. Let $N = n_\beta + n$ and define $n_* = n_\beta$ if the samples z_i, $(i = 1, ..., n_\beta)$, and z_k, $(k = 1, ..., n)$, are identical and $n_* = N$ if they are independent. The GEL estimation criterion is then

$$\hat{P}^{\theta}(\hat{\alpha}, \theta) = \sum_{i=1}^{n_*} \rho^{\theta}(\lambda' g_i^{\beta}(\hat{\alpha}, \beta))/n_*. \tag{2.5}$$

A GEL estimator for β_0 is obtained as the solution to the saddle point problem

$$\hat{\beta}_{GEL} = \arg\min_{\beta \in \mathcal{B}} \sup_{\lambda \in \hat{\Lambda}_{n_\beta}^{\beta}(\beta)} \hat{P}^{\theta}(\hat{\alpha}, \theta). \tag{2.6}$$

Let $\hat{\lambda}_{GEL} = \sup_{\lambda \in \hat{\Lambda}_{n_\beta}^{\beta}(\beta)} \hat{P}^{\theta}(\hat{\alpha}, \hat{\beta}_{GEL}, \lambda)$.

3 Higher Order Asymptotic Properties

Before detailing the various cases delineated in section 2, we discuss the asymptotic bias of estimators $\hat{\alpha}_{2S}$ or $\hat{\alpha}_{GEL}$ for the nuisance parameter α. We use the generic notation $\hat{\alpha}$ for $\hat{\alpha}_{2S}$ or $\hat{\alpha}_{GEL}$ where there is no possibility of confusion.

3.1 The Asymptotic Bias of the Nuisance Parameter Estimator

Let $g_j^\alpha = g_j^\alpha(\alpha_0)$, $G_j^\alpha(\alpha) = \partial g_j^\alpha(\alpha)/\partial\alpha'$, $G_j^\alpha = G_j^\alpha(\alpha_0)$ and

$$\begin{aligned} G^\alpha &= E[G_j^\alpha],\ \Omega^{\alpha\alpha} = E[g_j^\alpha g_j^{\alpha\prime}],\ \Sigma^{\alpha\alpha} = (G^{\alpha\prime}(\Omega^{\alpha\alpha})^{-1}G^\alpha)^{-1}, \\ H^\alpha &= \Sigma^{\alpha\alpha}G^{\alpha\prime}(\Omega^{\alpha\alpha})^{-1}, P^\alpha = (\Omega^{\alpha\alpha})^{-1} - (\Omega^{\alpha\alpha})^{-1}G^\alpha\Sigma^{\alpha\alpha}G^{\alpha\prime}(\Omega^{\alpha\alpha})^{-1}. \end{aligned}$$

Under conditions stated in Newey and Smith (2002, Theorems 3.3 and 3.4), both two-step GMM and GEL estimators for α admit stochastic expansions of the form

$$\hat{\alpha} = \alpha_0 + \tilde{\psi}^\alpha/\sqrt{n_\alpha} + (M_\alpha^\varphi)^{-1}[\tilde{A}^\varphi\tilde{\psi}^\varphi + \sum_{r=1}^{q_\varphi}\tilde{\psi}_r^\varphi M_r^\varphi\tilde{\psi}^\varphi/2]/n_\alpha + O_p(n_\alpha^{-3/2}),$$

where $\psi_j^\alpha = -H^\alpha g_j^\alpha$, $\psi_j^\varphi = -[H^{\alpha\prime}, P^\alpha]'g_j^\alpha$, $\tilde{\psi}^\alpha = \sum_{j=1}^{n_\alpha}\psi_j^\alpha/\sqrt{n_\alpha}$, $\tilde{\psi}^\varphi = \sum_{j=1}^{n_\alpha}\psi_j^\varphi/\sqrt{n_\alpha}$ and $\tilde{A}^\varphi = \sum_{j=1}^{n_\alpha}A_j^\varphi/\sqrt{n_\alpha}$. The matrix $(M_\alpha^\varphi)^{-1} = (\Sigma^{\alpha\alpha}, -H^\alpha)$ and the matrices M^φ and $\tilde{A}^\varphi$ are defined by analogy with $M_{\theta\theta}^\theta$ and $\tilde{A}^\theta$ given in eqs. (A.1) and (A.2) of Appendix A.

For GMM, to $O(n_\alpha^{-3/2})$,

$$\begin{aligned} Bias(\hat{\alpha}_{2S}) = {} & H^\alpha(-a_\alpha + E[G_j^\alpha H^\alpha g_j^\alpha])/n_\alpha - \Sigma^{\alpha\alpha}E[G_j^{\alpha\prime}P^\alpha g_j^\alpha]/n_\alpha \\ & + H^\alpha[g_j^\alpha g_j^{\alpha\prime}P^\alpha g_j^\alpha]/n_\alpha \\ & - H^\alpha(E[G_j^\alpha H_W^\alpha\Omega^{\alpha\alpha}P^\alpha g_j^\alpha] + E[g_j^\alpha tr(G_j^\alpha H_W^\alpha\Omega^{\alpha\alpha}P^\alpha)])/n_\alpha, \end{aligned}$$

where $H_W^\alpha = (G^{\alpha\prime}W^{-1}G^\alpha)^{-1}G^{\alpha\prime}W^{-1}$ and a_α is an m-vector such that

$$a_{\alpha s} \equiv tr(\Sigma^{\alpha\alpha}E[\partial^2 g_{js}^\alpha(\alpha_0)/\partial\alpha\partial\alpha'])/2, (s = 1, ..., m_\alpha), \tag{3.1}$$

where $g_{js}^\alpha(\alpha)$ denotes the sth element of $g_j^\alpha(\alpha)$. See Newey and Smith (2002, Theorem 4.1).

For GEL, to $O(n_\alpha^{-3/2})$,

$$\begin{aligned} Bias(\hat{\alpha}_{GEL}) = {} & H^\alpha(-a_\alpha + E[G_j^\alpha H^\alpha g_j^\alpha])/n_\alpha \\ & + [1 + (\rho_{vvv}^\varphi(0)/2)]H^\alpha E[g_j^\alpha g_j^{\alpha\prime}P^\alpha g_j^\alpha]/n_\alpha. \end{aligned}$$

See Newey and Smith (2002, Theorem 4.2). If $\rho^{\varphi}_{vvv}(0) = -2$, then the asymptotic bias of $\hat{\alpha}_{GEL}$ is identical to that of an infeasible GMM estimator with optimal linear combination of moment indicators $G^{\alpha\prime}(\Omega^{\alpha\alpha})^{-1}g_j^{\alpha}(\alpha)$, a condition which is satisfied by the EL estimator; see Newey and Smith (2002, Corollary 4.3). Moreover, this property is shared by any GEL estimator when third moments are zero, $E[g_{js}^{\alpha}g_j^{\alpha}g_j^{\alpha\prime}] = 0$, $(s = 1, ..., m_{\alpha})$; see Newey and Smith (2002, Corollary 4.4).

To describe the results, let $g_i^{\beta} = g_i^{\beta}(\alpha_0, \beta_0)$, $G_{\beta i}^{\beta}(\alpha, \beta) = \partial g_i^{\beta}(\alpha, \beta)/\partial\beta'$, $G_{\beta i}^{\beta} = G_{\beta i}^{\beta}(\alpha_0, \beta_0)$,

$$\begin{aligned}\Omega^{\beta\beta} &= E[g_i^{\beta}g_i^{\beta\prime}],\ G_{\beta}^{\beta} = E[G_{\beta i}^{\beta}],\ \Sigma^{\beta\beta} = (G_{\beta}^{\beta\prime}(\Omega^{\beta\beta})^{-1}G_{\beta}^{\beta})^{-1},\\ H^{\beta} &= \Sigma^{\beta}G_{\beta}^{\beta\prime}(\Omega^{\beta\beta})^{-1},\ P^{\beta} = (\Omega^{\beta\beta})^{-1} - (\Omega^{\beta\beta})^{-1}G_{\beta}^{\beta}\Sigma^{\beta\beta}G_{\beta}^{\beta\prime}(\Omega^{\beta\beta})^{-1}.\end{aligned}$$

We define a_{β} as an m-vector such that

$$a_{\beta r} = tr(\Sigma^{\beta\beta}E[\partial^2 g_{ir}^{\beta}/\partial\beta\partial\beta'])/2,\ (r = 1, ..., m_{\beta}).$$

Also let $G_{\alpha i}^{\beta}(\alpha, \beta) = \partial g_i^{\beta}(\alpha, \beta)/\partial\alpha'$, $G_{\alpha i}^{\beta} = G_{\alpha i}^{\beta}(\alpha_0, \beta_0)$, $G_{\alpha}^{\beta} = E[G_{\alpha i}^{\beta}]$ and

$$\Sigma_W^{\beta\beta} = (G_{\beta}^{\beta\prime}(W^{\beta\beta})^{-1}G_{\beta}^{\beta})^{-1}, H_W^{\beta} = \Sigma_W^{\beta\beta}G_{\beta}^{\beta\prime}(W^{\beta\beta})^{-1}.$$

3.2 Independent Samples

In this case, z_i, $(i = 1, ..., n_{\beta})$, x_j, $(j = 1, ..., n_{\alpha})$, and z_k, $(k = 1, ..., n)$, are independent i.i.d. samples of observations on the variables z and x. We assume that α is estimated by $\hat{\alpha}_{2S}$ or $\hat{\alpha}_{GEL}$ as described in section 2.

The precise form of the bias requires some additional notation. Let $a_{\beta\beta}^{\beta}$, $a_{\beta\alpha}^{\beta}$ and $a_{\alpha\alpha}^{\beta}$ be m_{β}-vectors such that

$$\begin{aligned}a_{\beta\beta r}^{\beta} &= tr(H^{\beta}G_{\alpha}^{\beta}\Sigma^{\alpha\alpha}G_{\alpha}^{\beta\prime}H^{\beta\prime}E[\partial^2 g_{ir}^{\beta}/\partial\beta\partial\beta'])/2,\ a_{\beta\alpha r}^{\beta} = -tr(H^{\beta}G_{\alpha}^{\beta}\Sigma^{\alpha\alpha}E[\partial^2 g_{ir}^{\beta}/\partial\alpha\partial\beta']),\\ a_{\alpha\alpha r}^{\beta} &= tr(\Sigma^{\alpha\alpha}E[\partial^2 g_{ir}^{\beta}/\partial\alpha\partial\alpha'])/2, (r = 1, ..., m_{\beta}).\end{aligned}$$

and $c_{\beta\beta}^{\beta}$ and $c_{\beta\alpha}^{\beta}$ are p_{β}-vectors with elements

$$\begin{aligned}c_{\beta\beta r}^{\beta} &= tr(E[\partial^2 g_i^{\beta\prime}/\partial\beta\partial\beta_r]P^{\beta}G_{\alpha}^{\beta}\Sigma^{\alpha\alpha}G_{\alpha}^{\beta\prime}H^{\beta\prime}),\\ c_{\beta\alpha r}^{\beta} &= -tr(E[\partial^2 g_i^{\beta\prime}/\partial\alpha\partial\beta_r]P^{\beta}G_{\alpha}^{\beta}\Sigma^{\alpha\alpha}),\ (r = 1, ..., p_{\beta}).\end{aligned}$$

For the two-step GMM estimator $\hat{\beta}_{2S}$, let

$$Bias_{\alpha_0}(\hat{\beta}_{2S}) = H^{\beta}(-a_{\beta} + E[G^{\beta}_{\beta i}H^{\beta}g^{\beta}_{i}])/n_{\beta} - \Sigma^{\beta\beta}E[G^{\beta\prime}_{\beta i}P^{\beta}g^{\beta}_{i}]/n_{\beta}.$$

This asymptotic bias corresponds to that for $\hat{\beta}_{2S}$ when α_0 and $\Omega^{\beta\beta}$ are known. For GEL estimation the samples z_i, $(i = 1, ..., n_{\beta})$, and z_k, $(k = 1, ..., n)$, are pooled. Hence,

$$\begin{aligned} Bias_{\alpha_0}(\hat{\beta}_{GEL}) &= H^{\beta}(-a_{\beta} + E[G^{\beta}_{\beta i}H^{\beta}g^{\beta}_{i}])/N \\ &\quad +[1 + (\rho^{\theta}_{vvv}(0)/2)]H^{\beta}E[g^{\beta}_{i}g^{\beta\prime}_{i}P^{\beta}g^{\beta}_{i}]/N, \end{aligned}$$

where $N = n + n_{\beta}$, which is the asymptotic bias for $\hat{\beta}_{GEL}$ after pooling when α_0 is known. See Newey and Smith (2002, Theorems 4.1 and 4.2).

The remainders in the following results are $O(\max[n^{-3/2}, n_{\alpha}^{-3/2}, n_{\beta}^{-3/2}])$ for GMM and $O(\max[N^{-3/2}, n_{\alpha}^{-3/2}])$ for GEL.

For GMM:

Theorem 3.1: *To $O(\max[n^{-3/2}, n_{\alpha}^{-3/2}, n_{\beta}^{-3/2}])$, if z_i, $(i = 1, ..., n_{\beta})$, x_j, $(j = 1, ..., n_{\alpha})$, and z_k, $(k = 1, ..., n)$, are independent samples, the asymptotic bias of the two-step GMM estimator is*

$$\begin{aligned} Bias(\hat{\beta}_{2S}) &= Bias_{\alpha_0}(\hat{\beta}_{2S}) - H^{\beta}G^{\beta}_{\alpha}Bias(\hat{\alpha}) \\ &\quad H^{\beta}(-a^{\beta}_{\beta\beta} - a^{\beta}_{\beta\alpha} - a^{\beta}_{\alpha\alpha})/n_{\alpha} - \Sigma^{\beta\beta}(-c^{\beta}_{\beta\beta} - c^{\beta}_{\beta\alpha})/n_{\alpha} \\ &\quad -H^{\beta}(E[G^{\beta}_{\beta i}H^{\beta}_{W}G^{\beta}_{\alpha}\Sigma^{\alpha\alpha}G^{\beta\prime}_{\alpha}P^{\beta}g^{\beta}_{i}] + E[g^{\beta}_{i}tr(G^{\beta}_{\beta i}H^{\beta}_{W}G^{\beta}_{\alpha}\Sigma^{\alpha\alpha}G^{\beta\prime}_{\alpha}P^{\beta})])/n \\ &\quad +H^{\beta}(E[G^{\beta}_{\alpha i}\Sigma^{\alpha\alpha}G^{\beta\prime}_{\alpha}P^{\beta}g^{\beta}_{i}] + E[g^{\beta}_{i}tr(G^{\beta}_{\alpha i}\Sigma^{\alpha\alpha}G^{\beta\prime}_{\alpha}P^{\beta})])/n_{\alpha}. \end{aligned}$$

As in Newey and Smith (2002), we may interpret the terms comprising the bias of the two step GMM estimator $\hat{\beta}_{2S}$. The first two terms of $Bias_{\alpha_0}(\hat{\beta}_{2S})$, which is the asymptotic bias for $\hat{\beta}_{2S}$ when α_0 and $\Omega^{\beta\beta}$ are known, are the bias that would arise from the (infeasible) optimal (variance minimising, Hansen, 1982) linear combination $G^{\beta\prime}_{\beta}(\Omega^{\beta\beta})^{-1}g^{\beta}(z, \alpha_0, \beta)$. The third term in $Bias_{\alpha_0}(\hat{\beta}_{2S})$ arises because of inefficient estimation of the Jacobian G^{β}_{β}. The second and third terms of $Bias(\hat{\beta}_{2S})$ reflect the presence of the nuisance parameter estimator $\hat{\alpha}$ in the (infeasible) linear combination $G^{\beta\prime}_{\beta}(\Omega^{\beta\beta})^{-1}g^{\beta}(z, \hat{\alpha}, \beta)$ whereas the fourth term arises because of the presence of $\hat{\alpha}$ in estimation of the Jacobian G^{β}_{β}. Likewise, the remaining terms are due to the presence of the nuisance parameter estimator $\hat{\alpha}$ used in the estimation of $\Omega^{\beta\beta}$. Overall therefore, the only role here for the preliminary two step GMM estimator $\tilde{\beta}$ in the estimation of $\Omega^{\beta\beta}$ is through $\hat{\alpha}$; cf. $\hat{\alpha}_{2S}$ above and Newey and Smith (2002). That is, if $g_k^\beta(\alpha,\beta) = g_k^\beta(\beta)$, $(k = 1, ..., n)$, these remaining terms vanish. If the GMM estimator is iterated at least once, H_W^β should be replaced by H^β.

We now turn to the bias formula for GEL based on the pooled samples z_i, $(i = 1, ..., n_\beta)$, and z_k, $(k = 1, ..., n)$.

Theorem 3.2: *To* $O(\max[N^{-3/2}, n_\alpha^{-3/2}])$, *where* $N = n_\beta + n$, *if* z_i, $(i = 1, ..., n_\beta)$, x_j, $(j = 1, ..., n_\alpha)$, *and* z_k, $(k = 1, ..., n)$, *are independent samples, the asymptotic bias of the GEL estimator is*

$$\begin{aligned} Bias(\hat{\beta}_{GEL}) &= Bias_{\alpha_0}(\hat{\beta}_{GEL}) - H^\beta G_\alpha^\beta Bias(\hat{\alpha}) \\ &+H^\beta(-a_{\beta\beta}^\beta - a_{\beta\alpha}^\beta - a_{\alpha\alpha}^\beta)/n_\alpha - \Sigma^{\beta\beta}(-c_{\beta\beta}^\beta - c_{\beta\alpha}^\beta)/n_\alpha \\ &+\Sigma^{\beta\beta}E[G_{\beta i}^{\beta\prime}P^\beta G_\alpha^\beta \Sigma^{\alpha\alpha}G_\alpha^{\beta\prime}P^\beta g_i^\beta]/n_\alpha \\ &+(\rho_{vvv}^\theta(0)/2)E[g_i^\beta g_i^{\beta\prime}P^\beta G_\alpha^\beta\Sigma^{\alpha\alpha}G_\alpha^{\beta\prime}P^\beta g_i^\beta]/n_\alpha \\ &-H^\beta(E[G_{\beta i}^\beta H^\beta G_\alpha^\beta \Sigma^{\alpha\alpha}G_\alpha^{\beta\prime}P^\beta g_i^\beta] + E[g_i^\beta tr(G_{\beta i}^{\beta\prime}P^\beta G_\alpha^\beta\Sigma^{\alpha\alpha}G_\alpha^{\beta\prime}H^{\beta\prime})])/n_\alpha \\ &+H^\beta(E[G_{\alpha i}^\beta\Sigma^{\alpha\alpha}G_\alpha^{\beta\prime}P^\beta g_i^\beta] + E[g_i^\beta tr(G_{\alpha i}^{\beta\prime}P^\beta G_\alpha^\beta\Sigma^{\alpha\alpha})])/n_\alpha. \end{aligned}$$

The first four terms are similar to those for GMM. The fifth and sixth terms arise because of the presence of the nuisance parameter estimator $\hat{\alpha}$ in the implicit estimation of $\Omega^{\beta\beta}$ and its inefficient estimation; see Newey and Smith (2002, Theorem 2.3). The remaining terms are similar to those for GMM except that H_W^β is replaced by H^β and would coincide if the GMM estimator were iterated at least once. If $G_\alpha^\beta = 0$, which ensures that $\hat{\beta}_{GEL}$ is first order efficient and occurs, for example, if $g_i^\beta(\alpha,\beta)$ is linear in α, there is no effect due to the implicit estimation of $\Omega^{\beta\beta}$ except through $Bias_{\alpha_0}(\hat{\beta}_{GEL})$ and, except for this term, $Bias(\hat{\beta}_{GEL})$ and $Bias(\hat{\beta}_{2S})$ coincide.

From Theorem 3.2, all GEL estimators have the same bias when third moments are zero as $Bias_{\alpha_0}(\hat{\beta}_{GEL})$ is the same for all GEL estimators in this case. See Newey and Smith (2002, Corollary 4.4).

Corollary 3.1: *To* $O(\max[N^{-3/2}, n_\alpha^{-3/2}])$, *where* $N = n_\beta + n$, *if* z_i, $(i = 1, ..., n_\beta)$,

x_j, $(j = 1, ..., n_\alpha)$, *and* z_k, $(k = 1, ..., n)$, *are independent samples and* $E[g^\beta_{ir} g^\beta_i g^{\beta\prime}_i] = 0$, $(r = 1, ..., m_\beta)$, *then all GEL estimators possess identical asymptotic bias.*

We now specialise these results for a standard sample-splitting scheme. Here the nuisance parameter vector α is not present. The remainders in the following results are $O(\max[n^{-3/2}, n_\beta^{-3/2}])$ for GMM and $O(N^{-3/2})$ for GEL. The sample-split two-step GMM estimator for β is one that satisfies

$$\hat{\beta}_{2S} = \arg\min_{\beta \in \mathcal{B}} \hat{g}^\beta(\beta)' \hat{\Omega}^{\beta\beta}(\tilde{\beta})^{-1} \hat{g}^\beta(\beta),$$

where $\hat{\Omega}^{\beta\beta}(\beta) \equiv \sum_{k=1}^n g^\beta_k(\beta) g^\beta_k(\beta)'/n$.

For GMM we have the following result:

Corollary 3.2: *In the absence of nuisance parameters, to* $O(\max[n^{-3/2}, n_\beta^{-3/2}])$, *if* z_i, $(i = 1, ..., n_\beta)$, *and* z_k, $(k = 1, ..., n)$, *are independent samples, the asymptotic bias of the two-step GMM estimator is*

$$\begin{aligned} Bias(\hat{\beta}_{2S}) &= Bias_{\alpha_0}(\hat{\beta}_{2S}) \\ &= H^\beta(-a_\beta + E[G^\beta_{\beta i} H^\beta g^\beta_i])/n_\beta - \Sigma^{\beta\beta} E[G^{\beta\prime}_{\beta i} P^\beta g^\beta_i]/n_\beta. \end{aligned}$$

This asymptotic bias result is that in Newey and Smith (2002) when $\Omega^{\beta\beta}$ is known. In particular, it is clear that because of independent sampling comprising the sample-split scheme an inefficient preliminary estimator for β_0 may be used with no effect on asymptotic bias. However, there would be implications for higher order variance.

We now turn to the bias formula for GEL which uses the pooled sample z_i, $(i = 1, ..., n_\beta)$, and z_k, $(k = 1, ..., n)$.

Corollary 3.3: *In the absence of nuisance parameters, to* $O(N^{-3/2})$, *where* $N = n_\beta + n$, *if* z_i, $(i = 1, ..., n_\beta)$, *and* z_k, $(k = 1, ..., n)$, *are independent samples, the asymptotic bias of the GEL estimator is*

$$\begin{aligned} Bias(\hat{\beta}_{GEL}) &= Bias_{\alpha_0}(\hat{\beta}_{GEL}) \\ &= H^\beta(-a_\beta + E[G^\beta_{\beta i} H^\beta g^\beta_i])/N + [1 + (\rho^\theta_{vvv}(0)/2)] H^\beta E[g^\beta_i g^{\beta\prime}_i P^\beta g^\beta_i]/N. \end{aligned}$$

In comparison with the GMM bias, we find that the Jacobian term drops out, i.e. there is no asymptotic bias from estimation of the Jacobian. As noted in Newey and Smith (2002), the absence of bias from the Jacobian is due to its efficient estimation in the first-order conditions. However, the last term reflects the implicit inefficient estimation of the variance matrix $\Omega^{\beta\beta}$; see Newey and Smith (2002, Theorem 2.3). The deleterious effect of this term relative to GMM will be offset at least partially by the use of the expanded pooled sample size N. However, in certain circumstances this term can be eliminated altogether.

The following corollary is immediate from Newey and Smith (2002, Corollary 4.3).

Corollary 3.4: *In the absence of nuisance parameters, to* $O(N^{-3/2})$, *where* $N = n_\beta + n$, *if* z_i, $(i = 1, ..., n_\beta)$, *and* z_k, $(k = 1, ..., n)$, *are independent samples, then*

$$Bias(\hat{\beta}_{EL}) = H^\beta(-a_\beta + E[G^\beta_{\beta i} H^\beta g^\beta_i])/N.$$

EL uses an efficient second moment estimator which leads to the above result; see Newey and Smith (2002, Theorem 2.3). Thus, for EL the bias is exactly the same as that for the infeasible optimal GMM estimator with moment functions $G^{\beta\prime}_\beta(\Omega^{\beta\beta})^{-1}g^\beta(z,\beta)$. This same property would be shared by any GEL estimator with $\rho^\theta_{vvv}(0) = -2$. It will also be shared by any GEL estimator when third moments are zero as detailed in Corollary 3.1 above.

3.3 Identical Samples

In this case, the samples z_i, $(i = 1, ..., n_\beta)$, and z_k, $(k = 1, ..., n)$, coincide. Hence, the estimator $\hat{\Omega}^{\beta\beta}(\alpha, \beta)$ for $\Omega^{\beta\beta}$ is based on the sample z_i, $(i = 1, ..., n_\beta)$. That is, $k = i$, $n = n_\beta$ and now $\hat{\Omega}^{\beta\beta}(\alpha, \beta) = \sum_{i=1}^{n_\beta} g^\beta_i(\alpha, \beta) g^\beta_i(\alpha, \beta)'$. Moreover, the nuisance parameter estimator $\hat{\alpha}$ is also based on the same sample z_i, $(i = 1, ..., n_\beta)$. That is, the samples z_i, $(i = 1, ..., n_\beta)$, and x_j, $(j = 1, ..., n_\alpha)$, also coincide. So $x = z$, $j = i$ and $n_\alpha = n_\beta$. The remainders in the following results are thus $O(n_\beta^{-3/2})$.

Let $g^{\beta.\alpha}_i = g^\beta_i - G^\beta_\alpha H^\alpha g^\alpha_i$,

$$\Omega^{\beta\beta.\alpha\alpha} = E[g^{\beta.\alpha}_i g^{\beta.\alpha\prime}_i], \Omega^{\beta\beta.\alpha} = E[g^\beta_i g^{\beta.\alpha\prime}_i], \Omega^{\alpha\beta.\alpha} = E[g^\alpha_i g^{\beta.\alpha\prime}_i].$$

Also let $a^\beta_{\beta\beta}$, $a^\beta_{\beta\alpha}$ and $a^\beta_{\alpha\alpha}$ be m_β-vectors such that

$$\begin{aligned} a^\beta_{\beta\beta r} &= tr(H^\beta \Omega^{\beta\beta.\alpha\alpha} H^{\beta\prime} E[\partial^2 g^\beta_{ir}/\partial\beta\partial\beta'])/2, \ a^\beta_{\beta\alpha r} = tr(H^\alpha \Omega^{\alpha\beta.\alpha} H^{\beta\prime} E[\partial^2 g^\beta_{ir}/\partial\beta\partial\alpha']), \\ a^\beta_{\alpha\alpha r} &= tr(\Sigma^{\alpha\alpha} E[\partial^2 g^\beta_{ir}/\partial\alpha\partial\alpha'])/2, (r = 1, ..., m_\beta), \end{aligned}$$

and $c^{\beta}_{\beta\beta}$ and $c^{\beta}_{\beta\alpha}$ are p_{β}-vectors with elements

$$\begin{aligned} c^{\beta}_{\beta\beta r} &= tr(H^{\beta}\Omega^{\beta\beta.\alpha\alpha}P^{\beta}E[\partial^2 g^{\beta}_i/\partial\beta'\partial\beta_r]), \\ c^{\beta}_{\beta\alpha r} &= tr(H^{\alpha}\Omega^{\alpha\beta.\alpha}P^{\beta}E[\partial^2 g^{\beta}_i/\partial\alpha'\partial\beta_r]), \ (r = 1, ..., p_{\beta}). \end{aligned}$$

For GMM we have the following result:

Theorem 3.3: *To $O(n_{\beta}^{-3/2})$, if the samples z_i, $(i = 1, ..., n_{\beta})$, x_j, $(j = 1, ..., n_{\alpha})$, and z_k, $(k = 1, ..., n)$, are identical, the asymptotic bias of the two-step GMM estimator is*

$$\begin{aligned} Bias(\hat{\beta}_{2S}) &= -H^{\beta}G^{\beta}_{\alpha}Bias(\hat{\alpha}) \\ &+H^{\beta}(-a^{\beta}_{\beta\beta} - a^{\beta}_{\beta\alpha} - a^{\beta}_{\alpha\alpha} + E[G^{\beta}_{\beta i}H^{\beta}g^{\beta.\alpha}_i] + E[G^{\beta}_{\alpha i}H^{\alpha}g^{\alpha}_i])/n_{\beta} \\ &-\Sigma^{\beta\beta}(-c^{\beta}_{\beta\beta} - c^{\beta}_{\beta\alpha} + E[G^{\beta\prime}_{\beta i}P^{\beta}g^{\beta.\alpha}_i])/n_{\beta} \\ &+H^{\beta}E[g^{\beta}_i g^{\beta\prime}_i P^{\beta}g^{\beta.\alpha}_i]/n_{\beta} \\ &-H^{\beta}(E[G^{\beta}_{\beta i}H^{\beta}_W\Omega^{\beta\beta.\alpha\alpha}P^{\beta}g^{\beta}_i] + E[g^{\beta}_i tr(G^{\beta}_{\beta i}H^{\beta}_W\Omega^{\beta\beta.\alpha\alpha}P^{\beta})])/n_{\beta} \\ &-H^{\beta}(E[G^{\beta}_{\alpha i}H^{\alpha}\Omega^{\alpha\beta.\alpha}P^{\beta}g^{\beta}_i] + E[g^{\beta}_i tr(G^{\beta}_{\alpha i}H^{\alpha}\Omega^{\alpha\beta.\alpha}P^{\beta})])/n_{\beta}. \end{aligned}$$

If $\tilde{\beta}$ is iterated at least once, H^{β}_W is replaced by H^{β}. The second line arises because of the presence of the nuisance parameter estimator $\hat{\alpha}$ in the (infeasible) linear combination $G^{\beta\prime}_{\beta}\Omega^{\beta\beta-1}g^{\beta}(z, \alpha, \beta)$ and the third is due to the estimation of the Jacobian G^{β}_{β}. The remaining terms reflect using $\hat{\alpha}$ and $\tilde{\beta}$. The penultimate and final lines reflect estimation of $\Omega^{\beta\beta}$ using respectively the preliminary estimator $\tilde{\beta}$ and the nuisance parameter estimator $\hat{\alpha}$.

For GEL:

Theorem 3.4: *To $O(n_{\beta}^{-3/2})$, if the samples z_i, $(i = 1, ..., n_{\beta})$, x_j, $(j = 1, ..., n_{\alpha})$, and z_k, $(k = 1, ..., n)$, are identical, the asymptotic bias of the GEL estimator is*

$$\begin{aligned} Bias(\hat{\beta}_{GEL}) &= -H^{\beta}G^{\beta}_{\alpha}Bias(\hat{\alpha}) \\ &+H^{\beta}(-a^{\beta}_{\beta\beta} - a^{\beta}_{\beta\alpha} - a^{\beta}_{\alpha\alpha} + E[G^{\beta}_{\beta i}H^{\beta}g^{\beta.\alpha}_i] + E[G^{\beta}_{\alpha i}H^{\alpha}g^{\alpha}_i])/n_{\beta} \\ &-\Sigma^{\beta\beta}(-c^{\beta}_{\beta\beta} - c^{\beta}_{\beta\alpha} + E[G^{\beta\prime}_{\beta i}P^{\beta}(\Omega^{\beta\beta} - \Omega^{\beta\beta.\alpha\alpha})P^{\beta}g^{\beta.\alpha}_i])/n_{\beta} \\ &+H^{\beta}(E[g^{\beta}_i g^{\beta\prime}_i P^{\beta}g^{\beta.\alpha}_i] + (\rho^{\theta}_{vvv}(0)/2)E[g^{\beta}_i g^{\beta\prime}_i P^{\beta}\Omega^{\beta\beta.\alpha\alpha}P^{\beta}g^{\beta}_i])/n_{\beta} \\ &-H^{\beta}(E[G^{\beta}_{\beta i}H^{\beta}\Omega^{\beta\beta.\alpha\alpha}P^{\beta}g^{\beta}_i] + E[g^{\beta}_i tr(G^{\beta}_{\beta i}H^{\beta}\Omega^{\beta\beta.\alpha\alpha}P^{\beta})])/n_{\beta} \\ &-H^{\beta}(E[G^{\beta}_{\alpha i}H^{\alpha}\Omega^{\alpha\beta.\alpha}P^{\beta}g^{\beta}_i] + E[g^{\beta}_i tr(G^{\beta}_{\alpha i}H^{\alpha}\Omega^{\alpha\beta.\alpha}P^{\beta})])/n_{\beta}. \end{aligned}$$

The terms in $Bias(\hat{\beta}_{GEL})$ are mostly identical to those for $\hat{\beta}_{2S}$. The major differences are the third line which reflects the inefficient estimation of the Jacobian term G^{β}_{β}. This term arises solely because of the presence of the nuisance parameter estimator $\hat{\alpha}$ and vanishes if the nuisance parameter is absent; see Newey and Smith (2002, Theorem 2.3). Other differences are, firstly, H^{β} in place of H^{β}_{W} in the penultimate line, a difference which is eliminated if two-step GMM is iterated once, and, secondly, the additional terms $\Sigma^{\beta\beta}E[G^{\beta\prime}_{\beta i}P^{\beta}\Omega^{\beta\beta.\alpha\alpha}P^{\beta}g^{\beta.\alpha}_{i}]$ and $(\rho^{\theta}_{vvv}(0)/2)E[g^{\beta}_{i}g^{\beta\prime}_{i}P^{\beta}\Omega^{\beta\beta.\alpha\alpha}P^{\beta}g^{\beta}_{i}]$ which arise through the implicit estimation of $\Omega^{\beta\beta}$ using both $\hat{\alpha}$ and $\hat{\beta}_{GEL}$.

From Theorem 3.4, all GEL estimators have the same bias when third moments are zero; cf. Corollary 3.1. See Newey and Smith (2002, Corollary 4.4).

Corollary 3.5: *To $O(n_{\beta}^{-3/2})$, if the samples z_i, $(i = 1, ..., n_{\beta})$, x_j, $(j = 1, ..., n_{\alpha})$, and z_k, $(k = 1, ..., n)$, coincide and $E[g^{\beta}_{ir}g^{\beta}_{i}g^{\beta\prime}_{i}] = 0$, $(r = 1, ..., m_{\beta})$, then all GEL estimators possess identical asymptotic bias.*

The above results in Theorems 3.3 and 3.4 may be specialised straightforwardly to deal with when z_i, $(i = 1, ..., n_{\beta})$, and x_j, $(j = 1, ..., n_{\alpha})$, are independent samples. In this case, $\Omega^{\beta\beta.\alpha\alpha} = \Omega^{\beta\beta} + G^{\beta}_{\alpha}\Sigma^{\alpha\alpha}G^{\beta\prime}_{\alpha}$, $\Omega^{\beta\beta.\alpha} = \Omega^{\beta\beta}$ and $\Omega^{\alpha\beta.\alpha} = -\Omega^{\alpha\alpha}H^{\alpha\prime}G^{\beta\prime}_{\alpha}$. Also, let $a^{\beta}_{\beta\beta}$, $a^{\beta}_{\beta\alpha}$, $a^{\beta}_{\alpha\alpha}$, $c^{\beta}_{\beta\beta}$ and $c^{\beta}_{\beta\alpha}$ be defined as in section 3.2; that is, $a^{\beta}_{\beta\beta}$, $a^{\beta}_{\beta\alpha}$ and $a^{\beta}_{\alpha\alpha}$ are m_{β}-vectors such that

$$\begin{aligned} a^{\beta}_{\beta\beta r} &= tr(H^{\beta}G^{\beta}_{\alpha}\Sigma^{\alpha\alpha}G^{\beta\prime}_{\alpha}H^{\beta\prime}E[\partial^2 g^{\beta}_{ir}/\partial\beta\partial\beta'])/2,\ a^{\beta}_{\beta\alpha r} = -tr(H^{\beta}G^{\beta}_{\alpha}\Sigma^{\alpha\alpha}E[\partial^2 g^{\beta}_{ir}/\partial\alpha\partial\beta']), \\ a^{\beta}_{\alpha\alpha r} &= tr(\Sigma^{\alpha\alpha}E[\partial^2 g^{\beta}_{ir}/\partial\alpha\partial\alpha'])/2, (r = 1, ..., m_{\beta}). \end{aligned}$$

and $c^{\beta}_{\beta\beta}$ and $c^{\beta}_{\beta\alpha}$ are p_{β}-vectors with elements

$$\begin{aligned} c^{\beta}_{\beta\beta r} &= tr(E[\partial^2 g^{\beta\prime}_{i}/\partial\beta\partial\beta_r]P^{\beta}G^{\beta}_{\alpha}\Sigma^{\alpha\alpha}G^{\beta\prime}_{\alpha}H^{\beta\prime}), \\ c^{\beta}_{\beta\alpha r} &= -tr(E[\partial^2 g^{\beta\prime}_{i}/\partial\alpha\partial\beta_r]P^{\beta}G^{\beta}_{\alpha}\Sigma^{\alpha\alpha}),\ (r = 1, ..., p_{\beta}). \end{aligned}$$

The remainders in the following corollaries are $O(\max[n_{\alpha}^{-3/2}, n_{\beta}^{-3/2}])$. Let

$$\begin{aligned} Bias_{\alpha_0}(\hat{\beta}_{2S}) &= H^{\beta}(-a_{\beta} + E[G^{\beta}_{\beta i}H^{\beta}g^{\beta}_{i}])/n_{\beta} - \Sigma^{\beta\beta}E[G^{\beta}_{\beta i}P^{\beta}g^{\beta}_{i}]/n_{\beta} \\ &\quad + H^{\beta}E[g^{\beta}_{i}g^{\beta}_{i}P^{\beta}g^{\beta}_{i}]/n_{\beta} \\ &\quad - H^{\beta}(E[G^{\beta}_{\beta i}H^{\beta}_{W}\Omega^{\beta\beta}P^{\beta}g^{\beta}_{i}] + E[g^{\beta}_{i}tr(G^{\beta}_{\beta i}H^{\beta}_{W}\Omega^{\beta\beta}P^{\beta})])/n_{\beta}, \end{aligned}$$

$$
\begin{aligned}
Bias_{\alpha_0}(\hat{\beta}_{GEL}) &= H^{\beta}(-a_{\beta} + E[G^{\beta}_{\beta i}H^{\beta}g^{\beta}_i])/n_{\beta} \\
&\quad +[1 + (\rho^{\theta}_{vvv}(0)/2)])H^{\beta}E[g^{\beta}_i g^{\beta}_i P^{\beta}g^{\beta}_i]/n_{\beta},
\end{aligned}
$$

which are the biases for $\hat{\beta}_{2S}$ and $\hat{\beta}_{GEL}$ when α_0 is known; see Newey and Smith (2002, Theorems 4.1 and 4.2).

Corollary 3.6: *To* $O(\max[n_{\alpha}^{-3/2}, n_{\beta}^{-3/2}])$, *if* z_i, $(i = 1, ..., n_{\beta})$, *and* x_j, $(j = 1, ..., n_{\alpha})$, *are independent samples and the samples* z_i, $(i = 1, ..., n_{\beta})$, *and* z_k, $(k = 1, ..., n)$, *are identical, the asymptotic bias of the two-step GMM estimator is*

$$
\begin{aligned}
Bias(\hat{\beta}_{2S}) &= Bias_{\alpha_0}(\hat{\beta}_{2S}) - H^{\beta}G^{\beta}_{\alpha}Bias(\hat{\alpha}) \\
&\quad +H^{\beta}(-a^{\beta}_{\beta\beta} - a^{\beta}_{\beta\alpha} - a^{\beta}_{\alpha\alpha})/n_{\alpha} - \Sigma^{\beta\beta}(-c^{\beta}_{\beta\beta} - c^{\beta}_{\beta\alpha})/n_{\alpha} \\
&\quad -H^{\beta}(E[G^{\beta\prime}_{\beta i}H^{\beta}_{W}G^{\beta}_{\alpha}\Sigma^{\alpha\alpha}G^{\beta\prime}_{\alpha}P^{\beta}g^{\beta}_i] + E[g^{\beta}_i tr(G^{\beta\prime}_{\beta i}H^{\beta}_{W}G^{\beta}_{\alpha}\Sigma^{\alpha\alpha}G^{\beta\prime}_{\alpha}P^{\beta})])/n_{\alpha} \\
&\quad +H^{\beta}(E[G^{\beta}_{\alpha i}\Sigma^{\alpha\alpha}G^{\beta\prime}_{\alpha}P^{\beta}g^{\beta}_i] + E[g^{\beta}_i tr(G^{\beta}_{\alpha i}\Sigma^{\alpha\alpha}G^{\beta\prime}_{\alpha}P^{\beta})])/n_{\alpha}.
\end{aligned}
$$

Corollary 3.7: *To* $O(\max[n_{\alpha}^{-3/2}, n_{\beta}^{-3/2}])$, *if* z_i, $(i = 1, ..., n_{\beta})$, *and* x_j, $(j = 1, ..., n_{\alpha})$, *are independent samples and the samples* z_i, $(i = 1, ..., n_{\beta})$, *and* z_k, $(k = 1, ..., n)$, *are identical, the asymptotic bias of the GEL estimator is*

$$
\begin{aligned}
Bias(\hat{\beta}_{GEL}) &= Bias_{\alpha_0}(\hat{\beta}_{GEL}) - H^{\beta}G^{\beta}_{\alpha}Bias(\hat{\alpha}) \\
&\quad +H^{\beta}(-a^{\beta}_{\beta\beta} - a^{\beta}_{\beta\alpha} - a^{\beta}_{\alpha\alpha})/n_{\alpha} - \Sigma^{\beta\beta}(-c^{\beta}_{\beta\beta} - c^{\beta}_{\beta\alpha})/n_{\alpha} \\
&\quad +\Sigma^{\beta\beta}E[G^{\beta\prime}_{\beta i}P^{\beta}G^{\beta}_{\alpha}\Sigma^{\alpha\alpha}G^{\beta\prime}_{\alpha}P^{\beta}g^{\beta}_i]/n_{\alpha} \\
&\quad +(\rho^{\theta}_{vvv}(0)/2)E[g^{\beta}_i g^{\beta\prime}_i P^{\beta}G^{\beta}_{\alpha}\Sigma^{\alpha\alpha}G^{\beta\prime}_{\alpha}P^{\beta}g^{\beta}_i])/n_{\alpha} \\
&\quad -H^{\beta}(E[G^{\beta}_{\beta i}H^{\beta}G^{\beta}_{\alpha}\Sigma^{\alpha\alpha}G^{\beta\prime}_{\alpha}P^{\beta}g^{\beta}_i] + E[g^{\beta}_i tr(G^{\beta}_{\beta i}H^{\beta}G^{\beta}_{\alpha}\Sigma^{\alpha\alpha}G^{\beta\prime}_{\alpha}P^{\beta})])/n_{\alpha} \\
&\quad +H^{\beta}(E[G^{\beta}_{\alpha i}\Sigma^{\alpha\alpha}G^{\beta\prime}_{\alpha}P^{\beta}g^{\beta}_i] + E[g^{\beta}_i tr(G^{\beta}_{\alpha i}\Sigma^{\alpha\alpha}G^{\beta\prime}_{\alpha}P^{\beta})])/n_{\alpha}.
\end{aligned}
$$

The representations given in Corollaries 3.6 and 3.7 are identical to those of Theorems 3.1 and 3.2 respectively. The only differences are in $Bias_{\alpha_0}(\hat{\beta}_{2S})$ and $Bias_{\alpha_0}(\hat{\beta}_{GEL})$. Here, because of the use of identical samples z_i, $(i = 1, ..., n_{\beta})$, and z_k, $(k = 1, ..., n)$, $Bias_{\alpha_0}(\hat{\beta}_{2S})$ additionally includes terms associated with the preliminary estimator $\tilde{\beta}$ and the estimation of $\Omega^{\beta\beta}$. For GEL, the only difference is the use of single sample n_{β} rather than the pooled sample $N = n_{\beta}+n$ when the samples z_i, $(i = 1, ..., n_{\beta})$, and z_k, $(k = 1, ..., n)$, are independent.

4 Simulation Experiments for Covariance Structure Models

Our investigation concerns models of covariance structure estimated on the same sample. Therefore, the asymptotic bias expressions in section 3.2 and, in particular, Theorems 3.3 and 3.4 apply. Altonji and Segal (1996) carried out an extensive analysis of the finite sample properties of GMM estimators for covariance structure models and found that the efficient two-step GMM estimator is severely downward biased in small samples for most distributions and in relatively large samples for "badly behaved" distributions. They argue that this poor performance is due to the correlation between the estimated second moments used to estimate the optimal weighting matrix and the moment indicators. Thus, as the theoretical results in section 3 reveal, both equally weighted GMM, which uses the identity matrix as weighting matrix, and efficient GMM estimation based on a sample-split estimator for the optimal weighting matrix produce parameter estimators with significantly improved properties in finite samples; see Theorem 3.3, Corollary 3.2 and also Horowitz (1998). The latter author also considered a bias-adjusted GMM estimator using the re-centred nonparametric bootstrap of Hall and Horowitz (1996) which is outlined below. This estimator, although biased in some cases, performed much better than the standard two-step GMM estimator.

The particular focus of attention of this section is GMM and GEL estimators for a common variance parameter constructed from a simulated panel data set in circumstances where the mean parameter is assumed unknown and is treated as a nuisance parameter. We initially consider the finite sample bias properties of the two-step GMM estimator, continuous updating estimator (CUE), exponential tilting (ET) and empirical likelihood (EL) estimators. We also examine analytical bias-adjustment methods for two-step GMM based on Theorem 3.3 and compare their finite sample properties with those of various forms of bootstrap bias-adjusted two-step GMM, both of which techniques achieve bias-adjustment of the two-step GMM estimator to the order of asymptotic approximation considered in this paper.

4.1 Bootstrap Bias-Adjustment

The generic form of bootstrap bias-adjustment for the two-step GMM estimator $\hat{\beta}_{2S}$ is as follows. The original data z_i, $(i = 1, ..., n_\beta)$, is sampled independently with replacement to yield a bootstrap sample of size n_β and a two-step GMM estimator thereby calculated from this bootstrap sample. This process is independently replicated. The bias of the two-step GMM estimator is estimated as the difference between the mean of the resultant bootstrap two-step GMM estimator empirical distribution and the two-step GMM estimator $\hat{\beta}_{2S}$. The bootstrap bias-adjusted two-step GMM estimator is then $\hat{\beta}_{2S}$ less the bias estimator.

We consider three forms of bootstrap bias-adjusted two-step GMM estimator. The first uses the standard non-parametric (NP) bootstrap. This resampling scheme applies equal weights $1/n_\beta$ to each observation z_i, $(i = 1, ..., n_\beta)$. That is, resampling is based on the empirical distribution function $F_{n_\beta}(z) = \sum_{i=1}^{n_\beta} 1(z_i \leq z)/n_\beta$, where $1(.)$ is an indicator function. Direct application of the NP bootstrap in the GMM framework seems to be unsatisfactory in many cases though. When the model is over-identified as in our experiments, while the population moment condition $E[g^\beta(z, \alpha_0, \beta_0)] = 0$ is satisfied, the estimated sample moments are typically non-zero, that is, there is typically no β such that $E_{F_{n_\beta}}[g^\beta(z, \hat{\alpha}, \beta)] = 0$ where $E_{F_{n_\beta}}[.]$ denotes expectation taken with respect to F_{n_β}. Therefore, F_{n_β} may be a poor approximation to the underlying distribution of the data and, hence, the NP bootstrap may not yield a substantial improvement over first-order asymptotic theory in standard applications of GMM. A second resampling scheme is the re-centred non-parametric (RNP) bootstrap; see Hall and Horowitz (1996). This method replaces the moment indicator $g^\beta(z, \hat{\alpha}, \beta)$ used in the GMM estimation criterion (2.4) by the re-centred moment indicator $g^{\beta *}(z, \hat{\alpha}, \beta) = g^\beta(z, \hat{\alpha}, \beta) - E_{F_{n_\beta}}[g^\beta(z, \hat{\alpha}, \hat{\beta}_{2S})]$. As $E_{F_{n_\beta}}[g^\beta(z, \hat{\alpha}, \hat{\beta}_{2S})] = \hat{g}^\beta(\hat{\alpha}, \hat{\beta}_{2S})$, this re-centring guarantees that the moment condition is satisfied with respect to F_{n_β}, that is, $E_{F_{n_\beta}}[g^{\beta *}(z, \hat{\alpha}, \hat{\beta}_{2S})] = 0$. Apart from the reformulation of the moment indicator, the RNP bootstrap is identical in execution to the NP bootstrap. The third bootstrap suggested by Brown, Newey and May (1997) employs an alternative empirical distribution to F_{n_β} for resampling which also ensures that the moment condition is satisfied. That is, the observations z_i, $(i = 1, ..., n_\beta)$, are assigned different rather than equal weights, the moment indicator $g^{\beta}(z, \hat{\alpha}, \beta)$ remaining unaltered. Given the two-step GMM estimator $\hat{\beta}_{2S}$, let $\hat{\lambda}_{2S} = \arg\sup_{\lambda \in \hat{\Lambda}^{\beta}_{n_\beta}(\hat{\beta}_{2S})} \hat{P}^{\theta}(\hat{\alpha}, \hat{\beta}_{2S}, \lambda)$, cf. (2.6). Each observation z_i is assigned the implied probability $\hat{\pi}_i^{2S} = \rho_v^{\theta}(k_\theta \hat{\lambda}_{2S}' g_i^{\beta}(\hat{\alpha}, \hat{\beta}_{2S})) / \sum_{j=1}^{n_\beta} \rho_v^{\theta}(k_\theta \hat{\lambda}_{2S}' g_j^{\beta}(\hat{\alpha}, \hat{\beta}_{2S}))$ associated with the two-step GMM estimator, $(i = 1, ..., n_\beta)$. The implied empirical distribution function $F_{n_\beta}^{GEL}(z) = \sum_{i=1}^{n_\beta} \hat{\pi}_i^{2S} 1(z_i \leq z)$ is thus obtained from the first step of a GEL estimation procedure and is denoted as (first-step GEL) FSGEL. From the first order conditions for GEL, the moment condition is satisfied with respect to $F_{n_\beta}^{GEL}$ as $\sum_{i=1}^{n_\beta} \hat{\pi}_i^{2S} g_i^{\beta}(\hat{\alpha}, \hat{\beta}_{2S}) = 0$ and, thus, $E_{F_{n_\beta}^{GEL}}[g^{\beta}(z, \hat{\alpha}, \hat{\beta}_{2S})] = 0$, where $E_{F_{n_\beta}}^{GEL}[.]$ denotes expectation taken with respect to $F_{n_\beta}^{GEL}$. We employ the EL criterion $\hat{P}^{\theta}(\hat{\alpha}, \hat{\beta}_{2S}, \lambda) = \sum_{i=1}^{n_\beta} \log(1 - \lambda' g_j^{\beta}(\hat{\alpha}, \hat{\beta}_{2S}))/n_\beta$ in our experiments. In the absence of nuisance parameters, the FSGEL bootstrap is asymptotically efficient relative to any bootstrap based on the empirical distribution function F_{n_β}, as shown by Brown, Newey and May (1997).

4.2 Analytical Bias-Adjustment

We also consider direct bias-adjustment of $\hat{\beta}_{2S}$ by subtraction of an estimator for $Bias(\hat{\beta}_{2S})$ given in Theorem 3.3; cf. Newey and Smith (2002, Theorem 5.1). We consider four forms of bias estimator. The first estimator for $Bias(\hat{\beta}_{2S})$, BCa, uses the empirical distribution function F_{n_β} for obtaining expectation estimators, i.e. functions of observation i are equally weighted by $1/n_\beta$, $(i = 1, ..., n_\beta)$. The second estimator, BCb, uses the FSGEL empirical distribution function $F_{n_\beta}^{GEL}$, i.e. functions of observation i are weighted by $\hat{\pi}_i^{2S}$, $(i = 1, ..., n_\beta)$. The third, BCc, uses F_{n_β} but with the true parameter values α_0 and β_0 substituted. The final estimator, BCd, employs the simulated counterpart of the expression for the asymptotic bias of $\hat{\beta}_{2S}$ given in Theorem 3.3.

4.3 Experimental Design

We consider an experimental design analyzed by Altonji and Segal (1996) where the objective is the estimation of a common population variance β_0 for a scalar random variable z_t, $(t = 1, ..., T)$, from observations on a balanced panel covering $T = 10$ time periods. Thus, $z = (z_1, ..., z_T)'$. We assume that n_β observations are available on z and that z_{ti}

is independent over t and i.i.d. over i. We consider the case where the mean α_0 of z is unknown. Hence, the results of section 3.2 apply. The nuisance parameter estimator is $\hat{\alpha} = (\hat{\alpha}_1, ..., \hat{\alpha}_T)'$, where the unbiased estimator $\hat{\alpha}_t = \sum_{i=1}^{n_\beta} z_{ti}/n_\beta$, $(t = 1, ..., T)$. The moment indicator vector is $g^\beta(z, \alpha, \beta) = m(z, \alpha) - \iota\beta$, where ι is a T-vector of units, $m(z, \alpha) = (m_1(z_1, \alpha_1), ..., m_T(z_T, \alpha_T))'$ and

$$m_t(z_t, \alpha_t) = n_\beta(z_t - \alpha_t)^2/(n_\beta - 1),\ (t = 1, ..., T).$$

Thus, $\hat{m}(\hat{\alpha}) = \sum_{i=1}^{n_\beta} m(z_i, \hat{\alpha})/n_\beta$ is an unbiased estimator for $\iota\beta_0$. Here $p_\beta = 1$, $m_\beta = T$ and $p_\alpha = m_\alpha = T$.

In this study, all observations z_{ti} are i.i.d. across both t and i although the common mean assumption is ignored in estimation. Although the elements of $\hat{m}(\hat{\alpha})$ are independent, the estimated variance matrix $\hat{\Omega}^{\beta\beta}(\hat{\alpha}, \tilde{\beta}) = \sum_{i=1}^{n_\beta} g_i^\beta(\hat{\alpha}, \tilde{\beta}) g_i^\beta(\hat{\alpha}, \tilde{\beta})'/n_\beta$ ignores this information. Seven different distributions for z_t, scaled to have mean $\alpha_0 = 0$ and variance $\beta_0 = 1$, were considered for two sample sizes $n_\beta = 100,\ 500$. In each experiment, 1000 replications were performed.

In this framework, the two-step GMM estimator is a weighted mean of the sample variances, $\hat{\beta}_{2S} = w'\hat{m}(\hat{\alpha}) = \sum_{i=1}^{n_\beta} w' m_i(\hat{\alpha})/n_\beta$, where $w = (\iota'\hat{\Omega}^{\beta\beta}(\hat{\alpha}, \tilde{\beta})^{-1}\iota)^{-1}\iota'\hat{\Omega}^{\beta\beta}(\hat{\alpha}, \tilde{\beta})^{-1}$. The preliminary estimator $\tilde{\beta}$ is obtained using equal weights ($w = \iota/T$). For GEL estimators, as $G_{\beta i}^\beta = -\iota$, it can be straightforwardly shown that $\hat{\beta}_{GEL} = n_\beta \sum_{i=1}^{n_\beta} \hat{\pi}_i \iota' m_i(\hat{\alpha})/T(n_\beta - 1)$ where $\hat{\pi}_i^{GEL} = \rho_v(\hat{\lambda}_{GEL}' g_i^\beta(\hat{\alpha}, \hat{\beta}_{GEL}))/\sum_{j=1}^{n_\beta} \rho_v(\hat{\lambda}_{GEL}' g_j^\beta(\hat{\alpha}, \hat{\beta}_{GEL}))$, $(i = 1, ..., n_\beta)$. The two-step GMM estimator ascribes equal weights over i whereas GEL applies the GEL implied probabilities $\hat{\pi}_i^{GEL}$. Over t, GMM assigns distinct weights given by the vector w while for GEL each time period receives an equal weight.

A number of important implications of this structure for the results of section 3.2 may be deduced. Firstly, as $G_{\alpha i}^\beta = -2n_\beta diag(z_{i1} - \alpha_1, ..., z_{iT} - \alpha_T)/(n_\beta - 1)$ and, thus, $G_\alpha^\beta = 0$, GMM or GEL estimators for β_0 are first order efficient. Secondly, as $G_{\beta i}^\beta = -\iota$ from the linearity of $g^\beta(z, \alpha, \beta)$ in β, substantial simplifications result in the asymptotic bias expressions of Theorems 3.3 and 3.4. In particular, it is evident from the asymptotic biases given in Theorems 3.3 and 3.4 that those for two-step and iterated GMM are identical and, moreover, that CUE also possesses an identical asymptotic bias.

To be more precise, for these experiments $a^{\beta}_{\beta\beta} = a^{\beta}_{\beta\alpha} = 0$ and $c^{\beta}_{\beta\beta} = c^{\beta}_{\beta\alpha} = 0$ from the linearity of $g^{\beta}(z, \alpha, \beta)$ in β. Also $a^{\beta}_{\alpha\alpha} = 2[n_{\beta}/(n_{\beta} - 1)]\iota\beta_0$. As $g^{\beta.\alpha}_i = g^{\beta}_i$, $\Omega^{\beta\beta.\alpha\alpha} = \Omega^{\beta\beta}$, $\Omega^{\beta\beta.\alpha} = \Omega^{\beta\beta}$ and $\Omega^{\alpha\beta.\alpha} = E[g^{\alpha}_i g^{\beta\prime}_i] = \Omega^{\alpha\beta}$. Therefore, from Theorems 3.3 and 3.4,

$$\begin{aligned} Bias(\hat{\beta}_{2S}) &= Bias_{\alpha_0}(\hat{\beta}_{2S}) \\ &\quad + H^{\beta}(-a^{\beta}_{\alpha\alpha} + E[G^{\beta}_{\alpha i} H^{\alpha} g^{\alpha}_i])/n_{\beta} \\ &\quad - H^{\beta}(E[G^{\beta}_{\alpha i} H^{\alpha} \Omega^{\alpha\beta} P^{\beta} g^{\beta}_i] + E[g^{\beta}_i tr(G^{\beta}_{\alpha i} H^{\alpha} \Omega^{\alpha\beta} P^{\beta})])/n_{\beta}, \end{aligned}$$

and

$$\begin{aligned} Bias(\hat{\beta}_{GEL}) &= Bias_{\alpha_0}(\hat{\beta}_{GEL}) \\ &\quad + H^{\beta}(-a^{\beta}_{\alpha\alpha} + E[G^{\beta}_{\alpha i} H^{\alpha} g^{\alpha}_i])/n_{\beta} \\ &\quad - H^{\beta}(E[G^{\beta}_{\alpha i} H^{\alpha} \Omega^{\alpha\beta} P^{\beta} g^{\beta}_i] + E[g^{\beta}_i tr(G^{\beta}_{\alpha i} H^{\alpha} \Omega^{\alpha\beta} P^{\beta})])/n_{\beta}. \end{aligned}$$

Therefore, there is no role for $Bias(\hat{\alpha})$. Moreover, $Bias(\hat{\beta}_{2S})$ and $Bias(\hat{\beta}_{GEL})$ only differ in $Bias_{\alpha_0}(\hat{\beta}_{2S})$ and $Bias_{\alpha_0}(\hat{\beta}_{GEL})$. Because $g^{\beta}(z, \alpha, \beta) = m(z, \alpha) - \iota\beta$ is linear in β and, thus, $G^{\beta}_{\beta i} = -\iota$ is non-stochastic, the asymptotic biases for $\hat{\beta}_{2S}$ and $\hat{\beta}_{GEL}$ when the nuisance parameter α_0 is known reduce to

$$\begin{aligned} Bias_{\alpha_0}(\hat{\beta}_{2S}) &= H^{\beta} E[g^{\beta}_i g^{\beta\prime}_i P^{\beta} g^{\beta}_i]/n_{\beta}, \\ Bias_{\alpha_0}(\hat{\beta}_{GEL}) &= [1 + (\rho^{\theta}_{vvv}(0)/2)] H^{\beta} E[g^{\beta}_i g^{\beta\prime}_i P^{\beta} g^{\beta}_i]/n_{\beta}. \end{aligned}$$

As there is no effect due to the preliminary estimator $\tilde{\beta}$, it is evident from $Bias(\hat{\beta}_{2S})$ that the asymptotic biases for the two-step GMM and iterated GMM estimators are identical. Moreover, from $Bias_{\alpha_0}(\hat{\beta}_{GEL})$, they also coincide with that of CUE as $\rho^{\theta}_{vvv}(0) = 0$. Furthermore, it is only the asymmetry of g^{β}_i which accounts for the differences in asymptotic biases between two-step GMM and other GEL estimators. Note that, apart from $-H^{\beta} a^{\beta}_{\alpha\alpha}/n_{\beta}$, the second and third lines in $Bias(\hat{\beta}_{2S})$ and $Bias(\hat{\beta}_{GEL})$ vanish if z_{ti} is symmetrically distributed; that is, $Bias(\hat{\beta}_{2S}) = Bias_{\alpha_0}(\hat{\beta}_{2S}) - H^{\beta} a^{\beta}_{\alpha\alpha}/n_{\beta}$ and $Bias(\hat{\beta}_{GEL}) = Bias_{\alpha_0}(\hat{\beta}_{GEL}) - H^{\beta} a^{\beta}_{\alpha\alpha}/n_{\beta}$. Furthermore, $Bias_{\alpha_0}(\hat{\beta}_{EL}) = 0$ and $Bias_{\alpha_0}(\hat{\beta}_{GEL}) = 0$ if $\rho^{\theta}_{vvv}(0) = -2$.

4.4 Results

The tables report estimated mean and median bias (as a percentage), 0.05 and 0.95 quantiles, standard error (SE), root mean squared error (RMSE) and median absolute error (MAE) of four asymptotically first-order equivalent methods for estimating moment condition models, two-step GMM (2S-GMM), CUE, ET and EL estimators.

Table 1 considers a sample size of $n_\beta = 100$. The results obtained for the two-step GMM estimator are very similar to those presented by Altonji and Segal (1996). As in their study, this estimator is clearly downward biased. This distortion is particularly marked for "badly-behaved" distributions, namely thicker-tailed symmetric (t_5) and long-tailed skewed (lognormal and exponential) distributions. As noted above, the asymptotic bias expressions for GMM and GEL involve further terms for asymmetric distributions. Note, however, that these expressions are not strictly valid for the t_5 distribution as moments of order greater than 4 do not exist. The worst case is given by the lognormal distribution, where the biases (MAE) are -0.415 and -0.430 (0.430). In this case the empirical 0.95 confidence interval does not cover the true value $\beta_0 = 1$.

Table 1 about here

Although, as noted above, the biases of GMM and CUE should be similar, Table 1 indicates that the results for CUE are in fact worse than for the two-step GMM estimator. Because the bias expressions for GMM and GEL only differ according to $Bias_{\alpha_0}(\hat{\beta}_{2S})$ and $Bias_{\alpha_0}(\hat{\beta}_{GEL})$, ET and EL estimators should display better finite sample properties relative to GMM and CUE. In particular, $Bias_{\alpha_0}(\hat{\beta}_{2S}) = 2Bias_{\alpha_0}(\hat{\beta}_{ET})$ and $Bias_{\alpha_0}(\hat{\beta}_{EL}) = 0$. While all methods have very similar standard errors (SE), the improvement for ET and EL in terms of both mean and median bias, root mean square error (RMSE) and mean absolute error (MAE) is clear. This is particularly marked for EL estimation. For ET, the improvements over GMM are rather more modest than those for EL as predicted by our theoretical results. However, although bias is not completely eliminated, especially for the skewed lognormal and exponential distributions, even for these cases, EL shows a marked improvement over two-step GMM.

Table 2 about here

Table 2 deals with the increased sample size $n_\beta = 500$. Overall, all estimators display less bias with reduced SE, RMSE and MAE. The general pattern across estimators revealed for the smaller sample size $n_\beta = 100$ is still apparent. CUE is somewhat worse than two-step GMM with ET delivering rather moderate improvements while EL dominates all other estimators in terms of mean and median biases, RMSE and MAE. For the skewed distributions, lognormal and exponential, EL offers substantially reduced bias, RMSE and MAE relative to other estimators including ET with very little or no increase in SE. For a number of the symmetric distributions, EL is able to eliminate bias more or less entirely.

Table 3 about here

The results reported in Table 3 with $n_\beta = 100$ use 100 bootstrap samples in each replication. In all cases, the bootstrap methods substantially reduce the bias of the two-step GMM estimator, although at the expense of a rather modest increase in SE. RMSE and MAE are also reduced, also quite substantially in the asymmetric cases for the RNP and FSGEL bootstrap methods. Clearly, the gain from bias reduction outweighs the increased contribution of SE to RMSE. The behaviour of these methods is not uniform, however, but overall the performances of RNP and FSGEL seem quite similar. It appears that RNP and FSGEL are rather better than NP which may be accounted for by the sample moments evaluated at the two-step GMM estimator being far from zero in these experiments. The performance of the feasible bias adjustment methods BCa and BCb is also quite encouraging leading to a substantial reduction in bias relative to $\hat{\beta}_{2S}$ in the "badly behaved" cases with BCb tending to dominate BCa. Like the bootstrap methods, SE increases somewhat for the analytical methods but again is less important compared to bias reduction for RMSE which in some cases is also reduced by a non-trivial amount. The results for BCc and BCd indicate that the theoretical expression for asymptotic bias in Theorem 3.3 accounts for the vast majority of finite sample bias. Comparing bootstrap and bias adjustment methods, BCb is rather similar to RNP and FSGEL in most cases in terms of bias reduction, RMSE and MAE. Therefore, BCb appears to be an efficacious rival to bootstrap methods.

Table 4 about here

Similar qualitative conclusions may be drawn from Table 4 for $n_\beta = 500$ with two-step GMM bias being more or less eliminated for a number of symmetric distributions. Again, for the "badly behaved" cases, bias is not eliminated entirely but is reduced substantially by RNP, FSGEL bootstrap bias-adjustment methods and the analytical approach BCb.

5 Conclusions

The context of this paper is the estimation of moment condition models in situations where the moment indicator depends on a nuisance parameter. The particular concern is the analysis of the higher-order bias of GMM and GEL estimators when a plug-in estimator is employed for the nuisance parameter. Such an environment covers a number of cases of interest including the use of generated regressors and sample-splitting methods. Expressions for the higher-order bias of these estimators is obtained in a general framework which allows specialisation to cases when the nuisance parameter is estimated from either an identical or an independent sample.

The efficacy of these asymptotic bias expressions is explored in a number of simulation experiments for covariance structure models. A rather pleasing conclusion from these experiments is that the mean and median bias, root mean squared error and mean absolute error properties of empirical likelihood represent a substantial improvement of those of two-step GMM, continuous updating and exponential tilting estimators with little or no increase in variance. Further experiments comparing various bootstrap bias-adjustment methods with those based on estimated analytical asymptotic bias expressions indicate that the less computationally intensive analytical methods are efficacious rivals to their bootstrap counterparts.

An interesting avenue for future research would be an exploration of the usefulness of the asymptotic bias expressions for bias-adjustment of GEL estimators such as continuous updating, exponential tilting and empirical likelihood.

Appendix A: Proofs

We find the asymptotic bias using a stochastic expansion for each estimator. Regularity conditions for the results given below may be obtained by suitable adaptation of those in Newey and Smith (2002). Lemmas A.1-A.3 generalise Newey and Smith (2002, Lemmas A4-A6) to the nuisance parameter context.

Lemma A.1 *Suppose the estimators* $\hat{\theta}$ *and* $\hat{\alpha}$ *and vector of functions* $m^{\theta}(z,\theta,\alpha)$ *satisfy (a)* $\hat{\theta}=\theta_0+O_p(\max[n^{-1/2},n_{\alpha}^{-1/2},n_{\beta}^{-1/2}])$, $\hat{\alpha}=\alpha_0+\tilde{\psi}^{\alpha}/\sqrt{n_{\alpha}}+Q^{\alpha}(\tilde{a}^{\varphi},\tilde{\psi}^{\varphi})/n_{\alpha}+O_p(n_{\alpha}^{-3/2})$, $\tilde{\psi}^{\alpha}=O_p(1)$, $Q^{\alpha}(\tilde{a}^{\varphi},\tilde{\psi}^{\varphi})=O_p(1)$; *(b)* $\hat{m}^{\theta}(\hat{\theta},\hat{\alpha})=\sum_{i=1}^{n_{\beta}}m^{\theta}(z_i,\hat{\theta},\hat{\alpha})/n_{\beta}=0$ *w.p.a.1 and* $\hat{m}^{\theta}(\theta_0,\alpha_0)=O_p(\max[n^{-1/2},n_{\alpha}^{-1/2},n_{\beta}^{-1/2}])$, $\tilde{A}^{\theta}=n_{\beta}^{1/2}[\partial\hat{m}^{\theta}(z,\theta_0,\alpha_0)/\partial\theta'-M^{\theta}]=O_p(\max[n^{-1/2},\ n_{\alpha}^{-1/2},n$ $\tilde{A}^{\theta}_{\alpha}=n_{\beta}^{1/2}[\partial\hat{m}^{\theta}(z,\theta_0,\alpha_0)/\partial\alpha'-M^{\theta}_{\alpha}]=O_p(\max[n^{-1/2},n_{\alpha}^{-1/2},n_{\beta}^{-1/2}])$, *where* $M^{\theta}=E[\partial m^{\theta}(z,\theta_0,\alpha_0)/\partial\theta']$ *and* $M^{\theta}_{\alpha}=E[\partial m(z;\theta_0,\alpha_0)/\partial\alpha']$; *(c)* $m^{\theta}(z,\theta,\alpha)$ *is two times continuously differentiable and for some* $d(z)$ *with* $E[d(z)]<\infty$ *on a neighbourhood of* (θ_0,α_0)

$$\|\partial^2 m(z,\theta,\alpha)/\partial(\theta,\alpha)_r\partial(\theta,\alpha)_s-\partial^2 m(z,\theta_0,\alpha_0)/\partial(\theta,\alpha)_r\partial(\theta,\alpha)_s\|\le d(z)\|(\theta,\alpha)-(\theta_0,\alpha_0)\|$$

on a neighbourhood of (θ_0,α_0); *(d)* $E[m^{\theta}(z,\theta_0,\alpha_0)]=0$ *and* M^{θ} *exists and is nonsingular. Let*

$$\begin{aligned}
M^{\theta}_{\theta\theta r} &= E[\partial^2 m(z,\theta_0,\alpha_0)/\partial\theta_r\partial\theta'], M^{\theta}_{\theta\alpha s}=E[\partial^2 m(z,\theta_0,\alpha_0)/\partial\alpha_s\partial\theta'],\\
M^{\theta}_{\alpha\theta r} &= E[\partial^2 m(z,\theta_0,\alpha_0)/\partial\theta_r\partial\alpha'], M^{\theta}_{\alpha\alpha s}=E[\partial^2 m(z,\theta_0,\alpha_0)/\partial\alpha_s\partial\alpha'],\\
\tilde{\psi}^{\theta} &= -n_{\beta}^{1/2}(M^{\theta})^{-1}\hat{m}^{\theta}(\theta_0,\alpha_0).
\end{aligned}$$

Then

$$\begin{aligned}
\hat{\theta} = {} & \theta_0+\tilde{\psi}^{\theta}/\sqrt{n_{\beta}}-(M^{\theta})^{-1}M^{\theta}_{\alpha}(\tilde{\psi}^{\alpha}/\sqrt{n_{\alpha}}+Q^{\alpha}(\tilde{a}^{\varphi},\tilde{\psi}^{\varphi})/n_{\alpha})\\
& -(M^{\theta})^{-1}[\tilde{A}^{\theta}(\tilde{\psi}^{\theta}/\sqrt{n_{\beta}}-M^{\theta-1}M^{\theta}_{\alpha}\tilde{\psi}^{\alpha}/\sqrt{n_{\alpha}})\sqrt{n_{\beta}}+\tilde{A}^{\theta}_{\alpha}\tilde{\psi}^{\alpha}/\sqrt{n_{\alpha}n_{\beta}}]\\
& -(M^{\theta})^{-1}[\sum_{r=1}^{q_{\theta}}e_r'[\tilde{\psi}^{\theta}/\sqrt{n_{\beta}}-(M^{\theta})^{-1}M^{\theta}_{\alpha}\tilde{\psi}^{\alpha}/\sqrt{n_{\alpha}}]M^{\theta}_{\theta\theta r}[\tilde{\psi}^{\theta}/\sqrt{n_{\beta}}-(M^{\theta})^{-1}M^{\theta}_{\alpha}\tilde{\psi}^{\alpha}/\sqrt{n_{\alpha}}]]/2\\
& -(M^{\theta})^{-1}\sum_{s=1}^{p_{\alpha}}e_s'\tilde{\psi}^{\varphi}M^{\theta}_{\theta\alpha s}[\tilde{\psi}^{\theta}/\sqrt{n_{\beta}}-(M^{\theta})^{-1}M^{\theta}_{\alpha}\tilde{\psi}^{\alpha}/\sqrt{n_{\alpha}}]/2\sqrt{n_{\alpha}}\\
& -(M^{\theta})^{-1}[\sum_{r=1}^{q_{\theta}}e_r'[\tilde{\psi}^{\theta}/\sqrt{n_{\beta}}-(M^{\theta})^{-1}M^{\theta}_{\alpha}\tilde{\psi}^{\alpha}/\sqrt{n_{\alpha}}]M^{\theta}_{\alpha\theta r}\tilde{\psi}^{\alpha}/\sqrt{n_{\alpha}}]/2
\end{aligned}$$

$$-(M^{\theta})^{-1}[\sum_{s=1}^{p_\alpha} e_s'\tilde{\psi}^{\alpha} M^{\theta}_{\alpha\alpha s}\tilde{\psi}^{\alpha}/n_\alpha]/2 + O_p(\max[n^{-3/2}, n_\alpha^{-3/2}, n_\beta^{-3/2}]).$$

Proof. Let $\hat{m}^{\theta}(\theta,\alpha) = \sum_{i=1}^{n_\beta} m_i^{\theta}(\theta,\alpha)/n_\beta$, $\hat{M}^{\theta}(\theta,\alpha) = \sum_{i=1}^{n_\beta}[\partial m_i^{\theta}(\theta,\alpha)/\partial\theta']/n_\beta$ and $\hat{M}^{\theta}_{\alpha}(\theta,\alpha) = \sum_{i=1}^{n_\beta}[\partial m_i^{\theta}(\theta,\alpha)/\partial\alpha']/n_\beta$. A Taylor expansion with Lagrange remainder gives

$$\begin{aligned} 0 &= \hat{m}^{\theta}(\theta_0,\alpha_0) + \hat{M}^{\theta}(\theta_0,\alpha_0)(\hat{\theta}-\theta_0) + \hat{M}^{\theta}_{\alpha}(\theta_0,\alpha_0)(\hat{\alpha}-\alpha_0) \\ &\quad +[\sum_{r=1}^{q_\theta}(\hat{\theta}_r-\theta_{0r})[\partial\hat{M}^{\theta}(\bar{\theta},\bar{\alpha})/\partial\theta_r](\hat{\theta}-\theta_0) + \sum_{s=1}^{p_\alpha}(\hat{\alpha}_s-\alpha_{0s})[\partial\hat{M}^{\theta}(\bar{\theta},\bar{\alpha})/\partial\alpha_s](\hat{\theta}-\theta_0) \\ &\quad +\sum_{r=1}^{q_\theta}(\hat{\theta}_r-\theta_{0r})[\partial\hat{M}^{\theta}_{\alpha}(\bar{\theta},\bar{\alpha})/\partial\theta_r](\hat{\alpha}-\alpha_0) + \sum_{s=1}^{p_\alpha}(\hat{\alpha}_s-\alpha_{0s})[\partial\hat{M}^{\theta}_{\alpha}(\bar{\theta},\bar{\alpha})/\partial\alpha_s](\hat{\alpha}-\alpha_0)]/2. \end{aligned}$$

Then adding and subtracting $M^{\theta}(\hat{\theta}-\theta_0)$ and solving gives

$$\begin{aligned} \hat{\theta} &= \theta_0 - (M^{\theta})^{-1}[\hat{m}^{\theta}(\theta_0,\alpha_0) + M^{\theta}_{\alpha}(\hat{\alpha}-\alpha_0)] \\ &\quad -(M^{\theta})^{-1}[(\hat{M}^{\theta}(\theta_0,\alpha_0) - M^{\theta})(\hat{\theta}-\theta_0) + (\hat{M}^{\theta}_{\alpha}(\theta_0,\alpha_0) - M^{\theta}_{\alpha})(\hat{\alpha}-\alpha_0)] \\ &\quad -(M^{\theta})^{-1}[\sum_{r=1}^{q_\theta}(\hat{\theta}_r-\theta_{0r})[\partial\hat{M}^{\theta}(\bar{\theta},\bar{\alpha})/\partial\theta_r](\hat{\theta}-\theta_0) + \sum_{s=1}^{p_\alpha}(\hat{\alpha}_s-\alpha_{0s})[\partial\hat{M}^{\theta}(\bar{\theta},\bar{\alpha})/\partial\alpha_s](\hat{\theta}-\theta_0) \\ &\quad +\sum_{r=1}^{q_\theta}(\hat{\theta}_r-\theta_{0r})[\partial\hat{M}^{\theta}_{\alpha}(\bar{\theta},\bar{\alpha})/\partial\theta_r](\hat{\alpha}-\alpha_0) + \sum_{s=1}^{p_\alpha}(\hat{\alpha}_s-\alpha_{0s})[\partial\hat{M}^{\theta}_{\alpha}(\bar{\theta},\bar{\alpha})/\partial\alpha_s](\hat{\alpha}-\alpha)]/2 \end{aligned}$$

so that $\hat{\theta} = \theta_0 + O_p(\max[n^{-1/2}, n_\alpha^{-1/2}, n_\beta^{-1/2}])$ and hence $\hat{\theta}-\theta_0 = -(M^{\theta})^{-1}[\hat{m}^{\theta}(\theta_0,\alpha_0) - M^{\theta}_{\alpha}(\hat{\alpha}-\alpha_0)] + O_p(\max[n^{-1}, n_\alpha^{-1}, n_\beta^{-1}])$. Note that replacing $\partial\hat{M}^{\theta}(\bar{\theta},\bar{\alpha})/\partial\theta_r$ by $M^{\theta}_{\theta\theta r}$, $\partial\hat{M}^{\theta}(\bar{\theta},\bar{\alpha})/\partial\alpha_s$ by $M^{\theta}_{\theta\alpha s}$, $\partial\hat{M}^{\theta}_{\alpha}(\bar{\theta},\bar{\alpha})/\partial\theta_r$ by $M^{\theta}_{\alpha\theta r}$ and $\partial\hat{M}^{\theta}_{\alpha}(\bar{\theta},\bar{\alpha})/\partial\alpha_s$ by $M^{\theta}_{\alpha\alpha s}$ introduces an error that is $O_p(\max[n^{-3/2}, n_\alpha^{-3/2}, n_\beta^{-3/2}])$ by hypothesis (c). Hence,

$$\begin{aligned} \hat{\theta} &= \theta_0 - (M^{\theta})^{-1}[\hat{m}^{\theta}(\theta_0,\alpha_0) + M^{\theta}_{\alpha}(\hat{\alpha}-\alpha_0)] \\ &\quad -(M^{\theta})^{-1}[(\hat{M}^{\theta}(\theta_0,\alpha_0) - M^{\theta})(\hat{\theta}-\theta_0) + (\hat{M}^{\theta}_{\alpha}(\theta_0,\alpha_0) - M^{\theta}_{\alpha})(\hat{\alpha}-\alpha_0)] \\ &\quad -(M^{\theta})^{-1}[\sum_{r=1}^{q_\theta}(\hat{\theta}_r-\theta_{0r})M^{\theta}_{\theta\theta r}(\hat{\theta}-\theta_0) + \sum_{s=1}^{p_\alpha}(\hat{\alpha}_s-\alpha_{0s})M^{\theta}_{\theta\alpha s}(\hat{\theta}-\theta_0) \\ &\quad +\sum_{r=1}^{q_\theta}(\hat{\theta}_r-\theta_{0r})M^{\theta}_{\alpha\theta r}(\hat{\alpha}-\alpha_0) + \sum_{s=1}^{p_\alpha}(\hat{\alpha}_s-\alpha_{0s})M^{\theta}_{\alpha\alpha s}(\hat{\alpha}-\alpha)]/2 \\ &\quad +O_p(\max[n^{-3/2}, n_\alpha^{-3/2}, n_\beta^{-3/2}]). \end{aligned}$$

Therefore, by recursive substitution, cf. Newey and Smith (2001, Lemma A4), the result is obtained. ■

Lemma A.2 *Suppose* $\hat{\alpha} = \alpha_0 + \tilde{\psi}^{\alpha}/\sqrt{n_\alpha} + Q^{\alpha}(\tilde{a}^{\varphi}, \tilde{\psi}^{\varphi})/n_\alpha + O_p(n_\alpha^{-3/2})$, *where* $\tilde{\psi}^{\alpha}$ *and* $Q^{\alpha}(\tilde{a}^{\varphi}, \tilde{\psi}^{\varphi})$ *are* $O_p(1)$. *Let* $P_W^{\beta} = (W^{\beta\beta})^{-1} - (W^{\beta\beta})^{-1} G_{\beta}^{\beta} \Sigma_W^{\beta\beta} G_{\beta}^{\beta\prime} (W^{\beta\beta})^{-1}$, $\tilde{\psi}^{\theta W} = -[H_W^{\beta\prime}, P_W^{\beta}]' \sum_{k=1}^{n} g_k^{\beta}/\sqrt{}$ $g_k^{\beta} = g_k^{\beta}(\alpha_0, \beta_0)$,

$$M^{\theta W} = -\begin{pmatrix} 0 & G_{\beta}^{\beta\prime} \\ G_{\beta}^{\beta} & W^{\beta\beta} \end{pmatrix}, (M^{\theta W})^{-1} = -\begin{pmatrix} -\Sigma_W^{\beta\beta} & H_W^{\beta} \\ H_W^{\beta\prime} & P_W^{\beta} \end{pmatrix}, M_{\alpha}^{\theta W} = -\begin{pmatrix} 0 \\ G_{\alpha}^{\beta} \end{pmatrix}.$$

Then for $\tilde{\lambda} = -(\hat{W}^{\beta\beta})^{-1} \hat{g}^{\beta}(\hat{\alpha}, \tilde{\beta})$, $\hat{\theta} = (\tilde{\beta}', \tilde{\lambda}')'$, *we have,*

$$\hat{\theta} = \theta_0 + \tilde{\psi}^{\theta W}/\sqrt{n} - (M^{\theta W})^{-1} M_{\alpha}^{\theta W} \tilde{\psi}^{\alpha}/\sqrt{n_\alpha} + O_p(\max[n^{-1}, n_\alpha^{-1}]).$$

Proof. Let $\theta = (\beta', \lambda')'$, $\lambda_0 = 0$, $m_k^{\theta}(\theta, \alpha) = -(\lambda' \partial g_k^{\beta}(\alpha, \beta)/\partial \beta', g_k^{\beta}(\alpha, \beta)' + \lambda' [W^{\beta\beta} + \xi^{\beta}(z)])'$ and $\hat{m}^{\theta}(\theta, \alpha) = \sum_{k=1}^{n} m_k^{\theta}(\theta, \alpha)/n$. The first-order conditions for $\tilde{\beta}$, the definition of $\tilde{\lambda}$ imply

$$0 = \hat{m}^{\theta}(\hat{\theta}, \hat{\alpha}) + [0, -\tilde{\lambda}'(O_p(n^{-1}))]'.$$

Hence, it follows from Lemma A.1 that $\hat{\theta} = \theta_0 + O_p(\max[n^{-1/2}, n_\alpha^{-1/2}])$. Therefore,

$$\hat{m}^{\theta}(\hat{\theta}, \hat{\alpha}) = O_p(n^{-1} \max[n^{-1/2}, n_\alpha^{-1/2}]).$$

A further application of Lemma A.1 gives the result. ■

Lemma A.3 *Suppose that* $\hat{\alpha} = \alpha_0 + \tilde{\psi}^{\alpha}/\sqrt{n_\alpha} + Q^{\alpha}(\tilde{a}^{\varphi}, \tilde{\psi}^{\varphi})/n_\alpha + O_p(n_\alpha^{-3/2})$, *where* $\tilde{\psi}^{\alpha}$ *and* $Q^{\alpha}(\tilde{a}^{\varphi}, \tilde{\psi}^{\varphi})$ *are* $O_p(1)$. *Let* $\Omega_k^{\beta\beta} = g_k^{\beta} g_k^{\beta\prime} - \Omega^{\beta\beta}$, $\tilde{\Omega}^{\beta\beta} = \sum_{k=1}^{n} \Omega_k^{\beta\beta}/\sqrt{n}$, $\bar{\Omega}_{\beta r} = E[\partial [g_k^{\beta} g_k^{\beta\prime}]/\partial \beta_r]$ *and* $\bar{\Omega}_{\alpha s} = E[\partial [g_k^{\beta} g_k^{\beta\prime}]/\partial \alpha_s]$. *Then*

$$\begin{aligned} \hat{\Omega}^{\beta\beta}(\hat{\alpha}, \tilde{\beta}) &= \Omega^{\beta\beta} + \tilde{\Omega}^{\beta\beta}/\sqrt{n} + \sum_{r=1}^{p_\beta} \bar{\Omega}_{\beta r} e_r' (\tilde{\psi}^{\theta W}/\sqrt{n} - (M^{\theta W})^{-1} M_{\alpha}^{\theta W} \tilde{\psi}^{\alpha}/\sqrt{n_\alpha}) \\ &\quad + \sum_{s=1}^{p_\alpha} \bar{\Omega}_{\alpha s} e_s' \tilde{\psi}^{\alpha}/\sqrt{n_\alpha} + O_p(\max[n^{-1}, n_\alpha^{-1}]). \end{aligned}$$

Proof. Similarly to the proof of Lemma A.4, expanding gives

$$\hat{\Omega}^{\beta\beta}(\hat{\alpha}, \tilde{\beta}) = \hat{\Omega}(\alpha_0, \beta_0) + \sum_{r=1}^{p_\beta} \bar{\Omega}_{\beta r}(\tilde{\beta}_r - \beta_{0r}) + \sum_{s=1}^{p_\alpha} \bar{\Omega}_{\alpha s}(\hat{\alpha}_s - \alpha_{0s}) + O_p(\max[n^{-1}, n_\alpha^{-1}]).$$

By Lemma A.1, $\tilde{\beta}_r - \beta_{0r} = e_r'(\tilde{\psi}^{\theta W}/\sqrt{n} - (M^{\theta W})^{-1} M_{\alpha}^{\theta W} \tilde{\psi}^{\alpha}/\sqrt{n_\alpha}) + O_p(\max[n^{-1}, n_\alpha^{-1}])$. The conclusion follows by substitution into the above equation. ■

Let $\hat{\Omega}^{\beta\beta} = \sum_{i=1}^{n_\beta} g_i^\beta g_i^{\beta\prime}/n_\beta$, $\hat{G}_\beta^\beta = \sum_{i=1}^{n_\beta} G_{\beta i}^\beta/n_\beta$, $\hat{G}_\alpha^\beta = \sum_{i=1}^{n_\beta} G_{\alpha i}^\beta/n_\beta$, $G_{\beta\beta i}^{\beta r} = \partial^2 g_i^\beta/\partial\beta_r\partial\beta'$, $G_{\beta\alpha i}^{\beta s} = \partial^2 g_i^\beta/\partial\alpha_s\partial\beta'$, $g_{\beta i}^{\beta r} = \partial g_i^\beta/\partial\beta_r$ and $g_{\alpha i}^{\beta s} = \partial g_i^\beta/\partial\alpha_s$.

We detail an expansion for GMM in the general case. Let $\theta = (\beta', \lambda')'$, $\theta_0 = (\beta_0', 0')'$, $\hat{\beta}$ be the two-step GMM estimator and

$$\hat{m}^\theta(\theta, \alpha) = -\begin{pmatrix} \hat{G}_\beta^\beta(\alpha, \beta)'\lambda \\ \hat{g}^\beta(\alpha, \beta) + (\Omega^{\beta\beta} + \tilde{\xi}^{\Omega^{\beta\beta}})\lambda \end{pmatrix},$$

where $\tilde{\xi}^{\Omega^{\beta\beta}} = \tilde{\Omega}^{\beta\beta}/\sqrt{n} + \sum_{r=1}^{p_\beta} \bar{\Omega}_{\beta r} e_r'(\tilde{\psi}^{\theta W}/\sqrt{n} - (M^{\theta W})^{-1} M_\alpha^{\theta W}\tilde{\psi}^\alpha/\sqrt{n_\alpha}) + \sum_{s=1}^{p_\alpha} \bar{\Omega}_{\alpha s} e_s'\tilde{\psi}^\alpha/\sqrt{n_\alpha}$. Also, let $\hat{\lambda} = -\hat{\Omega}^{\beta\beta}(\hat{\alpha}, \tilde{\beta})^{-1}\hat{g}^\beta(\hat{\alpha}, \hat{\beta})$. Then $\hat{\lambda} = O_p(\max[n_\alpha^{-1/2}, n_\beta^{-1/2}])$. The first-order conditions for GMM and Lemmas A.1-A.3 imply

$$0 = \hat{m}^\theta(\hat{\theta}, \hat{\alpha}) + [0, -\hat{\lambda}'(O_p(\max[n^{-1}, n_\alpha^{-1}]))]' = \hat{m}^\theta(\hat{\theta}, \hat{\alpha}) + O_p(\max[n_\alpha^{-1/2}, n_\beta^{-1/2}]\max[n^{-1}, n_\alpha^{-1}]).$$

Therefore, we can solve for $\hat{\theta}_{2S} - \theta_0$ as in the conclusion of Lemma A.1 using the definitions $\hat{m}^\theta(\theta_0, \alpha_0) = -(0', \hat{g}^\beta(\alpha_0, \beta_0)')'$,

$$M^\theta = -\begin{pmatrix} 0 & G_\beta^{\beta\prime} \\ G_\beta^\beta & \Omega^{\beta\beta} \end{pmatrix}, (M^\theta)^{-1} = -\begin{pmatrix} -\Sigma^{\beta\beta} & H^\beta \\ H^{\beta\prime} & P^\beta \end{pmatrix}, M_\alpha^\theta = -\begin{pmatrix} 0 \\ G_\alpha^\beta \end{pmatrix},$$

$$\tilde{A}^\theta = -n_\beta^{1/2}\begin{pmatrix} 0 & (\hat{G}_\beta^\beta - G_\beta^\beta)' \\ (\hat{G}_\beta^\beta - G_\beta^\beta) & \tilde{\xi}^{\Omega^{\beta\beta}} \end{pmatrix}, \tilde{A}_\alpha^\theta = -n_\beta^{1/2}\begin{pmatrix} 0 \\ (\hat{G}_\alpha^\beta - G_\alpha^\beta) \end{pmatrix},$$

$$M_{\theta\theta r}^\theta = -\begin{pmatrix} 0 & E[G_{\beta\beta i}^{\beta r}]' \\ E[G_{\beta\beta i}^{\beta r}] & 0 \end{pmatrix}, (r \le p_\beta), M_{\theta\theta, p_\beta + r}^\theta = -\begin{pmatrix} E[\partial^2 g_{ir}^\beta/\partial\beta\partial\beta'] & 0 \\ 0 & 0 \end{pmatrix}, (r \le m_\beta).$$

$$M_{\theta\alpha s}^\theta = -\begin{pmatrix} 0 & E[G_{\beta\alpha i}^{\beta s}]' \\ E[G_{\beta\alpha i}^{\beta s}] & 0 \end{pmatrix}, (s \le p_\alpha),$$

$$M_{\alpha\theta r}^\theta = -\begin{pmatrix} 0 \\ E[\partial^2 g_i^\beta/\partial\beta_r\partial\alpha'] \end{pmatrix}, (r \le p_\beta), M_{\alpha\theta, p_\beta + r}^\beta = -\begin{pmatrix} E[\partial^2 g_{ir}^\beta/\partial\beta\partial\alpha'] \\ 0 \end{pmatrix}, (r \le m_\beta).$$

$$M_{\alpha\alpha s}^\theta = -\begin{pmatrix} 0 \\ E[\partial^2 g_i^\beta/\partial\alpha_s\partial\alpha'] \end{pmatrix}, (s \le p_\alpha). \tag{A.1}$$

For a general expansion for GEL, we apply Lemma A.1. Let $\theta = (\beta', \lambda')'$, $\theta_0 = (\beta_0', 0')'$, $\hat{\theta}$ be the GEL estimator and

$$\hat{m}^\theta(\theta, \alpha) = \sum_{i=1}^{n_*} \rho_v^\theta(\lambda' g_i^\beta(\alpha, \beta)) \begin{pmatrix} G_{\beta i}^\beta(\alpha, \beta)'\lambda \\ g_i^\beta(\beta) \end{pmatrix} / n_*.$$

Therefore, using similar arguments to those in Newey and Smith (2002) we can solve for $\hat{\theta}_{GEL} - \theta_0$ as in the conclusion of Lemma A.1 by setting $n_\beta = n_*$, dropping n and with the definitions $\hat{m}^\theta(\theta_0, \alpha_0) = -(0', \hat{g}^\beta(\alpha_0, \beta_0)')'$,

$$M^\theta = -\begin{pmatrix} 0 & G_\beta^{\beta\prime} \\ G_\beta^\beta & \Omega^{\beta\beta} \end{pmatrix}, (M^\theta)^{-1} = -\begin{pmatrix} -\Sigma^{\beta\beta} & H^\beta \\ H^{\beta\prime} & P^\beta \end{pmatrix}, M_\alpha^\theta = -\begin{pmatrix} 0 \\ G_\alpha^\beta \end{pmatrix}$$

$$\tilde{A}^\theta = -n_*^{1/2} \begin{pmatrix} 0 & (\hat{G}_\beta^\beta - G_\beta^\beta)' \\ (\hat{G}_\beta^\beta - G_\beta^\beta) & \hat{\Omega}^{\beta\beta} - \Omega^{\beta\beta} \end{pmatrix}, A_\alpha^\theta(z_i) = -n_*^{1/2} \begin{pmatrix} 0 \\ (\hat{G}_\alpha^\beta - G_\alpha^\beta) \end{pmatrix}$$

$$\begin{aligned} M_{\theta\theta r}^\theta &= -\begin{pmatrix} 0 & E[G_{\beta\beta i}^{\beta r}]' \\ E[G_{\beta\beta i}^{\beta r}] & E[g_{\beta i}^{\beta r} g_i^{\beta\prime} + g_i^\beta g_{\beta i}^{\beta r\prime}] \end{pmatrix}, (r \le p_\beta), \\ M_{\theta\theta, p_\beta + r}^\theta &= -\begin{pmatrix} E[\partial^2 g_{ir}^\beta / \partial\beta\partial\beta'] & E[G_{\beta i}^{\beta\prime} e_r g_i^{\beta\prime} + g_{ir}^\beta G_{\beta i}^{\beta\prime}] \\ E[g_i^\beta e_r' G_{\beta i}^\beta + g_{ir}^\beta G_{\beta i}^\beta] & -\rho_{vvv}^\theta(0) E[g_{ir}^\beta g_i^\beta g_i^{\beta\prime}] \end{pmatrix}, (r \le m_\beta). \end{aligned}$$

$$M_{\theta\alpha s}^\theta = -\begin{pmatrix} 0 & E[G_{\beta\alpha i}^{\beta s}]' \\ E[G_{\beta\alpha i}^{\beta s}] & E[G_{\alpha i}^\beta e_s g_i^{\beta\prime} + g_i^\beta e_s' G_{\alpha i}^{\beta\prime}] \end{pmatrix}, (s \le p_\alpha),$$

$$\begin{aligned} M_{\alpha\theta r}^\theta &= -\begin{pmatrix} 0 \\ E[\partial^2 g_i^\beta(\beta_0, \alpha_0)/\partial\beta_r \partial\alpha'] \end{pmatrix}, (r \le p_\beta), \\ M_{\alpha\theta, p_\beta + r}^\beta &= -\begin{pmatrix} E[\partial^2 g_{ir}^\beta / \partial\beta\partial\alpha'] \\ E[g_i^\beta \partial g_{ir}^\beta / \partial\alpha'] + E[g_{ir}^\beta G_{\alpha i}^\beta] \end{pmatrix}, (r \le m_\beta). \end{aligned}$$

$$M_{\alpha\alpha s}^\theta = -\begin{pmatrix} 0 \\ E[\partial^2 g_i^\beta / \partial\alpha_s \partial\alpha'] \end{pmatrix}, (s \le p_\alpha). \tag{A.2}$$

Proof of Theorem 3.1: The matrices M^{θ}, $M^{\theta-1}$ are as defined in (A.1). Thus, $\tilde{\psi}^{\theta} = -n_{\beta}^{1/2}[H^{\beta\prime}, P^{\beta}]'\hat{g}^{\beta}$. For independent samples, $\tilde{\xi}^{\Omega\beta\beta}$ is uncorrelated with $\hat{g}^{\beta}$ as is $\tilde{A}^{\theta}_{\alpha}$ with $\tilde{\psi}^{\alpha}$. Thus,

$$
\begin{aligned}
Bias(\hat{\theta}_{2S}) \;=\;& \theta_0 - (M^{\theta})^{-1}M^{\theta}_{\alpha}Bias(\hat{\alpha}) \\
& -(M^{\theta})^{-1}E[\tilde{A}^{\theta}(\tilde{\psi}^{\theta}/\sqrt{n_{\beta}} - (M^{\theta})^{-1}M^{\theta}_{\alpha}\tilde{\psi}^{\alpha}/\sqrt{n_{\alpha}})]/\sqrt{n_{\beta}} \\
& -(M^{\theta})^{-1}\sum_{r=1}^{q_{\theta}} e_r'[E[\tilde{\psi}^{\theta}M^{\theta}_{\theta\theta r}\tilde{\psi}^{\theta}]/n_{\beta} + (M^{\theta})^{-1}M^{\theta}_{\alpha}E[\tilde{\psi}^{\alpha}M^{\theta}_{\theta\theta r}(M^{\theta})^{-1}M^{\theta}_{\alpha}\tilde{\psi}^{\alpha}]/n_{\alpha}]/2 \\
& +(M^{\theta})^{-1}\sum_{s=1}^{p_{\alpha}} e_s'E[\tilde{\psi}^{\alpha}M^{\theta}_{\theta\alpha s}(M^{\theta})^{-1}M^{\theta}_{\alpha}\tilde{\psi}^{\alpha}]/2n_{\alpha} \\
& +(M^{\theta})^{-1}[\sum_{r=1}^{q_{\theta}} e_r'(M^{\theta})^{-1}M^{\theta}_{\alpha}E[\tilde{\psi}^{\alpha}M^{\theta}_{\alpha\theta r}\tilde{\psi}^{\alpha}]/2n_{\alpha} \\
& -(M^{\theta})^{-1}[\sum_{s=1}^{p_{\alpha}} e_s'\tilde{\psi}^{\alpha}M^{\theta}_{\alpha\alpha s}\tilde{\psi}^{\alpha}/n_{\alpha}]/2 + O_p(\max[n^{-3/2}, n_{\alpha}^{-3/2}, n_{\beta}^{-3/2}]).
\end{aligned}
$$

Note that the penultimate two terms are identical. Now,

$$
\begin{aligned}
E[\tilde{A}^{\theta}\tilde{\psi}^{\theta}] &= \begin{pmatrix} E[G^{\beta\prime}_{\beta i}P^{\beta}g^{\beta}_i] \\ E[G^{\beta}_{\beta i}H^{\beta}g^{\beta}_i] \end{pmatrix}, \\
E[\tilde{A}^{\theta}(M^{\theta})^{-1}M^{\theta}_{\alpha}\tilde{\psi}^{\alpha}] &= \begin{pmatrix} 0 \\ -\sum_{r=1}^{p_{\beta}}\bar{\Omega}_{\beta r}P^{\beta}G^{\beta}_{\alpha}\Sigma^{\alpha\alpha}G^{\beta\prime}_{\alpha}H^{\beta\prime}_{W}e_r + \sum_{s=1}^{p_{\alpha}}\bar{\Omega}_{\alpha s}P^{\beta}G^{\beta}_{\alpha}\Sigma^{\alpha\alpha}e_s \end{pmatrix}.
\end{aligned}
$$

Let $(M^{\theta}_{\beta})^{-1} = (-\Sigma^{\beta\beta}, H^{\beta})$. By a similar analysis to that in Newey and Smith (2002, Proof of Theorem 4.1),

$$
\begin{aligned}
(M^{\theta}_{\beta})^{-1}\sum_{r=1}^{p_{\beta}} e_r'E[\tilde{\psi}^{\theta}M^{\theta}_{\theta\theta r}\tilde{\psi}^{\theta}] &= -H^{\beta}a_{\beta}. \\
(M^{\theta}_{\beta})^{-1}\sum_{r=p_{\beta}+1}^{q_{\theta}} e_r'E[\tilde{\psi}^{\theta}M^{\theta}_{\theta\theta r}\tilde{\psi}^{\theta}] &= 0. \\
(M^{\theta}_{\beta})^{-1}\sum_{r=1}^{p_{\beta}} e_r'(M^{\theta})^{-1}M^{\theta}_{\alpha}E[\tilde{\psi}^{\alpha}M^{\theta}_{\theta\theta r}(M^{\theta})^{-1}M^{\theta}_{\alpha}\tilde{\psi}^{\alpha}] &= \Sigma^{\beta\beta}c^{\beta}_{\beta\beta} - H^{\beta}a^{\beta}_{\beta\beta}. \\
(M^{\theta}_{\beta})^{-1}\sum_{r=p_{\beta}+1}^{m_{\beta}} e_r'(M^{\theta})^{-1}M^{\theta}_{\alpha}E[\tilde{\psi}^{\alpha}M^{\theta}_{\theta\theta r}(M^{\theta})^{-1}M^{\theta}_{\alpha}\tilde{\psi}^{\alpha}] &= \Sigma^{\beta\beta}c^{\beta}_{\beta\beta}. \\
-(M^{\theta}_{\beta})^{-1}\sum_{s=1}^{p_{\alpha}} e_s'E[\tilde{\psi}^{\alpha}M^{\theta}_{\theta\alpha s}(M^{\theta})^{-1}M^{\theta}_{\alpha}\tilde{\psi}^{\alpha}] &= \Sigma^{\beta\beta}c^{\beta}_{\beta\alpha} - H^{\beta}a^{\beta}_{\beta\alpha}. \\
(M^{\theta}_{\beta})^{-1}\sum_{s=1}^{p_{\alpha}} e_s'\tilde{\psi}^{\alpha}M^{\theta}_{\alpha\alpha s}\tilde{\psi}^{\alpha} &= -H^{\beta}a^{\beta}_{\alpha\alpha}.
\end{aligned}
$$

Therefore, as

$$\bar{\Omega}_{\beta r} = E[G^{\beta}_{\beta i} e_r g_i^{\beta.\alpha} + g_i^{\beta.\alpha} e_r' G^{\beta\prime}_{\beta i}], \bar{\Omega}_{\alpha s} = E[G^{\beta}_{\alpha i} e_s g_i^{\beta.\alpha} + g_i^{\beta.\alpha} e_s' G^{\beta\prime}_{\alpha i}],$$

and $Bias(\hat{\beta}_{2S}) = (I_{p_\beta}, 0) Bias(\hat{\theta}_{2S})$, after simplification and collecting terms the result of the theorem is obtained. ■

Proof of Theorem 3.2: From (A.2), because of independent sampling $\tilde{A}^{\theta}$ and $\tilde{A}^{\theta}_{\alpha}$ are uncorrelated with $\tilde{\psi}^{\alpha}$. Hence,

$$\begin{aligned}
\hat{\theta}_{GEL} \;=\; & -(M^{\theta})^{-1} M^{\theta}_{\alpha} Bias(\hat{\alpha}) \\
& -(M^{\theta})^{-1} E[\tilde{A}^{\theta} \tilde{\psi}^{\theta}]/N \\
& -(M^{\theta})^{-1} \sum_{r=1}^{q_\theta} e_r' [E[\tilde{\psi}^{\theta} M^{\theta}_{\theta\theta r} \tilde{\psi}^{\theta}] + (M^{\theta})^{-1} M^{\theta}_{\alpha} E[\tilde{\psi}^{\alpha} M^{\theta}_{\theta\theta r} (M^{\theta})^{-1} M^{\theta}_{\alpha} \tilde{\psi}^{\alpha}]]/2N \\
& +(M^{\theta})^{-1} \sum_{s=1}^{p_\alpha} e_s' E[\tilde{\psi}^{\alpha} M^{\theta}_{\theta\alpha s} (M^{\theta})^{-1} M^{\theta}_{\alpha} \tilde{\psi}^{\alpha}]/2n_\alpha \\
& +(M^{\theta})^{-1} \sum_{r=1}^{q_\theta} e_r' (M^{\theta})^{-1} M^{\theta}_{\alpha} E[\tilde{\psi}^{\alpha} M^{\theta}_{\alpha\theta r} \tilde{\psi}^{\alpha}]/2n_\alpha \\
& -(M^{\theta})^{-1} \sum_{s=1}^{p_\alpha} e_s' E[\tilde{\psi}^{\alpha} M^{\theta}_{\alpha\alpha s} \tilde{\psi}^{\alpha}]/2n_\alpha + O_p(\max[n^{-3/2}, n_\alpha^{-3/2}, n_\beta^{-3/2}]).
\end{aligned}$$

Note that the penultimate two terms are identical. Also, $Bias_{\alpha_0}(\hat{\theta}_{GEL}) = -(M^{\theta})^{-1}(E[\tilde{A}^{\theta}\tilde{\psi}^{\theta}] + \sum_{r=1}^{q_\theta} e_r' [E[\tilde{\psi}^{\theta} M^{\theta}_{\theta\theta r} \tilde{\psi}^{\theta}]/2)/N$; see Newey and Smith (2002, Proof of Theorem 4.2). Let $(M^{\theta}_{\beta})^{-1} = (-\Sigma^{\beta\beta}, H^{\beta})$. By a similar analysis to that in Newey and Smith (2002, Proof of Theorem 4.2),

$$\begin{aligned}
(M^{\theta}_{\beta})^{-1} \sum_{r=1}^{p_\beta} e_r' (M^{\theta})^{-1} M^{\theta}_{\alpha} E[\tilde{\psi}^{\alpha} M^{\theta}_{\theta\theta r} (M^{\theta})^{-1} M^{\theta}_{\alpha} \tilde{\psi}^{\alpha}] = \Sigma^{\beta\beta} c^{\beta}_{\beta\beta} - H^{\beta} a^{\beta}_{\beta\beta} \\
-H^{\beta}(E[G^{\beta}_{\beta i} H^{\beta} G^{\beta}_{\alpha} \Sigma^{\alpha\alpha} G^{\beta\prime}_{\alpha} P^{\beta} g_i^{\beta}] + E[g_i^{\beta} tr(G^{\beta\prime}_{\beta i} P^{\beta} G^{\beta}_{\alpha} \Sigma^{\alpha\alpha} G^{\beta\prime}_{\alpha} H^{\beta})].
\end{aligned}$$

$$\begin{aligned}
(M^{\theta}_{\beta})^{-1} \sum_{r=p_\beta+1}^{q_\theta} e_r' (M^{\theta})^{-1} M^{\theta}_{\alpha} E[\tilde{\psi}^{\alpha} M^{\theta}_{\theta\theta r} (M^{\theta})^{-1} M^{\theta}_{\alpha} \tilde{\psi}^{\alpha}] = \Sigma^{\beta\beta} c^{\beta}_{\beta\beta} + 2\Sigma^{\beta\beta} E[G^{\beta\prime}_{\beta i} P^{\beta} G^{\beta}_{\alpha} \Sigma^{\alpha\alpha} G^{\beta\prime}_{\alpha} P^{\beta} g_i^{\beta}] \\
-H^{\beta}(E[G^{\beta}_{\beta i} H^{\beta} G^{\beta}_{\alpha} \Sigma^{\alpha\alpha} G^{\beta\prime}_{\alpha} P^{\beta} g_i^{\beta}] + E[g_i^{\beta} tr(G^{\beta\prime}_{\beta i} P^{\beta} G^{\beta}_{\alpha} \Sigma^{\alpha\alpha} G^{\beta\prime}_{\alpha} H^{\beta})]) \\
+\rho^{\theta}_{vvv}(0) H^{\beta} E[g_i^{\beta} g_i^{\beta\prime} P^{\beta} G^{\beta}_{\alpha} \Sigma^{\alpha\alpha} G^{\beta}_{\alpha} P^{\beta} g_i^{\beta}].
\end{aligned}$$

$$\begin{aligned}
(M^{\theta}_{\beta})^{-1} \sum_{s=1}^{p_\alpha} e_s' E[\tilde{\psi}^{\alpha} M^{\theta}_{\theta\alpha s} (M^{\theta})^{-1} M^{\theta}_{\alpha} \tilde{\psi}^{\alpha}] = \Sigma^{\beta\beta} c^{\beta}_{\beta\alpha} - H^{\beta} a^{\beta}_{\beta\alpha} \\
-H^{\beta}(E[G^{\beta}_{\alpha i} \Sigma^{\alpha\alpha} G^{\beta\prime}_{\alpha} P^{\beta} g_i^{\beta}] + E[g_i^{\beta} tr(G^{\beta\prime}_{\alpha i} P^{\beta} G^{\beta}_{\alpha} \Sigma^{\alpha\alpha})]).
\end{aligned}$$

$$(M^{\theta}_{\beta})^{-1}\sum_{s=1}^{p_\alpha} e_s' E[\tilde{\psi}^{\alpha} M^{\theta}_{\alpha\alpha s}\tilde{\psi}^{\alpha}] = -H^{\beta} a^{\beta}_{\alpha\alpha}.$$

Therefore, simpifying and collecting terms gives the result of the theorem. ■

Proof of Corollary 3.2: Immediate as $G^{\beta}_{\alpha} = 0$, $E[\partial^2 g^{\beta}_{ir}/\partial\beta\partial\alpha'] = 0$ and $E[\partial^2 g^{\beta}_{kr}/\partial\beta\partial\alpha'] = 0$. ■

Proof of Corollary 3.3: Follows immediately as in Proof of Corollary 3.2 and from Newey and Smith (2002, Theorem 4.2). ■

Proof of Theorem 3.3: From (A.1), as $Bias(\hat{\beta}_{2S}) = (I_{p_\beta}, 0)Bias(\hat{\theta}_{2S})$,

$$\begin{aligned}
Bias(\hat{\beta}_{2S}) \;=\;& -H^{\beta}G^{\beta}_{\alpha}Bias(\hat{\alpha}) \\
& -\Sigma^{\beta\beta}E[G^{\beta\prime}_{\beta i}P^{\beta}g^{\beta.\alpha}_i]/n_\beta + H^{\beta}E[G^{\beta}_{\beta i}H^{\beta}g^{\beta.\alpha}_i]/n_\beta \\
& +H^{\beta}E[g^{\beta}_i g^{\beta\prime}_i P^{\beta}g^{\beta.\alpha}_i]/n_\beta \\
& -H^{\beta}(\sum_{r=1}^{p_\beta}\bar{\Omega}_{\beta r}P^{\beta}E[g^{\beta.\alpha}_i g^{\beta.\alpha\prime}_i]H^{\beta\prime}_{W}e_r + \sum_{s=1}^{p_\alpha}\bar{\Omega}_{\alpha s}P^{\beta}E[g^{\beta.\alpha}_i g^{\alpha\prime}_i]H^{\alpha\prime}e_s)/n_\beta \\
& +H^{\beta}E[G^{\beta}_{\alpha i}H^{\alpha}g^{\alpha}_i]/n_\beta \\
& +\sum_{r=1}^{p_\beta}(\Sigma^{\beta\beta}E[G^{\beta r}_{\beta\beta i}]'P^{\beta} - H^{\beta}E[G^{\beta r}_{\beta\beta i}]H^{\beta})E[g^{\beta.\alpha}_i g^{\beta.\alpha\prime}_i]H^{\beta\prime}e_r/2n_\beta \\
& +\sum_{r=1}^{m_\beta}\Sigma^{\beta\beta}E[\partial^2 g^{\beta}_{ir}/\partial\beta\partial\beta']H^{\beta}E[g^{\beta.\alpha}_i g^{\beta.\alpha\prime}_i]P^{\beta}e_r/2n_\beta \\
& +\sum_{s=1}^{p_\alpha}(\Sigma^{\beta\beta}E[G^{\beta s}_{\beta\alpha i}]'P^{\beta} - H^{\beta}E[G^{\beta s}_{\beta\alpha i}]H^{\beta})E[g^{\beta.\alpha}_i g^{\alpha}_i]H^{\alpha\prime}e_s/2n_\beta \\
& -\sum_{r=1}^{p_\beta}H^{\beta}E[\partial^2 g^{\beta}_i/\partial\beta_r\partial\alpha']H^{\alpha}E[g^{\alpha}_i g^{\beta.\alpha\prime}_i]H^{\beta\prime}e_r/2n_\beta \\
& +\sum_{r=1}^{m_\beta}\Sigma^{\beta\beta}E[\partial^2 g^{\beta}_{ir}/\partial\beta\partial\alpha']H^{\alpha}E[g^{\alpha}_i g^{\beta.\alpha\prime}_i]P^{\beta}e_r/2n_\beta \\
& -\sum_{s=1}^{p_\alpha}H^{\beta}E[\partial^2 g^{\beta}_i/\partial\alpha_s\partial\alpha']\Sigma^{\alpha\alpha}e_s/2n_\beta.
\end{aligned}$$

As

$$\bar{\Omega}_{\beta r} = E[G^{\beta}_{\beta i}e_r g^{\beta\prime}_i + g^{\beta}_i e_r' G^{\beta\prime}_{\beta i}], \bar{\Omega}_{\alpha s} = E[G^{\beta}_{\alpha i}e_s g^{\beta\prime}_i + g^{\beta}_i e_s' G^{\beta\prime}_{\alpha i}],$$

simplifying and collecting terms yields the result in Theorem 3.3. ■

Proof of Theorem 3.4: From (A.2), as $Bias(\hat{\beta}_{GEL}) = (I_{p_\beta}, 0)Bias(\hat{\theta}_{GEL})$,

$$Bias(\hat{\beta}_{GEL}) \;=\; -H^{\beta}G^{\beta}_{\alpha}Bias(\hat{\alpha})$$

$$
\begin{aligned}
&-\Sigma^{\beta\beta}E[G^{\beta\prime}_{\beta i}P^{\beta}g^{\beta.\alpha}_i]/n_\beta + H^{\beta}(E[G^{\beta}_{\beta i}H^{\beta}g^{\beta.\alpha}_i] + E[g^{\beta}_i g^{\beta\prime}_i P^{\beta}g^{\beta.\alpha}_i])/n_\beta \\
&+H^{\beta}E[G^{\beta}_{\alpha i}H^{\alpha}g^{\alpha}_i]/n_\beta \\
&-\sum_{r=1}^{p_\beta}(-\Sigma^{\beta\beta}E[G^{\beta r}_{\beta\beta i}]'P^{\beta} \\
&+H^{\beta}(E[G^{\beta r}_{\beta\beta i}]H^{\beta} + E[g^{\beta r}_i g^{\beta\prime}_i + g^{\beta}_i g^{\beta r\prime}_i]P^{\beta}))E[g^{\beta.\alpha}_i g^{\beta.\alpha\prime}_i]H^{\beta\prime}e_r/2n_\beta \\
&+\sum_{r=1}^{m_\beta}\Sigma^{\beta\beta}(E[\partial^2 g^{\beta}_{ir}(\alpha_0)/\partial\beta\partial\beta']H^{\beta} + E[G^{\beta\prime}_{\beta i}e_r g^{\beta\prime}_i + g^{\beta}_{ir}G^{\beta\prime}_{\beta i}]P^{\beta})E[g^{\beta.\alpha}_i g^{\beta.\alpha}_i]P^{\beta}e_r/2n_\beta \\
&-\sum_{r=1}^{m_\beta}H^{\beta}(E[g^{\beta}_i e_r' G^{\beta}_{\beta i} + g^{\beta}_{ir}G^{\beta}_{\beta i}]H^{\beta} \\
&-\rho^{\theta}_{vvv}(0)E[g^{\beta}_{ir}g^{\beta}_i g^{\beta\prime}_i]P^{\beta})E[g^{\beta.\alpha}_i g^{\beta.\alpha}_i]P^{\beta}e_r/2n_\beta \\
&-\sum_{s=1}^{p_\alpha}[-\Sigma^{\beta\beta}E[G^{\beta s}_{\beta\alpha i}]'P^{\beta} \\
&+H^{\beta}(E[G^{\beta s}_{\beta\alpha i}]H^{\beta} + E[g^{\beta s}_{\alpha i}g^{\beta\prime}_i + g^{\beta}_i g^{\beta s\prime}_{\alpha i}]P^{\beta})]E[g^{\beta.\alpha}_i g^{\alpha\prime}_i]H^{\alpha\prime}e_s/n_\beta \\
&-H^{\beta}\sum_{s=1}^{p_\alpha}E[\partial^2 g^{\beta}_i/\partial\alpha_s\partial\alpha']\Sigma^{\alpha\alpha}e_s/2n_\beta.
\end{aligned}
$$

Simplifying and collecting terms gives the result in Theorem 3.4. ■

Appendix B: Some Notation

We use the generic notation e_r and e_s to indicate unit vectors of dimension indicated by context.

B.1 System-α

$$
\begin{aligned}
g^{\alpha}_j(\alpha) &\equiv g^{\alpha}(x_j,\alpha),\ (j=1,...,n_\beta),\ \hat{g}^{\alpha}(\alpha) \equiv \sum_{j=1}^{n_\alpha} g^{\alpha}_j(\alpha)/n_\alpha, \\
\hat{\Omega}^{\alpha\alpha}(\alpha) &\equiv \sum_{j=1}^{n_\alpha} g^{\alpha}_j(\alpha)g^{\alpha}_j(\alpha)'/n_\alpha.
\end{aligned}
$$

B.2 System-β

$$
\begin{aligned}
g^{\beta}_i(\alpha,\beta) &\equiv g^{\beta}(z_i,\alpha,\beta),\ (i=1,...,n_\beta),\ \hat{g}^{\beta}(\alpha,\beta) \equiv \sum_{i=1}^{n_\beta} g^{\beta}_i(\alpha,\beta)/n_\beta, \\
g^{\beta}_k(\alpha,\beta) &\equiv g^{\beta}(z_k,\alpha,\beta), (k=1,...,n),\ \hat{\Omega}^{\beta\beta}(\alpha,\beta) \equiv \sum_{k=1}^{n} g^{\beta}_k(\alpha,\beta)g^{\beta}_k(\alpha,\beta)'/n.
\end{aligned}
$$

B.3 Asymptotic Bias System-α

$$\begin{aligned}
g_j^\alpha &= g_j^\alpha(\alpha_0), G_j^\alpha(\alpha) = \partial g_j^\alpha(\alpha)/\partial\alpha', \\
G_j^\alpha &= G_j^\alpha(\alpha_0), (j = 1, ..., n_\alpha), \\
G^\alpha &= E[G_j^\alpha],\ \Omega^{\alpha\alpha} = E[g_j^\alpha g_j^{\alpha\prime}],\ \Sigma^{\alpha\alpha} = (G^{\alpha\prime}(\Omega^{\alpha\alpha})^{-1}G^\alpha)^{-1}, \\
H^\alpha &= \Sigma^{\alpha\alpha}G^{\alpha\prime}(\Omega^{\alpha\alpha})^{-1}, P^\alpha = (\Omega^{\alpha\alpha})^{-1} - (\Omega^{\alpha\alpha})^{-1}G^\alpha\Sigma^{\alpha\alpha}G^{\alpha\prime}(\Omega^{\alpha\alpha})^{-1}.
\end{aligned}$$

$$a_{\alpha s} \equiv tr(\Sigma^{\alpha\alpha}E[\partial^2 g_{js}^\alpha/\partial\alpha\partial\alpha'])/2, (s = 1, ..., m_\alpha). \tag{B.1}$$

B.4 Asymptotic Bias System-β

$$\begin{aligned}
g_i^\beta &= g_i^\beta(\alpha_0, \beta_0), G_{\beta i}^\beta(\alpha, \beta) = \partial g_i^\beta(\alpha, \beta)/\partial\beta', \\
G_{\beta i}^\beta &= G_{\beta i}^\beta(\alpha_0, \beta_0), (i = 1, ..., n_\beta), \\
\Omega^{\beta\beta} &= E[g_i^\beta g_i^{\beta\prime}],\ G_\beta^\beta = E[G_{\beta i}^\beta],\ \Sigma^{\beta\beta} = (G_\beta^{\beta\prime}(\Omega^{\beta\beta})^{-1}G_\beta^\beta)^{-1}, \\
H^\beta &= \Sigma^\beta G_\beta^{\beta\prime}(\Omega^{\beta\beta})^{-1},\ P^\beta = (\Omega^{\beta\beta})^{-1} - (\Omega^{\beta\beta})^{-1}G_\beta^\beta\Sigma^{\beta\beta}G_\beta^{\beta\prime}(\Omega^{\beta\beta})^{-1}.
\end{aligned}$$

$$a_{\beta r} \equiv tr(\Sigma^{\beta\beta}E[\partial^2 g_{ir}^\beta/\partial\beta\partial\beta'])/2, (r = 1, ..., m_\beta). \tag{B.2}$$

$$\begin{aligned}
G_{\alpha i}^\beta(\alpha, \beta) &= \partial g_i^\beta(\alpha, \beta)/\partial\alpha', G_{\alpha i}^\beta = G_{\alpha i}^\beta(\alpha_0, \beta_0), G_\alpha^\beta = E[G_{\alpha i}^\beta] \\
\Sigma_W^{\beta\beta} &= (G_\beta^{\beta\prime}(W^{\beta\beta})^{-1}G_\beta^\beta)^{-1}, H_W^\beta = \Sigma_W^{\beta\beta}G_\beta^{\beta\prime}(W^{\beta\beta})^{-1}.
\end{aligned}$$

B.5 Independent Samples

$$\begin{aligned}
a_{\beta\beta r}^\beta &= tr(H^\beta G_\alpha^\beta\Sigma^{\alpha\alpha}G_\alpha^{\beta\prime}H^{\beta\prime}E[\partial^2 g_{ir}^\beta/\partial\beta\partial\beta'])/2,\ a_{\beta\alpha r}^\beta = -tr(H^\beta G_\alpha^\beta\Sigma^{\alpha\alpha}E[\partial^2 g_{ir}^\beta/\partial\alpha\partial\beta']), \\
a_{\alpha\alpha r}^\beta &= tr(\Sigma^{\alpha\alpha}E[\partial^2 g_{ir}^\beta/\partial\alpha\partial\alpha'])/2, (r = 1, ..., m_\beta).
\end{aligned}$$

$$\begin{aligned}
c_{\beta\beta r}^\beta &= tr(E[\partial^2 g_i^{\beta\prime}/\partial\beta\partial\beta_r]P^\beta G_\alpha^\beta\Sigma^{\alpha\alpha}G_\alpha^{\beta\prime}H^{\beta\prime}), \\
c_{\beta\alpha r}^\beta &= -tr(E[\partial^2 g_i^{\beta\prime}/\partial\alpha\partial\beta_r]P^\beta G_\alpha^\beta\Sigma^{\alpha\alpha}),\ (r = 1, ..., p_\beta).
\end{aligned}$$

B.6 Identical Samples

$$g_i^{\beta.\alpha} = g_i^\beta - G_\alpha^\beta H^\alpha g_i^\alpha, (i = 1, ..., n_\beta),$$
$$\Omega^{\beta\beta.\alpha\alpha} = E[g_i^{\beta.\alpha} g_i^{\beta.\alpha\prime}], \Omega^{\beta\beta.\alpha} = E[g_i^\beta g_i^{\beta.\alpha\prime}], \Omega^{\alpha\beta.\alpha} = E[g_i^\alpha g_i^{\beta.\alpha\prime}]$$

$$a_{\beta\beta r}^\beta = tr(H^\beta \Omega^{\beta\beta.\alpha\alpha} H^{\beta\prime} E[\partial^2 g_{ir}^\beta / \partial\beta\partial\beta'])/2,\ a_{\beta\alpha r}^\beta = tr(H^\alpha \Omega^{\alpha\beta.\alpha} H^{\beta\prime} E[\partial^2 g_{ir}^\beta / \partial\beta\partial\alpha']),$$
$$a_{\alpha\alpha r}^\beta = tr(\Sigma^{\alpha\alpha} E[\partial^2 g_{ir}^\beta / \partial\alpha\partial\alpha'])/2, (r = 1, ..., m_\beta),$$

$$c_{\beta\beta r}^\beta = tr(H^\beta \Omega^{\beta\beta.\alpha\alpha} P^\beta E[\partial^2 g_i^\beta / \partial\beta'\partial\beta_r]),$$
$$c_{\beta\alpha r}^\beta = tr(H^\alpha \Omega^{\alpha\beta.\alpha} P^\beta E[\partial^2 g_i^\beta / \partial\alpha'\partial\beta_r]),\ (r = 1, ..., p_\beta).$$

References

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.

Brown, B.W., W.K. Newey and S. May (1997): "Efficient Bootstrapping for GMM", *mimeo*, M.I.T.

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

Hall, P., and J.L. Horowitz (1996): "Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators", *Econometrica*, 50, 1029-1054.

Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica*, 64, 891-916.

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.

Heckman, J.J. (1979): "Sample Selection Bias as a Specification Error", *Econometrica*, 47, 153-162.

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.

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

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. and R.J. Smith (2002): "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *mimeo*, M.I.T. and University of Warwick.

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

Pagan, A. (1984): "Econometric Issues in the Analysis of Regressions with Generated Regressors," *International Economic Review*, 25, 221-247.

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

Rothenberg, T.J. (1996): "Empirical Likelihood Parameter Estimation Under Moment Restrictions," seminar notes, Harvard/M.I.T. and University of 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 Criteria for Moment Condition Models", *mimeo*, University of Bristol. Revision of paper presented at Cowles Foundation Econometrics Conference on "New Developments in Time Series Econometrics", Yale University, 1999.

Table 1: Covariance Structure Models: $n_\beta = 100$

Estimator	Bias		Quantiles		SE	RMSE	MAE
	Mean	Median	0.05	0.95			
			t_5				
2S-GMM	-.111	-.116	0.789	0.998	.065	.129	.116
CUE	-.125	-.128	0.765	0.990	.069	.143	.128
ET	-.094	-.098	0.805	1.021	.067	.115	.099
EL	-.065	-.069	0.835	1.057	.067	.094	.073
			t_{10}				
2S-GMM	-.059	-.060	0.856	1.026	.053	.079	.062
CUE	-.066	-.067	0.845	1.022	.055	.086	.068
ET	-.046	-.048	0.866	1.042	.054	.071	.053
EL	-.028	-.030	0.886	1.063	.055	.062	.043
			Normal				
2S-GMM	-.036	-.034	0.889	1.041	.047	.059	.041
CUE	-.040	-.038	0.881	1.039	.049	.063	.044
ET	-.026	-.025	0.896	1.051	.048	.055	.038
EL	-.015	-.012	0.905	1.063	.048	.050	.035
			Uniform				
2S-GMM	-.007	-.008	0.946	1.043	.029	.030	.021
CUE	-.008	-.009	0.945	1.042	.030	.031	.021
ET	-.005	-.007	0.948	1.045	.030	.030	.020
EL	-.003	-.004	0.950	1.048	.030	.030	.020
			Lognormal				
2S-GMM	-.415	-.430	0.434	0.777	.111	.429	.430
CUE	-.481	-.490	0.332	0.727	.125	.497	.490
ET	-.396	-.408	0.429	0.807	.120	.414	.408
EL	-.303	-.317	0.513	0.927	.131	.331	.317
			Exponential				
2S-GMM	-.141	-.146	0.722	1.004	.087	.166	.147
CUE	-.162	-.166	0.680	0.996	.097	.189	.166
ET	-.108	-.110	0.751	1.043	.088	.140	.113
EL	-.058	-.061	0.803	1.097	.087	.105	.076
			Bimodal				
2S-GMM	-.009	-.009	0.945	1.036	.028	.029	.020
CUE	-.010	-.010	0.944	1.035	.028	.030	.021
ET	-.006	-.005	0.948	1.040	.028	.029	.020
EL	-.002	-.001	0.951	1.044	.028	.028	.019

Table 2: Covariance Structure Models: $n_\beta = 500$

Estimator	Bias		Quantiles		SE	RMSE	MAE
	Mean	Median	0.05	0.95			
			t_5				
2S-GMM	-.041	-.042	0.904	1.013	.034	.053	.042
CUE	-.042	-.043	0.903	1.012	.034	.054	.043
ET	-.029	-.029	0.917	1.024	.033	.044	.031
EL	-.016	-.016	0.929	1.039	.034	.038	.026
			t_{10}				
2S-GMM	-.016	-.016	0.945	1.024	.025	.029	.021
CUE	-.016	-.016	0.945	1.024	.025	.030	.021
ET	-.010	-.010	0.952	1.030	.024	.026	.018
EL	-.004	-.005	0.957	1.036	.025	.025	.017
			Normal				
2S-GMM	-.008	-.008	0.959	1.027	.021	.022	.015
CUE	-.008	-.008	0.959	1.027	.021	.022	.015
ET	-.005	-.005	0.962	1.030	.020	.021	.014
EL	-.001	-.001	0.965	1.034	.021	.021	.014
			Uniform				
2S-GMM	-.002	-.002	0.976	1.019	.013	.013	.009
CUE	-.002	-.002	0.976	1.019	.013	.013	.009
ET	-.001	-.001	0.977	1.019	.013	.013	.009
EL	-.001	-.001	0.977	1.019	.013	.013	.009
			Lognormal				
2S-GMM	-.225	-.227	0.652	0.917	.082	.239	.227
CUE	-.231	-.233	0.634	0.912	.085	.246	.233
ET	-.178	-.182	0.705	0.965	.079	.194	.182
EL	-.118	-.124	0.757	1.034	.081	.143	.125
			Exponential				
2S-GMM	-.041	-.042	0.894	1.029	.040	.057	.044
CUE	-.042	-.043	0.892	1.028	.040	.058	.045
ET	-.024	-.025	0.914	1.043	.039	.046	.032
EL	-.006	-.007	0.929	1.059	.039	.040	.029
			Bimodal				
2S-GMM	-.002	-.001	0.977	1.018	.012	.013	.009
CUE	-.002	-.001	0.976	1.018	.012	.013	.009
ET	-.001	-.000	0.978	1.019	.012	.012	.008
EL	-.000	.001	0.979	1.020	.012	.012	.008

Table 3: Covariance Structure Models: Bias-Corrected and Bootstrap GMM Estimators: n_β = 100

Estimator	Bias		Quantiles		SE	RMSE	MAE
	Mean	Median	0.05	0.95			
			t_5				
2S-GMM	-.111	-.116	0.789	0.998	.065	.129	.116
NP	-.073	-.079	0.808	1.061	.076	.105	.084
RNP	-.049	-.056	0.834	1.084	.077	.091	.068
FSGEL	-.044	-.050	0.845	1.089	.075	.086	.065
BCa	-.060	-.066	0.828	1.065	.072	.094	.072
BCb	-.049	-.054	0.841	1.081	.073	.088	.067
BCc	-.067	-.073	0.817	1.065	.076	.101	.079
BCd	-.016	-.021	0.884	1.093	.065	.067	.047
			t_{10}				
2S-GMM	-.059	-.060	0.856	1.026	.053	.079	.062
NP	-.026	-.028	0.881	1.072	.060	.065	.046
RNP	-.017	-.020	0.890	1.079	.059	.061	.044
FSGEL	-.011	-.013	0.899	1.084	.058	.059	.040
BCa	-.018	-.020	0.891	1.076	.057	.060	.043
BCb	-.015	-.017	0.895	1.079	.057	.059	.043
BCc	-.022	-.024	0.882	1.077	.061	.065	.045
BCd	-.002	-.003	0.914	1.083	.053	.053	.036
			Normal				
2S-GMM	-.036	-.034	0.889	1.041	.047	.059	.041
NP	-.008	-.007	0.911	1.074	.050	.051	.036
RNP	-.005	-.004	0.916	1.076	.050	.050	.035
FSGEL	-.001	.000	0.920	1.078	.049	.049	.033
BCa	-.004	-.004	0.918	1.076	.049	.049	.034
BCb	-.003	-.002	0.918	1.076	.049	.049	.034
BCc	-.007	-.007	0.910	1.079	.053	.053	.038
BCd	.002	.003	0.926	1.078	.047	.047	.033
			Uniform				
2S-GMM	-.007	-.008	0.946	1.043	.029	.030	.021
NP	.006	.004	0.958	1.055	.030	.030	.020
RNP	.005	.004	0.959	1.055	.030	.030	.020
FSGEL	.007	.006	0.961	1.057	.030	.030	.020
BCa	.005	.004	0.958	1.055	.030	.030	.020
BCb	.005	.004	0.958	1.055	.029	.030	.020
BCc	.005	.003	0.954	1.058	.032	.032	.022
BCd	.006	.005	0.959	1.055	.029	.030	.020
			Lognormal				
2S-GMM	-.415	-.430	0.434	0.777	.111	.429	.430
NP	-.380	-.403	0.429	0.887	.145	.407	.403
RNP	-.230	-.282	0.511	1.128	.453	.508	.289
FSGEL	-.264	-.290	0.531	1.024	.158	.308	.292
BCa	-.352	-.371	0.465	0.889	.135	.377	.371
BCb	-.278	-.302	0.524	0.991	.152	.317	.303
BCc	-.369	-.393	0.449	0.874	.137	.394	.393
BCd	-.096	-.111	0.753	1.096	.111	.147	.121
			Exponential				
2S-GMM	-.141	-.146	0.722	1.004	.087	.166	.147
NP	-.089	-.095	0.744	1.096	.108	.140	.107
RNP	-.060	-.066	0.771	1.125	.105	.122	.085
FSGEL	-.042	-.048	0.799	1.136	.102	.110	.077
BCa	-.080	-.086	0.764	1.092	.099	.128	.097
BCb	-.059	-.065	0.788	1.115	.098	.114	.082
BCc	-.089	-.096	0.746	1.092	.104	.137	.106
BCd	-.026	-.031	0.838	1.119	.087	.091	.060
			Bimodal				
2S-GMM	-.009	-.009	0.945	1.036	.028	.029	.020
NP	.006	.006	0.958	1.051	.029	.029	.021
RNP	.006	.006	0.959	1.052	.028	.029	.020
FSGEL	.008	.008	0.963	1.053	.028	.029	.020
BCa	.007	.007	0.960	1.052	.028	.029	.021
BCb	.007	.006	0.960	1.052	.028	.029	.021
BCc	.006	.006	0.955	1.055	.031	.031	.022
BCd	.008	.008	0.962	1.052	.028	.029	.020

Table 4: Covariance Structure Models: Bias-Corrected and Bootstrap GMM Estimators: n_β = 500

Estimator	Bias		Quantiles		SE	RMSE	MAE
	Mean	Median	0.05	0.95			
			t_5				
2S-GMM	-.041	-.042	0.904	1.013	.034	.053	.042
NP	-.020	-.020	0.921	1.042	.038	.042	.029
RNP	-.014	-.015	0.927	1.050	.039	.041	.028
FSGEL	-.014	-.015	0.927	1.048	.037	.040	.028
BCa	-.017	-.018	0.924	1.043	.037	.041	.028
BCb	-.015	-.016	0.926	1.045	.037	.040	.028
BCc	-.018	-.019	0.923	1.043	.037	.041	.029
BCd	-.004	-.004	0.942	1.050	.034	.034	.023
			t_{10}				
2S-GMM	-.016	-.016	0.945	1.024	.025	.029	.021
NP	-.003	-.003	0.955	1.039	.026	.026	.018
RNP	-.002	-.002	0.957	1.040	.026	.026	.018
FSGEL	-.002	-.001	0.958	1.040	.026	.026	.018
BCa	-.002	-.002	0.957	1.039	.026	.026	.018
BCb	-.002	-.002	0.958	1.039	.026	.026	.018
BCc	-.002	-.003	0.956	1.040	.026	.026	.018
BCd	.000	.000	0.961	1.040	.025	.025	.017
			Normal				
2S-GMM	-.008	-.008	0.959	1.027	.021	.022	.015
NP	.000	.001	0.966	1.035	.021	.021	.014
RNP	.001	.001	0.966	1.035	.021	.021	.014
FSGEL	.001	.001	0.966	1.036	.021	.021	.014
BCa	.001	.001	0.967	1.036	.021	.021	.013
BCb	.001	.001	0.967	1.036	.021	.021	.013
BCc	.001	.000	0.967	1.036	.021	.021	.014
BCd	.001	.001	0.968	1.036	.021	.021	.013
			Uniform				
2S-GMM	-.002	-.002	0.976	1.019	.013	.013	.009
NP	.001	.001	0.979	1.022	.013	.013	.009
RNP	.001	.001	0.979	1.021	.013	.013	.009
FSGEL	.001	.001	0.980	1.022	.013	.013	.009
BCa	.001	.001	0.979	1.021	.013	.013	.008
BCb	.001	.001	0.979	1.021	.013	.013	.008
BCc	.001	.001	0.979	1.022	.013	.013	.009
BCd	.001	.001	0.979	1.022	.013	.013	.008
			Lognormal				
2S-GMM	-.225	-.227	0.652	0.917	.082	.239	.227
NP	-.161	-.166	0.674	1.027	.108	.194	.168
RNP	-.107	-.118	0.724	1.109	.123	.163	.129
FSGEL	-.121	-.128	0.720	1.068	.106	.161	.131
BCa	-.161	-.166	0.691	1.007	.097	.188	.166
BCb	-.132	-.138	0.724	1.038	.097	.164	.139
BCc	-.164	-.169	0.687	1.005	.098	.191	.170
BCd	-.044	-.046	0.833	1.098	.082	.093	.067
			Exponential				
2S-GMM	-.041	-.042	0.894	1.029	.040	.057	.044
NP	-.012	-.013	0.914	1.065	.044	.046	.032
RNP	-.009	-.011	0.919	1.066	.044	.045	.031
FSGEL	-.007	-.009	0.923	1.069	.043	.044	.030
BCa	-.011	-.013	0.917	1.062	.043	.044	.030
BCb	-.009	-.011	0.921	1.064	.043	.044	.030
BCc	-.012	-.013	0.916	1.062	.043	.045	.031
BCd	-.003	-.004	0.932	1.067	.040	.040	.027
			Bimodal				
2S-GMM	-.002	-.001	0.977	1.018	.012	.013	.009
NP	.002	.002	0.980	1.021	.013	.013	.008
RNP	.002	.002	0.980	1.021	.013	.013	.008
FSGEL	.002	.002	0.981	1.022	.012	.013	.008
BCa	.002	.003	0.980	1.022	.012	.013	.008
BCb	.002	.003	0.980	1.022	.012	.013	.008
BCc	.002	.003	0.980	1.022	.013	.013	.009
BCd	.002	.003	0.981	1.022	.012	.013	.009