

Smith, Richard J.

**Working Paper**

Weak instruments and empirical likelihood: A discussion of the papers by D. W. K. Andrews and J. H. Stock and Y. Kitamura

cemmap working paper, No. CWP13/05

**Provided in Cooperation with:**

The Institute for Fiscal Studies (IFS), London

*Suggested Citation:* Smith, Richard J. (2005) : Weak instruments and empirical likelihood: A discussion of the papers by D. W. K. Andrews and J. H. Stock and Y. Kitamura, cemmap working paper, No. CWP13/05, Centre for Microdata Methods and Practice (cemmap), London, <https://doi.org/10.1920/wp.cem.2005.1305>

This Version is available at:

<https://hdl.handle.net/10419/79367>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

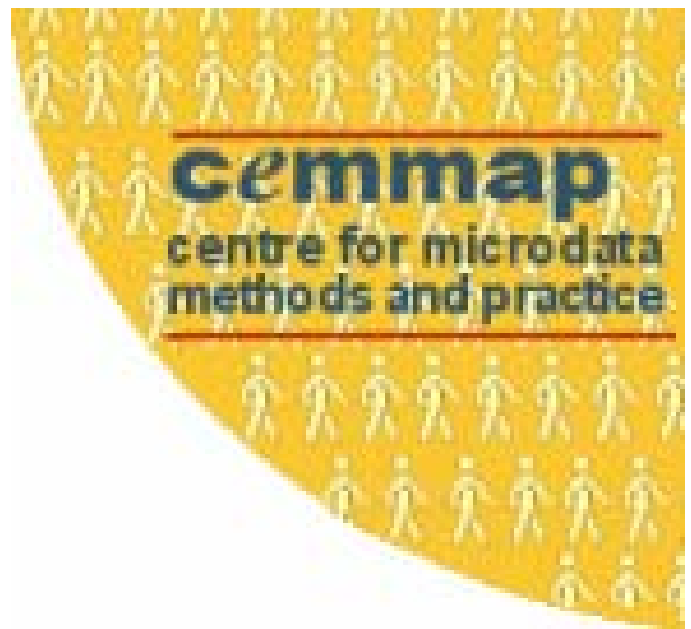
Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



WEAK INSTRUMENTS AND EMPIRICAL LIKELIHOOD:  
A DISCUSSION OF THE PAPERS BY D.W.K. ANDREWS  
AND J.H. STOCK AND Y. KITAMURA

---

*Richard J. Smith*

THE INSTITUTE FOR FISCAL STUDIES  
DEPARTMENT OF ECONOMICS, UCL  
**cemmap** working paper CWP13/05

WEAK INSTRUMENTS AND EMPIRICAL LIKELIHOOD:  
A DISCUSSION OF THE PAPERS BY D.W.K. ANDREWS  
AND J.H. STOCK AND Y. KITAMURA\*

Richard J. Smith

cemmap

U.C.L and I.F.S.

and

Faculty of Economics

University of Cambridge

First Draft: August 2005

This Revision: October 2005

**Abstract**

Initially this discussion briefly reviews the contributions of Andrews and Stock and Kitamura, henceforth AS and K respectively. Because the breadth of material covered by AS and K is so vast, we concentrate only on a few topics. Generalized empirical likelihood (GEL) provides the focus for the discussion. By defining an appropriate set of nonlinear moment conditions, GEL estimation yields objects which mirror in an asymptotic sense those which form the basis of the exact theory in AS allowing the definition of asymptotically pivotal test statistics appropriate for weakly identified models, the acceptance regions of which may then be inverted to provide asymptotically valid confidence interval estimators for the parameters of interest. The general minimum distance approach of Corcoran (1998) which parallels the information theoretic development of EL in K is briefly reviewed. A new class of estimators mirroring Schennach (2004) is suggested which shares the same asymptotic bias properties of EL and possess a well-defined limit distribution under misspecification.

**JEL Classification:** C13, C30

**Keywords:** Empirical Likelihood, Generalized Empirical Likelihood, Weak Identification, Minimum Distance, Asymptotic Bias, Higher Order Efficiency, Misspecification.

---

\*This paper reports the Invited Discussion of the Symposium on Weak Instruments and Empirical Likelihood held at the Econometric Society World Congress, U.C.L., London (2005). I am grateful for insightful discussions with G.H. Hillier which helped me formulate the weak identification section of my discussion. Neither he nor anyone else bears responsibility for any misconceptions that remain. The research for this Invited Discussion benefitted from the financial support of a 2002 Leverhulme Major Research Fellowship.

# 1 Introduction

These two papers represent the fruition of important and thorough investigations undertaken by the authors of their respective fields of enquiry. I feel that they will add considerably to our understanding of these topics. Before describing the contents of my discussion I initially and briefly outline the contributions of both sets of authors.

Andrews and Stock (2005), henceforth referred to as AS, continues the programme of research initiated with the papers by Moreira (2001, 2003) through Andrews, Moriera and Stock (2004), henceforth AMS. Like those contributions, this paper is primarily concerned with the weak instrument problem for the classical two variable linear simultaneous equations model with normally distributed reduced form errors and known error variance matrix. The particular advantage of using a well-understood classical framework for analysis is that results here as elsewhere should have important implications and conclusions for estimators and statistics in more general settings enabling specific recommendations for practice. Apart from reviewing and detailing existing results, this paper provides a comprehensive treatment of the many weak instrumental variables problem for this model. Generally speaking with weak instruments standard point estimators such as 2SLS and LIML are no longer consistent and have non-standard limiting distributions which cannot be consistently estimated. Therefore recourse is typically made to tests based on unconditionally or conditionally pivotal statistics. Acceptance regions associated with these tests may then be inverted to provide valid confidence interval estimators for the parameters of interest. I now briefly summarise their findings and conclusions.

AMS obtains the power envelope for two-sided similar tests of the structural parameter *via* consideration of point optimal tests. The power envelope changes little between similar and nonsimilar tests. A new test class is obtained which maximises weighted average power. However, the conditional likelihood ratio (CLR) test due to Moreira (2003), also a similar test, comes close to reaching the power envelope as does the CLR test with estimated reduced form error variance matrix. Apart from surveying extant results in the literature, AS again mostly confines attention to invariant similar tests and extends the analysis of AMS for many weak instruments. Let  $\lambda_\pi$  denote the concentration parameter,  $k$  the number of instruments and  $n$  the sample size. AS characterise the various situations under consideration by the limit of the ratio  $\lambda_\pi/k^\tau \rightarrow r_\tau$ ,  $r_\tau \in [0, \infty)$ ,  $\tau \in (0, \infty)$ , where  $k \rightarrow \infty$ ,  $n \rightarrow \infty$ . Briefly, (a)  $\tau \in (0, 1/2)$ , there is no test with non-trivial power, (b)  $\tau = 1/2$ , the Anderson-Rubin (AR), Lagrange multiplier (LM) and likelihood ratio (LR) statistics all have non-trivial power, (c)  $\tau > 1/2$ , the AR statistic has trivial power whereas the LM and LR statistics are asymptotically equivalent and have non-trivial power. AS also obtain the asymptotic power envelopes using a least favourable distribution approach to circumvent the difficulty of the composite null hypothesis, thus, enabling the application of classical Neyman-Pearson theory. They find that for  $\tau = 1/2$  the CLR test is close to the asymptotically efficient power envelope and that for  $\tau > 1/2$  the LM and CLR tests are asymptotically equivalent. As a consequence, tests based on the CLR statistic are to be recommended as a useful and powerful tool in weakly identified models.

Kitamura (2005), henceforth K, provides an extensive overview of empirical likelihood (EL), see Owen (2001). K demonstrates the well known result that EL is a nonparametric maximum likelihood estimator

but, of particular note, also reinterprets EL as a generalized minimum contrast (GMC) estimator using an information-theoretic treatment based on Fenchel duality, see Borwein and Lewis (1991). This GMC interpretation of EL is particularly useful when considering issues of estimation and inference in the presence of misspecification, allowing a generalization of the analysis of likelihood ratio test statistics of Vuong (1989) to the nested and non-nested moment restrictions environment. Both unconditional and conditional moment settings are treated.

A broader question concerns why the more computationally complex EL should be entertained instead of GMM [Hansen (1982)]. It is now commonly appreciated that EL possesses some desirable higher order properties. The asymptotic bias of EL is that of an infeasible GMM estimator when Jacobian and moment indicator variance matrices are known. Furthermore, bias-corrected EL is higher order efficient. See Newey and Smith (2004), henceforth NS. A particularly innovative approach taken by K is the application of the theory of large deviations. Recent work co-authored with Otsu shows that a minimax EL estimator achieves the asymptotic minimax lower bound. The EL criterion function statistic also provides an asymptotically optimal test, see Kitamura (2001). This statistic has the added advantage of Bartlett correctability. K provides some simulation evidence on the efficacy of these procedures, minimax EL generally appearing superior and good coverage probabilities for the EL criterion function statistics. It is worth noting that when heteroskedastic models are considered EL is generally competitive, being internally self-studentised, as compared with homoskedastic environments. A number of other applications of EL are also briefly surveyed by K including time series moment condition models, conditional moment restrictions and weak instruments among many others.

Because the breadth of the topics covered by AS and K is so vast, I must necessarily confine my discussion to a limited number of topics. To provide some focus, I use generalized empirical likelihood (GEL) as an organisational tool. Section 2 briefly summarises the first order theory concerning GEL and defines some objects needed later.

Weak identification is the subject of section 3. By considering a model specified by nonlinear moment conditions, an appropriate GEL estimation problem allows consideration of objects which mirror in an asymptotic sense those which form the basis of the exact theory in AMS and AS. As a consequence, we define asymptotically pivotal statistics, the acceptance regions of which may then be inverted to provide asymptotically valid confidence interval estimators for the parameters of interests. The resultant statistics are compared and contrasted with those already extant in the literature and some new statistics are also defined which may warrant further investigation.

Paralleling the information theoretic development of EL in K, section 4 briefly discusses the contribution of Corcoran (1998) which provides a general minimum distance (MD) approach to estimation and is the empirical counterpart to the analysis given by K. By comparing first order conditions as in Newey and Smith (2001), similarly to K, it is immediately apparent that although the MD class has many members in common with GEL they do not coincide.

Schennach (2004) casts some doubt on the efficacy of EL in misspecified situations and proves that an alternative estimator which embeds exponential tilting (ET) empirical probabilities in the EL criterion not only has desirable asymptotic properties under misspecification but when bias-corrected is also

higher order efficient, sharing the higher order bias and variance properties of EL when the moment restrictions are correct. Section 5 suggests an equivalent approach based on GEL rather than ET empirical probabilities. We show that the resultant estimator has the same asymptotic bias as EL and hazard that its higher order variance also coincides with that of EL. Furthermore, this estimator should also have useful properties for misspecified moment conditions as discussed in section 6.

## 2 GEL

We consider a model defined by a finite number of non-linear moment restrictions. That is, the model has a true parameter  $\beta_0$  satisfying the moment condition

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

where  $g(z, \beta)$  an  $m$ -vector of functions of the data observation  $z$  and the parameter  $\beta$ , a  $p$ -vector of parameters, where  $m \geq p$  and  $E[\cdot]$  denotes expectation taken with respect to the distribution of  $z$ . We assume throughout this discussion that  $z_i$ , ( $i = 1, \dots, n$ ), are i.i.d. observations on the data vector  $z$ .

Let  $g_i(\beta) \equiv g(z_i, \beta)$ ,  $\hat{g}(\beta) \equiv n^{-1} \sum_{i=1}^n g_i(\beta)$  and  $\hat{\Omega}(\beta) \equiv n^{-1} \sum_{i=1}^n g_i(\beta)g_i(\beta)'$ .

The class of GEL estimators, [NS, Smith (1997, 2001)], is based on  $\rho(v)$ , a function of a scalar  $v$  that is concave on its domain, an open interval  $\mathcal{V}$  containing zero, and, without loss of generality, normalized with  $\rho_1 = \rho_2 = -1$  where  $\rho_j(v) = \partial^j \rho(v) / \partial v^j$  and  $\rho_j = \rho_j(0)$ , ( $j = 0, 1, 2, \dots$ ). Let  $\hat{\Lambda}_n(\beta) = \{\lambda : \lambda' g_i(\beta) \in \mathcal{V}, i = 1, \dots, n\}$ . The GEL estimator is the solution to a saddle point problem

$$\hat{\beta} = \arg \min_{\beta \in \mathcal{B}} \sup_{\lambda \in \hat{\Lambda}_n(\beta)} \sum_{i=1}^n \rho(\lambda' g_i(\beta)), \quad (2.1)$$

where  $\mathcal{B}$  denotes the parameter space. Both EL and exponential tilting (ET) estimators are special cases of GEL with  $\rho(v) = \ln(1 - v)$  and  $\mathcal{V} = (-\infty, 1)$ , [Qin and Lawless (1994), Imbens (1997) and Smith (1997)] and  $\rho(v) = -\exp(v)$ , [Kitamura and Stutzer (1997), Imbens, Spady and Johnson (1998) and Smith (1997)], respectively, as is the continuous updating estimator (CUE), [Hansen, Heaton, and Yaron (1996)], if  $\rho(v)$  is quadratic as shown by NS (2004, Theorem 2.1, p.223).<sup>1</sup>

We adopt the following assumptions from NS. Let  $G \equiv E[\partial g(z, \beta_0) / \partial \beta']$ ,  $\Omega \equiv E[g(z, \beta_0)g(z, \beta_0)']$  and  $\mathcal{N}$  denote a neighborhood of  $\beta_0$ .

**Assumption 2.1** (a)  $\beta_0 \in \mathcal{B}$  is the unique solution to  $E[g(z, \beta)] = 0$ ; (b)  $\mathcal{B}$  is compact; (c)  $g(z, \beta)$  is continuous at each  $\beta \in \mathcal{B}$  with probability one; (d)  $E[\sup_{\beta \in \mathcal{B}} \|g(z, \beta)\|^\alpha] < \infty$  for some  $\alpha > 2$ ; (e)  $\Omega$  is nonsingular; (f)  $\rho(v)$  is twice continuously differentiable in a neighborhood of zero.

**Assumption 2.2** (a)  $\beta_0 \in \text{int}(\mathcal{B})$ ; (b)  $g(z, \beta)$  is continuously differentiable in  $\mathcal{N}$  and  $E[\sup_{\beta \in \mathcal{N}} \|\partial g(z, \beta) / \partial \beta'\|] < \infty$ ; (c)  $\text{rank}(G) = p$ .

---

<sup>1</sup>The CUE is analogous to GMM and is given by  $\hat{\beta}_{CUE} = \arg \min_{\beta \in \mathcal{B}} \hat{g}(\beta)' \hat{\Omega}(\beta)^- \hat{g}(\beta)$ , where  $A^-$  denotes any generalized inverse of a matrix  $A$  satisfying  $AA^-A = A$ . The two-step GMM estimator is  $\hat{\beta}_{GMM} = \arg \min_{\beta \in \mathcal{B}} \hat{g}(\beta)' \hat{\Omega}(\tilde{\beta})^{-1} \hat{g}(\beta)$ , where  $\tilde{\beta}$  is some preliminary consistent estimator for  $\beta_0$ .

Assumption 2.1 is sufficient for the consistency of  $\hat{\beta}$  for  $\beta_0$  whereas taken together with Assumption 2.2 the large sample normality of  $\hat{\beta}$  and  $\hat{\lambda}$  may be shown. See NS, Theorems 3.1 and 3.2.

Let  $\hat{\lambda} \equiv \hat{\lambda}(\hat{\beta})$  where  $\hat{\lambda}(\beta) \equiv \arg \sup_{\lambda \in \hat{\Lambda}_n(\beta)} \sum_{t=1}^T \rho(\lambda' g_i(\beta))/n$ .

**Theorem 2.1** *If Assumptions 2.1 and 2.2 are satisfied then*

$$n^{1/2} \begin{pmatrix} \hat{\beta} - \beta_0 \\ \hat{\lambda} \end{pmatrix} \xrightarrow{d} N(0, \text{diag}(\Sigma, P)), 2n[\sum_{i=1}^n \rho(\hat{\lambda}' g_i(\hat{\beta}))/n - \rho_0] \xrightarrow{d} \chi^2(m-p),$$

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

Given  $\rho(v)$ , empirical probabilities for the observations may also be described

$$\pi_i(\beta) \equiv \frac{\rho_1(\hat{\lambda}(\beta)' g_i(\beta))}{\sum_{j=1}^n \rho_1(\hat{\lambda}(\beta)' g_j(\beta))}, (i = 1, \dots, n); \quad (2.2)$$

cf. Back and Brown (1993). The GEL empirical probabilities  $\pi_i(\beta)$ , ( $i = 1, \dots, n$ ), sum to one by construction, satisfy the sample moment condition  $\sum_{i=1}^n \pi_i(\beta) g_i(\beta) = 0$  when the first order conditions for  $\hat{\lambda}(\beta)$  hold, and are positive when  $\hat{\lambda}(\hat{\beta})' \hat{g}_i(\hat{\beta})$  is small uniformly in  $i$ . From Brown and Newey (1998),  $\sum_{i=1}^n \pi_i(\hat{\beta}) a(z_i, \hat{\beta})$  is a semiparametrically efficient estimator of  $E[a(z, \beta_0)]$ .

### 3 Weak Identification

This section addresses weak identification when the moment indicators are nonlinear in  $\beta$ . The set-up used here is based on Guggenberger and Smith (2005a), henceforth GS.

Assumption 2.1 (a) implies that  $\beta_0$  is strongly identified, a conclusion that Assumption 2.2 (c) makes explicit. Therefore, we will need to revise these assumptions appropriately to address weak identification of  $\beta_0$ . As is now well documented, standard GMM and GEL estimators in the weakly identified context are inconsistent and have limiting representations which depend on parameters which cannot be consistently estimated rendering their use for estimation and inference purposes currently infeasible. See Stock and Wright (2000) for results on GMM and GS for GEL. In particular, the limit normal distributions of Theorem 2.1 will no longer hold. As a consequence, and similarly to AS, recent research for the nonlinear case has sought acceptance regions of tests for  $\beta = \beta_0$  based on asymptotically pivotal statistics which may then be inverted to provide well-defined interval estimates for  $\beta_0$ .

We will require some additional notation. Let  $G(z, \beta) \equiv \partial g(z, \beta) / \partial \beta'$  and  $G_i(\beta) \equiv G(z_i, \beta)$ .

Our interest will concern tests for the hypothesis  $H_0 : \beta = \beta_0$ . To make transparent the relation between the analysis in AS and that for GEL based procedures, we treat the Jacobian matrix  $G$  as  $mp$  additional parameters to be estimated with associated moment conditions

$$E[G(z, \beta_0) - G] = 0. \quad (3.1)$$

The resultant GEL criterion which incorporates the hypothesis  $H_0 : \beta = \beta_0$  is then  $\sum_{i=1}^n \rho(\lambda' g_i(\beta_0) + \mu' \text{vec}(G_i(\beta_0) - G))/n$  with the  $mp$ -vector  $\mu$  of auxiliary parameters associated with the additional moment constraints (3.1). It is straightforward to see that the auxiliary parameter  $\mu$  is estimated as

identically zero. Thus, the auxiliary parameter estimator  $\tilde{\lambda} = \hat{\lambda}(\beta_0)$ . Moreover, the corresponding GEL estimator for  $G$  is given by

$$\tilde{G} = \sum_{i=1}^n \pi_i(\beta_0) G_i(\beta_0), \quad (3.2)$$

where the empirical probabilities  $\pi_i(\beta_0)$ , ( $i = 1, \dots, n$ ), are defined in (2.2).<sup>2</sup>

To describe the weakly identified set-up and to detail the limiting properties of the estimators, we adapt Assumptions  $\Theta$ , ID,  $\rho$  and  $M_\theta$  of GS.

**Assumption 3.1** (a)  $\beta_0 \in \text{int}(\mathcal{B})$ ; (b)  $\mathcal{B}$  is compact.

**Assumption 3.2**  $E[\hat{g}(\beta)] = n^{-1/2}m(\beta)$ , where  $m(\beta)$  is a continuous function of  $\beta \in \mathcal{B}$  and  $m(\beta_0) = 0$ .

**Assumption 3.3**  $\rho(v)$  is twice continuously differentiable in a neighborhood of zero.

Assumption 3.2 encapsulates weak identification of  $\beta_0$  in the nonlinear moment restrictions setting. Next we detail the necessary moment assumptions.

**Assumption 3.4** (a)  $E[\sup_{\beta \in \mathcal{B}} \|g(z, \beta)\|^\alpha] < \infty$  for some  $\alpha > 2$ ; (b)  $\Omega(\beta)$  is nonsingular; (c)  $g(z, \beta)$  is continuously differentiable in  $\mathcal{N}$  and  $E[\sup_{\beta \in \mathcal{N}} \|\partial g_i(\beta)/\partial \beta'\|^\alpha] < \infty$  for some  $\alpha > 2$ ; (d)  $V(\beta) \equiv \text{var}[(g(z, \beta)', (\text{vec} G(z, \beta))')']$  is positive definite.

Partition  $V(\beta)$  conformably with  $g(z, \beta)$  and  $\text{vec}(G(z, \beta))$  as

$$V(\beta) \equiv \begin{pmatrix} \Omega(\beta) & \Delta_G(\beta)' \\ \Delta_G(\beta) & \Delta_{GG}(\beta) \end{pmatrix}.$$

The hypotheses of GS Assumption  $M_\theta$  are therefore satisfied. For example, from Assumptions 3.4 (a) and (c), by an i.i.d. WLLN,  $\hat{\Omega}(\beta) \xrightarrow{P} \Omega(\beta)$  and  $n^{-1} \sum_{i=1}^n g_i(\beta)(\text{vec} G_i(\beta))' \xrightarrow{P} \Delta_G(\beta)$ . Furthermore, by an i.i.d. CLT, on  $\mathcal{N}$ ,  $n^{-1/2} \sum_{i=1}^n ((g_i(\beta) - E[g_i(\beta)])', (\text{vec}(G_i(\beta) - E[G_i(\beta)]))')' \xrightarrow{d} N(0, V(\beta))$ . Assumption 3.4 (c) ensures that  $\partial E[\hat{g}(\beta)]/\partial \beta' = E[\partial \hat{g}(\beta)/\partial \beta']$  on  $\mathcal{N}$  and, thus, from Assumption 3.2,  $E[\hat{G}(\beta)] = n^{-1/2}M(\beta)$  where  $M(\beta) \equiv \partial m(\beta)/\partial \beta'$ .

The following theorem is the counterpart to Theorem 2.1 above for the weakly identified context.<sup>3</sup>

**Theorem 3.1** Under Assumptions 3.1-3.4,

$$n^{1/2} \begin{pmatrix} \tilde{\lambda} \\ \text{vec}(\tilde{G}) \end{pmatrix} \xrightarrow{d} N((0', (\text{vec}(M))')', \text{diag}(\Omega^{-1}, \Delta_{GG} - \Delta_G' \Omega^{-1} \Delta_G)),$$

where  $M \equiv M(\beta_0)$ ,  $\Delta_{GG} \equiv \Delta_{GG}(\beta_0)$  and  $\Delta_G \equiv \Delta_G(\beta_0)$ .

<sup>2</sup>When  $G$  is strongly identified, the Jacobian estimator  $\tilde{G}$  is an efficient estimator for  $G$  under  $H_0 : \beta = \beta_0$ ; see Brown and Newey (1998).

<sup>3</sup>If Assumption 3.2 were modified as  $E[\hat{g}(\beta)] = n^{-\tau}m(\beta)$ ,  $\tau \in (0, 1]$ , Theorem 3.1 would need to be altered appropriately, cf. AS, Theorem 1. Thus, (a)  $\tau \in (0, 1/2)$ ,  $n^{1/2} \text{vec}(\tilde{G}) \xrightarrow{d} N(0, \Delta_{GG} - \Delta_G' \Omega^{-1} \Delta_G)$  and all of the tests discussed below have trivial asymptotic power, (b)  $\tau = 1/2$ , the results are as stated in Theorems 3.1 and 3.2, (c)  $\tau > 1/2$ ,  $n^{1-\tau} \tilde{G} \xrightarrow{P} M(\beta_0)$  and the first order large sample properties of the tests are as in the strongly identified case when  $\tau = 1$ .



To aid comparison with AMS and AS, we will assume that  $p = 1$  in the remainder of this section unless otherwise indicated. AS concentrated their search for optimal tests of the hypothesis  $H_0 : \beta = \beta_0$  on invariant and similar tests. Invariance restricts attention to statistics based on the random matrix  $Q$  whereas similarity requires consideration of tests defined in terms of  $Q$  conditional on  $Q_T$ . See AS, sections 7.2 and 7.3.

First, note that  $n^{1/2}\tilde{\lambda} = -\Omega^{-1}n^{1/2}\hat{g} + o_p(1)$ . As a consequence, in large samples, it follows immediately from Theorem 3.1 that the normalised vectors

$$\begin{aligned}\tilde{S} &\equiv \Omega^{-1/2}n^{1/2}\hat{g} \xrightarrow{d} Z_S \sim N(0, I_m), \\ \tilde{T} &\equiv (\Delta_{GG} - \Delta'_G\Omega^{-1}\Delta_G)^{-1/2}n^{1/2}\tilde{G} \xrightarrow{d} Z_T \sim N(M, I_m),\end{aligned}$$

and are mutually asymptotically independent.<sup>4</sup> Because they are constructed from analogous objects apposite for the nonlinear setting, the random vectors  $\tilde{S}$  and  $\tilde{T}$  parallel  $S$  and  $T$  respectively in AS in the construction of asymptotically pivotal statistics for tests of  $H_0 : \beta = \beta_0$ ; cf. AS, eq. (7.13). Therefore, to make the analogy with AS explicit, define the random matrix  $\tilde{Q}$  as

$$\tilde{Q} \equiv (\tilde{S}, \tilde{T})'(\tilde{S}, \tilde{T}) = \begin{pmatrix} \tilde{S}'\tilde{S} & \tilde{S}'\tilde{T} \\ \tilde{T}'\tilde{S} & \tilde{T}'\tilde{T} \end{pmatrix} = \begin{pmatrix} \tilde{Q}_S & \tilde{Q}_{ST} \\ \tilde{Q}_{TS} & \tilde{Q}_T \end{pmatrix}.$$

The matrices  $\tilde{Q}$  and  $\tilde{Q}_T$  thus mirror the maximal invariant  $Q$  and  $Q_T$  in AS, equation (7.14). It follows from Theorem 3.1 that

$$\tilde{Q} \xrightarrow{d} (Z_S, Z_T)'(Z_S, Z_T)$$

which is noncentral Wishart distributed with variance matrix  $I_2$  and noncentrality parameter  $(0, M)'(0, M)$ . The asymptotic counterparts to  $Q$  and  $Q_T$  are thus  $(Z_S, Z_T)'(Z_S, Z_T)$  and  $Z_T'Z_T$ .<sup>5</sup>

Therefore, statistics corresponding to the Anderson-Rubin (AR), Lagrange multiplier (LM) and likelihood ratio (LR) statistics in AS, eq. (7.15), also see AMS, eq. (3.4), may likewise be defined as

$$\begin{aligned}\widetilde{AR} &= \tilde{Q}_S, L\widetilde{M} = \tilde{Q}_{ST}^2/\tilde{Q}_T, \\ \widetilde{LR} &= \frac{1}{2} \left( \tilde{Q}_S - \tilde{Q}_T + \sqrt{(\tilde{Q}_S - \tilde{Q}_T)^2 + 4\tilde{Q}_{ST}^2} \right).\end{aligned}$$

That is,

$$\begin{aligned}\widetilde{AR} &\equiv n\hat{g}'\Omega^{-1}\hat{g}, \\ L\widetilde{M} &\equiv (n\tilde{G}'\Delta^{-1/2}\Omega^{-1/2}\hat{g})^2/(n\tilde{G}'\Delta^{-1}\tilde{G}), \\ \widetilde{LR} &= \frac{n}{2} \left( \widetilde{AR} - n\tilde{G}'\Delta^{-1}\tilde{G} \right. \\ &\quad \left. + \sqrt{(\widetilde{AR} - \tilde{G}'\Delta^{-1}\tilde{G})^2 + 4L\widetilde{M}(\tilde{G}'\Delta^{-1}\tilde{G})} \right).\end{aligned}$$

---

<sup>4</sup>For expositional purposes only we will assume that the variance matrices  $\Omega$  and  $\Delta_{GG} - \Delta'_G\Omega^{-1}\Delta_G$  are known. As noted above, their components may be consistently estimated using the outer product form.

<sup>5</sup>It is interesting to note that when  $G$  is strongly identified  $\tilde{G}'\Omega^{-1}\tilde{G}$  is an estimator for the semiparametric counterpart of the information matrix in fully parametric models. Conditioning on such objects or their asymptotic representations, i.e.  $\tilde{G}$  or  $Z_T$  here, follows a long tradition in statistics. See, for example, Reid (1995).

**Theorem 3.2** *Let Assumptions 3.1-3.4 hold. Then, under  $H_0 : \beta = \beta_0$ , conditional on  $Z_T$ ,  $\widetilde{AR}$  and  $\widetilde{LM}$  converge in distribution to chi-square random variables with  $m$  and one degrees of freedom respectively and  $\widetilde{LR}$  converges in distribution to a random variable whose distribution is characterised by*

$$\frac{1}{2} \left( \chi^2(1) + \chi^2(m-1) - Z_T' Z_T + \sqrt{(\chi^2(1) + \chi^2(m-1) - Z_T' Z_T)^2 + 4\chi^2(1)(Z_T' Z_T)} \right),$$

where  $\chi^2(1)$  and  $\chi^2(m-1)$  denote independent chi-square random variables with one and  $m-1$  degrees of freedom respectively.

The various statistics suggested in the literature may be related to  $\widetilde{AR}$ ,  $\widetilde{LM}$  and  $\widetilde{LR}$ . Asymptotic equivalence, denoted by  $\stackrel{a}{=}$ , is under  $H_0 : \beta = \beta_0$ .

First, analogues of  $\widetilde{AR}$  may be constructed using the GEL criterion function statistic and the auxiliary parameter estimator  $\tilde{\lambda}$ , viz.

$$\begin{aligned} \mathcal{GELR}(\beta_0) &\equiv 2 \sum_{i=1}^n (\rho(\tilde{\lambda}' g_i(\beta_0)) - \rho_0) \\ &\stackrel{a}{=} n \tilde{\lambda}' \Omega \tilde{\lambda} \\ &\stackrel{a}{=} n \hat{g}' \Omega^{-1} \hat{g}. \end{aligned}$$

Stock and Wright (2000) suggested the CUE version of  $\widetilde{AR}$ . The quadratic form statistics in  $\tilde{\lambda}$  and  $\tilde{g}$  for CUE are given in Kleibergen (2005) for CUE, in which case they coincide, and for GEL by GS. GS also suggest the GEL criterion function statistic  $\mathcal{GELR}(\beta_0)$ . Caner (2003) also describes similar statistics based on ET.

Secondly, LM and score statistics that have appeared in the literature are adaptations of statistics suggested in Newey and West (1987). These statistics use a slightly different normalisation to that for  $\widetilde{LM}$  given above although their limiting distribution remains that of a chi-square random variable with one degree of freedom, viz.

$$\begin{aligned} \mathcal{LM}(\beta_0) &\equiv (n \tilde{G}' \Omega^{-1} \tilde{g})^2 / (n \tilde{G}' \Omega^{-1} \tilde{G}) \\ &\stackrel{a}{=} (n \tilde{G}' \tilde{\lambda})^2 / (n \tilde{G}' \Omega^{-1} \tilde{G}) \\ &\equiv \mathcal{S}(\beta_0). \end{aligned}$$

Kleibergen (2005) details the Lagrange multiplier  $\mathcal{LM}(\beta_0)$  statistic for CUE, which again coincides with the score statistic  $\mathcal{S}(\beta_0)$ . GS describe  $\mathcal{LM}(\beta_0)$  and  $\mathcal{S}(\beta_0)$  for GEL whereas Caner (2003) gives their ET counterparts. Otsu (2003) suggests an alternative statistic based on the GEL criterion which is related and asymptotically equivalent to  $\mathcal{LM}(\beta_0)$  and  $\mathcal{S}(\beta_0)$ . Let  $\tilde{G}_i(\beta_0) \equiv \pi_i(\beta_0) G_i(\beta_0)$  and define  $\tilde{\xi} \equiv \arg \max_{\xi \in \hat{\Xi}_n(\beta_0)} \sum_{i=1}^n \rho(\xi' \tilde{G}_i(\beta_0)' \Omega^{-1} g_i(\beta_0)) / n$  where  $\hat{\Xi}_n(\beta) \equiv \{\xi : \xi' \tilde{G}_i(\beta)' \Omega^{-1} g_i(\beta) \in \mathcal{V}, i = 1, \dots, n\}$ . The statistic is then given by  $2 \sum_{i=1}^n (\rho(\tilde{\xi}' \tilde{G}_i(\beta_0)' \Omega^{-1} g_i(\beta_0)) - \rho_0)$  but requires two maximizations, one for  $\tilde{\lambda}$  in  $\tilde{G}_i(\beta_0)$  and the other for  $\tilde{\xi}$ . As noted in Guggenberger and Smith (2005b), the latter maximization may be simply avoided by the substitution of either  $-(\tilde{G}' \Omega^{-1} \tilde{G})^{-1} \tilde{G}' \Omega^{-1} \hat{g}(\beta_0)$  or  $(\tilde{G}' \Omega^{-1} \tilde{G})^{-1} \tilde{G}' \tilde{\lambda}$  for  $\tilde{\xi}$ . The function  $\rho(v)$  used to obtain  $\tilde{\lambda}$  in  $\tilde{G}$  may be different from that defining the statistic as long as both satisfy Assumption 3.3.

Finally, numerous analogues of  $\widetilde{LR}$  may also be described. For example,

$$\begin{aligned}\mathcal{LR}(\beta_0) \equiv & \frac{1}{2} \left( \mathcal{GELR}(\beta_0) - n\tilde{G}'\Delta^{-1}\tilde{G} \right. \\ & \left. + \sqrt{(\mathcal{GELR}(\beta_0) - n\tilde{G}'\Delta^{-1}\tilde{G})^2 + 4\widetilde{LM}(n\tilde{G}'\Delta^{-1}\tilde{G})} \right),\end{aligned}$$

in which  $\widetilde{LR}$  has been replaced by the GEL criterion function statistic  $\mathcal{GELR}(\beta_0)$  (or  $n\tilde{\lambda}'\Omega\tilde{\lambda}$ ). Similarly,  $\widetilde{LM}$  might be replaced by  $\mathcal{LM}(\beta_0)$  or  $\mathcal{S}(\beta_0)$  with  $\Delta$  substituted by  $\Omega$ . A CUE version of  $\mathcal{LR}(\beta_0)$  was proposed in Kleibergen (2005, eq. (31), p.1113) in which  $\mathcal{LM}(\beta_0)$  and  $\Omega$  replace  $\widetilde{LM}$  and  $\Delta$ . The limiting distribution of  $\mathcal{LR}(\beta_0)$  with  $\mathcal{GELR}(\beta_0)$  or  $n\tilde{\lambda}'\Omega\tilde{\lambda}$  substituted for  $n\hat{g}'\Omega^{-1}\hat{g}$  remains that for  $\widetilde{LR}$  given in Theorem 3.2 above. If  $\widetilde{LM}$  is replaced by  $\mathcal{LM}(\beta_0)$  or  $\mathcal{S}(\beta_0)$  and  $\Delta$  by  $\Omega$ , then the limiting representation must be altered by substituting likewise  $Z_T'\Delta^{1/2}\Omega^{-1}\Delta^{1/2}Z_T$  for  $Z_T'Z_T$ . Note that if  $n\tilde{G}'\Delta^{-1}\tilde{G} \xrightarrow{p} \infty$ , corresponding to the strong identification of  $G$ , then  $\mathcal{LR}(\beta_0) - \mathcal{LM}(\beta_0) \xrightarrow{p} 0$ , confirming the asymptotic equivalence of these statistics in this circumstance, cf. AS, section 7.4.

The large sample representation for  $\widetilde{LR}$ ,  $\mathcal{LR}(\beta_0)$  or any of its analogues may easily be consistently estimated by simulation, thus enabling an asymptotically valid interval estimator for  $\beta_0$  to be obtained by inversion of the acceptance region of the LR-type test for  $H_0 : \beta = \beta_0$ . Given  $\tilde{G}$ , realisations of  $\mathcal{LR}(\beta_0)$  based on its limiting representation are given from simulation of independent chi-square random variables with one and  $m-1$  degrees of freedom respectively. An estimator of the asymptotic conditional distribution function of  $\mathcal{LR}(\beta_0)$  given  $Z_T$  may then be simply obtained. Cf. Kleibergen (2005, p.1114).

When  $p > 1$ , Kleibergen (2005, section 5.1, pp.1113-5) suggests replacing  $n\tilde{G}'\Omega^{-1}\tilde{G}$  in the CUE version of  $\mathcal{LR}(\beta_0)$ , after substitution of  $\Omega$  and  $\mathcal{LM}(\beta_0)$  for  $\Delta$  and  $\widetilde{LM}$  respectively, by a statistic which incorporates  $H_0 : \beta = \beta_0$  appropriate for testing  $rk[G] = p-1$  against  $rk[G] = p$  based on  $\tilde{G}$ . Examples of such statistics are given in Cragg and Donald (1996, 1997), Kleibergen and Paap (2005) and Robin and Smith (2000).

## 4 GMC and GEL

K provides an information-theoretic characterisation of GMC estimators which includes EL as a special case. Corcoran (1998) formulated a general class of MD estimators which are the empirical counterparts of those GMC estimators detailed in K. NS, see also Newey and Smith (2001), compared GEL with the MD type of estimator discussed by Corcoran (1998) which helps explain the form of the probabilities in equation (2.2) and connects their results with the existing literature.

Let  $h(\pi)$  be a convex function of a scalar  $\pi$  that measures the discrepancy between  $\pi$  and the empirical probability  $1/n$  of a single observation, that can depend on  $n$ . Consider the optimization problem

$$\min_{\pi_1, \dots, \pi_n, \beta} \sum_{i=1}^n h(\pi_i), \text{ s.t. } \sum_{i=1}^n \pi_i g_i(\beta) = 0, \sum_{i=1}^n \pi_i = 1. \quad (4.3)$$

and the resultant MD estimator is defined by

$$\hat{\beta}_{MD} = \arg \min_{\beta \in \mathcal{B}, \pi_1, \dots, \pi_n} \sum_{i=1}^n h(\pi_i), \text{ s.t. } \sum_{i=1}^n \pi_i g_i(\beta) = 0, \sum_{i=1}^n \pi_i = 1.$$

Like GEL, this class also includes as special cases EL, ET and CUE, where  $h(\pi)$  is  $-\ln(\pi)$ ,  $\pi \ln(\pi)$  and  $[(n\pi)^2 - 1]/2n$  respectively together with members of the Cressie-Read (1984) family of power divergence criteria discussed below. When the solutions  $\hat{\pi}_1, \dots, \hat{\pi}_n$  of this problem are nonnegative, they can be interpreted as empirical probabilities that minimize the discrepancy with the empirical measure subject to the moment conditions.

To relate MD and GEL estimators we compare their first-order conditions. For an  $m$ -vector of Lagrange multipliers  $\hat{\alpha}_{MD}$  associated with the first constraint and a scalar  $\hat{\mu}_{MD}$  for the second in (4.3), the MD first order conditions for  $\hat{\pi}_i$  are  $h_\pi(\hat{\pi}_i) = -\hat{\alpha}'_{MD} g_i(\hat{\beta}_{MD}) + \hat{\mu}_{MD}$  which, if  $h_\pi(\cdot)$  is one-to-one, may be solved for  $\hat{\pi}_i$ , ( $i = 1, \dots, n$ ). Substituting into the first-order conditions for  $\hat{\beta}_{MD}$ ,  $\hat{\alpha}_{MD}$  and  $\hat{\mu}_{MD}$  gives

$$\begin{aligned} \sum_{i=1}^n h_\pi^{-1}(-\hat{\alpha}'_{MD} g_i(\hat{\beta}_{MD}) + \hat{\mu}_{MD}) G_i(\hat{\beta}_{MD})' \hat{\alpha}_{MD} &= 0, \\ \sum_{i=1}^n h_\pi^{-1}(-\hat{\alpha}'_{MD} g_i(\hat{\beta}_{MD}) + \hat{\mu}_{MD}) g_i(\hat{\beta}_{MD}) &= 0, \end{aligned} \quad (4.4)$$

and  $\sum_{i=1}^n h_\pi^{-1}(-\hat{\alpha}'_{MD} g_i(\hat{\beta}_{MD}) + \hat{\mu}_{MD}) = 1$ . For comparison, the GEL first-order conditions are

$$\begin{aligned} \sum_{i=1}^n \rho_1(\hat{\lambda}' g_i(\hat{\beta})) G_i(\hat{\beta})' \hat{\lambda} &= 0, \\ \sum_{i=1}^n \rho_1(\hat{\lambda}' g_i(\hat{\beta})) g_i(\hat{\beta}) &= 0. \end{aligned} \quad (4.5)$$

In general, the first order conditions for GEL and MD are different, and hence so are the estimators of  $\beta$ . However, if  $h_\pi^{-1}(\cdot)$  is homogenous, the Lagrange multiplier  $\hat{\mu}_{MD}$  can be factored out of (4.4). Then the first-order conditions equations (4.4) and (4.5) coincide for

$$\hat{\lambda} = \hat{\alpha}_{MD} / \hat{\mu}_{MD}. \quad (4.6)$$

In this case the GEL saddle point problem is a dual of the MD one, in the sense that  $\hat{\lambda}$  is a ratio of Lagrange multipliers (4.6) from MD. If  $h_\pi^{-1}(\cdot)$  is not homogenous, MD and GEL estimators are different. In general, though, for large  $n$ , the GEL class is obtained from a much smaller dimensional optimization problem than MD and is consequently computationally less complex. Duality also justifies the GEL empirical probabilities  $\pi_i(\beta)$  (2.2) as MD estimates which may thus be used to efficiently estimate the distribution of the data by  $\hat{\mathcal{P}}\{z \leq c\} = \sum_{i=1}^n \pi_i(\hat{\beta}) 1(z_i \leq c)$ .

A particular example of the relationship between MD and GEL occurs when  $h(\cdot)$  is a member of the Cressie-Read (1984) power divergence criteria in which  $h(\pi) = [\gamma(\gamma+1)]^{-1}[(n\pi)^{\gamma+1} - 1]/n$ . We interpret expressions as limits for  $\gamma = 0$  or  $\gamma = -1$ . In this case  $h_\pi^{-1}(\cdot)$  is homogenous and, hence, for each MD estimator there is a dual GEL estimator in this case. The following is Theorem 2.2, p.224, in NS.<sup>6</sup>

**Theorem 4.1** *If  $g(z, \beta)$  is continuously differentiable in  $\beta$ , for some scalar  $\gamma$*

$$\rho(v) = -(1 + \gamma v)^{(\gamma+1)/\gamma} / (\gamma + 1), \quad (4.7)$$

---

<sup>6</sup>Duality between MD and GEL estimators occurs for EL when  $\gamma = -1$ , for ET when  $\gamma = 0$  and for CUE when  $\gamma = 1$  as well as for all the other members of the Cressie-Read (1984) family.

the solutions to equation (4.3) and (2.1) occur in the interior of  $\mathcal{B}$ ,  $\hat{\lambda}$  exists, and  $\sum_{i=1}^n \rho_2(\hat{\lambda}' \hat{g}_i) \hat{g}_i \hat{g}_i'$  is nonsingular, then the first order conditions for GEL and MD coincide for  $\hat{\beta} = \hat{\beta}_{MD}$ ,  $\pi_i(\hat{\beta}) = \hat{\pi}_i^{MD}$ , ( $i = 1, \dots, n$ ), and  $\hat{\lambda} = \hat{\alpha}_{MD}/(\gamma \hat{\mu}_{MD})$  for  $\gamma \neq 0$  and  $\hat{\lambda} = \hat{\alpha}_{MD}$  for  $\gamma = 0$ .

## 5 Asymptotic Bias

Schennach (2004) recently reconsidered EL and examined an alternative estimator, exponentially tilted empirical likelihood (ET\_EL), which embeds the ET implied probabilities in the EL criterion function. ET\_EL has been considered elsewhere by Jing and Wood (1996) and Corcoran (1998, section 4, pp.971-972). Although the ET\_EL criterion is not Bartlett correctable, see Jing and Wood (1996), Schennach (2004) proves that ET\_EL possesses the same higher order bias and variance properties as EL and, hence, like EL, is higher order efficient among bias-corrected estimators.

Rather than, as Schennach (2004) suggests, embedding the ET implied probabilities in the EL criterion, we substitute the GEL implied probabilities. Given a suitable choice for  $\rho(v)$ , even with unbounded  $g_i(\beta)$ , the implied probabilities  $\pi_i(\beta)$  (2.2) will always be positive, for example, members of the Cressie-Read family for which  $\gamma \leq 0$ .<sup>7</sup> As a consequence, the GEL\_EL criterion is defined as

$$\begin{aligned} \log L_{GEL\_EL}(\beta) &= \sum_{i=1}^n \log \pi_i(\beta) / n \\ &= \sum_{i=1}^n \log \rho_1(\hat{\lambda}(\beta)' g_i(\beta)) / n - \log \sum_{j=1}^n \rho_1(\hat{\lambda}(\beta)' g_j(\beta)). \end{aligned}$$

We show that GEL\_EL shares the same bias properties as EL.

The following assumption mirrors the hypotheses of Schennach (2004, Theorem 5).

**Assumption 5.1** (a)  $E[\sup_{\beta \in \mathcal{N}} \|g_i(\beta)\|^4] < \infty$  and  $E[\sup_{\beta \in \mathcal{N}} \|G_i(\beta)\|^2] < \infty$ ; (b) for each  $\beta \in \mathcal{N}$ ,  $\|\partial g(z, \beta) / \partial \beta' - \partial g(z, \beta_0) / \partial \beta'\| \leq b(z) \|\beta - \beta_0\|$  such that  $E[b(z)] < \infty$ ; (d)  $\rho(v)$  is four times continuously differentiable with Lipschitz fourth derivative in a neighborhood of zero.

The next theorem follows as a consequence.

**Theorem 5.1** Let Assumptions 2.1, 2.2 and 5.1 hold. Then

$$\hat{\beta}_{GEL\_EL} - \hat{\beta}_{EL} = O_p(n^{-3/2}).$$

An immediate consequence of Theorem 5.1 is that the GEL\_EL and EL estimators share the same asymptotic bias. Hence, we adopt Assumption 3 in NS; viz.

**Assumption 5.2** There is  $b(z)$  with  $E[b(z)^6] < \infty$  such that for  $0 \leq j \leq 4$  and all  $z$ ,  $\nabla^j g(z, \beta)$  exists on a neighborhood  $\mathcal{N}$  of  $\beta_0$ ,  $\sup_{\beta \in \mathcal{N}} \|\nabla^j g(z, \beta)\| \leq b(z)$ , and for each  $\beta \in \mathcal{N}$ ,  $\|\nabla^4 g(z, \beta) -$

---

<sup>7</sup>Negative  $\pi_i(\beta)$  may be avoided by incorporating a shrinkage factor; viz.

$$\pi_i^*(\beta) = \frac{1}{1 + \varepsilon_n} \pi_i(\beta) + \frac{\varepsilon_n}{1 + \varepsilon_n} \frac{1}{n},$$

where  $\varepsilon_n = \min[\min_{1 \leq i \leq n} \pi_i(\beta), 0]$ . See Bonnal and Renault (2004).

$\nabla^4 g(z, \beta_0) \leq b(z) \|\beta - \beta_0\|$ ,  $\rho(v)$  is four times continuously differentiable with Lipschitz fourth derivative in a neighborhood of zero.

Let  $H \equiv \Sigma G' \Omega^{-1}$  and  $a(\beta_0)$  be an  $m$ -vector such that

$$a_j(\beta_0) \equiv \text{tr}(\Sigma E[\partial^2 g_{ij}(\beta_0)/\partial \beta \partial \beta'])/2, (j = 1, \dots, m), \quad (5.8)$$

where  $g_{ij}(\beta)$  denotes the  $j$ th element of  $g_i(\beta)$ , and  $e_j$  the  $j$ th unit vector. Therefore,

**Theorem 5.2** *If Assumptions 2.1, 2.2 and 5.2 are satisfied, then to  $O(n^{-2})$*

$$\text{Bias}(\hat{\beta}_{\text{GEL\_EL}}) = H(-a(\beta_0) + E[G_i(\beta_0)H g_i(\beta_0)])/n.$$

Cf. NS, Theorem 4.1, p.228, and Corollary 4.3, p.229.

Given the  $O_p(n^{-3/2})$  equivalence between EL and GEL\_EL estimators, an open question remains concerning whether GEL\_EL is also higher order efficient, sharing the same higher order variance as EL and ET\_EL.

## 6 Misspecification

Schennach (2004, Theorem 1) also proves that, when the moment condition model is misspecified, EL is no longer root- $n$  consistent for its pseudo true value (PTV). The difficulty for EL under misspecification arises because its influence function is unbounded, a property shared by other members of the Cressie-Read family for which  $\gamma < 0$ , whereas that for ET is, see Imbens, Spady and Johnson (1998, p.337). Schennach (2004, Theorem 10) shows that ET\_EL is, however, root- $n$  consistent for its PTV which may be advantageous for the properties of ET *vis-à-vis* EL. We provide a similar result below for GEL\_EL estimators defined for GEL criteria with bounded influence functions under misspecification, that is, if there exists no  $\beta \in \mathcal{B}$  such that  $E[g(z, \beta)] = 0$ .

Following Schennach (2004), we reformulate the estimation problem as a just-identified GMM system to facilitate the derivation of the large sample properties of GEL\_EL by the introduction of the additional auxiliary scalar and  $m$ -vector parameters  $\rho_1$  and  $\mu$ . Computationally, of course, this reparameterisation is unnecessary to obtain the GEL\_EL estimators.

For brevity, we write  $g_i \equiv g_i(\beta)$ ,  $G_i \equiv G_i(\beta)$ ,  $\rho_{1i} \equiv \rho_{1i}(\lambda' g_i)$  and  $\rho_{2i} \equiv \rho_{2i}(\lambda' g_i)$ , ( $i = 1, \dots, n$ ). The GEL\_EL and auxiliary parameter estimators may then be obtained *via* the following lemma; cf. Schennach (2004, Lemma 9).

**Lemma 6.1** *The GEL\_EL and auxiliary parameter estimators  $\hat{\beta}_{\text{GEL\_EL}}$  and  $\hat{\lambda}_{\text{GEL\_EL}}$  are given as appropriate subvectors of  $\hat{\theta} = (\hat{\rho}_1, \hat{\mu}', \hat{\lambda}, \hat{\beta})'$  which is the solution to*

$$\sum_{i=1}^n \psi(z_i, \hat{\theta})/n = 0,$$

where

$$\psi(z_i, \theta) \equiv \begin{pmatrix} \rho_{1i} - \rho_1 \\ \rho_{1i}g_i \\ (\rho_{2i}g_i g'_i)\mu + ((\rho_1\rho_{2i}/\rho_{1i}) - \rho_{2i})g_i \\ -\rho_{1i}G'_i\mu - \rho_{2i}G'_i\lambda g'_i\mu + ((\rho_1\rho_{2i}/\rho_{1i}) - \rho_{2i})G'_i\lambda \end{pmatrix}.$$

Likewise, see Schennach (2004, equation (30)), the structure of the moment indicator vector  $\psi(z_i, \theta)$  becomes more transparent when re-expressed as

$$\psi(z_i, \theta) = \begin{pmatrix} \rho_{1i} - \rho_1 \\ \partial(\rho_{1i}g'_i\mu + \rho_1 \log(\rho_{1i} \exp(-\rho_{1i}))) / \partial \kappa \\ \partial(\rho_{1i}g'_i\mu + \rho_1 \log(\rho_{1i} \exp(-\rho_{1i}))) / \partial \lambda \\ \partial(\rho_{1i}g'_i\mu + \rho_1 \log(\rho_{1i} \exp(-\rho_{1i}))) / \partial \beta \end{pmatrix}.$$

Let  $\lambda(\beta)$  be the unique solution of  $E[\rho_1(\lambda'g(z, \beta))g(z, \beta)] = 0$  which exists by the concavity of  $\rho(\cdot)$ . We also define  $\mathcal{N}$  as a neighbourhood of the PTV  $\beta_*$ , see Assumption 6.1 (b) below, and  $P(\beta) \equiv E[\rho(\lambda(\beta)'(g(z, \beta) - E[g(z, \beta)]))]$ . The following assumption adapts Schennach (2004, Assumption 3) for GEL-EL. Let  $g^j(z, \beta)$  be the  $j$ th element of  $g(z, \beta)$ , ( $j = 1, \dots, m$ ).

**Assumption 6.1** (a)  $\mathcal{B}$  is compact; (b)  $P(\beta)$  is minimised at the unique PTV  $\beta_* \in \text{int}(\mathcal{B})$ ; (c)  $g(z, \beta)$  is continuous at each  $\beta \in \mathcal{B}$  with probability one; (d)  $\Lambda(\beta)$  is a compact set such that  $\lambda(\beta) \in \Lambda(\beta)$  and  $E[\sup_{\beta \in \mathcal{B}} \sup_{\lambda \in \Lambda(\beta)} |\rho(\lambda'g(z, \beta))|] < \infty$ ; (e)  $g(z, \beta)$  is twice continuously differentiable on  $\mathcal{N}$ ; (f) there is  $b(z)$  with  $E[\sup_{\beta \in \mathcal{N}} \sup_{\lambda \in \Lambda(\beta)} \prod_{j=1}^3 |\rho_j(\lambda'g(z, \beta))|^{k_j} b(z)^k] < \infty$  for  $0 \leq k \leq 4$  such that  $\|g(z, \beta)\|$ ,  $\|\partial g(z, \beta)/\partial \beta'\|$ ,  $\|\partial^2 g^j(z, \beta)/\partial \beta \partial \beta'\| < b(z)$ , ( $j = 1, \dots, m$ ), and  $k_3 = 0$  unless  $-1 \leq k_1 \leq 0$ , when  $k_2 = 0$ ,  $k_3 = 1$ , if  $-2 \leq k_1 \leq -1$ ,  $k_2 = 2$  and if  $0 \leq k_1, k_2 \leq 2$ ,  $1 \leq k_1 + k_2 \leq 2$ .

The just-identified GMM system based on the moment indicator vector  $\psi(z, \theta)$  allows the use of standard results on the large sample behaviour of GMM estimators to be employed, e.g. Newey and McFadden (1994, Theorem 3.4, p.2148).

**Theorem 6.1** Let  $G_\theta^* = E[\partial \psi(z, \theta_*)/\partial \theta']$  and  $\Omega_\theta^* = E[\psi(z, \theta_*)\psi(z, \theta_*)']$ . If Assumption 6.1 satisfied and  $G_\theta^*$  is nonsingular, then  $n^{1/2}(\hat{\theta} - \theta_*) \xrightarrow{d} N(0, (G_\theta^*)^{-1}\Omega_\theta^*(G_\theta^*)'^{-1})$ .

## Appendix: Proofs

**Proof of Theorem 3.2:** Given Theorem 3.1, it is straightforward to see that, under  $H_0 : \beta = \beta_0$ ,  $\widetilde{AR}$  converges in distribution to  $Z'_S Z_S$  and, thus, has a limiting  $\chi^2(m)$  distribution. Similarly,  $n\tilde{G}'\Delta^{-1/2}\Omega^{-1/2}\tilde{g}$  converges in distribution to  $Z'_T Z_S$ . Given  $Z_T$ ,  $Z'_T Z_S \sim N(0, Z'_T Z_T)$  and, thus,  $Z'_T Z_S/(Z'_T Z_T)^{1/2}$  is standard normally distributed and independent of  $Z'_T Z_T$ . Therefore, as  $n\tilde{G}'\Delta^{-1}\tilde{G} \xrightarrow{d} Z'_T Z_T$ ,  $\widetilde{LR}$  has a limiting  $\chi^2(1)$  distribution under  $H_0 : \beta = \beta_0$ . Finally,  $\widetilde{LR}$  converges in distribution to  $(Z'_S Z_S - Z'_T Z_T + \sqrt{(Z'_S Z_S - Z'_T Z_T)^2 + 4(Z'_T Z_S)^2})/2$ . Write  $Z'_S Z_S = (Z'_T Z_S)^2/Z'_T Z_T + (Z'_S Z_S - (Z'_T Z_S)^2/Z'_T Z_T)$  which are independent  $\chi^2(1)$  and  $\chi^2(m-1)$  random variates respectively independent

of  $Z_T'Z_T$ . Therefore, conditionally on  $Z_T$ ,  $\widetilde{LR}$  has a limiting distribution described by

$$\frac{1}{2} \left( \chi^2(1) + \chi^2(m-1) - Z_T'Z_T + \sqrt{(\chi^2(1) + \chi^2(m-1) - Z_T'Z_T)^2 + 4\chi^2(1)(Z_T'Z_T)} \right).$$

■

**Proof of Theorem 5.1:** Let  $\hat{g}_i = g_i(\hat{\beta})$ ,  $\hat{G}_i = G_i(\hat{\beta})$ ,  $\hat{\rho}_{1i} = \rho_1(\hat{\lambda}'\hat{g}_i)$  and  $\hat{\rho}_{2i} = \rho_2(\hat{\lambda}'\hat{g}_i)$ ,  $(i = 1, \dots, n)$ . The first order conditions defining the GEL-EL estimator are

$$\begin{aligned} 0 &= \frac{\partial \hat{\lambda}(\hat{\beta})'}{\partial \beta} \sum_{i=1}^n \left( \frac{1}{n\hat{\rho}_{1i}} - \frac{1}{\sum_{j=1}^n \hat{\rho}_{1j}} \right) \hat{\rho}_{2i}\hat{g}_i \\ &\quad + \sum_{i=1}^n \left( \frac{1}{n\hat{\rho}_{1i}} - \frac{1}{\sum_{j=1}^n \hat{\rho}_{1j}} \right) \hat{\rho}_{2i}\hat{G}_i'\hat{\lambda}, \end{aligned} \quad (\text{A.1})$$

cf. Schennach (2004, eq. (52)). As  $\rho(v)$  is continuously differentiable to the fourth order, a similar argument to Schennach (2004, eq. (55)) yields

$$\begin{aligned} \hat{\rho}_{ji} &= -1 + \rho_{j+1}\hat{\lambda}'\hat{g}_i + \frac{1}{2}\rho_{j+2}(\hat{\lambda}'\hat{g}_i)^2 + O_p(n^{-3/2}) \|\hat{g}_i\|^3, \\ \sum_{k=1}^n \hat{\rho}_{jk}/n &= -1 + O_p(n^{-1}), (j = 1, 2). \end{aligned}$$

Hence, noting  $\hat{g} = n^{-1} \sum_{i=1}^n \hat{g}_i = O_p(n^{-1/2})$  and  $\hat{\lambda} = O_p(n^{-1/2})$ ,

$$\begin{aligned} \hat{\rho}_{2i} - \hat{\rho}_{1i} &= (\rho_3 + 1)\hat{\lambda}'\hat{g}_i + \frac{1}{2}(\rho_4 - \rho_3)(\hat{\lambda}'\hat{g}_i)^2 + O_p(n^{-3/2}) \|\hat{g}_i\|^3, \\ \frac{\hat{\rho}_{2i} - \hat{\rho}_{1i}}{\hat{\rho}_{1i}} &= -(\rho_3 + 1)\hat{\lambda}'\hat{g}_i - \frac{1}{2}(\rho_4 - \rho_3)(\hat{\lambda}'\hat{g}_i)^2 + (\rho_3 + 1)(\hat{\lambda}'\hat{g}_i)^2 \\ &\quad + O_p(n^{-3/2}) \|\hat{g}_i\|^3, \\ \frac{n(\hat{\rho}_{2i} - \hat{\rho}_{1i})}{\sum_{k=1}^n \hat{\rho}_{1k}} &= -(\rho_3 + 1)\hat{\lambda}'\hat{g}_i - \frac{1}{2}(\hat{\rho}_4 - \hat{\rho}_3)(\hat{\lambda}'\hat{g}_i)^2 \\ &\quad + O_p(n^{-3/2})(\|\hat{g}_i\| + \|\hat{g}_i\|^2 + \|\hat{g}_i\|^3). \end{aligned}$$

Therefore, after cancellation, substitution of these expressions in (A.1) yields

$$\begin{aligned} \sum_{i=1}^n \left( \frac{1}{n\hat{\rho}_{1i}} - \frac{1}{\sum_{j=1}^n \hat{\rho}_{1j}} \right) \rho_{2i}g_i &= \frac{1}{n} \sum_{i=1}^n (1 - n\pi_i)g_i + (\rho_3 + 1) \frac{1}{n} \sum_{i=1}^n (\lambda'g_i)^2 g_i \\ &\quad + O_p(n^{-3/2}) \frac{1}{n} \sum_{i=1}^n \|\hat{g}_i\|^3 \hat{g}_i \\ &= \hat{g} + (\rho_3 + 1) \frac{1}{n} \sum_{i=1}^n (\hat{\lambda}'\hat{g}_i)^2 \hat{g}_i + O_p(n^{-3/2}), \end{aligned} \quad (\text{A.2})$$

as  $\sum_{i=1}^n \hat{\pi}_i \hat{g}_i = 0$ , where  $\hat{\pi}_i = \pi_i(\hat{\beta})$ ,  $(i = 1, \dots, n)$ . Furthermore,

$$\begin{aligned} \sum_{i=1}^n \left( \frac{1}{n\hat{\rho}_{1i}} - \frac{1}{\sum_{j=1}^n \hat{\rho}_{1j}} \right) \hat{\rho}_{2i}\hat{G}_i'\hat{\lambda} &= \frac{1}{n} \sum_{i=1}^n (1 - n\hat{\pi}_i + O_p(n^{-1}) \|\hat{g}_i\|^2) \hat{G}_i'\hat{\lambda} \\ &= \frac{1}{n} \sum_{i=1}^n (1 - n\hat{\pi}_i) \hat{G}_i'\hat{\lambda} + O_p(n^{-3/2}) \\ &= -\frac{1}{n} \sum_{i=1}^n (\hat{\lambda}'\hat{g}_i) \hat{G}_i'\hat{\lambda} + O_p(n^{-3/2}), \end{aligned} \quad (\text{A.3})$$



as  $n\hat{\pi}_i = 1 + \hat{\lambda}'\hat{g}_i + O_p(n^{-1}) \|\hat{g}_i\|^2$ , ( $i = 1, \dots, n$ ); cf. Schennach (2004, eq. (57)).

Now, similarly to Schennach (2004, eq. (56)), from the first order conditions determining  $\hat{\lambda}(\beta)$ , substituting for  $n\hat{\pi}_i$ , ( $i = 1, \dots, n$ ),

$$\begin{aligned} \sum_{i=1}^n \hat{\pi}_i \hat{g}_i &= \hat{g} + \frac{1}{n} \sum_{i=1}^n \hat{g}_i \hat{g}_i' \hat{\lambda} + O_p(n^{-1}) \frac{1}{n} \sum_{i=1}^n \|\hat{g}_i\|^2 \hat{g}_i \\ &= \hat{g} + \sum_{i=1}^n \hat{\pi}_i \hat{g}_i \hat{g}_i' \hat{\lambda} + \frac{1}{n} \sum_{i=1}^n (1 - n\hat{\pi}_i) \hat{g}_i \hat{g}_i' \hat{\lambda} + O_p(n^{-1}) \\ &= \hat{g} + \tilde{\Omega} \hat{\lambda} + O_p(n^{-1}), \end{aligned}$$

where  $\tilde{\Omega} = \sum_{i=1}^n \hat{\pi}_i \hat{g}_i \hat{g}_i'$ . Hence, as  $\tilde{\Omega} = O_p(1)$  and is p.d. w.p.a.1,

$$\hat{\lambda} = -\tilde{\Omega}^{-1} \hat{g} + O_p(n^{-1}). \quad (\text{A.4})$$

Therefore, from (A.3),

$$\sum_{i=1}^n \left( \frac{1}{n\hat{\rho}_{1i}} - \frac{1}{\sum_{j=1}^n \hat{\rho}_{1j}} \right) \hat{\rho}_{2i} \hat{G}_i' \hat{\lambda} = \hat{g}' \tilde{\Omega}^{-1} \frac{1}{n} \sum_{i=1}^n \hat{g}_i \hat{G}_i' \hat{\lambda} + O_p(n^{-3/2}), \quad (\text{A.5})$$

as  $n^{-1} \sum_{i=1}^n \hat{g}_i \hat{G}_i' \hat{\lambda} = O_p(n^{-1/2})$ .

The total differential of the first order conditions determining  $\hat{\lambda}(\beta)$  at  $\hat{\beta}$  is

$$\sum_{i=1}^n \hat{\rho}_{2i} \hat{g}_i \hat{g}_i' d\hat{\lambda} + \sum_{i=1}^n [\hat{\rho}_{1i} I_m + \hat{\rho}_{2i} \hat{g}_i \hat{\lambda}'] \hat{G}_i d\beta = 0.$$

Now,

$$\begin{aligned} \sum_{i=1}^n \left[ \frac{\hat{\rho}_{1i}}{\sum_{j=1}^n \hat{\rho}_{1j}} I + \frac{\hat{\rho}_{2i}}{\sum_{j=1}^n \hat{\rho}_{1j}} \hat{g}_i \hat{\lambda}' \right] \hat{G}_i &= \sum_{i=1}^n \hat{\pi}_i \hat{G}_i + \sum_{i=1}^n \hat{\pi}_i \hat{g}_i \hat{\lambda}' \hat{G}_i \\ &\quad - \frac{1}{n} \sum_{i=1}^n [(\rho_3 + 1) \hat{\lambda}' \hat{g}_i + O_p(n^{-1}) \|\hat{g}_i\|^2] \hat{g}_i \hat{\lambda}' \hat{G}_i \\ &= \tilde{G} + \sum_{i=1}^n \hat{\pi}_i \hat{g}_i \hat{\lambda}' \hat{G}_i - (\rho_3 + 1) \frac{1}{n} \sum_{i=1}^n \hat{g}_i \hat{g}_i' \hat{\lambda} \hat{\lambda}' \hat{G}_i + O_p(n^{-3/2}) \\ &= \tilde{G} + \sum_{i=1}^n \hat{\pi}_i \hat{g}_i \hat{\lambda}' \hat{G}_i + O_p(n^{-1}), \end{aligned}$$

where  $\tilde{G} = \sum_{i=1}^n \hat{\pi}_i \hat{G}_i$ . Therefore,

$$\begin{aligned} \frac{\partial \hat{\lambda}(\hat{\beta})}{\partial \beta'} &= -\tilde{\Omega}^{-1} \sum_{i=1}^n \left[ \frac{\hat{\rho}_{1i}}{\sum_{j=1}^n \hat{\rho}_{1j}} I_m + \frac{\hat{\rho}_{2i}}{\sum_{j=1}^n \hat{\rho}_{1j}} \hat{g}_i \hat{\lambda}' \right] \hat{G}_i \\ &\quad - \tilde{\Omega}^{-1} \sum_{i=1}^n \frac{\hat{\rho}_{2i} - \hat{\rho}_{1i}}{\sum_{k=1}^n \hat{\rho}_{1k}} \hat{g}_i \hat{g}_i' \frac{\partial \hat{\lambda}(\hat{\beta})}{\partial \beta'} \\ &= -\tilde{\Omega}^{-1} \tilde{G} - \tilde{\Omega}^{-1} \sum_{i=1}^n \hat{\pi}_i \hat{g}_i \hat{\lambda}' \hat{G}_i + O_p(n^{-1}) \\ &\quad + (\rho_3 + 1) \tilde{\Omega}^{-1} \frac{1}{n} \sum_{i=1}^n (\hat{\lambda}' \hat{g}_i + O_p(n^{-1}) \|\hat{g}_i\|^2) \hat{g}_i \hat{g}_i' \frac{\partial \hat{\lambda}(\hat{\beta})}{\partial \beta'} \end{aligned} \quad (\text{A.6})$$

$$\begin{aligned}
&= -\tilde{\Omega}^{-1}\tilde{G} - \tilde{\Omega}^{-1} \sum_{i=1}^n \hat{\pi}_i \hat{g}_i \hat{\lambda}' \hat{G}_i + O_p(n^{-1}) \\
&\quad - (\rho_3 + 1) \tilde{\Omega}^{-1} \frac{1}{n} \sum_{i=1}^n [(\hat{\lambda}' \hat{g}_i) \hat{g}_i \hat{g}_i' + O_p(n^{-1}) \|\hat{g}_i\|^4] \tilde{\Omega}^{-1} [\tilde{G} + O_p(n^{-1/2})] \\
&= -\tilde{\Omega}^{-1}\tilde{G} - \tilde{\Omega}^{-1} \sum_{i=1}^n \hat{\pi}_i \hat{g}_i \hat{\lambda}' \hat{G}_i \\
&\quad - (\rho_3 + 1) \tilde{\Omega}^{-1} \frac{1}{n} \sum_{i=1}^n (\hat{\lambda}' \hat{g}_i) \hat{g}_i \hat{g}_i' \tilde{\Omega}^{-1} \tilde{G} + O_p(n^{-1});
\end{aligned}$$

cf. Schennach (2004, eq. (54)).

Therefore, after substitution of eqs. (A.2) and (A.6) into the first term of the first order conditions eq. (A.1) defining the GEL\_EL estimator,

$$\begin{aligned}
\frac{\partial \hat{\lambda}(\hat{\beta})'}{\partial \beta} \sum_{i=1}^n \left( \frac{1}{n \hat{\rho}_{1i}} - \frac{1}{\sum_{j=1}^n \hat{\rho}_{1j}} \right) \hat{\rho}_{2i} \hat{g}_i &= -\tilde{G}' \tilde{\Omega}^{-1} \hat{g} - (\rho_3 + 1) \tilde{G}' \tilde{\Omega}^{-1} \frac{1}{n} \sum_{i=1}^n \hat{g}_i (\hat{\lambda}' \hat{g}_i)^2 \\
&\quad - \sum_{i=1}^n \hat{\pi}_i \hat{G}_i' \hat{\lambda} \hat{g}_i' \tilde{\Omega}^{-1} \hat{g} \\
&\quad - (\rho_3 + 1) \tilde{G}' \tilde{\Omega}^{-1} \frac{1}{n} \sum_{i=1}^n \hat{g}_i (\hat{\lambda}' \hat{g}_i) \hat{g}_i' \tilde{\Omega}^{-1} \hat{g} + O_p(n^{-3/2}) \\
&= -\tilde{G}' \tilde{\Omega}^{-1} \hat{g} - \sum_{i=1}^n \hat{\pi}_i \hat{G}_i' \hat{\lambda} \hat{g}_i' \tilde{\Omega}^{-1} \hat{g} + O_p(n^{-3/2}),
\end{aligned} \tag{A.7}$$

where the second equality follows from eq. (A.4). Combining eqs. (A.5) and (A.7) in eq. (A.1) yields

$$0 = -\tilde{G}' \tilde{\Omega}^{-1} \hat{g} + O_p(n^{-3/2}).$$

As  $n(\hat{\pi}_i - \hat{\pi}_i^{EL}) = O_p(n^{-1}) \|\hat{g}_i\|^2$ , ( $i = 1, \dots, n$ ),  $\tilde{G} = \tilde{G}_{EL} + O_p(n^{-1})$  and  $\tilde{\Omega} = \tilde{\Omega}_{EL} + O_p(n^{-1})$ , where  $\tilde{G}_{EL} = \sum_{i=1}^n \hat{\pi}_i^{EL} \hat{G}_i$ ,  $\tilde{\Omega}_{EL} = \sum_{i=1}^n \hat{\pi}_i^{EL} \hat{g}_i \hat{g}_i'$  and  $\hat{\pi}_i^{EL}$ , ( $i = 1, \dots, n$ ), denote the EL probabilities evaluated at the GEL\_EL estimator  $\hat{\beta}$ . Therefore, as  $\hat{g} = O_p(n^{-1/2})$ , the GEL\_EL estimator  $\hat{\beta}$  satisfies the same first order conditions as the EL estimator (to  $O_p(n^{-3/2})$ ),  $\tilde{G}_{EL}' \tilde{\Omega}_{EL}^{-1} \hat{g} = O_p(n^{-3/2})$ , and so

$$\hat{\beta} - \hat{\beta}_{EL} = O_p(n^{-3/2}).$$

■

**Proof of Lemma 6.1:** Let  $\hat{\rho}_{1i}(\beta) = \rho_1(\hat{\lambda}(\beta) g_i(\beta))$  and  $\hat{\rho}_{2i}(\beta) = \rho_2(\hat{\lambda}(\beta) g_i(\beta))$ , ( $i = 1, \dots, n$ ). The total differential of the first order conditions determining  $\hat{\lambda}(\beta)$  at  $\hat{\beta}$  is given by

$$\sum_{i=1}^n \hat{\rho}_{2i}(\beta) g_i(\beta) g_i(\beta)' d\hat{\lambda}(\beta) + \sum_{i=1}^n [\hat{\rho}_{1i}(\beta) I_m + \hat{\rho}_{2i}(\beta) g_i(\beta) \hat{\lambda}(\beta)'] G_i(\beta) d\beta = 0,$$

from which the derivative matrix  $\partial \hat{\lambda}(\beta) / \partial \beta'$  may be derived. Therefore, the first order conditions defining the GEL\_EL and auxiliary parameter estimators are

$$0 = - \sum_{i=1}^n G_i' [\rho_{1i} I_m + \rho_{2i} \lambda g_i'] \left( \sum_{i=1}^n \rho_{2i} g_i g_i' \right)^{-1} \sum_{i=1}^n \left( \frac{1}{n \rho_{1i}} - \frac{1}{\sum_{j=1}^n \rho_{1j}} \right) \rho_{2i} g_i$$

$$\begin{aligned}
& + \sum_{i=1}^n \left( \frac{1}{n\rho_{1i}} - \frac{1}{\sum_{j=1}^n \rho_{1j}} \right) \rho_{2i} G'_i \lambda \\
0 &= \sum_{i=1}^n \rho_{1i} g_i.
\end{aligned}$$

Write the additional scalar and  $m$ -vector parameter estimators  $\rho_1$  and  $\mu$  as

$$\rho_1 \equiv \sum_{i=1}^n \rho_{1i}/n, \mu \equiv - \left( \sum_{i=1}^n \rho_{2i} g_i g'_i / n \right)^{-1} \sum_{i=1}^n ((\rho_1 \rho_{2i} / \rho_{1i}) - \rho_{2i}) g_i / n.$$

The first order conditions determining  $\rho_1$ ,  $\mu$ ,  $\lambda$  and  $\beta$  then become

$$\begin{aligned}
0 &= \sum_{i=1}^n \begin{pmatrix} \rho_{1i} - \rho_1 \\ \rho_{1i} g_i \\ (\rho_{2i} g_i g'_i) \mu + ((\rho_1 \rho_{2i} / \rho_{1i}) - \rho_{2i}) g_i \\ \rho_{1i} G'_i \mu + \rho_{2i} G'_i \lambda g'_i \mu - ((\rho_1 \rho_{2i} / \rho_{1i}) - \rho_{2i}) G'_i \lambda \end{pmatrix} \\
&= \sum_{i=1}^n \begin{pmatrix} \rho_{1i} - \rho_1 \\ \partial(\rho_{1i} g'_i \mu - \rho_1 \log(\rho_{1i} \exp(-\rho_{1i}))) / \partial \kappa \\ \partial(\rho_{1i} g'_i \mu - \rho_1 \log(\rho_{1i} \exp(-\rho_{1i}))) / \partial \lambda \\ \partial(\rho_{1i} g'_i \mu - \rho_1 \log(\rho_{1i} \exp(-\rho_{1i}))) / \partial \beta \end{pmatrix}.
\end{aligned}$$

■

**Proof of Theorem 6.1:** The proof is an adaptation for GEL\_EL of that of Schennach (2004, Theorem 10). We first demonstrate consistency of  $\hat{\beta}$  for  $\beta_*$  and  $\hat{\lambda}(\hat{\beta})$  for  $\lambda_* \equiv \lambda_*(\beta_*)$ .

Now, from Assumptions 6.1 (a) and (b), by Newey and McFadden (1994, Lemma 2.4, p.2129),  $\hat{P}(\beta, \lambda) \equiv \sum_{i=1}^n \rho(\lambda'(g_i(\beta) - E[g(z, \beta)])) \xrightarrow{P} P(\beta, \lambda) \equiv E[\rho(\lambda'(g(z, \beta) - E[g(z, \beta)]))]$  uniformly  $(\beta, \lambda) \in \mathcal{B} \times \Lambda(\beta)$ . By a similar argument to that in Schennach (2004, Proof of Theorem 10),  $\sup_{\beta \in \mathcal{B}} \|\bar{\lambda}(\beta) - \lambda(\beta)\| \rightarrow 0$ . Likewise  $\sup_{\beta \in \mathcal{B}} \|\hat{\lambda}(\beta) - \lambda(\beta)\| \xrightarrow{P} 0$ .

Next,  $\sup_{\beta \in \mathcal{B}} \|\hat{P}(\beta, \hat{\lambda}(\beta)) - P(\beta, \lambda(\beta))\| \rightarrow 0$  by Assumptions 6.1 (c) and (d) and the concavity of  $\rho(\cdot)$ . As  $P(\beta, \lambda(\beta))$  is uniquely minimised at  $\beta_*$  from Assumption 6.1 (b),  $\hat{\beta} \xrightarrow{P} \beta_*$  and, thus,  $\hat{\lambda}(\hat{\beta}) \xrightarrow{P} \lambda(\beta_*)$ .

Finally,  $\hat{\rho}_1$  and  $\hat{\mu}$  are explicit continuous functions of  $\hat{\lambda}$  and  $\hat{\beta}$ . Let  $g_* \equiv g(z, \beta_*)$ . Hence,

$$\begin{aligned}
\hat{\rho}_1 &\xrightarrow{P} \rho_{1*} \equiv E[\rho_1(\lambda'_* g_*)], \\
\hat{\mu} &\xrightarrow{P} \mu_* \equiv -(E[\rho_2(\lambda'_* g_*) g_* g_*'])^{-1} \\
&\quad \times E[(\rho_{1*} \rho_2(\lambda'_* g_*) / \rho_1(\lambda'_* g_*)) - \rho_2(\lambda'_* g_*) g_*],
\end{aligned}$$

noting that  $E[\rho_2(\lambda'_* g_*) g_* g_*']$  is nonsingular follows as  $\Psi_*$  is n.s..

To show the asymptotic normality of the GEL\_EL and auxiliary parameter estimators, from Newey and McFadden (1994, Theorem 3.4, p.2148), we need to establish (a)  $E[\sup_{\theta \in \Theta} \|\partial \psi(z, \theta) / \partial \theta'\|] < \infty$  and (b)  $E[\psi(z, \theta_*) \psi(z, \theta_*)']$  exists. First, for (a), the normed derivative matrix  $\|\partial \psi(z, \theta) / \partial \theta'\|$  (apart from multiplicative factors that are bounded) consists of terms like

$$\prod_{j=1}^3 |\rho_j(\lambda' g(z, \beta))|^{k_j} \|g(z, \beta)\|^{k_{g0}} \|\partial g(z, \beta) / \partial \beta'\|^{k_{g1}} \|\partial^2 g^l(z, \beta) / \partial \beta \partial \beta'\|^{k_{g2}}, \quad (l = 1, \dots, m),$$

where  $1 \leq k_{g0} + k_{g1} + k_{g2} \leq 3$ , which is bounded by  $\prod_{j=1}^3 |\rho_j(\lambda'g(z, \beta))|^{k_j} b(z)^k$ ,  $1 \leq k \leq 3$ . The indices  $k_j$ , ( $j = 1, 2, 3$ ), obey (a)  $k_3 = 0$  unless  $-1 \leq k_1 \leq 0$  when  $k_2 = 0$ ,  $k_3 = 1$ , (b) if  $-2 \leq k_1 \leq -1$ ,  $k_1 + k_2 = 0$ , (c) if  $0 \leq k_1 \leq 1$ ,  $k_1 + k_2 = 1$  where  $k_1, k_2 \geq 0$ . Therefore, for some positive constant  $C$ ,

$$\begin{aligned} E[\sup_{\theta \in \Theta} \|\partial\psi(z, \theta)/\partial\theta'\|] &\leq CE[\sup_{\theta \in \Theta} \prod_{j=1}^3 |\rho_j(\lambda'g(z, \beta))|^{k_j} b(z)^k] \\ &= CE[\sup_{\beta \in \mathcal{B}} \sup_{\lambda \in \Lambda(\beta)} \prod_{j=1}^3 |\rho_j(\lambda'g(z, \beta))|^{k_j} b(z)^k] < \infty. \end{aligned}$$

Likewise, the normed matrix  $\|\psi(z, \theta)\psi(z, \theta)'\|$  has terms of the form

$$\prod_{j=1}^2 |\rho_j(\lambda'g(z, \beta))|^{k_j} \|g(z, \beta)\|^{k_{g0}} \|\partial g(z, \beta)/\partial\beta'\|^{k_{g1}} \|\partial^2 g^l(z, \beta)/\partial\beta\partial\beta'\|^{k_{g2}}, (l = 1, \dots, m),$$

where  $0 \leq k_{g0} + k_{g1} + k_{g2} \leq 4$ , which is bounded by  $\prod_{j=1}^2 |\rho_j(\lambda'g(z, \beta))|^{k_j} b(z)^k$ ,  $0 \leq k \leq 4$ . The indices  $k_j$ , ( $j = 1, 2$ ), obey (a) if  $-2 \leq k_1 \leq -1$ ,  $k_2 = 2$ , (b) if  $k_1 = 0$ ,  $1 \leq k_2 \leq 2$ , (c) if  $1 \leq k_1 \leq 2$ ,  $k_1 + k_2 = 2$  where  $k_2 \geq 0$ . That  $E[\psi(z, \theta_*)\psi(z, \theta_*)']$  exists then follows as above. ■

## References

- Andrews, D.W.K., Moreira, M.J., and J.H. Stock (2004): "Optimal Invariant Similar Tests for Instrumental Variables Regression." Cowles Foundation Discussion Paper No. 1476.
- Andrews, D.W.K., and J.H. Stock (2005): "Inference with Weak Instruments." Invited paper presented at the Econometric Society World Congress, U.C.L., London.
- Back, K., and D. Brown (1993): "Implied Probabilities in GMM Estimators", *Econometrica*, 61, 971-976.
- Bonnal, H., and E. Renault (2003): "On the Efficient Use of the Informational Content of Estimating Equations: Implied Probabilities and Maximum Euclidean Likelihood", working paper, Département de Sciences Économiques, Université de Montréal.
- Borwein, J.M., and A.S. Lewis (1991): "Duality Relations for Entropy-Like Minimisation Problems", *SIAM Journal of Control and Optimisation*, 29, 325-338.
- Brown, B.W. and W.K. Newey (1998): "Efficient Semiparametric Estimation of Expectations," *Econometrica*, 66, 453-464.
- Caner, M. (2003): "Exponential Tilting with Weak Instruments: Estimation and Testing", working paper, University of Pittsburg.
- Corcoran, S.A. (1998): "Bartlett Adjustment of Empirical Discrepancy Statistics", *Biometrika*, 85, 967-972.

- Cragg, J.G., and S.G. Donald (1996): “On the Asymptotic Properties of LDU-based Tests of the Rank of a Matrix”, *Journal of the American Statistical Association*, 91, 1301-1309.
- Cragg, J.G., and S.G. Donald (1997): “Inferring the Rank of a Matrix”, *Journal of Econometrics*, 76, 223-250.
- Cressie, N., and T. Read (1984): “Multinomial Goodness-of-Fit Tests”, *Journal of the Royal Statistical Society Series B*, 46, 440-464.
- Guggenberger, P., and R.J. Smith (2005a): “Generalized Empirical Likelihood Estimators and Tests under Partial, Weak and Strong Identification”, *Econometric Theory*, 21, 667-709.
- Guggenberger, P., and R.J. Smith (2005b): “Generalized Empirical Likelihood Tests in Time Series Models With Potential Identification Failure”, working paper, U.C.L.A..
- Hansen, L. P. (1982): “Large Sample Properties of Generalized Method of Moments Estimators”, *Econometrica*, 50, 1029-1054.
- 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.
- 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.
- Jing, B.-Y., and A.T.A. Wood (1996): “Exponential Empirical Likelihood is Not Bartlett Correctable”, *Annals of Statistics*, 24, 365-369.
- Kitamura, Y. (2001): “Asymptotic Optimality of Empirical Likelihood for Testing Moment Restriction”, *Econometrica*, 69, 1661-1672.
- Kitamura, Y. (2005): “Empirical Likelihood Methods in Econometrics: Theory and Practice.” Invited paper presented at the Econometric Society World Congress, U.C.L., London.
- Kitamura, Y., and M. Stutzer (1997): “An Information-Theoretic Alternative to Generalized Method of Moments Estimation”, *Econometrica*, 65, 861-874.
- Kleibergen, F. (2005): “Testing Parameters in GMM Without Assuming that They Are Identified”, *Econometrica*, 73, 1103-1123.
- Kleibergen, F., and R. Paap (2005): “Generalized Reduced Rank Tests using the Singular Value Decomposition.” Forthcoming *Journal of Econometrics*.
- Moreira, M.J. (2001): “Tests with Correct Size When Instruments Can Be Arbitrarily Weak,” working paper, University of California, Berkeley.

- Moreira, M.J. (2003): "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica*, 71, 1027-1048.
- Newey, W.K. and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in Engle, R. and D. McFadden, eds., *Handbook of Econometrics*, Vol. 4, New York: North Holland.
- Newey, W.K., and R.J. Smith (2001): "Asymptotic Bias and Equivalence of GMM and GEL Estimators." Working Paper No. 01/517, University of Bristol.
- Newey, W.K., and R.J. Smith (2004): "Higher Order Properties of GMM and Generalized Empirical Likelihood." *Econometrica*, 72, 219-255.
- Newey, W.K., and K.D. West (1987): "Hypothesis Testing with Efficient Method of Moments Estimation," *International Economic Review*, 28, 777-787.
- Otsu, T. (2003): "Generalized Empirical Likelihood Inference under Weak Identification." Forthcoming *Econometric Theory*.
- Owen, A. (2001): *Empirical Likelihood*, New York: Chapman and Hall.
- Qin, J. and Lawless, J. (1994): "Empirical Likelihood and General Estimating Equations", *Annals of Statistics*, 22, 300-325.
- Reid, N. (1995): "The Roles of Conditioning in Inference", *Statistical Science*, 10, 138-157.
- Robin, J.-M., and R.J. Smith (2000): "Tests of Rank", *Econometric Theory*, 16, 151-175.
- Schennach, S.M. (2004): "Exponentially Tilted Empirical Likelihood", working paper, University of Chicago.
- Smith, R. J. (1997): "Alternative Semi-Parametric Likelihood Approaches to Generalized Method of Moments Estimation", *Economic Journal*, 107, 503-519.
- Smith, R. J. (2001): "GEL Methods for Moment Condition Models", working paper, University of Bristol. Revised version CWP 19/04, cemmap, I.F.S. and U.C.L. <http://cemmap.ifs.org.uk/wps/cwp0419.pdf>
- Stock, J.H., and J.H. Wright (2000): "GMM with Weak Identification", *Econometrica*, 68, 1055-1096.
- Vuong, Q. (1989): "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses", *Econometrica*, 57, 307-333.