GEL Criteria for Moment Condition Models

Richard J. Smith

THE INSTITUTE FOR FISCAL STUDIES
DEPARTMENT OF ECONOMICS, UCL
cemmap working paper CWP19/04

GEL Criteria for Moment Condition Models*

Richard J. Smith
Department of Economics
University of Warwick

This Revision: December 2004

Abstract

GEL methods which generalize and extend previous contributions are defined and analysed for moment condition models specified in terms of weakly dependent data. These procedures offer alternative one-step estimators and tests that are asymptotically equivalent to their efficient two-step GMM counterparts. The basis for GEL estimation is *via* a smoothed version of the moment indicators using kernel function weights which incorporate a bandwidth parameter. Examples for the choice of bandwidth parameter and kernel function are provided. Efficient moment estimators based on implied probabilities derived from the GEL method are proposed, a special case of which is estimation of the stationary distribution of the data. The paper also presents a unified set of test statistics for over-identifying moment restrictions and combinations of parametric and moment restriction hypotheses.

JEL Classification: C13, C30
Keywords: GMM, Generalized Empirical Likelihood, Efficient Moment Estimation, Overidentifying Moments, Parametric Restrictions, Additional Moment Restrictions.

*I am grateful for the constructive comments and criticisms by a Co-Editor and referee on an earlier draft. Earlier versions of this paper were presented at the London School of Economics, the *Seminaire Malinvaud*, CREST-INSEE, Paris, the Montréal Econometrics Seminar, the Tinbergen Seminar, University of Amsterdam, and the Cowles Foundation Conference on "New Developments in Time Series Econometrics", Yale University, for which Y. Kitamura was discussant. This revision has also benefited from collaboration with W.K. Newey on joint work and comments by P. Guggenberger. Of course, the aforementioned are not culpable for any remaining errors. I am also grateful for financial support of this research from the ESRC (Grant No. R000237386) and a 2002 Leverhulme Major Research Fellowship.

1 Introduction

This paper provides a unified treatment of *generalized empirical likelihood* (GEL) methods for moment condition models defined using weakly dependent data *via* smoothing the moment indicators using kernel function based weights which incorporate a bandwidth parameter. These procedures generalize and extend earlier contributions, including those of Kitamura and Stutzer (1997) and Smith (1997, 2000). Efficient GEL estimators are provided which are asymptotically equivalent to efficient two-step *generalized method of moments* (GMM) estimators [Hansen (1982)].[1] New estimators for the Jacobian and limiting variance matrices of the moment indicators are proposed. Examples of particular choices of bandwidth parameter and kernel function are discussed. Efficient moment estimators based on implied probabilities derived from the GEL method are also obtained, of which efficient estimators of the stationary distribution of the data are a special case. This approach generalizes that suggested by Back and Brown (1993) to GEL and Brown and Newey (1998) to weakly dependent data. The paper also presents a unified set of test statistics for over-identifying moment restrictions and combinations of moment and parametric restriction hypotheses expressed in mixed form, which admit most forms of restrictions as special cases. These test statistics also extend existing treatments.

The estimation of moment condition models is not only of theoretical interest but has substantial empirical importance. In particular, it is now widely recognized that the most commonly used efficient two-step *generalized method of moments* (GMM) estimator [Hansen (1982)] may be severely biased for the sample sizes typically encountered in applications. See, for example, the Special Section, July 1996, of the *Journal of Business and Economic Statistics*. A number of alternative efficient estimators have been proposed to ameliorate bias. Hansen, Heaton, and Yaron (1996) suggested the *continuous updating estimator* (CUE). Other estimators include *empirical likelihood* (EL) [Imbens (1997), Qin and Lawless (1994)], and *exponential tilting* (ET) [Imbens, Spady and Johnson (1998),

[1] Let T denote the sample size. The relevant optimality concept for estimation throughout this paper is that of minimum asymptotic variance among root-T consistent and asymptotic normal GMM estimators based on a given set of unconditional moment restrictions. See Newey and McFadden (1994, section 5.2, pp.2164-2165) and below Theorem 2.3.

Kitamura and Stutzer (1997)]. All of these estimators are members of the GEL class considered here and outlined in Smith (1997) as are estimators based on the Cressie and Read (1984) family of power divergence criteria. In a random sampling setting, Newey and Smith (2004), henceforth NS, compare the asymptotic higher order bias of two-step (and iterated) GMM estimators to estimators in the class of GEL estimators. Their results account for the poor bias properties of two-step and iterated GMM estimators which arise through the estimation of the Jacobian and efficient metric in the GMM criterion function, the latter suggested as a cause of bias by the Monte Carlo experiments conducted by Altonji and Segal (1996). The former source of asymptotic bias is absent for GEL estimators and, in particular, the EL estimator behaves like the infeasible optimal GMM estimator.

Given these encouraging findings for the GEL class of estimators, a primary aim of this paper is to synthesise and extend this class of estimators to weakly dependent data. To deal with the time series nature of the data, a smoothed version of the moment indicators forms the basis of the suggested estimation procedure rather than the moment indicators themselves as in standard GMM estimation; cf. Kitamura and Stutzer (1997).[2] The GEL method offers attractive alternative one-step efficient estimators, not requiring explicit calculation or estimation of the efficient metric, that are asymptotically equivalent to those based on efficient two-step GMM. Efficient moment estimators are also proposed. Moreover, because of their quasi-likelihood construction, the elucidation of classical-type test statistics for over-identifying moment conditions, additional moment conditions and parametric restrictions is relatively straightforward. An additional emphasis of this paper concerns issues of specification. In particular, this paper discusses specification test statistics based on GEL criteria rather than the more typical approach of using a quadratic form in estimated sample analogues of the assumed or implicit population moment conditions; see Hansen (1982) and Newey (1985b). The tests presented here mimic in a rather obvious way standard classical tests.

[2] An alternative approach for EL estimation in the time series context is suggested in Kitamura (1997) using blockwise EL which should be adaptable for the GEL criteria considered here. Kitamura (1997) also shows the Bartlett correctability of blockwise EL.

Section 2 introduces GEL criteria for time series data which are formed by incorporating a kernel weighted sample version of the moment indicators. The parameter vector and weights associated with these smoothed moment indicators are the respective objects of interest for estimation and inference, the former from an economic-theoretic standpoint and the latter for tests of specification. The GEL estimation procedure is then described and the limiting distribution of the estimators is obtained. Consistent estimators for the Jacobian and moment indicator limiting variance matrix are detailed together with suggestions for the choice of bandwidth parameter and kernel function. Section 3 describes moment estimators which optimally incorporate the moment information and, therefore, dominate more traditional estimated sample average forms based on the empirical distribution function. A special case of the moment estimator is one for the stationary distribution of the data which, therefore, is more efficient than the empirical distribution function. Section 4 is concerned with deriving classical-type tests for over-identifying moment restrictions; cf. Hansen (1982). Section 5 presents a unified treatment of classical-type tests for additional moment restrictions, cf. Newey (1985b), and parametric contraints expressed in mixed form, see Gourieroux and Monfort (1989), which are sufficiently general to include other forms of parametric constraint of interest. Section 6 concludes. Proofs of the results are given in the Appendices.

The following abbreviations are used throughout the paper: w.p.a.1: with probability approaching one; $\overset{a}{=}$: differs by no more than an $o_p(1)$ term; $\overset{p}{\to}$: converges in probability to; $\overset{d}{\to}$: converges in distribution to; $\|.\|$: the matrix norm defined by $\|A\| = \sqrt{\lambda_{\max}(A'A)}$ where $\lambda_{\max}(\cdot)$ is the maximum eigenvalue of $\cdot$; p.d.: positive definite; n.d.: negative definite; p.s.d.: positive semi-definite; f.c.r.: full column rank.

2 Generalized Empirical Likelihood

Let z_t, $(t = 1, ..., T)$, denote observations on a finite dimensional stationary and strongly mixing process $\{z_t\}_{t=1}^{\infty}$. Consider the moment indicator $g(z_t, \beta)$, an m-vector of known functions of the data observation z_t and the p-vector β of unknown parameters which are the object of inferential interest, where $m \geq p$. It is assumed that the true parameter vector β_0 uniquely satisfies the moment condition

$$E[g(z_t, \beta_0)] = 0, \tag{2.1}$$

where $E[\cdot]$ denotes expectation taken with respect to the unknown distribution of z_t. Typically, (2.1) will arise from conditional moment restrictions. In such cases, z_t may also include lagged endogenous and current and lagged values of exogenous variables.

For economy of notation, we define $g_t(\beta) = g(z_t, \beta)$, $(t = 1, ..., T)$, and $\hat{g}(\beta) = T^{-1} \sum_{t=1}^{T} g_t(\beta)$.

2.1 Moment Indicators

Standard generalized method of moments (GMM) criteria are defined directly as quadratic forms in terms of the moment indicators $g_t(\beta)$, $(t = 1, ..., T)$, and their sample average $\hat{g}(\beta)$. In contradistinction, we work here with the smoothed counterparts

$$g_{tT}(\beta) = \frac{1}{S_T} \sum_{s=t-T}^{t-1} k\left(\frac{s}{S_T}\right) g_{t-s}(\beta),\ (t = 1, ..., T), \tag{2.2}$$

where S_T is a bandwidth parameter and $k(\cdot)$ a kernel function. As described in section 2.6 below the kernel weights $\frac{1}{S_T} k\left(\frac{s}{S_T}\right)$ in (2.2) give rise to weights similar in nature to those used in heteroskedastic and autocorrelation consistent (HAC) covariance matrix estimation; see *inter alia* Andrews (1991) and Newey and West (1987). Theorem 2.1 below and the following discussion provide some intuition for the necessity of smoothing $g_t(\beta)$, $(t = 1, ..., T)$, to achieve asymptotic efficiency; see also Kitamura and Stutzer (1997). Kitamura and Stutzer (1997) and Smith (1997) employ particular choices for the bandwidth parameter S_T and kernel function $k(\cdot)$, the former also using a special case of the class of GEL criteria defined below. Suitable choices of S_T and $k(\cdot)$ which ensure that GEL estimators are first order asymptotically equivalent to efficient GMM estimators are discussed in section 2.6 below.

Let $k_j = \int_{-\infty}^{\infty} k(a)^j da$, $j = 1, 2$.

The large sample properties of consistency and asymptotic normality of GMM estimators rely on a uniform weak law (UWL) of large numbers and central limit theorem

(CLT) defined in terms of the sample average $\hat{g}(\beta)$. Similar results in terms of the sample average $\hat{g}_T(\beta) = T^{-1}\sum_{t=1}^{T} g_{tT}(\beta)$ of the smoothed moment indicators (2.2) are required for the asymptotic properties of GEL estimators. In particular,

$$\sup_{\beta\in\mathcal{B}} \|\hat{g}_T(\beta) - k_1 E[g_t(\beta)]\| = o_p(1), \tag{2.3}$$

where $\mathcal{B}$ denotes the parameter space. In addition,

$$T^{1/2}[\hat{g}_T(\beta) - E[\hat{g}_T(\beta)]] \xrightarrow{d} N(0, (k_1^2)\Omega(\beta)), \tag{2.4}$$

where

$$\Omega(\beta) = \lim_{T\to\infty} var[T^{1/2}\hat{g}(\beta)].$$

The limiting variance matrix of $T^{1/2}\hat{g}(\beta_0)$, $\Omega = \Omega(\beta_0)$, is assumed p.d.. Detailed statements of and proofs for (2.3) and (2.4) are provided by Lemmas A.1 and A.2 respectively in Appendix A.

2.2 GEL Criteria

Let $\rho(\cdot)$ be a function that is concave on its domain $\mathcal{V}$, an open interval containing zero. It will be convenient to impose a normalization on $\rho(\cdot)$. Let $\rho_j(\cdot) = \partial^j\rho(\cdot)/\partial v^j$ and $\rho_j = \rho_j(0)$, $(j = 0, 1, 2, ...)$. We normalize so that $\rho_1 = \rho_2 = -1$. As long as $\rho_1 \neq 0$ and $\rho_2 < 0$, which we will assume to be true, this normalization can always be imposed by replacing $\rho(\cdot)$ by $[-\rho_2/\rho_1^2]\rho([\rho_1/\rho_2]\cdot)$, which does not affect the estimator of β. It is satisfied by the $\rho(\cdot)$ given below for EL, ET and CUE.

We introduce a m-vector of auxiliary parameters λ, each element of which is associated with a corresponding element of the smoothed indicator $g_{tT}(\beta)$ of (2.2). The class of GEL criteria considered here is then defined as

$$\hat{P}(\beta, \lambda) = \sum_{t=1}^{T}[\rho(k\lambda' g_{tT}(\beta)) - \rho_0]/T. \tag{2.5}$$

The normalisation

$$k = \frac{k_1}{k_2} \tag{2.6}$$

has no effect on the GEL estimator for β but makes the scale of the estimator of the auxiliary parameters λ comparable for different choices of kernel function $k(\cdot)$.

We restrict the auxiliary parameters $\lambda \in \Lambda_T$ in order that w.p.a.1 $k\lambda' g_{tT}(\beta)$ is in the domain $\mathcal{V}$ of $\rho(\cdot)$ for all $\lambda \in \Lambda_T$, $\beta \in \mathcal{B}$, and $1 \leq t \leq T$. It suffices for the theory here that Λ_T places bounds on λ that shrink with T slower than $(T/S_T^2)^{-1/2}$ which is the rate of convergence of the GEL estimator for λ; see Assumption 2.4 (b) and Theorem 2.2 below.

The GEL criterion (2.5) may be interpreted as an adaptation of the approach taken in Chesher and Smith (1997) to the moment conditions context. Chesher and Smith (1997) analyses likelihood ratio test statistics for implied moment conditions in a fully parametric likelihood setting with the likelihood obtained by augmenting the null hypothesis parametric density multiplicatively by a carrier function $h(\cdot)$ of a weighted version of the moment indicators underpinning the implied moment conditions as in (2.5).[3] In the GMM context, however, there is no explicit knowledge of the underlying density function for $\{z_t\}_{t=1}^{\infty}$, the only parametric information being contained in the moment conditions (2.1). Replacing the parametric density function in Chesher and Smith's (1997) procedure by the non-parametric empirical measures $d\mu_t = T^{-1}$, $(t = 1, ..., T)$, circumvents this difficulty. Viewed in this light, the function $\rho(\cdot)$ is minus the logarithm of the carrier function, $-\log h(\cdot)$, in Chesher and Smith's (1997) framework.

The GEL class admits a number of special cases which have been the focus of recent attention in the statistics and econometrics literature. The EL estimator is a GEL estimator with $\rho(v) = \log(1 - v)$, see Imbens (1997), Qin and Lawless (1994) and Smith (2000), and the ET estimator of Imbens, Spady, and Johnson (1998) is also GEL with $\rho(v) = -\exp(v)$. The CUE of Hansen, Heaton and Yaron (1996) is obtained if $\rho(\cdot)$ is quadratic; see Theorem 2.1, p.223, of NS. Similarly, the criterion suggested by Kitamura and Stutzer (1997) based on (2.2) is $-\log\left(T^{-1}\sum_{t=1}^{T}\exp(k\lambda' g_{tT}(\beta))\right)$ with $k(\cdot)$ the truncated kernel, see Example 2.1 of section 2.5, which, as $\log(1+x) \doteq x$, is approximately $1 - T^{-1}\sum_{t=1}^{T}\exp(\lambda' g_{tT}(\beta))$. Hence, to the orders of magnitude considered in this paper,

[3] Many tests of specification in classical settings may be formulated as tests for implied moment conditions; see *inter alia* Newey (1985a) and Tauchen (1985).

Kitamura and Stutzer's (1997) criterion is equivalent to GEL with $\rho(v) = -\exp(v)$, that is, the ET criterion. Moreover, the ET estimator is identical to that based on the criterion $-\log\left(T^{-1}\sum_{t=1}^{T}\exp(k\lambda' g_{tT}(\beta))\right)$ since it is a monotonic transformation of the ET criterion. More generally, members of the Cressie-Read (1984) power divergence family of discrepancies discussed by Imbens, Spady, and Johnson (1998) are included in the GEL class with $\rho(v) = -\gamma^2(1+v)^{(\gamma+1)/\gamma}/(\gamma+1)$. In this case, the GEL optimisation problem is a dual of that arising from the Cressie-Read (1984) family. GEL estimators are also related to minimum discrepancy (MD) estimators considered by Corcoran (1998) but do not necessarily coincide unless the first inverse derivative of the MD function is homogenous. An advantage of the GEL class over MD is that GEL estimators are obtained from a much smaller dimensional optimization problem than that for MD which increases with T.[4] Moreover, the ability to estimate the distribution of the data is not lost for GEL as detailed in section 3. See NS for further discussion.

2.3 GEL Estimation

GEL estimators are obtained as the solution to a saddle point problem. Firstly, the GEL criterion $\hat{P}(\beta,\lambda)$ is maximised for given β. That is, $\hat{\lambda}(\beta) = \arg\sup_{\lambda\in\Lambda_T}\hat{P}(\beta,\lambda)$ and satisfies the GEL first order conditions[5]

$$T^{-1}\sum_{t=1}^{T}\rho_1(k\hat{\lambda}(\beta)' g_{tT}(\beta))g_{tT}(\beta) = 0. \tag{2.7}$$

Consequently, (2.7) indicates that the moment conditions are satisfied in the sample. Secondly, the GEL estimator $\hat{\beta}$ is the minimiser of the profile GEL criterion $\hat{P}(\beta,\hat{\lambda}(\beta))$

$$\hat{\beta} = \arg\min_{\beta\in\mathcal{B}}\hat{P}(\beta,\hat{\lambda}(\beta)) = \arg\min_{\beta\in\mathcal{B}}\sup_{\lambda\in\Lambda_T}\hat{P}(\beta,\lambda). \tag{2.8}$$

[4] The Cressie-Read (1984) family is parameterised through the single parameter γ. Higher order asymptotic expansions for GEL estimators and test statistics involve derivatives evaluated at zero of $\rho(\cdot)$ at least to the third order which confers greater flexibility on GEL criteria, a feature also possessed by MD criteria. For example, see NS for an analysis of asymptotic bias and higher order efficiency in the i.i.d. context.

[5] Appendix D details the second order derivatives of $\hat{P}(\cdot,\cdot)$. In particular, if $\sum_{t=1}^{T} g_{tT}(\beta)g_{tT}(\beta)'$ is p.d., $\hat{\lambda}(\beta)$ is a unique maximiser of $\hat{P}(\beta,\cdot)$.

Writing $\hat{\lambda} = \hat{\lambda}(\hat{\beta})$, $\hat{\beta}$ satisfies the GEL first order conditions

$$T^{-1}\sum_{t=1}^{T}\rho_1(k\hat{\lambda}'g_{tT}(\hat{\beta}))G_{tT}(\hat{\beta})'\hat{\lambda} = 0, \tag{2.9}$$

where $G_{tT}(\beta) = \partial g_{tT}(\beta)/\partial\beta'$, $(t = 1, ..., T)$. Hence, the solutions $\hat{\beta}$ and $\hat{\lambda}$ define a saddle point of the GEL criterion $\hat{P}(\beta, \lambda)$.[6] As is evident from the GEL first order conditions (2.7) and (2.9) the introduction of the auxiliary parameters λ renders the first order conditions determining the GEL estimator $\hat{\beta}$ and $\hat{\lambda}$ as corresponding to a just-identified GMM problem. In the just-identified case $m = p$, $\hat{\lambda} = 0$ from (2.9) and, thus, (2.7) reduces to the familiar GMM first order conditions for a just-identified problem.

A re-interpretation of the GEL first order conditions (2.7) and (2.9) aids understanding of the efficiency of the GEL estimator $\hat{\beta}$ and why GEL might be expected to be less biased than efficient GMM.

Let $G_t(\beta) = \partial g_t(\beta)/\partial\beta'$, $(t = 1, ..., T)$, $\hat{G}(\beta) = \sum_{t=1}^{T} G_t(\beta)/T$ and $\hat{\Omega}(\tilde{\beta})$ denote a consistent estimator of Ω constructed, for example, as in Andrews (1991) or section 2.5 below with $\tilde{\beta}$ an initial $T^{1/2}$-consistent estimator for β_0. The GMM first order conditions imply

$$\hat{G}(\hat{\beta}_{GMM})'\hat{\Omega}(\tilde{\beta})^{-1}\hat{g}(\hat{\beta}_{GMM}) = 0, \tag{2.10}$$

where $\hat{\beta}_{GMM}$ denotes the GMM estimator.

An analogous expression may also be obtained for any GEL estimator $\hat{\beta}$. Define $p(v) = [\rho_1(v) + 1]/v, v \neq 0$ and $p(0) = -1$. Also, let $\hat{p}_t = p(k\hat{\lambda}'\hat{g}_{tT})/\sum_{s=1}^{T} p(k\hat{\lambda}'\hat{g}_{sT})$ and $\hat{\pi}_t = \pi_t(\hat{\beta}, \hat{\lambda})$ be as defined in (3.1) below.

Theorem 2.1 *The GEL first order conditions imply*

$$[\sum_{t=1}^{T}\hat{\pi}_t G_{tT}(\hat{\beta})]'[S_T\sum_{t=1}^{T}\hat{p}_t g_{tT}(\hat{\beta})g_{tT}(\hat{\beta})']^{-1}\hat{g}_T(\hat{\beta}) = 0, \tag{2.11}$$

where $\hat{p}_t = \hat{\pi}_t$ *for EL and* $\hat{p}_t = 1/T$ *for CUE.*

---

[6] Imbens, Spady and Johnson (1998) detail a robust method for the computation of $\hat{\lambda}$ and $\hat{\beta}$ for the Cressie-Read power divergence family which may be suitably adapted for the GEL class. A consistent estimator for β_0 to initiate an iterative procedure to locate $\hat{\lambda}$ and $\hat{\beta}$ is any GMM estimator, optimal or otherwise.

Let $G_t = G_t(\beta_0)$ and $G = E[G_t]$. Theorem 2.1 mirrors Theorem 2.3, p.224, in NS for the random sampling case. It is straightforward to show that under the assumptions given below $T^{1/2}[\hat{g}_T(\hat{\beta}) - k_1\hat{g}(\hat{\beta})] \overset{a}{=} 0$. Hence, when comparing the GMM and GEL first order conditions (2.10) and (2.11), we see that each approximately sets a particular linear combination of $\hat{g}(\beta)$ equal to zero. Furthermore, as described in section 3, $\hat{\pi}_t$ (and $\hat{p}_t$ similarly) behaves like the empirical measure $d\mu_t = T^{-1}$, $(t = 1, ..., T)$, i.e. $T^{-1}(1 + o_p(1))$, see (B.5) in the proof of Theorem 3.1. Therefore, $\sum_{t=1}^{T} \hat{\pi}_t G_{tT}(\hat{\beta}) \xrightarrow{p} k_1 G$ and $S_T \sum_{t=1}^{T} \hat{p}_t g_{tT}(\hat{\beta}) g_{tT}(\hat{\beta})' \xrightarrow{p} k_2\Omega$, cf. section 2.5 below. GMM thus consistently estimates the Jacobian term G using the sample average $\hat{G}(\hat{\beta}_{GMM})$ whereas GEL uses the re-weighted smoothed derivative estimator $\sum_{t=1}^{T} \hat{\pi}_t G_{tT}(\hat{\beta})/(k_1)$. The estimators for the variance matrix Ω are correspondingly $\hat{\Omega}(\tilde{\beta})$ and $S_T \sum_{t=1}^{T} \hat{p}_t g_{tT}(\hat{\beta}) g_{tT}(\hat{\beta})'/(k_2)$. Efficient GMM and GEL estimators therefore approximately solve the same first order conditions by setting the optimal linear combination $G'\Omega^{-1}\hat{g}(\beta)$ equal to zero.

All GEL estimators implicitly use an efficient estimator $\sum_{t=1}^{T} \hat{\pi}_t G_{tT}(\hat{\beta})/(k_1)$ of the Jacobian term; see Theorem 3.1. It is also interesting to note that EL uses a similar weighting scheme in the estimation of Ω whereas CUE uses the sample average, and other GEL estimators use other weighted averages. Theorem 3.1 demonstrates that efficient moment estimators are asymptotically uncorrelated with $\hat{g}_T(\hat{\beta})$ or $\hat{g}(\hat{\beta})$. As noted by NS for random sampling, correlations between corresponding terms in the first order conditions are an important source of bias. Similarly to NS, therefore, one might expect that GEL will also be less prone to bias than GMM when the data are weakly dependent.

2.4 Asymptotic Theory for GEL Estimators

To describe the asymptotic results for $\hat{\beta}$ and $\hat{\lambda}$, let

$$\begin{aligned} \Sigma &= (G'\Omega^{-1}G)^{-1}, H = \Sigma G'\Omega^{-1}, \\ P &= \Omega^{-1} - \Omega^{-1}G\Sigma G'\Omega^{-1}. \end{aligned}$$

We firstly detail some regularity conditions sufficient for a consistency result. These assumptions are quite standard and are similar to those given in Andrews (1991) and

Kitamura and Stutzer (1997).

Assumption 2.1 *The process* $\{z_t\}_{t=1}^{\infty}$ *is a finite dimensional stationary and strong mixing with mixing coefficients* $\sum_{j=1}^{\infty} j^2\alpha(j)^{(\nu-1)/\nu} < \infty$ *for some* $\nu > 1$.

Hence, $\{g_t(\beta)\}_{t=1}^{\infty}$ satisfies Assumption 2.1 and is therefore ergodic which ensures the uniform convergence of certain sample averages to their population counterparts; see, for example, (2.3).

The next assumption introduces standard conditions on the bandwidth parameter S_T and kernel function $k(\cdot)$ which ensure that S_T obeys conditions similar to those described in Andrews (1991, Theorem 1 (a), p.827). Let

$$\bar{k}(x) = \begin{cases} \sup_{y\geq x}|k(y)| & \text{if } x \geq 0 \\ \sup_{y\leq x}|k(y)| & \text{if } x < 0 \end{cases}$$

and $K(\lambda) = (2\pi)^{-1}\int k(x)\exp(-\iota x\lambda)dx$ denote the spectral window generator of the kernel $k(\cdot)$.

Assumption 2.2 **(a)** $S_T \to \infty$, $S_T/T^2 \to 0$ *and* $S_T = O(T^{\frac{1}{2}-\eta})$ *for some* $\eta > 0$; **(b)** $k(\cdot) : \mathcal{R} \to [-k_{\max}, k_{\max}]$, $k_{\max} < \infty$, $k(0) \neq 0$, $k_1 \neq 0$, *and is continuous at* 0 *and almost everywhere;* **(c)** $\int_{(-\infty,\infty)} \bar{k}(x)dx < \infty$; **(d)** $|K(\lambda)| \geq 0$ *for all* $\lambda \in \mathcal{R}$.

Assumptions 2.2 (b) and (c) ensure $k_2 > 0$. Assumptions 2.2 (b) and (c) also guarantee that the induced kernel $k^*(\cdot)$

$$k^*(a) = \frac{1}{k_2}\int_{-\infty}^{\infty} k(b-a)k(b)db, \tag{2.12}$$

which arises implicitly in GEL estimation based on (2.5), is a member of the p.s.d. class of kernels $\mathcal{K}_2$ used in HAC covariance matrix estimation [Andrews (1991, p.822)] which is defined in (2.15) of section 2.5 below; see Lemma C.3 in Appendix C. Assumption 2.2 (c) is required to ensure that certain normalised sums defined in terms of the kernel $k(\cdot)$ converge appropriately to their integral representation counterparts; see Jansson (2002).

While the next assumption in part states regularity conditions which are usual for the consistency of GMM estimators, the existence of higher moments is required for GEL estimators.

Assumption 2.3 **(a)** *$\beta_0 \in \mathcal{B}$ is the unique solution to $E[g_t(\beta)] = 0$;* **(b)** *$\mathcal{B}$ is compact;* **(c)** *$g_t(\beta)$ is continuous at each $\beta \in \mathcal{B}$ with probability one;* **(d)** *$E[\sup_{\beta\in\mathcal{B}} \|g_t(\beta)\|^\alpha] < \infty$ for some $\alpha > \max\left(4\nu, \frac{1}{\eta}\right)$;* **(e)** *$\Omega(\beta)$ is finite and p.d. for all $\beta \in \mathcal{B}$.*

Assumption 2.1 together with Assumption 2.3 (d) ensures $\{g_t(\beta) - E[g_t(\beta)]\}_{t=1}^\infty$ will satisfy the hypotheses of Lemma 1, p.824, of Andrews (1991).

Assumption 2.4 **(a)** *$\rho(\cdot)$ is twice continuously differentiable and concave on its domain, an open interval $\mathcal{V}$ containing 0, $\rho_1 = \rho_2 = -1$;* **(b)** *$\lambda \in \Lambda_T$ where $\Lambda_T = \{\lambda : \|\lambda\| \leq D(T/S_T^2)^{-\zeta}\}$ for some $D > 0$ with $\frac{1}{2} > \zeta > \frac{1}{2\alpha\eta}$.*

Assumption 2.4 (b) specifies bounds on λ which shrink slower than the stochastic order of the auxiliary parameter estimator $\hat{\lambda}$ stated in Theorem 2.2 below. When combined with the existence of higher than second moments in the previous assumption and the restriction in Assumption 2.3 (d) on α, this condition leads to the argument $k\lambda' g_{tT}(\beta)$ being in the domain $\mathcal{V}$ of $\rho(\cdot)$ w.p.a.1 for all β and $1 \leq t \leq T$.

The above conditions lead to a consistency result.

Theorem 2.2 *If Assumptions 2.1-2.4 are satisfied then $\hat{\beta} \xrightarrow{p} \beta_0$ and $\hat{\lambda} \xrightarrow{p} 0$. Moreover, $\|\hat{\lambda}\| = O_p[(T/S_T^2)^{-1/2}]$ and $\|\hat{g}_T(\hat{\beta})\| = O_p(T^{-1/2})$.*

For asymptotic normality we need additional regularity conditions.

Assumption 2.5 **(a)** *$\beta_0 \in int(\mathcal{B})$;* **(b)** *$g(\cdot, \beta)$ is differentiable in a neighborhood $\mathcal{N}$ of β_0 and $E[\sup_{\beta\in\mathcal{N}} \|\partial g_t(\beta)/\partial\beta'\|^{\alpha/(\alpha-1)}] < \infty$;* **(c)** *$rank(G) = p$.*

Theorem 2.3 *Let Assumptions 2.1-2.5 hold. Then*

$$T^{1/2}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, \Sigma),\ (T/S_T^2)^{1/2}\hat{\lambda} \xrightarrow{d} N(0, P),$$

and the GEL estimator $\hat{\beta}$ and the auxiliary parameter estimator $\hat{\lambda}$ are asymptotically uncorrelated.

The asymptotic variance matrix $\Sigma = (G'\Omega^{-1}G)^{-1}$ is the efficiency lower bound for GMM estimators based on the quadratic form $T\hat{g}(\beta)'\hat{W}\hat{g}(\beta)$ where $\hat{W}$ is p.s.d., $\hat{W} \overset{a}{=} W$ and W is p.d.; see Newey and McFadden (1994, section 5.2, pp.2164-2165). The lack of asymptotic correlation between $\hat{\beta}$ and $\hat{\lambda}$ underlines that the moment conditions (2.1) are used efficiently in the estimation of β and, therefore, that $\hat{\beta}$ is asymptotically equivalent to efficient GMM estimators. Furthermore, $(T/S_T^2)^{1/2}\hat{\lambda} \overset{a}{=} -\Omega^{-1}T^{1/2}\hat{g}(\hat{\beta})$. Hence, the auxiliary parameter estimator $\hat{\lambda}$ may be used to assess the validity or otherwise of the moment conditions (2.1); see section 4.

2.5 Estimation of Ω and G

The Hessian of the optimised GEL criterion (2.5) given in Appendix C provides a basis for the consistent estimation of Ω and G.

Let $\hat{g}_{tT} = g_{tT}(\hat{\beta})$ and $\hat{G}_{tT} = G_{tT}(\hat{\beta})$, $(t = 1, ..., T)$.

The (λ, β)-block $\partial^2\hat{P}(\hat{\beta}, \hat{\lambda})/\partial\lambda\partial\beta'$ (D.1) of the Hessian provides a basis for the consistent estimation of G.

Theorem 2.4 *(Consistent Estimation of G.) Under Assumptions 2.1-2.5, the re-scaled (λ, β)-block of the Hessian, $-(k_2/k_1^2)(\partial^2\hat{P}(\hat{\beta}, \hat{\lambda})/\partial\lambda\partial\beta')$, is a consistent estimator of G.*

By adapting UWL Lemma A.1, $\sum_{t=1}^T G_{tT}(\hat{\beta})/(Tk_1)$, as well as $\sum_{t=1}^T G_t(\hat{\beta})/T$, is also a consistent estimator of G.

The (λ, λ)-block $\partial^2\hat{P}(\hat{\beta}, \hat{\lambda})/\partial\lambda\partial\lambda'$ (D.2) of the Hessian is $k^2\sum_{t=1}^T \rho_2(k\hat{\lambda}'\hat{g}_{tT})\hat{g}_{tT}\hat{g}_{tT}'/T$, which is n.d. if $\sum_{t=1}^T \hat{g}_{tT}\hat{g}_{tT}'$ is p.d..

Theorem 2.5 *(Consistent Estimation of Ω.) Let Assumptions 2.1-2.5 be satisfied. Then the re-scaled (λ, λ)-block of the Hessian, $-(k_2/k_1^2)S_T(\partial^2\hat{P}(\hat{\beta}, \hat{\lambda})/\partial\lambda\partial\lambda')$, is a consistent estimator of Ω.*

The proof of Theorem 2.5 shows that $S_T\sum_{t=1}^T \rho_2(k\hat{\lambda}'\hat{g}_{tT})\hat{g}_{tT}\hat{g}_{tT}'/(Tk_2) \overset{a}{=} -\hat{\Omega}_T(\beta_0)$ where

$$\hat{\Omega}_T(\beta) = S_T\sum_{t=1}^T g_{tT}(\beta)g_{tT}(\beta)'/(Tk_2). \tag{2.13}$$

Lemma A.3 in Appendix A, a subsidiary result of particular importance in the proofs of Theorems 2.2, 2.3 and 2.5, establishes the validity of Lemma 2.1 in Smith (2005), i.e. $\hat{\Omega}_T(\beta_0) \xrightarrow{p} \Omega$, the proof of which was omitted there. Theorem 2.1 of Smith (2005) then shows that evaluation of $\hat{\Omega}_T(\beta)$ at a $T^{1/2}$-consistent estimator for β_0 offers an alternative to those p.s.d. consistent estimators for Ω described *inter alia* in Andrews (1991) and Newey and West (1987). Therefore, $S_T \sum_{t=1}^{T} \hat{g}_{tT}\hat{g}_{tT}'/(Tk_2) \xrightarrow{p} \Omega$.[7]

Alternative estimators for Ω and G are obtained if the empirical measure $d\mu_t = T^{-1}$ is replaced by the GEL implied probability $\pi_t(\hat{\beta}, \hat{\lambda})$, $(t = 1, ..., T)$, defined in (3.1) below. The scaling constants k_1 and k_2 may also be replaced by their respective sample counterparts $\hat{k}_j = \sum_{s=1-T}^{T-1} k\left(\frac{s}{S_T}\right)^j / S_T$, $(j = 1, 2)$.

2.6 Kernel $k(\cdot)$ and Bandwidth Parameter S_T Choices

Consider the class of symmetric kernels $\mathcal{K}_1$ defined by

$$\begin{aligned}\mathcal{K}_1 &= \{k^*(\cdot) : \mathcal{R} \to [-1, 1] | k^*(0) = 1, k^*(-a) = k^*(a) \forall x \in \mathcal{R}, \int_{[0,\infty)} \bar{k}^*(a)da < \infty, \\ &\quad k^*(\cdot) \text{ continuous at 0 and almost everywhere}\}. \qquad (2.14)\end{aligned}$$

where $\bar{k}^*(a) = \sup_{b \geq |a|} |k^*(b)|$; see, for example, Andrews (1991) and Andrews and Monahan (1992).[8] The p.s.d. class $\mathcal{K}_2$ is then defined as in Andrews (1991, p.822) by

$$\mathcal{K}_2 = \{k^*(\cdot) \in \mathcal{K}_1 : K^*(\lambda) \geq 0 \text{ for all } \lambda \in \mathcal{R}\}, \qquad (2.15)$$

where $K^*(\lambda) = (2\pi)^{-1} \int k^*(x) \exp(-\iota x \lambda) dx$ is the spectral window generator of the kernel $k^*(\cdot)$.

[7]Theorems 3.1 and 3.2 in Smith (2005) demonstrate that under the hypotheses of Theorems 2.2 and 2.3, two-step and iterated GMM estimators based on the criterion $\hat{g}_T(\beta)'\hat{\Omega}_T(\tilde{\beta})^{-1}\hat{g}_T(\beta)$, where $\tilde{\beta}$ is an initial $T^{1/2}$-consistent estimator for β_0, are consistent estimators for β_0, asymptotically equivalent to the GEL estimator $\hat{\beta}$ and, thus, asymptotically efficient.

[8]Neither the square integrability condition $\int_{-\infty}^{\infty} k^*(x)^2 dx < \infty$ in Andrews (1991, (2.6), p.821) nor the stronger absolute integrability condition $\int_{-\infty}^{\infty} |k^*(x)| dx < \infty$ in Andrews and Monahan (1992, (2.5), p.955) is sufficient for the consistency results claimed in those papers; see Jansson (2002). The condition $\int_{[0,\infty)} \bar{k}^*(x)dx < \infty$ ensures that particular summations used in those papers converge appropriately; see Lemma 1 of Jansson (2002).

In the proof of Lemma A.3, it is required that the infeasible estimator

$$\Omega_T(\beta_0) = \sum_{s=1-T}^{T-1} k_T^*\left(\frac{s}{S_T}\right) C_T(s)/k_2 \tag{2.16}$$
$$\xrightarrow{p} \Omega,$$

where $k_T^*\left(\frac{s}{S_T}\right) = \frac{1}{S_T}\sum_{t=\max[1-T,1-T+s]}^{\min[T-1,T-1+s]} k\left(\frac{t-s}{S_T}\right) k\left(\frac{t}{S_T}\right)/k_2$ and the infeasible sample covariances $C_T(s) = \sum_{t=\max[1,1-s]}^{\min[T,T-s]} g_{t+s}(\beta_0)g_t(\beta_0)'/T$, $C_T(-s) = C_T(s)'$, $(s = 1-T, ..., T-1)$. Lemma C.2 in Appendix A shows that $k_T^*(a) = k^*(a) + o(1)$ uniformly where $k^*(\cdot)$ is the induced kernel defined in (2.12).[9] Moreover, Lemma C.3 proves that if Assumptions 2.2 (b) and (c) are satisfied then $k^*(\cdot)$ belongs to the p.s.d. class $\mathcal{K}_2$ (2.15), cf. Andrews (1991, p.822).

Therefore choices for the bandwidth parameter S_T and kernel function $k(\cdot)$ should satisfy the conditions set out above in Assumption 2.2; see also Andrews (1991, Theorem 1 (a), p.827). In particular, we require $S_T = o(T^{1/2})$, see Assumption 2.2 (a), and that $k(\cdot)$ implies $k^*(\cdot) \in \mathcal{K}_2$, see Assumptions 2.2 (b) and (c). The spectral window generator for $k^*(\cdot)$ is related to that of $k(\cdot)$ by $K^*(\lambda) = (2\pi)^{-1}\int \exp(-ia\lambda)k^*(a)da = 2\pi\left|K(\lambda)\right|^2$; see the proof of Lemma C.3. This relationship between $K(\cdot)$ and $K^*(\cdot)$ allows the kernel $k(\cdot)$ employed in (2.2) to be straighforwardly deduced from suitable choices for $k^*(\cdot)$ as the examples given below attest. The optimal rate for S_T will depend on the particular kernel $k(\cdot)$ (and, thus, implicit $k^*(\cdot)$) chosen; see Andrews (1991, section 5, pp.830-832).

The following examples detail the optimal rate for S_T and the kernel $k(\cdot)$ corresponding to particular choices for $k^*(\cdot)$. The quadratic spectral kernel $k^*(\cdot)$ of Example 2.3 below is the optimal kernel in a truncated asymptotic mean squared error sense for HAC consistent estimation of Ω; see Andrews (1991, Theorem 2, p.829).

Example 2.1: Bartlett Kernel.

[9] The estimator $\Omega_T(\beta_0)$ (2.16) belongs the general class of quadratic estimators [Grenander and Rosenblatt (1984, Section 4.1)] and has Toeplitz weight matrix. As $\hat{\Omega}_T(\beta_0)$ (2.13) and $\Omega_T(\beta_0)$ are asymptotically equivalent, i.e. $\hat{\Omega}_T(\beta_0) - \Omega_T(\beta_0) \overset{a}{=} 0$, it might be expected that the estimators for Ω suggested in section 2.5 would inherit the desirable asymptotic mean squared error properties of standard lag kernel estimators [Grenander and Rosenblatt (1984, Section 4.2)].

Consider the truncated kernel $k(x) = 1$, $|x| \leq 1$, and 0, $|x| > 1$, $k_1 = 2$ and $k_2 = 2$. Hence, defining the bandwidth parameter $S_T = (2m_T + 1)/2$,

$$g_{tT}(\beta) = (2m_T + 1)^{-1} \sum_{s=\max[t-T,-m_T]}^{\min[t-1,m_T]} g_{t-s}(\beta),\ (t = 1, ..., T).$$

The truncated kernel has spectral window generator $K(\lambda) = \pi^{-1}[(\sin \lambda)/\lambda]$. It is immediate that the induced kernel is the Bartlett kernel $k^*(x) = 1 - |x/2|$, $|x| \leq 2$, and 0, $|x| > 2$, as its spectral window generator is $K^*(\lambda) = (2\pi)^{-1}[(\sin \lambda/2)/(\lambda/2)]^2$. Hence, the optimal bandwidth parameter rate is $m_T = O(T^{1/3})$, see Andrews (1991, (5.3), p.830). Cf. Kitamura and Stutzer (1997).

Example 2.2: Parzen Kernel.

For the Bartlett kernel $k(x) = 1 - |x|$, $|x| \leq 1$, and 0, $|x| > 1$, $k_1 = 1$ and $k_2 = 2/3$. Hence, again defining the bandwidth parameter $S_T = (2m_T + 1)/2$,

$$g_{tT}(\beta) = (2m_T + 1)^{-1} \sum_{s=\max[t-T,-m_T]}^{\min[t-1,m_T]} \left(1 - \frac{2}{2m_T + 1}\right) g_{t-s}(\beta),\ (t = 1, ..., T).$$

The spectral window generator for the Parzen kernel is $K^*(\lambda) = (8\pi/3)^{-1}[(\sin \lambda/4)/(\lambda/4)]^2$. Therefore, it follows that the induced kernel corresponding to the Bartlett kernel is the Parzen kernel $k^*(x) = 1 - 6(x/2)^2 + 6\,|x/2|^3$, $|x| \leq 1$, $2(1 - |x/2|)^3$, $1 < |x| \leq 2$ and 0, $|x| > 2$. The optimal bandwidth parameter rate is $m_T = O(T^{1/5})$, see Andrews (1991, (5.3), p.830).

Example 2.3: Quadratic Spectral Kernel.

The quadratic spectral kernel is defined by

$$k^*(x) = \frac{3}{(ax)^2}\left(\frac{\sin ax}{ax} - \cos ax\right),\ x \in \mathcal{R},$$

where $a = 6\pi/5$; see Andrews (1991, (2.7), p.821). Hence, using 3.741.2 and 3.714.3, p.414, of Gradshteyn and Ryzhik (1980), hereafter GR, the spectral window generator of the quadratic spectral kernel is

$$K^*(\lambda) = \begin{cases} \frac{3}{4a}\left(1 - \left(\frac{\lambda}{a}\right)^2\right), & |\lambda| \leq a \\ 0, & |\lambda| > a \end{cases}.$$

Let the kernel $k(\cdot)$ have spectral window generator $K(\lambda) = K^*(\lambda)^{1/2}$, that is,

$$K(\lambda) = \begin{cases} \left(\frac{3}{4a}\right)^{1/2}\left(1-\left(\frac{\lambda}{a}\right)^2\right)^{1/2}, & |\lambda| \le a \\ 0, & |\lambda| > a \end{cases}.$$

Therefore, using the inverse Fourier transform and GR (3.752.2, p.419),

$$k(x) = \left(\frac{5\pi}{8}\right)^{1/2}\frac{1}{x}J_1\left(\frac{6\pi x}{5}\right), \; x \in \mathcal{R},$$

where the Bessel function, see GR (8.402, p.951),

$$J_\nu(z) = \frac{z^\nu}{2^\nu}\sum_{k=0}^{\infty}(-1)^k\frac{z^{2k}}{2^{2k}\Gamma(k+1)\Gamma(\nu+k+1)}.$$

Thus, $k_1 = (5\pi/2)^{1/2}$, GR (6.561.14, p.684), and $k_2 = 2\pi$, GR (6.574.2, p.692). The optimal bandwidth parameter rate is $S_T = O(T^{1/5})$, see Andrews (1991, (5.3), p.830).

3 Efficient Moment Estimation

Given the conditions on the function $\rho(\cdot)$ in Assumption 2.4 (a) and as the argument $k\hat{\lambda}'g_{tT}(\hat{\beta}) \in \mathcal{V}$ w.p.a.1, $(t = 1, ..., T)$, from Lemma A.4, the ratios

$$\pi_t(\hat{\beta}, \hat{\lambda}) = \frac{\rho_1(k\hat{\lambda}'g_{tT}(\hat{\beta}))}{\sum_{s=1}^T \rho_1(k\hat{\lambda}'g_{sT}(\hat{\beta}))}, \; (t = 1, ..., T), \tag{3.1}$$

may be thought of as implied probabilities being bounded between zero and unity w.p.a.1 and also summing to unity; cf. Back and Brown (1993). The ratios $\pi_t(\hat{\beta}, \hat{\lambda})$, $(t = 1, ..., T)$, are thus empirical measure counterparts to the expectation operator in (2.1) that ensure that the moment conditions are satisfied in the sample; see (2.7).[10] In contrast, the empirical measures $d\mu_t = T^{-1}$, $(t = 1, ..., T)$, from which the empirical distribution function (EDF) $\hat{\mu}(z) = \sum_{t=1}^T 1(z_t \le z)d\mu_t$ is constructed, where $1(\cdot)$ denotes an indicator function, underpin the calculation of sample mean-like quantities.[11]

[10] In a random sampling setting, Brown and Newey (2002) independently propose similar empirical measures to (3.1) in which the GMM estimator is substituted in (2.5) and, thus, (2.7) and optimisation is with respect to λ alone. See also NS, p.223.

[11] The empirical measures $d\mu_t$, $(t = 1, ..., T)$, are nonparametric maximum likelihood estimators which maximise the non-parametric log-likelihood $\sum_{t=1}^T \log d\mu_t$ subject to the constraints $0 < d\mu_t < 1$, $(t = 1, ..., T)$, and $\sum_{t=1}^T d\mu_t = 1$.

Let α_0 denote a r-vector of moments of interest derived from the stationary distribution of the process $\{z_t\}_{t=1}^{\infty}$. Suppose that $E[a(z, \beta_0)] = \alpha_0$, where $a(z, \beta)$ is a known vector of functions of the data observation vector z and parameter vector β. The information contained in the moment conditions $E[g_t(\beta_0)] = 0$ (2.1) may then be exploited using (3.1) to provide an efficient estimator of the moment vector α_0 which therefore dominates the EDF based estimator $\sum_{t=1}^{T} a(z_t, \hat{\beta})/T$.

Let $a_t(\beta) = a(z_t, \beta)$, $(t = 1, ..., T)$. We use a simple adaptation of Back and Brown's (1993) method for the GEL context; cf. Brown and Newey (1998) for i.i.d. data. Define the additional moment indicator vector $a(z, \beta) - \alpha$ with its smoothed counterpart as $a_{tT}(\beta) - k_{tT}\alpha$ where $a_{tT}(\beta) = \frac{1}{S_T}\sum_{s=t-T}^{t-1} k\left(\frac{s}{S_T}\right) a_t(\beta)$ and the normalisation factor $k_{tT} = \frac{1}{S_T}\sum_{s=t-T}^{t-1} k\left(\frac{s}{S_T}\right)$, $(t = 1, ..., T)$, cf. (2.2). We associate the auxiliary parameter vector φ with the smoothed moment indicator $a_{tT}(\beta) - k_{tT}\alpha$, $(t = 1, ..., T)$, and incorporate this indicator into the GEL criterion (2.5), i.e. $\hat{P}(\alpha, \beta, \varphi, \lambda) = \sum_{t=1}^{T}[\rho(k(\lambda' g_{tT}(\beta) + \varphi'(a_{tT}(\beta) - k_{tT}\alpha))) - \rho_0]/T$. From the first order conditions, cf. (2.9), optimisation of $\hat{P}(\alpha, \beta, \varphi, \lambda)$ over the parameter vectors (α, β) and auxiliary parameters (φ, λ) results in the additional auxiliary parameter φ being estimated as identically zero since the derivative matrix of the augmented smoothed moment indicator is block triangular between α and β. Therefore, the GEL estimators $\hat{\beta}$ and $\hat{\lambda}$ defined in section 2.3 are solutions to this optimisation problem. Moreover, the corresponding GEL estimator for α is given by

$$\hat{\alpha} = \frac{1}{\sum_{t=1}^{T} k_{tT}\pi_t(\hat{\beta}, \hat{\lambda})}\sum_{t=1}^{T}\pi_t(\hat{\beta}, \hat{\lambda}) a_{tT}(\hat{\beta}). \tag{3.2}$$

Let $A(\beta) = E[\partial a_t(\beta)/\partial\beta']$ and $\Xi(\beta) = \lim_{T\to\infty} var[T^{1/2}\hat{a}(\beta)]$, where $\hat{a}(\beta) = \sum_{t=1}^{T} a_t(\beta)/T$. Also define $A = A(\beta_0)$ and $\Xi = \Xi(\beta_0)$. The following assumption modifies Assumptions 2.3-2.5 appropriately for $a(z, \beta)$.

Assumption 3.1 **(a)** *$a_t(\beta)$ is continuous at each $\beta \in \mathcal{B}$ with probability one;* **(b)** *$E[\sup_{\beta\in\mathcal{B}} \|a_t(\beta)\|^{\alpha}] < \infty$ for some $\alpha > \max\left(4\nu, \frac{1}{\eta}\right)$;* **(c)** *$\Xi(\beta)$ is finite and p.d. for all $\beta \in \mathcal{B}$;* **(d)** *$a(\cdot, \beta)$ is differentiable in a neighborhood $\mathcal{N}$ of β_0 and $E[\sup_{\beta\in\mathcal{N}} \|\partial a_t(\beta)/\partial\beta'\|^{\alpha/(\alpha-1)}] < \infty$.*

The next result on the asymptotic properties of the GEL moment estimator $\hat{\alpha}$ (3.2) follows as a consequence.

Theorem 3.1 *(Consistency and Limit Distribution of the GEL Moment Estimator $\hat{\alpha}$.) If Assumptions 2.1-2.5 and 3.1 are satisfied, then $\hat{\alpha} \xrightarrow{p} \alpha_0$ and has limiting distribution given by*

$$T^{1/2}(\hat{\alpha} - \alpha_0) \xrightarrow{d} N(0, \Psi - BPB'),$$

where $\Psi = \Xi - AHB' - BH'A' + A\Sigma A'$ and $B = \sum_{s=-\infty}^{\infty} E[a_t(\beta_0)g_{t-s}(\beta_0)']$. The GEL moment estimator $\hat{\alpha}$ and $\hat{g}_T(\hat{\beta})$ (or $\hat{g}(\hat{\beta})$) are asymptotically uncorrelated.

It is clear both from the proof of Theorem 3.1 and the above derivation that the GEL moment estimator $\hat{\alpha}$ has an asymptotic variance identical to that of the efficient GMM estimator for α_0 based on the unconditional moment conditions (2.1) and $E[a(z, \beta_0)] = \alpha_0$. The GEL moment estimator $\hat{\alpha}$ thus efficiently incorporates the moment information (2.1). Under the conditions of Theorem 3.1, the sample average moment estimator $\hat{a}(\hat{\beta})$ has a limiting distribution described by $T^{1/2}(\hat{a}(\hat{\beta}) - \alpha_0) \xrightarrow{d} N(0, \Psi)$. Hence, the GEL moment estimator $\hat{\alpha}$ clearly dominates the EDF based $\hat{a}(\hat{\beta})$. Likewise, the conclusions of Theorem 3.1 hold without alteration for any first order equivalent estimator substituted for the GEL estimator $\hat{\beta}$, for example, a two-step efficient GMM estimator, where optimisation of (2.5) would now take place solely in terms of λ; cf. Brown and Newey (2002).

Consistent estimators for the additional components Ξ, A and B in the limiting variance matrix $\Psi - BPB'$ are straightforwardly obtained in a similar manner to those for Ω and G described in section 2.5. Let $\hat{g}_{tT} = g_{tT}(\hat{\beta})$, $\hat{a}_{tT} = a_{tT}(\hat{\beta})$, $\hat{A}_t = \partial a_t(\hat{\beta})/\partial\beta'$ and $\hat{A}_{tT} = \partial a_{tT}(\hat{\beta})/\partial\beta'$, $(t = 1, ..., T)$. Similarly to Theorem 2.4, $-(k_2/k_1^2)$ times the (φ, α)-block of the Hessian provides a consistent estimator for A. From UWL Lemma A.1, $\sum_{t=1}^{T} A_{tT}(\hat{\beta})/(Tk_1)$, as well as $\sum_{t=1}^{T} A_t(\hat{\beta})/T$, is also a consistent estimator of A. The (φ, φ) block of the Hessian of the optimised GEL criterion is $k^2 \sum_{t=1}^{T} \rho_2(k\hat{\lambda}'\hat{g}_{tT})(\hat{a}_{tT} - k_{tT}\hat{\alpha})(\hat{a}_{tT} - k_{tT}\hat{\alpha})'/T$, which as $S_T \sum_{t=1}^{T} \rho_2(k\hat{\lambda}'\hat{g}_{tT})(\hat{a}_{tT} - k_{tT}\hat{\alpha})(\hat{a}_{tT} - k_{tT}\hat{\alpha})'/T \xrightarrow{p} -k_2\Xi$, cf. the proof of Theorem 2.5, provides a basis for consistent estimation of Ξ. Likewise, $S_T \sum_{t=1}^{T}(\hat{a}_{tT} - k_{tT}\hat{\alpha})(\hat{a}_{tT} - k_{tT}\hat{\alpha})'/T \xrightarrow{p} k_2\Xi$. The (λ, φ)-block of the Hessian is given by $k^2 \sum_{t=1}^{T} \rho_2(k\hat{\lambda}'\hat{g}_{tT})(\hat{a}_{tT} - k_{tT}\hat{\alpha})\hat{g}_{tT}'/T$. Consistent estimators for B may be derived by noting $S_T \sum_{t=1}^{T} \rho_2(k\hat{\lambda}'\hat{g}_{tT})(\hat{a}_{tT} - k_{tT}\hat{\alpha})\hat{g}_{tT}'/T \overset{a}{=} S_T \sum_{t=1}^{T} \rho_2(k\hat{\lambda}'\hat{g}_{tT})\hat{a}_{tT}\hat{g}_{tT}'/T \xrightarrow{p} -k_2 B$ and $S_T \sum_{t=1}^{T}(\hat{a}_{tT} - k_{tT}\hat{\alpha})\hat{g}_{tT}'/T \overset{a}{=} S_T \sum_{t=1}^{T} \hat{a}_{tT}\hat{g}_{tT}'/T \xrightarrow{p} k_2 B$.

A special case of the above analysis concerns efficient estimation of the stationary distribution $\mu(z) = \mathcal{P}\{z_t \leq z\}$ of the process $\{z_t\}_{t=1}^{\infty}$. Let $a_t(\beta) = 1(z_t \leq z)$ and $1_{tT}(z) = \frac{1}{S_T} \sum_{s=t-T}^{t-1} k\left(\frac{s}{S_T}\right) 1(z_t \leq z)$, $(t = 1, ..., T)$, where $1(\cdot)$ denotes the indicator function. Here $\alpha_0 = \mu(z)$. Therefore, from (3.2), the GEL c.d.f estimator for $\mu(z)$ is

$$\hat{\mu}_T(z) = \frac{1}{\sum_{t=1}^{T} k_{tT}\pi_t(\hat{\beta}, \hat{\lambda})} \sum_{t=1}^{T} \pi_t(\hat{\beta}, \hat{\lambda}) 1_{tT}(z). \tag{3.3}$$

If $k(a) \geq 0$ for all a, the GEL c.d.f. estimator $\hat{\mu}_T(z)$ is a proper c.d.f.; that is, $0 \leq \hat{\mu}_T(z) \leq 1$ for all z and is increasing in z. More generally, $\hat{\mu}_T(z)$ may not be proper for particular realisations z_t, $(t = 1, ..., T)$ but is w.p.a.1; see the proof of Theorem 3.1. The next result follows immediately from Theorem 3.1.

Corollary 3.1 *(Limit Distribution of the GEL c.d.f. estimator $\hat{\mu}_T(z)$.) If Assumptions 2.1-2.5 are satisfied, $\hat{\mu}_T(z) \xrightarrow{p} \mu(z)$, where $\mu(z) = \mathcal{P}\{z_t \leq z\}$, and has limiting distribution given by*

$$T^{1/2} (\hat{\mu}_T(z) - \mu(z)) \xrightarrow{d} N(0, \sigma^2 - b'Pb),$$

where $\sigma^2 = \sum_{s=-\infty}^{\infty}(E[1(z_t \leq z)1(z_{t-s} \leq z)] - \mu(z)^2)$ and $b = \sum_{s=-\infty}^{\infty} E[1(z_t \leq z)g_{t-s}(\beta_0)]$.

The estimator $\hat{\mu}_T(z)$ is an efficient estimator for $\mu(z)$. Clearly $\hat{\mu}_T(z)$ dominates the EDF $\hat{\mu}(z)$ which has limiting distribution $T^{1/2} (\hat{\mu}(z) - \mu(z)) \xrightarrow{d} N(0, \sigma^2)$. If the GEL criterion chosen is that for the CUE, that is, if $\rho(\cdot)$ is quadratic, see section 2.2, the structure of the GEL c.d.f. estimator $\hat{\mu}_T(z)$ is very similar to that described by Back and Brown (1993). If the moment indicators $g_t(\beta_0)$, $(t = 1, 2, ...)$, are uncorrelated as occurs, for example, in the random sampling context, the kernel $k(\cdot)$ is the Dirac delta function, $k(a) = 1$ if $a = 0$ and 0 otherwise, in which case $\hat{\mu}_T(z)$ is identical in form to that given by Brown and Newey (2002, p.509) which forms the basis of their suggestion for bootstrapping GMM test statistics for over-identifying moment conditions.

4 Over-Identifying Moments

This section discusses alternative test statistics based on GEL for gauging the validity of the over-identifying moment conditions (2.1). The structure of these statistics is classical in construction, some of which resemble those described in Imbens, Spady and Johnson (1998), Kitamura and Stutzer (1997) and Smith (1997, 2000). As will be seen below, these statistics are first order equivalent to that suggested by Hansen (1982) based on the optimised GMM criterion, i.e.

$$\mathcal{J} = T\hat{g}(\hat{\beta})'\dot{\Omega}^{-1}\hat{g}(\hat{\beta}) \tag{4.1}$$

where $\dot{\Omega}$ is a preliminary p.s.d. consistent estimator for Ω and $\hat{\beta}$ is an efficient two-step GMM estimator. Under suitable conditions such as those described in section 2.4 $\mathcal{J}$ may be shown to be asymptotically chi-square with $(m-p)$ degrees of freedom.

Firstly, consider the likelihood ratio-like (LR) statistic based on the optimized GEL criterion (2.5)

$$\mathcal{LR} = 2(T/S_T)\hat{P}(\hat{\beta},\hat{\lambda})/(k_1^2/k_2). \tag{4.2}$$

Now, interpreting the GEL criterion as a quasi-likelihood, cf. Chesher and Smith (1997) and the discussion in section 2.2, $\hat{P}(\beta,0)=\rho(0)$ corresponds to the imposition of the parametric restriction that the auxiliary parameter $\lambda=0$. The hypothesis $\lambda=0$ may be regarded as the dual of the moment conditions $E[g_t(\beta_0)]=0$ (2.1). For the Cressie-Read family of power divergence criteria, this is explicitly the case as λ is a Lagrange multiplier which ensures that the moment conditions are satisfied in the sample, see (2.7). Thus, $\mathcal{LR}$ (4.2) is a likelihood ratio-like statistic for testing the hypothesis $\lambda=0$ which also directly examines the validity of the moment conditions (2.1).

The duality between over-identifying moments and the parametric restriction $\lambda=0$ suggests other classical-like statistics. A GEL LM-type statistic for (2.1) is defined directly in terms of the auxiliary parameter estimator $\hat{\lambda}$

$$\mathcal{LM} = (T/S_T^2)\hat{\lambda}'\hat{\Omega}\hat{\lambda}, \tag{4.3}$$

where $\hat{\Omega}$ is a p.s.d. consistent estimator for Ω, for example, as in section 2.5, with Ω a generalized inverse for the asymptotic variance matrix P of $(T/S_T^2)^{1/2}\hat{\lambda}$ given in Theorem 2.3. A score-like statistic encounters the difficulty that β is no longer identified if $\lambda = 0$ but this is simply circumvented by evaluating the scores (2.7) and (2.9) at $\lambda = 0$ and the GEL estimator $\hat{\beta}$ which yields $-(\hat{g}_T(\hat{\beta})', 0')'$. Hence, the GEL score statistic is defined as

$$\mathcal{S} = T\hat{g}_T(\hat{\beta})'\hat{\Omega}^{-1}\hat{g}_T(\hat{\beta})/(k_1^2); \tag{4.4}$$

cf. the optimal GMM statistic $\mathcal{J}$ (4.1).[12]

Theorem 4.1 *(Limit Distribution of the GEL statistics $\mathcal{LR}$, $\mathcal{LM}$ and $\mathcal{S}$ for Over-Identifying Moments.) If Assumptions 2.1-2.5 are satisfied, the GEL statistics $\mathcal{LR}$, $\mathcal{LM}$ and $\mathcal{S}$ are asymptotically equivalent and have a limiting distribution described by*

$$\mathcal{LR}, \mathcal{W}, \mathcal{S} \xrightarrow{d} \chi^2(m-p).$$

Because the GEL estimator $\hat{\beta}$ is first order equivalent to optimal two-step GMM estimators, it obeys in an asymptotic sense the corresponding GMM first order conditions. More precisely, $T^{1/2}\hat{g}_T(\hat{\beta}) \overset{a}{=} (k_1)T^{1/2}\hat{g}(\hat{\beta})$; see (2.4) and Lemma A.2. Hence, an equivalent score-type statistic may be based on $T\hat{g}(\hat{\beta})'\hat{\Omega}^{-1}\hat{g}(\hat{\beta})$, which is the GMM statistic $\mathcal{J}$ (4.1) using the GEL estimator $\hat{\beta}$. Moreover, the classical-type GEL statistics (4.2), (4.3), (4.4) and $\mathcal{J}$ are first order equivalent as $(T/S_T^2)^{1/2}\hat{\lambda} \overset{a}{=} -\Omega^{-1}T^{1/2}\hat{g}(\hat{\beta})$, i.e., $\mathcal{LR}$, $\mathcal{LM}$, $\mathcal{S} \overset{a}{=} \mathcal{J}$.

Although not discussed here, other first order equivalent tests based on the $\mathcal{C}(\alpha)$ principle may also be defined in a parallel fashion; cf. *inter alia* Neyman (1959) and Smith (1987).

5 Specification Tests

This section is concerned with tests for the validity of additional information on β_0. Let $\theta = (\alpha', \beta')'$ and $\theta_0' = (\alpha_0', \beta_0')'$ where α is a q-vector of additional parameters. To provide

[12]The optimised normalised form $T\hat{g}_T(\beta)'\hat{\Omega}_T(\beta)^{-}\hat{g}_T(\beta)/(k_1^2)$ of the alternative GMM criterion suggested in Smith (2005, section 3) provides another test of the moment conditions (2.1) which is first order equivalent to $\mathcal{LR}$, $\mathcal{LM}$ and $\mathcal{S}$ under the hypotheses of Theorem 4.1.

sufficient generality which covers both parametric hypotheses and additional moment conditions, we consider constraints in mixed form, see Gouriéroux and Monfort (1989),

$$E[h(z_t, \theta_0)] = 0,\ \ r(\theta_0) = 0, \tag{5.1}$$

where both the s-vector of moment indicators $h(\cdot, \cdot)$ and the r-vector of parametric constraints $r(\cdot)$ depend on α as well as β.[13]

Let $h_t(\theta) = h(z_t, \theta)$, $(t = 1, ..., T)$. A number of special cases are covered by (5.1). The exclusion of either α or β from $h_t(\cdot)$ is permitted. Moreover, the mixed form for the parametric function $r(\cdot)$ is sufficiently general to include other types of constraints as special cases; *viz.* freedom equation, $r(\theta_0) = \beta_0 - \beta(\alpha_0) = 0$ [Seber (1964)], constraint equation, $r(\theta_0) = (r_\alpha(\alpha_0)', r_\beta(\beta_0)')' = 0$ [Aitchison (1962) and Sargan (1980)], and restrictions in mixed implicit and constraint equation form, $r(\theta_0) = (r_\theta(\theta_0)', r_\alpha(\alpha_0)', r_\beta(\beta_0)')' = 0$ [Szroeter (1983)], encountered in simultaneous equations models. Furthermore, the constraints (5.1) and the test statistics defined below are easily adapted for either additional moment restrictions $E[h(z_t, \theta_0)] = 0$ or parametric restrictions $r(\theta_0) = 0$.

Following section 2.2, an appropriate GEL criterion similar in form to (2.5) which incorporates (5.1) is defined as

$$\hat{P}(\theta, \eta, \mu) = \sum_{t=1}^{T} [\rho(k(\varphi' q_{tT}(\theta) + k_1 \mu' r(\theta))) - \rho_0]/T, \tag{5.2}$$

where $\varphi = (\lambda', \psi')'$ and $q_{tT}(\theta) = (g_{tT}(\beta)', h_{tT}(\theta)')'$ with $h_{tT}(\theta) = \frac{1}{S_T} \sum_{s=t-T}^{t-1} k\left(\frac{s}{S_T}\right) h_{t-s}(\theta)$, $(t = 1, ..., T)$. The corresponding GEL and auxiliary saddle point estimators are denoted by $\tilde{\theta}$, $\tilde{\varphi}$ and $\tilde{\mu}$.

We need to modify Assumptions 2.3-2.5 appropriately for the results of this section. Let $q_t(\theta) = (g_t(\theta)', h_t(\theta)')'$, $(t = 1, ..., T)$, $\hat{q}(\theta) = \sum_{t=1}^{T} q_t(\theta)/T$ and

$$\Sigma(\theta) = \lim_{T \to \infty} var[T^{1/2} \hat{q}(\theta)].$$

We also define $\Sigma = \Sigma(\theta_0)$, $H_\alpha = E[\partial h(z_t; \theta_0)/\partial \alpha']$ and $R_\alpha = \partial r(\theta_0)/\partial \alpha'$.

[13] Without loss of generality, it is assumed that parametric restrictions in freedom equation form, $\delta_0 = \delta(\alpha_0)$, have been substituted out.

Assumption 5.1 **(a)** $\theta_0 = (\alpha_0', \beta_0')' \in \Theta$, $\Theta = \mathcal{A} \times \mathcal{B}$, *is the unique solution to* $E[q_t(\theta)] = 0$ *and* $r(\theta) = 0$; **(b)** $\mathcal{A}$ *and* $\mathcal{B}$ *are compact;* **(c)** $q_t(\theta)$ *and* $r(\theta)$ *are continuous at each* $\theta \in \Theta$ *with probability one;* **(d)** $E[\sup_{\theta \in \Theta} \|q_t(\theta)\|^\alpha] < \infty$ *for some* $\alpha > \max\left(4\nu, \frac{1}{\eta}\right)$; **(e)** $\Sigma(\theta)$ *is finite and p.d. for all* $\theta \in \Theta$.

Assumption 5.2 **(a)** $\rho(\cdot)$ *is twice continuously differentiable and concave on its domain, an open interval* $\mathcal{V}$ *containing* 0, $\rho_1 = \rho_2 = -1$; **(b)** $\varphi \in \Delta_T$ *where* $\Delta_T = \{\varphi : \|\varphi\| \leq D(T/S_T^2)^{-\zeta}\}$ *for some* $D > 0$ *with* $\frac{1}{2} > \zeta > \frac{1}{2\alpha\eta}$.

Assumption 5.3 **(a)** $\theta_0 \in int(\Theta)$; **(b)** $q(\cdot; \theta)$ *is differentiable in a neighborhood* $\mathcal{N}$ *of* θ_0 *and* $E[\sup_{\theta \in \mathcal{N}} \|\partial q_t(\theta)/\partial\theta'\|^{\alpha/(\alpha-1)}] < \infty$; **(c)** $r(\cdot)$ *is continuously differentiable in a neighborhood* $\mathcal{N}$ *of* θ_0 *and* $\sup_{\theta \in \mathcal{N}} \|\partial r(\theta)/\partial\theta'\| < \infty$; **(d)** $rank(G) = p$, $rank(R) = r$ *and* $rank((H_\alpha', R_\alpha')') = q$.

The rank conditions of Assumption 5.3 (d) are sufficient to guarantee the local independence of the constraints (2.1) and (5.1) and the local identifiability of θ_0. Furthermore, $(Q', R')'$ is f.c.r., which together with Assumption 5.1 (e), implies $Q'\Sigma^{-1}Q + R'R$ p.d.

As a preliminary, we firstly detail the limiting properties of the GEL and auxiliary parameter estimators $\tilde{\theta}$, $\tilde{\varphi}$ and $\tilde{\mu}$ mirroring Theorems 2.2 and 2.3.

Theorem 5.1 *(Consistency and Limiting Distribution of the GEL estimators* $\tilde{\theta}$, $\tilde{\eta}$ *and* $\tilde{\mu}$.*) If Assumptions 2.1, 2.2 and 5.1-5.3 are satisfied, then* $\tilde{\theta} \xrightarrow{p} \theta_0$, $\tilde{\varphi} \xrightarrow{p} 0$ *and* $\tilde{\mu} \xrightarrow{p} 0$ *and*

$$T^{1/2}(\tilde{\theta} - \theta_0) \xrightarrow{d} N(0, K),$$

$$(T/S_T^2)^{1/2} \begin{pmatrix} \tilde{\varphi} \\ \tilde{\mu} \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma^{-1} - \Sigma^{-1}QKQ'\Sigma^{-1} & -\Sigma^{-1}QMR'(RMR')^{-1} \\ -(RMR')^{-1}RMQ'\Sigma^{-1} & (RMR')^{-1} - I_r \end{pmatrix} \right),$$

where $M = (Q'\Sigma^{-1}Q + R'R)^{-1}$, $K = M - MR'(RMR')^{-1}RM$, $Q = E[\partial q_t(\theta_0)/\partial\theta']$ *and* $R = \partial r(\theta_0)/\partial\theta'$. *Moreover, the GEL estimator* $\tilde{\theta}$ *and auxiliary parameter estimators* $\tilde{\varphi}$ *and* $\tilde{\mu}$ *are asymptotically uncorrelated.*

As in section 4, classical-like statistics may be constructed for the additional moment conditions and parametric restrictions (5.1) by considering GEL-based tests for the parametric hypothesis $\psi = 0$ and $\mu = 0$ within the GEL criterion $\hat{P}(\theta, \varphi, \mu)$ (5.2). The approach due to Newey (1985b) would set up a conditional moment test for (5.1) based on the difference between the normalised optimised GMM criterion constructed from the sample moments $\hat{q}(\theta)$, using as metric the inverse of a consistent estimator for Σ, and $\mathcal{J}$ of (4.1); cf. (5.6) below.

Firstly, consider the difference of LR-like statistics based on the optimized GEL criteria (2.5) and (5.2)

$$\mathcal{LR}_a = 2(T/S_T)(\hat{P}(\tilde{\theta}, \tilde{\varphi}, \tilde{\mu}) - \hat{P}(\hat{\beta}, \hat{\lambda}))/(k_1^2/k_2). \tag{5.3}$$

Note that $\hat{P}(\tilde{\theta}, \tilde{\varphi}, \tilde{\mu}) = \hat{P}(\tilde{\theta}, \tilde{\varphi}, 0)$ as $r(\tilde{\theta}) = 0$. Similarly to the discussion below (4.2), the optimised criterion $\hat{P}(\hat{\beta}, \hat{\lambda})$ corresponds to the imposition of the parametric constraints $\psi = 0$ and $\mu = 0$. Thus, $\mathcal{LR}_a$ is a LR-like statistic for testing this parametric hypothesis which also directly examines the validity of (5.1).

Secondly, a LM-type statistic for $\psi = 0$ and $\mu = 0$ or (5.1) is defined in a standard way *via* the auxiliary parameter estimators $\tilde{\psi}$ and $\tilde{\mu}$; *viz.*

$$\mathcal{LM}_a = (T/S_T^2) \begin{pmatrix} \tilde{\psi} \\ \tilde{\mu} \end{pmatrix}' \left(S'_{\psi,\mu} \begin{pmatrix} \tilde{\Sigma} & \tilde{Q} & 0 \\ \tilde{Q}' & 0 & \tilde{R}' \\ 0 & \tilde{R} & 0 \end{pmatrix}^{-1} S_{\psi,\mu} \right)^{-1} \begin{pmatrix} \tilde{\psi} \\ \tilde{\mu} \end{pmatrix}, \tag{5.4}$$

where $S_{\psi,\mu}$ is a $(m+s+r+p+q, s+r)$ selection matrix such that $S'_{\psi,\mu}(\varphi', \theta', \mu')' = (\psi', \mu')'$, $\tilde{\Sigma}$ is a p.s.d. consistent estimator for Σ and $\tilde{Q}$ and $\tilde{R}$ are consistent estimators for Q and R defined similarly to those for Ω and G described in section 2.5 using the GEL and auxiliary parameter estimators $\tilde{\theta}$, $\tilde{\varphi}$ and $\tilde{\mu}$.

Thirdly, cf. section 4, α is no longer identified if $\psi = 0$ and $\mu = 0$. Evaluation of the scores for (5.2), cf. (2.7) and (2.9), at $\hat{\theta} = (\tilde{\alpha}', \hat{\beta}')'$, $\hat{\varphi} = S_g \hat{\lambda}$ and $\hat{\mu} = 0$ avoids this difficulty where S_g is a selection matrix such that $S'_g \varphi = \lambda$. That is, the score becomes $\sum_{t=1}^T \rho_1(k\hat{\lambda}' g_{tT}(\hat{\beta}))(h_{tT}(\hat{\theta})', k_1 r(\hat{\theta})')'$. Hence, as $\sum_{t=1}^T \rho_1(k\hat{\lambda}' g_{tT}(\hat{\beta})) g_{tT}(\hat{\beta}) = 0$ from (2.7), a GEL score-type statistic is given by

$$\mathcal{S}_a = (k_1)^{-2} T^{-1} \sum_{t=1}^T \rho_1(k\hat{\lambda}' g_{tT}(\hat{\beta})) \begin{pmatrix} q_{tT}(\hat{\theta}) \\ k_1 r(\hat{\theta}) \end{pmatrix}' \tag{5.5}$$

$$\times \begin{pmatrix} \tilde{\Sigma}^{-1} - \tilde{\Sigma}^{-1}\tilde{Q}\tilde{K}\tilde{Q}'\tilde{\Sigma}^{-1} & -\tilde{\Sigma}^{-1}\tilde{Q}\tilde{M}\tilde{R}'(\tilde{R}\tilde{M}\tilde{R}')^{-1} \\ -(\tilde{R}\tilde{M}\tilde{R}')^{-1}\tilde{R}\tilde{M}\tilde{Q}'\tilde{\Sigma}^{-1} & (\tilde{R}\tilde{M}\tilde{R}')^{-1} - I_r \end{pmatrix} \sum_{t=1}^{T} \rho_1(k\hat{\lambda}' g_{tT}(\hat{\beta})) \begin{pmatrix} q_{tT}(\hat{\theta}) \\ k_1 r(\hat{\theta}) \end{pmatrix},$$

where $\tilde{M} = (\tilde{Q}'\tilde{\Sigma}^{-1}\tilde{Q} + \tilde{R}'\tilde{R})^{-1}$ and $\tilde{K} = \tilde{M} - \tilde{M}\tilde{R}'(\tilde{R}\tilde{M}\tilde{R}')^{-1}\tilde{R}\tilde{M}$.

Therefore

Theorem 5.2 *(Limiting Distribution of GEL Statistics $\mathcal{LR}_a$, $\mathcal{LM}_a$ and $\mathcal{S}_a$ for Additional Moment Restrictions and Parametric Hypotheses.) Let Assumptions 2.1, 2.2 and 5.1-5.3 hold. Then the GEL statistics $\mathcal{LR}_a$, $\mathcal{LM}_a$ and $\mathcal{S}_a$ are asymptotically equivalent and have a limiting distribution described by*

$$\mathcal{LR}_a, \mathcal{LM}_a, \mathcal{S}_a \xrightarrow{d} \chi^2(s + r - q).$$

Let $\hat{q}_T(\theta) = \sum_{t=1}^{T} q_{tT}(\theta)/T$ and S_β denote a selection matrix such that $S_\beta'\theta = \beta$. From the Proofs of Theorems 2.3 and 3.1 in Appendix B,

$$\begin{aligned} &T^{-1/2}\sum_{t=1}^{T} \rho_1(k\hat{\lambda}' g_{tT}(\hat{\beta})) q_{tT}(\hat{\theta}) \stackrel{a}{=} -T^{1/2}\hat{q}_T(\tilde{\theta}) + \Sigma S_g \Omega^{-1} T^{1/2}\hat{g}_T(\hat{\beta}) \\ &-k_1 Q S_\beta T^{1/2}(\hat{\beta} - \tilde{\beta}), \\ &T^{-1/2}\sum_{t=1}^{T} \rho_1(k\hat{\lambda}' g_{tT}(\hat{\beta})) k_1 r(\hat{\theta}) \stackrel{a}{=} -k_1 R S_\beta T^{1/2}(\hat{\beta} - \tilde{\beta}). \end{aligned}$$

Therefore, as $(\Sigma^{-1} - \Sigma^{-1}QKQ'\Sigma^{-1})Q = \Sigma^{-1}QMR'(RMR')^{-1}R$ and $(RMR')^{-1}RMQ'\Sigma^{-1}Q = ((RMR')^{-1} - I_r)R$,

$$\begin{aligned} \mathcal{S}_a &\stackrel{a}{=} T(\hat{q}_T(\tilde{\theta}) - \Sigma S_g \Omega^{-1}\hat{g}_T(\hat{\beta}))'(\Sigma^{-1} - \Sigma^{-1}QKQ'\Sigma^{-1})(\hat{q}_T(\tilde{\theta}) - \Sigma S_g \Omega^{-1}\hat{g}_T(\hat{\beta}))/(k_1^2) \\ &\stackrel{a}{=} T(\hat{q}_T(\tilde{\theta}) - \Sigma S_g \Omega^{-1}\hat{g}_T(\hat{\beta}))'\Sigma^{-1}(\hat{q}_T(\tilde{\theta}) - \Sigma S_g \Omega^{-1}\hat{g}_T(\hat{\beta}))/(k_1^2) \\ &\stackrel{a}{=} T(\hat{q}_T(\tilde{\theta})'\tilde{\Sigma}^{-1}\hat{q}_T(\tilde{\theta}) - \hat{g}_T(\hat{\beta})'\hat{\Omega}^{-1}\hat{g}_T(\hat{\beta}))/(k_1^2) \\ &\stackrel{a}{=} T(\hat{q}(\tilde{\theta})'\tilde{\Sigma}^{-1}\hat{q}(\tilde{\theta}) - \hat{g}(\hat{\beta})'\hat{\Omega}^{-1}\hat{g}(\hat{\beta})), \end{aligned} \tag{5.6}$$

noting $T^{1/2}\hat{q}_T(\tilde{\theta}) \stackrel{a}{=} (I_{m+s} - QKQ'\Sigma^{-1})T^{1/2}\hat{q}_T(\theta_0)$, $T^{1/2}\hat{g}_T(\hat{\beta}) \stackrel{a}{=} (I_m - GH)T^{1/2}\hat{g}_T(\beta_0)$ and $Q'S_g\Omega^{-1}T^{1/2}\hat{g}_T(\hat{\beta}) \stackrel{a}{=} 0$. Hence, $\mathcal{S}_a$ is first order equivalent to optimal GMM statistics for (5.1) based on the GEL estimators $(\hat{\beta}, \hat{\lambda})$ and $(\tilde{\theta}, \tilde{\varphi}, \tilde{\mu})$; cf. Newey (1985b). Although the first and second expressions for $\mathcal{S}_a$ in (5.6) are p.s.d., neither the third nor the last need be even if the estimator $\hat{\Omega}$ is the (m, m) top left diagonal block of $\tilde{\Sigma}$.

Asymptotically equivalent classical-like statistics are straightforwardly defined for tests of the full vector of constraints, (2.1) and (5.1), which have a limiting chi-squared distribution with $(m+s+r)-(p+q)$ degrees of freedom; cf. section 4.

Other statistics asymptotically equivalent to the above GEL-based statistics may be defined. For example, a minimum chi-squared statistic is given by

$$\mathcal{MC}_a = (T/S_T^2)(\tilde{\varphi}-\hat{\varphi})'\tilde{\Sigma}(\tilde{\varphi}-\hat{\varphi}). \tag{5.7}$$

Asymptotically equivalent p.s.d. score-type statistics which only use the GEL estimators $\tilde{\theta}$, $\tilde{\varphi}$ and $\tilde{\mu}$ are

$$\begin{aligned} & T\hat{q}_T(\tilde{\theta})'(\tilde{\Sigma}^{-1} - S_g\tilde{P}S_g')\tilde{\Sigma}(\tilde{\Sigma}^{-1} - S_g\tilde{P}S_g')\hat{q}_T(\tilde{\theta})/(k_1^2) \\ = \ & T(\hat{q}_T(\tilde{\theta})'\tilde{\Sigma}^{-1}\hat{q}_T(\tilde{\theta}) - \hat{g}_T(\tilde{\beta})'\tilde{P}\hat{g}_T(\tilde{\beta}))/(k_1^2) \\ & \stackrel{a}{=} T(\hat{q}(\tilde{\theta})'\tilde{\Sigma}^{-1}\hat{q}(\tilde{\theta}) - \hat{g}(\tilde{\beta})'\tilde{P}\hat{g}(\tilde{\beta})), \end{aligned}$$

where $\tilde{P} = \tilde{\Omega}^{-1} - \tilde{\Omega}^{-1}\tilde{G}\tilde{\Sigma}\tilde{G}'\tilde{\Omega}$, $\tilde{\Sigma} = (\tilde{G}'\tilde{\Omega}^{-1}\tilde{G})^{-1}$, $\tilde{G}$ a consistent estimator for G based on $\tilde{\theta}$, $\tilde{\varphi}$ and $\tilde{\mu}$ and $\tilde{\Omega}$ is the (m,m) top left diagonal block of $\tilde{\Sigma}$.

Similarly to section 2.5, let $\tilde{\Sigma}(\theta) = S_T\sum_{t=1}^T q_{tT}(\theta)q_{tT}(\theta)'/(Tk_2)$. Two-step efficient, iterated and continuous updating estimators are provided by the alternative GMM Lagrangean criterion $\hat{q}_T(\theta)'\tilde{\Sigma}(\theta)^{-1}\hat{q}_T(\theta) - \mu' r(\theta)$. From the first order conditions, under the hypotheses of Theorem 5.1, $T^{1/2}(\tilde{\theta}-\theta_0) \stackrel{a}{=} -KQ'\Sigma^{-1}T^{1/2}\hat{q}_T(\theta_0)$ and $T^{1/2}\tilde{\mu} \stackrel{a}{=} (RMR')^{-1}RMQ'\Sigma^{-1}T^{1/2}\hat{q}_T(\theta_0)$ where $\tilde{\theta}$ and $\tilde{\mu}$ denote the corresponding GMM and Lagrange multiplier estimators; cf. proof of Theorem 5.1 in Appendix B. These relationships yield the limiting results $T^{1/2}(\tilde{\theta}-\theta_0) \xrightarrow{d} N(0,(k_1^2)K)$, $T^{1/2}\tilde{\mu} \xrightarrow{d} N(0,(k_1^2)((RMR')^{-1}-I_r))$ and $\tilde{\theta}$ and $\tilde{\mu}$ asymptotically uncorrelated. Under the hypotheses of Theorem 5.2, cf. (5.6), the normalised difference of GMM criteria $T(\hat{q}_T(\tilde{\theta})'\tilde{\Sigma}(\theta)^{-1}\hat{q}_T(\tilde{\theta})-\hat{g}_T(\hat{\beta})'\tilde{\Omega}(\beta)^{-1}\hat{g}_T(\hat{\beta}))/(k_1^2)$ based on a common $T^{1/2}$-consistent estimator for θ in $\tilde{\Sigma}(\theta)$ and $\tilde{\Omega}(\beta)$, the (m,m) top left diagonal block of $\tilde{\Sigma}(\theta)$, is p.s.d. and asymptotically equivalent to $\mathcal{LR}_a$, $\mathcal{LM}_a$, $\mathcal{S}_a$ and $\mathcal{MC}_a$; cf. Newey (1985b) and Smith (2005).

The proof of Theorem 5.2 demonstrates the GEL-based statistics $\mathcal{LR}_a$, $\mathcal{LM}_a$, $\mathcal{S}_a$ and $\mathcal{MC}_a$ are all first order equivalent. It also immediately follows because $\mathcal{LR}_a$ is expressed as the difference of likelihood ratio-like statistics that equivalent statistics may be obtained as the difference of appropriately defined Lagrange multiplier-like and score-like statistics. However, these statistics may not possess positive support although common estimator choices for Σ, Q and R may ameliorate this problem. Furthermore, given the discussion in section 4 concerning the equivalence of those GEL-based statistics with the GMM statistic $\mathcal{J}$ of (4.1), the statistics of this section are equivalent to the difference of estimated GMM criteria and, as noted above, to the GMM statistic for the additional constraints (5.1); cf. (5.6). A final point is that, if the moment and parametric constraints, (2.1) and (5.1), hold, all of the statistics of this section are asymptotically independent of the over-identifying moment tests of section 4, a property also displayed by classical tests for a sequence of nested hypotheses; see *inter alia* Aitchison (1962) and Sargan (1980).

6 Summary and Conclusions

This paper analyses a class of GEL criteria for the one-step estimation of models specified by moment conditions defined in terms of weakly dependent data. This class includes EL, ET, CUE and Cressie-Read power divergence criteria as special cases. The resultant GEL estimators are asymptotically equivalent to two-step efficient GMM estimators. An efficient moment estimator is also described, a special case of which is the stationary distribution of the data. The latter application may potentially be of use in the development of tests for the distributional form in fully parametric models which also imply the moment conditions underpinning the GEL criterion. The structure of GEL criteria parallels conventional likelihood. Thus, likelihood ratio-, Lagrange multiplier- and score-like statistics are obtained for testing over-identifying moment conditions. This analysis is extended to tests of a combination of additional moment conditions and parametric constraints expressed in mixed form which are sufficiently general to admit as special cases most forms of moment and parametric restrictions of practical interest.

The finite sample behaviour of GEL estimators and GEL-based statistics and choices of GEL function have not been studied in this paper. The GEL implied probabilities offer the possibility of improved inference with weakly dependent data using bootstrap samples along the lines suggested by Brown and Newey (2002) in the random sampling context. However, the exploration of this topic lies outside the scope of the paper but is the subject of current research. Given the parallels with conventional likelihood, Edgeworth expansions offer a feasible method for the improvement of the quality of first order asymptotic approximations and the ability to detect circumstances in which these approximations are likely to be poor. This research agenda is also currently under investigation.

Appendix A: Preliminary Lemmata

Throughout these Appendices, C will denote a generic positive constant that may be different in different uses, and C, CS, H, J and T Chebychev, Cauchy-Schwarz, Hölder, Jensen and triangle inequalities respectively. Unless otherwise stated, UWL and CLT refer to Lemma A.1 and Lemma A.2 below respectively.

Let $k_T(a) = k((s-1)/S_T)$, $(s-1)/S_T \leq a < s/S_T$, if $s \leq 0$, $k(s/S_T)$, $(s-1)/S_T < a \leq s/S_T$, if $s > 0$. Also let $[\cdot]$ denote the integer part of $\cdot$.

Let $g(\beta) = E[g_t(\beta)]$.

Lemma A.1 *(UWL) If Assumptions 2.1, 2.2 and 2.3 (b)-(d) with* $\alpha = 1$ *are satisfied, then*

$$\sup_{\beta \in B} \|\hat{g}_T(\beta) - k_1 g(\beta)\| = o_p(1).$$

Proof: By T

$$\sup_{\beta \in \mathcal{B}} \|\hat{g}_T(\beta) - k_1 g(\beta)\| \leq \sup_{\beta \in \mathcal{B}} \|\hat{g}_T(\beta) - E[\hat{g}_T(\beta)]\| + \sup_{\beta \in \mathcal{B}} \|E[\hat{g}_T(\beta)] - k_1 g(\beta)\|. \qquad (A.1)$$

Consider the first term in eq. (A.1)

$$\begin{aligned} \hat{g}_T(\beta) - E[\hat{g}_T(\beta)] &= T^{-1} \sum_{t=1}^{T} \frac{1}{S_T} \sum_{s=t-T}^{t-1} k\left(\frac{s}{S_T}\right) (g_{t-s}(\beta) - E[g_{t-s}(\beta)]) \qquad (A.2) \\ &= T^{-1} \sum_{t=1}^{T} (g_t(\beta) - g(\beta)) \frac{1}{S_T} \sum_{s=1-t}^{T-t} k\left(\frac{s}{S_T}\right). \end{aligned}$$

Now,

$$\left| \frac{1}{S_T} \sum_{s=1-t}^{T-t} k\left(\frac{s}{S_T}\right) \right| \leq \frac{1}{S_T} \sum_{s=1-T}^{T-1} \left| k\left(\frac{s}{S_T}\right) \right|. \qquad (A.3)$$

Using the change of variable $s = [S_T a]$, as $T/S_T \to \infty$ by Assumption 2.2 (a),

$$\frac{1}{S_T} \sum_{s=1-T}^{T-1} \left| k\left(\frac{s}{S_T}\right) \right| \leq \lim_{T \to \infty} \frac{1}{S_T} \sum_{s=1-T}^{T-1} \left| k\left(\frac{s}{S_T}\right) \right| \qquad (A.4)$$

$$
\begin{aligned}
&= \lim_{T\to\infty} \int_{(1-T)/S_T}^{(T-1)/S_T} |k_T(a)|\, da + \lim_{T\to\infty} \frac{1}{S_T} |k(0)| \\
&\leq \lim_{T\to\infty} \int_{(1-T)/S_T}^{(T-1)/S_T} \bar{k}(a) da + o(1) \\
&= \int_{-\infty}^{\infty} \bar{k}(a) da + o(1).
\end{aligned}
$$

Then, by Assumption 2.2 (c), from eqs. (A.3) and (A.4),

$$
\left| \frac{1}{S_T} \sum_{s=1-t}^{T-t} k\left(\frac{s}{S_T}\right) \right| = O(1) \tag{A.5}
$$

uniformly t. Hence, we may rewrite eq. (A.2) as

$$
\hat{g}_T(\beta) - E[\hat{g}_T(\beta)] = O(1)(\hat{g}(\beta) - g(\beta)).
$$

Using UWL Lemma 2.4, p.2129, of Newey and McFadden (1994) for stationary and mixing (and, thus, ergodic) processes, $\sup_{\beta\in\mathcal{B}} \|\hat{g}(\beta) - g(\beta)\| \xrightarrow{p} 0$ by Assumptions 2.1 and 2.3 (d). Therefore,

$$
\sup_{\beta\in\mathcal{B}} \|\hat{g}_T(\beta) - E[\hat{g}_T(\beta)]\| \xrightarrow{p} 0. \tag{A.6}
$$

For the second term in eq. (A.1), we firstly note that

$$
\hat{g}_T(\beta) = \frac{1}{S_T} \sum_{s=1-T}^{T-1} k\left(\frac{s}{S_T}\right) T^{-1} \sum_{t=\max[1,1-s]}^{\min[T,T-s]} g_t(\beta). \tag{A.7}
$$

Next, using the stationarity of $\{g_t(\beta)\}_{t=1}^{\infty}$, by Lemma C.1

$$
\begin{aligned}
E[\hat{g}_T(\beta)] &= \frac{1}{S_T} \sum_{s=1-T}^{T-1} \left(1 - \frac{|s|}{T}\right) k\left(\frac{s}{S_T}\right) E[g_t(\beta)] \qquad (A.8) \\
&= \left(\frac{1}{S_T} \sum_{s=1-T}^{T-1} k\left(\frac{s}{S_T}\right) + o(1) \right) g(\beta)
\end{aligned}
$$

where the remainder term is uniform s. From eq. (A.8), by Assumption 2.3 (d),

$$
\|E[\hat{g}_T(\beta)] - k_1 g(\beta)\| \leq \left| \frac{1}{S_T} \sum_{s=1-T}^{T-1} k\left(\frac{s}{S_T}\right) - k_1 \right| \sup_{\beta\in\mathcal{B}} \|g(\beta)\| + o(1).
$$

Now, using the change of variable $s = [S_T a]$,

$$
\begin{aligned}
\lim_{T\to\infty} \frac{1}{S_T} \sum_{s=1-T}^{T-1} k\left(\frac{s}{S_T}\right) &= \lim_{T\to\infty} \int_{(1-T)/S_T}^{(T-1)/S_T} k_T(a) da + \frac{1}{S_T} k(0) \qquad (A.9) \\
&= \int_{-\infty}^{\infty} k(a) da + o(1) = k_1 + o(1).
\end{aligned}
$$

Therefore, by Assumption 2.3 (d),

$$\begin{aligned}\|E[\hat{g}_T(\beta)] - k_1 g(\beta)\| &\leq o(1)E[\sup_{\beta\in\mathcal{B}} \|g_t(\beta)\|] + o(1) \\ &= o(1)\end{aligned} \tag{A.10}$$

uniformly β. The conclusion follows from eqs. (A.6) and (A.10). ■

Let $g_t^*(\beta) = g_t(\beta) - g(\beta)$, $g_{tT}^*(\beta) = \frac{1}{S_T}\sum_{s=t-T}^{t-1} k\left(\frac{s}{S_T}\right) g_{t-s}^*(\beta)$, $(t = 1, ..., T)$. Write $g_t^* = g_t^*(\beta)$ and $g_{tT}^* = g_{tT}^*(\beta)$, $(t = 1, ..., T)$.

Lemma A.2 *(CLT) If Assumptions 2.1, 2.2 and 2.3 (b)-(e) are satisfied,*

$$T^{1/2}[\hat{g}_T(\beta) - E[\hat{g}_T(\beta)]] \xrightarrow{d} N(0, (k_1^2)\Omega(\beta)).$$

Proof: Let $\hat{g}_T^*(\beta) = T^{-1}\sum_{t=1}^T g_{tT}^*(\beta)$. Similarly to eq. (A.7), the normalised sample average

$$T^{1/2}\hat{g}_T^*(\beta) = \frac{1}{S_T}\sum_{s=1-T}^{T-1} k\left(\frac{s}{S_T}\right) T^{-1/2} \sum_{t=\max[1,1-s]}^{\min[T,T-s]} g_t^*.$$

The difference between $\sum_{t=\max[1,1-s]}^{\min[T,T-s]} g_t^*$ and $\hat{g}^*(\beta) = \sum_{t=1}^T g_t^*$ consists of $|s|$ terms. By C, using White (1984, Lemma 6.19, p.153),

$$\begin{aligned}\mathcal{P}\left\{\left|\sum_{t=1}^{|s|} g_t^*/T\right| \geq \varepsilon\right\} &\leq E\left[\left|\sum_{t=1}^{|s|} g_t^*\right|^2\right]/(T\varepsilon)^2 \\ &= |s|\, O(T^{-2})\end{aligned}$$

where the $O(T^{-2})$ term is independent of s. Therefore, by Lemma C.1 and eq. (A.9),

$$\begin{aligned}T^{1/2}\hat{g}_T^*(\beta) &= \frac{1}{S_T}\sum_{s=1-T}^{T-1} k\left(\frac{s}{S_T}\right) T^{1/2}(\hat{g}^*(\beta) + |s|\, O_p(T^{-2})) \\ &= \frac{1}{S_T}\sum_{s=1-T}^{T-1} k\left(\frac{s}{S_T}\right) T^{1/2}\hat{g}^*(\beta) + O_p(T^{-1/2}) \\ &= (k_1 + o(1))T^{1/2}\hat{g}^*(\beta) + O_p(T^{-1/2}) \\ &\xrightarrow{d} N(0, (k_1^2)\Omega(\beta))\end{aligned} \tag{A.11}$$

using, for example, CLT Theorem 5.19, p.124, of White (1984). ■

Let

$$\hat{\Omega}_T^*(\beta) = \left(\frac{1}{S_T} \sum_{t=1-T}^{T-1} k\left(\frac{t}{S_T}\right)^2 \right)^{-1} S_T \sum_{t=1}^{T} g_{tT}^*(\beta) g_{tT}^*(\beta)'/T.$$

Lemma A.3 *If Assumptions 2.1, 2.2 and 2.3 (d) and (e) are satisfied, then* $\hat{\Omega}_T^*(\beta) \xrightarrow{p} \Omega(\beta)$.

Proof: The numerator of $\hat{\Omega}_T^*(\beta)$ may be written as

$$\begin{aligned} S_T \sum_{t=1}^{T} g_{tT}^* g_{tT}^{*\prime}/T &= \sum_{t=1}^{T} \frac{1}{S_T} \sum_{s=1}^{T} k\left(\frac{t-s}{S_T}\right) g_s^* \sum_{r=1}^{T} k\left(\frac{t-r}{S_T}\right) g_r^{*\prime}/T \\ &= \sum_{r=1}^{T} \frac{1}{S_T} \sum_{s=1-r}^{T-r} g_{r+s}^* g_r^{*\prime} \left[T^{-1} \sum_{t=1}^{T} k\left(\frac{t-(s+r)}{S_T}\right) k\left(\frac{t-r}{S_T}\right) \right] \\ &= \sum_{s=1-T}^{T-1} T^{-1} \sum_{r=\max[1,1-s]}^{\min[T,T-s]} g_{r+s}^* g_r^{*\prime} \left[\frac{1}{S_T} \sum_{t=1-r}^{T-r} k\left(\frac{t-s}{S_T}\right) k\left(\frac{t}{S_T}\right) \right] \\ &= \sum_{s=1-T}^{T-1} \left[\frac{1}{S_T} \sum_{t=\max[1-T,1-T+s]}^{\min[T-1,T-1+s]} k\left(\frac{t-s}{S_T}\right) k\left(\frac{t}{S_T}\right) \right] T^{-1} \sum_{r=\max[1,1-s,1-t]}^{\min[T,T-s,T-t]} g_{r+s}^* g_r^{*\prime}. \end{aligned}$$

Define the infeasible sample autocovariance estimators $C_T^*(s) = \sum_{r=\max[1,1-s]}^{\min[T,T-s]} g_{r+s}^* g_r^{*\prime}/T$, $C_T^*(-s) = C_T^*(s)'$, $(s = 1-T, ..., T-1)$. The difference between $\sum_{r=\max[1,1-s,1-t]}^{\min[T,T-s,T-t]} g_{r+s}^* g_r^{*\prime}/T$ and $C_T^*(s)$ consists of no more than $|t|$ terms. By C, using White (1984, Lemma 6.19, p.153),

$$\begin{aligned} \mathcal{P}\left\{ \left| \frac{1}{T} \sum_{r=1}^{|t|} (g_{r+s}^* g_r^{*\prime} - \Gamma^*(s)) \right| \geq \varepsilon \right\} &\leq E\left[\left| \sum_{r=1}^{|t|} (g_{r+s}^* g_r^{*\prime} - \Gamma^*(s)) \right|^2 \right] /(T\varepsilon)^2 \\ &= |t|\, O(T^{-2}) \end{aligned}$$

uniformly s and the $O(T^{-2})$ term is independent of t, where $\Gamma^*(s) = E[g_{r+s}^* g_r^{*\prime}]$. Therefore,

$$\left\| \frac{1}{T} \sum_{r=\max[1,1-s,1-t]}^{\min[T,T-s,T-t]} g_{r+s}^* g_r^{*\prime} - C_T^*(s) \right\| \leq \frac{|t|}{T} \|\Gamma^*(s)\| + |t|\, O_p(T^{-2})$$

uniformly s. Now, by Lemma C.1,

$$\lim_{T\to\infty}\left|\frac{1}{S_T}\sum_{t=\max[1-T,1-T+s]}^{\min[T-1,T-1+s]}\frac{|t|}{T}k\left(\frac{t-s}{S_T}\right)k\left(\frac{t}{S_T}\right)\right| \leq \lim_{T\to\infty}\frac{1}{S_T}\sum_{t=1-T}^{T-1}\frac{|t|}{T}\left|k\left(\frac{t-s}{S_T}\right)k\left(\frac{t}{S_T}\right)\right|$$
$$\leq \sup_a \bar{k}(a)\lim_{T\to\infty}\frac{1}{S_T}\sum_{t=1-T}^{T-1}\frac{|t|}{T}\left|k\left(\frac{t}{S_T}\right)\right| = 0.$$

Hence, by Lemma C.2, as $\lim_{T\to\infty}\sum_{s=1-T}^{T-1}\|\Gamma^*(s)\| < \infty$ [see Assumption A, p.823, and Lemma 1, p.824, in Andrews (1991)],

$$\begin{aligned}\hat{\Omega}_T(\beta) &= \sum_{s=1-T}^{T-1}\left[\frac{1}{S_T}\sum_{t=\max[1-T,1-T+s]}^{\min[T-1,T-1+s]}k\left(\frac{t-s}{S_T}\right)k\left(\frac{t}{S_T}\right)\right]C_T^*(s)/k_2 + o_p(1)\\ &= \sum_{s=1-T}^{T-1}(k^*\left(\frac{s}{S_T}\right) + o(1))C_T^*(s) + o_p(1)\\ &= \sum_{s=1-T}^{T-1}k^*\left(\frac{s}{S_T}\right)C_T^*(s) + o_p(1),\end{aligned}$$

where the remainder terms are uniform in s, since $\lim_{T\to\infty} E[\sum_{s=1-T}^{T-1}C_T^*(s)] = \Omega(\beta)$ and $\lim_{T\to\infty} var[\sum_{s=1-T}^{T-1}C_T^*(s)] = O(1)$ by standard results on the inconsistency of the periodogram. Therefore, as $k^*(\cdot) \in \mathcal{K}_2$ by Lemma C.3, it follows that $\hat{\Omega}_T(\beta) \xrightarrow{p} \Omega(\beta)$. ■

Let $b_t = \sup_{\beta\in\mathcal{B}}\|g_t(\beta)\|$.

Lemma A.4 *If Assumptions 2.1 and 2.3 (d) are satisfied then* $\sup_{\beta\in\mathcal{B},\lambda\in\Lambda_T,1\leq t\leq T}|\lambda' g_{tT}(\beta)| \xrightarrow{p} 0$. *Also w.p.a.1* $\Lambda_T \subseteq \hat{\Lambda}_T(\beta)$ *where* $\hat{\Lambda}_T(\beta) = \{\lambda : k\lambda' g_{tT}(\beta) \in \mathcal{V},\ (t = 1, ..., T)\}$.

Proof: By Assumption 2.3 (d) it follows by M that $\max_{1\leq t\leq T} b_t = O_p(T^{1/\alpha})$. Then by CS,

$$\sup_{\beta\in\mathcal{B},\lambda\in\Lambda_T,1\leq t\leq T}|\lambda' g_{tT}(\beta)| \leq D(T/S_T^2)^{-\zeta}\sup_{\beta\in\mathcal{B},1\leq t\leq T}\|g_{tT}(\beta)\|.$$

From eq. (A.5)

$$\begin{aligned}\sup_{\beta\in\mathcal{B}}\|g_{tT}(\beta)\| &\leq b_t\frac{1}{S_T}\sum_{s=t-T}^{t-1}\left|k\left(\frac{s}{S_T}\right)\right|\\ &\leq b_t O(1)\end{aligned}$$

where the remainder term is uniform t. Therefore,

$$\sup_{\beta \in \mathcal{B}, \lambda \in \Lambda_T, 1 \le t \le T} |\lambda' g_{tT}(\beta)| \le (T/S_T^2)^{-\zeta} O_p(T^{1/\alpha}) \xrightarrow{p} 0.$$

giving the first conclusion. W.p.a.1 $k\lambda' g_{tT}(\beta) \in \mathcal{V}$ for all $\beta \in \mathcal{B}$ and $\lambda \in \Lambda_T$. ■

We now give two preliminary lemmas which will prove useful in the proofs of Theorems 2.2 and 2.3.

Lemma A.5 *Let Assumptions 2.1-2.4 be satisfied. Then there is a C such that w.p.a.1,*

$$\frac{1}{S_T} \sup_{\lambda \in \Lambda_T} \hat{P}(\beta_0, \lambda) \le C \|\hat{g}_T(\beta_0)\|^2.$$

Proof: Let $g_{tT} = g_{tT}(\beta_0)$, $\bar{\lambda} = \arg\max_{\lambda \in \Lambda_T} \hat{P}(\beta_0, \lambda)$ and $\dot{\lambda} = \tau \bar{\lambda}$, $0 \le \tau \le 1$. Then, by Lemma A.4,

$$\max_{1 \le t \le T} |\rho_2(k\dot{\lambda}' g_{tT}) - \rho_2(0)| \xrightarrow{p} 0.$$

Hence, by Lemma A.3, $S_T \sum_{t=1}^T \rho_2(k\dot{\lambda}' g_{tT}) g_{tT} g_{tT}'/T \xrightarrow{p} -k_2 \Omega$. It then follows by Ω p.d. that w.p.a.1 $S_T \sum_{t=1}^T \rho_2(k\dot{\lambda}' g_{tT}) g_{tT} g_{tT}'/T \le -C I_m$ in the p.s.d. sense.

By a second-order Taylor expansion with Lagrange remainder, we have w.p.a.1,

$$\begin{aligned}
\frac{1}{S_T} \sup_{\lambda \in \Lambda_T} \hat{P}(\beta_0, \lambda) &= \frac{1}{S_T} \hat{P}(\beta_0, \bar{\lambda}) \\
&= -k(\bar{\lambda}/S_T)' \hat{g}_T(\beta_0) + k^2 (\bar{\lambda}/S_T)' \left[S_T \sum_{t=1}^T \rho_2(\dot{\lambda}' g_{tT}) g_{tT} g_{tT}'/T \right] (\bar{\lambda}/S_T)/2 \\
&\le -k(\bar{\lambda}/S_T)' \hat{g}_T(\beta_0) - C k^2 (\bar{\lambda}/S_T)' (\bar{\lambda}/S_T)/2 \\
&\le \sup_{\lambda \in \mathcal{R}^m} [-\lambda' \hat{g}_T(\beta_0) - C\lambda'\lambda/2] = C \|\hat{g}_T(\beta_0)\|^2.
\end{aligned}$$

■

Lemma A.6 *If Assumptions 2.1-2.4 hold then $\hat{\beta} \xrightarrow{p} \beta_0$ and $\left\| \hat{g}_T(\hat{\beta}) \right\| = O_p(T^{-1/2})$.*

Proof: Let $\hat{g}_{tT} = g_{tT}(\hat{\beta})$ and $\delta_T = D(T/S_T^2)^{-\zeta}$ for ζ and D as in Assumption 2.4 (b). Also let $\bar{\lambda} = -\hat{g}_T(\hat{\beta})\delta_T/\|\hat{g}_T(\hat{\beta})\|$. So $\bar{\lambda} \in \Lambda_T$. Write $g_{tT} = g_{tT}(\beta)$, $(t = 1, ..., T)$.

Now, $g_{tT}g_{tT}' = g_{tT}^* g_{tT}^{*\prime} + g_{tT}^* E[g_{tT}'] + E[g_{tT}]g_{tT}^{*\prime} + E[g_{tT}]E[g_{tT}']$, $(t = 1, ..., T)$. By Lemma A.3, $S_T \sum_{t=1}^T g_{tT}^* g_{tT}^{*\prime}/T \xrightarrow{p} k_2\Omega(\beta)$. From eq. (A.5), by Assumption 2.3 (d), $E[g_{tT}] = E[g_t(\beta)]O(1) = O(1)$ uniformly β and t. Using CLT, $S_T \sum_{t=1}^T g_{tT}^* E[g_{tT}']/T = (T/S_T^2)^{-1/2}T^{1/2}\hat{g}_T^*(\beta)O(1) = o_p(1)$ by Assumption 2.2 (a) where $\hat{g}_T^*(\beta) = \sum_{t=1}^T g_{tT}^*/T$. Similarly, $\sum_{t=1}^T E[g_{tT}]E[g_{tT}']/T = g(\beta)g(\beta)'O(1) = O(1)$. Hence, $\sum_{t=1}^T \hat{g}_{tT}\hat{g}_{tT}'/T = O_p(1)$. Let $\dot{\lambda} = \tau\bar{\lambda}$, $0 \leq \tau \leq 1$. It then follows as $\max_{\lambda\in\Lambda_T, 1\leq t\leq T} |\lambda'\hat{g}_{tT}| \xrightarrow{p} 0$ from Lemma A.4 that $\sum_{t=1}^T [\rho_2(k\dot{\lambda}'\hat{g}_{tT}) - \rho_2(0)]\hat{g}_{tT}\hat{g}_{tT}'/T \xrightarrow{p} 0$. Therefore, w.p.a.1 $\sum_t \rho_2(k\dot{\lambda}'\hat{g}_{tT})\hat{g}_{tT}\hat{g}_{tT}'/T \geq -CI_m$ in the p.s.d. sense. So by a second-order Taylor expansion

$$\begin{aligned}\hat{P}(\hat{\beta}, \bar{\lambda}) &\geq -k\bar{\lambda}'\hat{g}_T(\hat{\beta}) - k^2 C\bar{\lambda}'\bar{\lambda} \\ &= k\|\hat{g}_T(\hat{\beta})\|\delta_T - k^2 C\delta_T^2\end{aligned}$$

w.p.a.1. Noting that $\hat{P}(\hat{\beta}, \bar{\lambda}) \leq \sup_{\lambda\in\Lambda_T} \hat{P}(\hat{\beta}, \lambda) \leq \sup_{\lambda\in\Lambda_T} \hat{P}(\beta_0, \lambda)$, it follows by Lemma A.5 that w.p.a.1, $(k\|\hat{g}_T(\hat{\beta})\|\delta_T - k^2C\delta_T^2)/S_T \leq C\|\hat{g}_T(\beta_0)\|^2$. Solving for $\|\hat{g}_T(\hat{\beta})\|$ then gives

$$k\|\hat{g}_T(\hat{\beta})\| \leq C\|\hat{g}_T(\beta_0)\|^2/(k\delta_T/S_T) + kC\delta_T = O_p(\delta_T),$$

as $\|\hat{g}_T(\beta_0)\|^2 = O_p(T^{-1})$ by CLT and δ_T^2/S_T is of higher order than T^{-1}.

As $\|\hat{g}_T(\hat{\beta})\| = O_p(\delta_T)$, $\hat{g}_T(\hat{\beta}) \xrightarrow{p} 0$. By UWL, $\sup_{\beta\in\mathcal{B}} \|\hat{g}_T(\beta) - k_1 g(\beta)\| \xrightarrow{p} 0$ and $g(\beta)$ is continuous. Then T gives $g(\hat{\beta}) \xrightarrow{p} 0$. Since $g(\beta) = 0$ has a unique zero at β_0, $\|g(\beta)\|$ must be bounded away from zero outside any neighborhood of β_0. Therefore, $\hat{\beta}$ must be inside any neighborhood of β_0 w.p.a.1, i.e. $\hat{\beta} \xrightarrow{p} \beta_0$.

Therefore, $S_T \sum_{t=1}^T \hat{g}_{tT}\hat{g}_{tT}'/T = O_p(1)$, cf. Kitamura and Stutzer (1997, Proof of Theorem 1, p.871). As $S_T \sum_{t=1}^T [\rho_2(k\dot{\lambda}'\hat{g}_{tT}) - \rho_2(0)]\hat{g}_{tT}\hat{g}_{tT}'/T \xrightarrow{p} 0$, then w.p.a.1 $S_T \sum_{t=1}^T \rho_2(k\dot{\lambda}'\hat{g}_{tT})\hat{g}_{tT}\hat{g}_{tT}'/T \geq -CI_m$ in the p.s.d. sense. Hence,

$$\begin{aligned}\frac{1}{S_T}\hat{P}(\hat{\beta}, \bar{\lambda}) &\geq -k(\bar{\lambda}/S_T)'\hat{g}_T(\hat{\beta}) - k^2C(\bar{\lambda}/S_T)'(\bar{\lambda}/S_T) \\ &= k\|\hat{g}_T(\hat{\beta})\|(\delta_T/S_T) - k^2C(\delta_T/S_T)^2,\end{aligned}$$

w.p.a.1. By a similar argument to that above, $k\|\hat{g}_T(\hat{\beta})\|(\delta_T/S_T) - k^2C(\delta_T/S_T)^2 \leq C\|\hat{g}_T(\beta_0)\|^2$ and

$$\|\hat{g}_T(\hat{\beta})\| \leq C\|\hat{g}_T(\beta_0)\|^2/(k\delta_T/S_T) + kC(\delta_T/S_T) = O_p(\delta_T/S_T). \tag{A.12}$$

Now, for any $\varepsilon_T \to 0$, re-define $\bar{\lambda} = -S_T\varepsilon_T\hat{g}_T(\hat{\beta})$. Note that $\bar{\lambda} = o_p(\delta_T)$ by eq. (A.12), so that $\bar{\lambda} \in \Lambda_T$ w.p.a.1. Then,

$$k\varepsilon_T\|\hat{g}_T(\hat{\beta})\|^2(1 - \varepsilon_T C) \leq C\|\hat{g}_T(\beta_0)\|^2 = O_p(T^{-1}).$$

Since, for all T large enough, $1 - \varepsilon_T C$ is bounded away from zero, it follows that $\varepsilon_T\|\hat{g}_T(\hat{\beta})\|^2 = O_p(T^{-1})$. The conclusion then follows by a standard result from probability theory, that if $\varepsilon_T Y_T = O_p(T^{-1})$ for all $\varepsilon_T \to 0$, then $Y_T = O_p(T^{-1})$. ■

Appendix B: Proofs of Theorems

Proof of Theorem 2.1: Let $\hat{G}_{tT} = G_{tT}(\hat{\beta})$ and $\hat{g}_{tT} = g_{tT}(\hat{\beta})$. By eq. (2.7) and the definition of $p(v)$,

$$\begin{aligned} 0 &= \sum_{t=1}^{T}\rho_1(k\hat{\lambda}'\hat{g}_{tT})\hat{g}_{tT} = \sum_{t=1}^{T}[\rho_1(k\hat{\lambda}'\hat{g}_{tT}) + 1]\hat{g}_{tT} - T\hat{g}_T(\hat{\beta}) \\ &= k\sum_{t=1}^{T}p(k\hat{\lambda}'\hat{g}_{tT})\hat{g}_{tT}\hat{g}_{tT}'\hat{\lambda} - T\hat{g}_T(\hat{\beta}). \end{aligned}$$

Solving for $\hat{\lambda}$, substituting into eq. (2.9), and multiplying by $k\sum_{s=1}^{T}p(k\hat{\lambda}'\hat{g}_{sT})/(TS_T)$ gives the first result. Note that for EL $p(v) = [-(1-v)^{-1}+1]/v = -(1-v)^{-1} = \rho_1(v)$ and for CUE $p(v) = [-(1+v)+1]/v = -1$ is constant. ■

Proof of Theorem 2.2: The first and third results follow from Lemma A.6. Let $\hat{g}_{tT} = g_{tT}(\hat{\beta})$. By Lemma A.4, for any $\dot{\lambda} = \tau\hat{\lambda}$, $0 \leq \tau \leq 1$, as $\hat{\lambda} \in \Lambda_T$, $\max_{1\leq t\leq T}|\dot{\lambda}'\hat{g}_{tT}| \xrightarrow{p} 0$ and, thus, $\max_{1\leq t\leq T}|\rho_2(k\dot{\lambda}'\hat{g}_{tT}) - \rho_2(0)| \xrightarrow{p} 0$. Hence, by a second-order Taylor expansion, as $\hat{P}(\hat{\beta}, 0) = 0$ and $S_T\sum_{t=1}^{T}\rho_2(\dot{\lambda}'\hat{g}_{tT})\hat{g}_{tT}\hat{g}_{tT}'/T \leq -CI_m$,

$$0 \leq S_T^{-1}\hat{P}(\hat{\beta}, \hat{\lambda})$$

$$= \quad -kS_T^{-1}\hat\lambda'\hat g_T(\hat\beta) + k^2 S_T^{-2}\hat\lambda'[S_T\sum_{t=1}^{T}\rho_2(\dot\lambda'\hat g_{tT})\hat g_{tT}\hat g_{tT}'/T]\hat\lambda$$
$$\leq \quad \|kS_T^{-1}\hat\lambda\|\|\hat g_T(\hat\beta)\| - C\|kS_T^{-1}\hat\lambda\|^2,$$

w.p.a.1. Dividing through by $\|kS_T^{-1}\hat\lambda\|$ and solving gives $\|kS_T^{-1}\hat\lambda\| \leq C\|\hat g_T(\hat\beta)\| = O_p(T^{-1/2})$ from Lemma A.6. Hence, $\|\hat\lambda\| = O_p[(T/S_T^2)^{-1/2}]$ and, thus, $\hat\lambda \xrightarrow{p} 0$ by Assumption 2.2 (a). ■

Proof of Theorem 2.3: By Theorem 2.2, w.p.a.1 the constraint on λ is not binding, and by β_0 in the interior of $\mathcal{B}$ neither is the constraint $\beta \in \mathcal{B}$. Therefore, the first order conditions of eqs. (2.7) and (2.9) are satisfied w.p.a.1. Then by a mean-value expansion of the former of these conditions we have, for $\hat g_{tT} = g_{tT}(\hat\beta)$, $\hat\theta_T = (\hat\beta', (\hat\lambda/S_T)')'$ and $\theta_0 = (\beta_0', 0')'$,

$$0 \quad = \quad -T^{1/2}\begin{pmatrix} 0 \\ \hat g_T(\beta_0) \end{pmatrix} + \bar M T^{1/2}(\hat\theta_T - \theta_0), \tag{B.1}$$
$$\bar M \quad = \quad \begin{pmatrix} 0 & \sum_{t=1}^T \rho_1(k\hat\lambda'\hat g_{tT})G_{tT}(\hat\beta)'/T \\ \sum_{t=1}^T \rho_1(k\bar\lambda'\hat g_{tT})G_{tT}(\bar\beta)/T & S_T\sum_{t=1}^T k\rho_2(k\bar\lambda'\hat g_{tT})g_{tT}(\bar\beta)\hat g_{tT}'/T \end{pmatrix},$$

where $\bar\beta$ and $\bar\lambda$ are mean-values that may differ from row to row of the matrix $\bar M$.

As $\bar\lambda = O_p[(T/S_T^2)^{-1/2}]$ by Theorem 2.2, it follows from Assumptions 2.2 (a) and 2.3 (d) by an argument like that for the proof of Lemma A.4 that for $\tilde\lambda$ equal to $\hat\lambda$ or $\bar\lambda$,

$$\max_{1\leq t\leq T}|\tilde\lambda'\hat g_{tT}| \leq \|\tilde\lambda\| \max_{1\leq t\leq T}\|\hat g_{tT}\| = O_p[(T/S_T^2)^{-1/2}T^{1/\alpha}] \xrightarrow{p} 0.$$

Therefore,

$$\max_{1\leq t\leq T}|\rho_1(k\tilde\lambda'\hat g_{tT}) - \rho_1(0)| \xrightarrow{p} 0, \ \max_{1\leq t\leq T}|\rho_2(k\bar\lambda'\hat g_{tT}) - \rho_2(0)| \xrightarrow{p} 0.$$

Similar arguments to the proof of UWL applied to the off-diagonal components of $\bar M$ show that $\sum_{t=1}^T \rho_1(k\bar\lambda'\hat g_{tT})G_{tT}(\bar\beta)/T \xrightarrow{p} -k_1 G$ and $\sum_{t=1}^T \rho_1(k\hat\lambda'\hat g_{tT})G_{tT}(\hat\beta)'/T \xrightarrow{p} -k_1 G'$.

From Lemma A.6 a first order Taylor expansion of $T^{1/2}\hat g_T(\hat\beta)$ about β_0 yields $O_p(1) = T^{1/2}\hat g_T(\beta_0) + \hat G_T(\bar\beta)T^{1/2}(\hat\beta - \beta_0)$ where $\hat G_T(\beta) = \sum_{t=1}^T G_{tT}(\beta)/T$ and $\bar\beta$ lies on the line segment joining $\hat{\beta}$ and β_0 and may differ from row to row. An application of UWL adapted for $\hat{G}_T(\beta)$ shows that $\hat{G}_T(\beta) \xrightarrow{p} k_1 G$. Hence, by Assumption 2.5 (c), as $T^{1/2}\hat{g}_T(\beta_0) = O_p(1)$ from CLT, $T^{1/2}(\hat{\beta} - \beta_0) = O_p(1)$. By H, from Assumption 2.5 (b), eq. (A.2) in the proof of Theorem 2.1 in Smith (2005) may be replaced by

$$\begin{aligned} \sup_{|s|\geq 1} \left\| \sum_{r=\max[1,1-s]}^{\min[T,T-s]} g_{r+s,k}(\bar{\beta}) \partial g_r(\bar{\beta})/\partial\beta'/T \right\| &\leq \left(\sum_{r=1}^{T} \sup_{\beta\in\mathcal{N}} |g_{r,k}(\beta)|^{\alpha} /T \right)^{\frac{1}{\alpha}} \\ &\quad \times \left(\sum_{r=1}^{T} \sup_{\beta\in\mathcal{N}} \|\partial g_r(\beta)/\partial\beta'\|^{\frac{\alpha}{\alpha-1}} /T \right)^{\frac{\alpha-1}{\alpha}} \\ &= O_p(1), \end{aligned} \qquad \text{(B.2)}$$

$(k = 1, ..., m)$. Therefore, because $\hat{\beta}$ is $T^{1/2}$-consistent, it follows from Smith (2005, Theorem 2.1) using Lemma A.3 that $S_T \sum_{t=1}^T k\rho_2(k\bar{\lambda}'\hat{g}_{tT}) g_{tT}(\bar{\beta})\hat{g}_{tT}'/T \xrightarrow{p} -k_1\Omega$.

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

$$M = -k_1 \begin{pmatrix} 0 & G' \\ G & \Omega \end{pmatrix}, M^{-1} = -(k_1)^{-1} \begin{pmatrix} -\Sigma & H \\ H' & P \end{pmatrix}.$$

As $\bar{M}$ is p.d. w.p.a.1, inverting and solving eq. (B.1), as $T^{1/2}\hat{g}_T(\beta_0) = O_p(1)$ from eq. (A.11),

$$\begin{aligned} T^{1/2}(\hat{\theta}_T - \theta_0) &= -\bar{M}^{-1}(0, -T^{1/2}\hat{g}_T(\beta_0)')' = -M^{-1}(0, -T^{1/2}\hat{g}_T(\beta_0)')' + o_p(1) \\ &= -(k_1)^{-1}(H', P)'T^{1/2}\hat{g}_T(\beta_0) + o_p(1). \end{aligned} \qquad \text{(B.3)}$$

The conclusions of the theorem then follow from eq. (B.3) and CLT. ■

Proof of Theorem 2.4: As $\rho_2(k\hat{\lambda}'\hat{g}_{tT}) \xrightarrow{p} \rho_2(0)$, $S_T \sum_{t=1}^T (\rho_2(k\hat{\lambda}'\hat{g}_{tT} - \rho_2(0))\hat{g}_{tT}(\hat{\beta})(\hat{\lambda}'\hat{G}_{tT})/T \xrightarrow{p} 0$. Similarly to the proof of Lemma A.3 and Smith (2005, Theorem 2.1), $S_T \sum_{t=1}^T \hat{g}_{tT,k}\hat{G}_{tT}/T = O_p(1)$, $(k = 1, ..., m)$. Because $\hat{\lambda} = O_p[(T/S_T^2)^{-1/2}]$, the (λ, β)-block of the Hessian (D.1) may thus be written as $k \sum_{t=1}^T \rho_1(k\hat{\lambda}'\hat{g}_{tT})G_{tT}(\hat{\beta})/T + o_p(1)$. As $\rho_1(k\hat{\lambda}'\hat{g}_{tT}) \xrightarrow{p} \rho_1(0)$, $k\sum_{t=1}^T [\rho_1(k\hat{\lambda}'\hat{g}_{tT}) - \rho_1(0)]G_{tT}(\hat{\beta})/T \xrightarrow{p} 0$. Adapting UWL, $\sum_{t=1}^T G_{tT}(\hat{\beta})/T \xrightarrow{p} k_1 G$ from which the result is proved. ■

Proof of Theorem 2.5: By Lemma A.4, from Assumptions 2.4 (a) and (b), $\rho_2(k\hat{\lambda}'\hat{g}_{tT}) \xrightarrow{p} \rho_2(0)$. Thus, $S_T \sum_{t=1}^{T} [\rho_2(k\hat{\lambda}'\hat{g}_{tT}) - \rho_2(0)]\hat{g}_{tT}\hat{g}_{tT}'/T \xrightarrow{p} 0$. Therefore, because $\hat{\beta}$ is $T^{1/2}$-consistent by Theorem 2.3, it follows by Smith (2005, Theorem 2.1) using (B.2) and Lemma A.3 that $S_T \sum_{t=1}^{T} \hat{g}_{tT}\hat{g}_{tT}'/T \xrightarrow{p} k_2\Omega$. ■

Proof of Theorem 3.1: Let $\hat{g}_{tT} = g_{tT}(\hat{\beta})$, $\hat{a}_{tT} = a_{tT}(\hat{\beta})$ and $\hat{\pi}_t = \pi_t(\hat{\beta}, \hat{\lambda})$, $(t = 1, ..., T)$. A mean value expansion of $\hat{\pi}_t$ around $\lambda = 0$ yields

$$\hat{\pi}_t = T^{-1} + T^{-1} \left(\frac{k\rho_2(k\dot{\lambda}'\hat{g}_{tT})\hat{\lambda}'\hat{g}_{tT}}{T^{-1}\sum_{s=1}^{T}\rho_1(k\dot{\lambda}'\hat{g}_{sT})} - \frac{k\rho_1(k\dot{\lambda}'\hat{g}_{tT})(T^{-1}\sum_{s=1}^{T}\rho_2(k\dot{\lambda}'\hat{g}_{sT})\hat{g}_{sT}')\hat{\lambda}}{(T^{-1}\sum_{s=1}^{T}\rho_1(k\dot{\lambda}'\hat{g}_{sT}))^2} \right) \quad \text{(B.4)}$$

where $\dot{\lambda} = \tau\hat{\lambda}$, $0 \leq \tau \leq 1$. By Lemma A.4, $\max_{1\leq t\leq T} |\rho_1(k\dot{\lambda}'\hat{g}_{tT}) - \rho_1(0)| \xrightarrow{p} 0$ and $\max_{1\leq t\leq T} |\rho_2(k\dot{\lambda}'\hat{g}_{tT}) - \rho_2(0)| \xrightarrow{p} 0$. Thus, from Assumption 2.4 (a), $\sum_{s=1}^{T} \rho_1(k\dot{\lambda}'\hat{g}_{sT})/T \xrightarrow{p} -1$ and $\sum_{s=1}^{T} \rho_2(k\dot{\lambda}'\hat{g}_{sT})/T \xrightarrow{p} -1$. As $T^{1/2}\hat{g}_T(\hat{\beta}) = O_p(1)$, $\hat{\lambda} = O_p[(T/S_T^2)^{-1/2}]$ by Theorem 2.2 and $\max_{1\leq t\leq T} |\hat{\lambda}'\hat{g}_{tT}| \xrightarrow{p} 0$, from eq. (B.4),

$$\begin{aligned} \hat{\pi}_t &= T^{-1} + T^{-1}(k + o_p(1))\hat{\lambda}'\hat{g}_{tT} + O_p(S_T/T)) && \text{(B.5)} \\ &= T^{-1}(1 + o_p(1)) \end{aligned}$$

uniformly t. Also, using Lemma C.1 and eq. (A.9), substituting for $\hat{\pi}_t$ from eq. (B.5),

$$\begin{aligned} \sum_{t=1}^{T} k_{tT}\hat{\pi}_t &= (1 + o_p(1))\frac{1}{T}\sum_{t=1}^{T}\frac{1}{S_T}\sum_{s=t-T}^{t-1} k\left(\frac{s}{S_T}\right) && \text{(B.6)} \\ &= (1 + o_p(1))\frac{1}{S_T}\sum_{s=1-T}^{T-1}\left(1 - \frac{|s|}{T}\right) k\left(\frac{s}{S_T}\right) \\ &= k_1 + o_p(1). \end{aligned}$$

Similarly,

$$\sum_{t=1}^{T} \hat{\pi}_t\hat{a}_{tT} = (1 + o_p(1))\hat{a}_T(\hat{\beta}) \quad \text{(B.7)}$$

where $\hat{a}_T(\hat{\beta}) = \sum_{t=1}^T \hat{a}_{tT}/T$. Hence, substituting eqs. (B.6) and (B.7) into eq. (3.2), the first conclusion follows from

$$\begin{aligned}\hat{\alpha} &= \frac{1}{(k_1 + o_p(1))}(1 + o_p(1))\hat{a}_T(\hat{\beta}) \\ &= \alpha_0 + o_p(1)\end{aligned}$$

as $\hat{a}_T(\hat{\beta}) \xrightarrow{p} k_1\alpha_0$ by UWL.

For the second conclusion, from eq. (3.2), substituting the first order Taylor expansion (B.5) and (B.6),

$$\begin{aligned}T^{1/2}(\hat{\alpha} - \alpha_0) &= \frac{1}{\sum_{t=1}^T k_{tT}\hat{\pi}_t} T^{1/2}\sum_{t=1}^T \hat{\pi}_t(\hat{a}_{tT} - k_{tT}\alpha_0) \\ &= \frac{1}{(k_1 + o_p(1))}\left(T^{-1/2}\sum_{t=1}^T(\hat{a}_{tT} - k_{tT}\alpha_0)\right. \\ &\quad + \left((k + o_p(1))S_T\sum_{t=1}^T(\hat{a}_{tT} - k_{tT}\alpha_0)\hat{g}'_{tT}/T\right)(T/S_T^2)^{1/2}\hat{\lambda} \\ &\quad \left. + O_p[(T/S_T^2)^{-1/2}]\sum_{t=1}^T(\hat{a}_{tT} - k_{tT}\alpha_0)/T\right).\end{aligned} \tag{B.8}$$

By UWL $\sum_{t=1}^T(\hat{a}_{tT} - k_{tT}\alpha_0)/T = o_p(1)$ as $\sum_{t=1}^T k_{tT}/T = k_1 + o_p(1)$ from eq. (B.6). Hence, the third term in eq. (B.8) is $o_p[(T/S_T^2)^{-1/2}] = o_p(1)$ by Assumption 2.2 (a). As $\hat{\beta}$ is $T^{1/2}$-consistent from Theorem 2.3, by an argument like that in the proofs of Theorem 2.3 and 2.5 above, $S_T\sum_{t=1}^T(\hat{a}_{tT} - k_{tT}\alpha_0)\hat{g}'_{tT}/T \xrightarrow{p} B$. Hence, the second term becomes $B(T/S_T^2)^{1/2}\hat{\lambda} + o_p(1)$. Therefore, from eqs. (B.3) and (B.8),

$$\begin{aligned}T^{1/2}(\hat{\alpha} - \alpha_0) &= T^{-1/2}\frac{1}{k_1}\sum_{t=1}^T(a_{tT}(\beta_0) - k_{tT}\alpha_0) \\ &\quad + AT^{1/2}(\hat{\beta} - \beta_0)(1 + o_p(1)) + B(T/S_T^2)^{1/2}\hat{\lambda} + o_p(1) \\ &= \frac{1}{k_1}\begin{pmatrix} I_r & -AH - B'P \end{pmatrix} T^{-1/2}\sum_{t=1}^T\begin{pmatrix} a_{tT}(\beta_0) - k_{tT}\alpha_0 \\ g_{tT}(\beta_0) \end{pmatrix} + o_p(1).\end{aligned} \tag{B.9}$$

By an adaptation of CLT,

$$T^{-1/2}\sum_{t=1}^T\begin{pmatrix} a_{tT}(\beta_0) - k_{tT}\alpha_0 \\ g_{tT}(\beta_0)) \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, (k_1)^2\begin{pmatrix} \Xi & B \\ B' & \Omega \end{pmatrix}\right). \tag{B.10}$$

Therefore the second conclusion follows immediately from eq. (B.9).

As $T^{1/2}\hat{g}_T(\hat{\beta}) = (k_1)T^{1/2}\hat{g}(\hat{\beta}) + o_p(1) = (k_1)\Omega P T^{1/2}\hat{g}(\beta_0) + o_p(1)$, the asymptotic correlation between $\hat{\alpha}$ and $\hat{g}(\hat{\beta})$ is given from eqs. (B.9) and (B.10) by

$$\begin{pmatrix} I_r & -AH - B'P \end{pmatrix} \begin{pmatrix} \Xi & B \\ B' & \Omega \end{pmatrix} \begin{pmatrix} 0 \\ P\Omega \end{pmatrix} = 0$$

from which the final conclusion is obtained. ■

Proof of Corollary 3.1: Immediate from Theorem 3.1. ■

Proof of Theorem 4.1: Define $\bar{P}_G = I_m - GH$. An expansion of $\hat{g}_T(\hat{\beta})$ about β_0 gives

$$\begin{aligned} T^{1/2}\hat{g}_T(\hat{\beta}) &= \bar{P}_G T^{1/2}\hat{g}_T(\beta_0) + o_p(1) \\ &= -(k_1)\Omega(T/S_T^2)^{1/2}\hat{\lambda} + o_p(1). \end{aligned} \tag{B.11}$$

Let $\hat{g}_{tT} = g_{tT}(\hat{\beta})$, $(t = 1, ..., T)$. Expanding $\hat{P}(\hat{\beta}, \hat{\lambda})$ about $\lambda = 0$,

$$\begin{aligned} 2(T/S_T)\hat{P}(\hat{\beta}, \hat{\lambda}) &= -2k(T/S_T^2)^{1/2}\hat{\lambda}'T^{1/2}\hat{g}_T(\hat{\beta}) + k^2(T/S_T^2)\hat{\lambda}'(S_T\sum_{t=1}^{T}\rho_2(k\bar{\lambda}'\hat{g}_{tT})\hat{g}_{tT}\hat{g}_{tT}'/T)\hat{\lambda} \\ &= -2(k_1/k_2)(T/S_T^2)^{1/2}\hat{\lambda}'T^{1/2}\hat{g}_T(\hat{\beta}) - (k_1^2/k_2)(T/S_T^2)\hat{\lambda}'\Omega\hat{\lambda} + o_p(1) \\ &= T\hat{g}_T(\hat{\beta})'\Omega^{-1}\hat{g}_T(\hat{\beta})/(k_2) + o_p(1). \end{aligned}$$

It follows as in Hansen (1982) from CLT that $T\hat{g}_T(\hat{\beta})'\Omega^{-1}\hat{g}_T(\hat{\beta}) \xrightarrow{d} (k_1)^2\chi^2(m-p)$ from which the conclusions for $\mathcal{LR}$ and $\mathcal{S}$ follow. The result for $\mathcal{LM}$ is obtained directly from the above expansion and (B.11). ■

Proof of Theorem 5.1: Consider the first order conditions, cf. eqs. (2.7) and (2.9), determining $\tilde{\theta}$ and $\tilde{\varphi}$:

$$\sum_{t=1}^{T}\rho_1(k(\tilde{\varphi}'q_{tT}(\tilde{\theta}) + k_1\tilde{\mu}'r(\tilde{\theta})))\begin{pmatrix} q_{tT}(\tilde{\theta}) \\ Q_{tT}(\tilde{\theta})'\tilde{\varphi} + k_1 R(\tilde{\theta})'\tilde{\mu} \\ r(\tilde{\theta}) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \tag{B.12}$$

where $Q_{tT}(\theta) = E[\partial q_{tT}(\theta)/\partial\theta']$ and $R(\theta) = \partial r(\theta)/\partial\theta'$. It is immediate from eq. (B.12) that the constrained GEL estimator $\tilde{\theta}$ satisfies the parametric constraints, *viz.* $r(\tilde{\theta}) = 0$. Hence, a similar proof to that of Theorem 2.2 establishes that, if Assumptions 2.1, 2.2, 5.1 and 5.2 hold, $\tilde{\theta} \xrightarrow{p} \theta_0$ and $\tilde{\varphi} \xrightarrow{p} 0$. Therefore, from eq. (B.12), as $\max_{1\leq t\leq T}\left|\rho_1(k\tilde{\varphi}' q_{tT}(\tilde{\theta})) - \rho_1(0)\right| \xrightarrow{p} 0$, using a UWL similar to Lemma A.1, $\tilde{\mu} \xrightarrow{p} 0$ by Assumptions 5.3 (c) and (d). Arguments like those in the proof of Theorem 2.3 give

$$\begin{aligned} T^{1/2}\hat{q}_T(\theta_0) + k_1\Sigma(T/S_T^2)^{1/2}\tilde{\varphi} + k_1 Q T^{1/2}(\tilde{\theta} - \theta_0) &= o_p(1), \quad \text{(B.13)} \\ Q'(T/S_T^2)^{1/2}\tilde{\varphi} + R'(T/S_T^2)^{1/2}\tilde{\mu} &= o_p(1), \\ RT^{1/2}(\tilde{\theta} - \theta_0) &= o_p(1). \end{aligned}$$

Let $H_\beta = E[\partial h_t(\theta_0)/\partial\beta']$ and $R_\beta = \partial r(\theta_0)/\partial\beta'$. Because G, $(Q_\alpha', R_\alpha')' \equiv (0', H_\alpha', R_\alpha')'$ and $(Q_\beta', R_\beta')' \equiv (G', H_\beta', R_\beta')'$ are full column rank by Assumption 5.3 (d), $Q'\Sigma^{-1}Q + R'R$ is p.d.. Hence, from eq. (B.13),

$$(T/S_T^2)^{1/2}\tilde{\mu} = -(RMR')^{-1}RMQ'(T/S_T^2)^{1/2}\tilde{\varphi} + o_p(1) \quad \text{(B.14)}$$

and, thus,

$$KQ'(T/S_T^2)^{1/2}\tilde{\varphi} = o_p(1). \quad \text{(B.15)}$$

Therefore, premultiplying the first equation in eq. (B.13) by $KQ'\Sigma^{-1}$ and using eq. (B.15),

$$KQ'\Sigma^{-1}T^{1/2}\hat{q}_T(\theta_0) + k_1 KQ'\Sigma^{-1}QT^{1/2}(\tilde{\theta} - \theta_0) = o_p(1).$$

Now, $Q'\Sigma^{-1}Q = M^{-1} - R'R$ and $KQ'\Sigma^{-1}Q = KM^{-1}$. Hence,

$$KQ'\Sigma^{-1}T^{1/2}\hat{q}_T(\theta_0) + k_1 T^{1/2}(\tilde{\theta} - \theta_0) = o_p(1),$$

using eq. (B.13), and, as $T^{1/2}\hat{q}_T(\theta_0) = k_1 T^{1/2}\hat{q}(\theta_0) + o_p(1)$,

$$T^{1/2}(\tilde{\theta} - \theta_0) = -KQ'\Sigma^{-1}T^{1/2}\hat{q}(\theta_0) + o_p(1). \quad \text{(B.16)}$$

Substituting eq. (B.16) back into eq. (B.13),

$$(T/S_T^2)^{1/2}\tilde{\varphi} = -(\Sigma^{-1} - \Sigma^{-1}QKQ'\Sigma^{-1})T^{1/2}\hat{q}(\theta_0) + o_p(1), \quad \text{(B.17)}$$

and, thus, from eq. (B.14),

$$(T/S_T^2)^{1/2}\tilde{\mu} = (RMR')^{-1}RMQ'\Sigma^{-1}T^{1/2}\hat{q}(\theta_0) + o_p(1), \qquad \text{(B.18)}$$

as $RMQ'\Sigma^{-1}QK = 0$. The result follows immediately from eqs. (B.16)-(B.18) as $T^{1/2}\hat{q}(\theta_0) \xrightarrow{d} N(0,\Sigma)$ by a CLT similar to Lemma A.2. ■

Proof of Theorem 5.2: Let $\bar{P}_G = I_m - GH$ and S_g denote a selection matrix such that $S_g' q_{tT}(\theta) = g_{tT}(\beta)$, $(t = 1, ..., T)$. Similarly to the proof of Theorem 4.1,

$$\begin{aligned} 2(T/S_T)\hat{P}(\tilde{\theta},\tilde{\varphi},\tilde{\mu}) &= T\hat{q}_T(\tilde{\theta})'\Sigma^{-1}\hat{q}_T(\tilde{\theta})/(k_2) + o_p(1) \\ &= T\hat{q}_T(\theta_0)'(\Sigma^{-1} - \Sigma^{-1}QKQ'\Sigma^{-1})\hat{q}_T(\theta_0)/(k_2) + o_p(1). \end{aligned}$$

Hence,

$$\begin{aligned} 2(T/S_T)(\hat{P}(\tilde{\theta},\tilde{\varphi},\tilde{\mu}) - \hat{P}(\hat{\beta},\hat{\lambda}))/(k_1^2/k_2) &= T\hat{q}(\theta_0)'(\Sigma^{-1} - \Sigma^{-1}QKQ'\Sigma^{-1} - S_g\bar{P}_G'\Omega^{-1}\bar{P}_G S_g')\hat{q}(\theta_0) \\ &\quad + o_p(1). \end{aligned} \qquad \text{(B.19)}$$

As $\Omega = S_g'\Sigma S_g$ and $S_g'Q = (G, 0)$,

$$\begin{aligned} &\Sigma(\Sigma^{-1} - \Sigma^{-1}QKQ'\Sigma^{-1} - S_g\bar{P}_G'\Omega^{-1}\bar{P}_G S_g')\Sigma(\Sigma^{-1} - \Sigma^{-1}QKQ'\Sigma^{-1} - S_g\bar{P}_G'\Omega^{-1}\bar{P}_G S_g')\Sigma \\ &= \Sigma(\Sigma^{-1} - \Sigma^{-1}QKQ'\Sigma^{-1} - S_g\bar{P}_G'\Omega^{-1}\bar{P}_G S_g')\Sigma. \end{aligned}$$

Therefore, the result for $\mathcal{LR}_a$ follows from Rao and Mitra (1971, Theorem 9.2.1, p.171) with degrees of freedom given by

$$\begin{aligned} tr[\Sigma(\Sigma^{-1} - \Sigma^{-1}QKQ'\Sigma^{-1} - S_g\bar{P}_G'\Omega^{-1}\bar{P}_G S_g')] &= tr[\Sigma(\Sigma^{-1} - \Sigma^{-1}QKQ'\Sigma^{-1})] - tr[\Sigma S_g\bar{P}_G'\Omega^{-1}\bar{P}_G S_g'] \\ &= tr[I_{m+s}] - tr[QKQ'\Sigma^{-1}] - tr[\Sigma S_g\bar{P}_G'\Omega^{-1}\bar{P}_G S_g'] \\ &= (m+s) - (p+q-r) - (m-p) = s + r - q. \end{aligned}$$

As $(T/S_T^2)^{1/2}\hat{\lambda} = -\Omega^{-1}\bar{P}_G T^{1/2}\hat{g}(\theta_0) + o_p(1)$, from eq. (B.17),

$$(T/S_T^2)^{1/2}(\tilde{\varphi} - \hat{\varphi}) = -(\Sigma^{-1} - \Sigma^{-1}QKQ'\Sigma^{-1} - S_g\Omega^{-1}\bar{P}_G S_g')T^{1/2}\hat{q}(\theta_0) + o_p(1). \qquad \text{(B.20)}$$

Eq. (B.20) yields the intermediate result that the GEL minimum chi-squared statistic $\mathcal{MC}_a$ eq. (5.7)

$$\begin{aligned}\mathcal{MC}_a &= (T/S_T^2)(\tilde{\varphi}-\hat{\varphi})'\hat{\Sigma}(\tilde{\varphi}-\hat{\varphi}) \\ &= T\hat{q}(\theta_0)'(\Sigma^{-1}-\Sigma^{-1}QKQ'\Sigma^{-1}-S_g\bar{P}_G'\Omega^{-1}\bar{P}_GS_g')T^{1/2}\hat{q}(\theta_0)+o_p(1)\end{aligned} \tag{B.21}$$

and is asymptotically equivalent to $\mathcal{LR}_a$ from eq. (B.19).

We now consider the statistics $\mathcal{LM}_a$ and $\mathcal{S}_a$. Firstly, a Taylor expansion for the score eq. (B.12) at $(\tilde{\theta},\tilde{\varphi},\tilde{\mu})$ around $(\hat{\theta},\hat{\varphi},0)$ yields

$$\begin{aligned}T^{1/2}\sum_{t=1}^{T}\rho_1(k\hat{\lambda}'g_{tT}(\hat{\beta}))\begin{pmatrix} q_{tT}(\hat{\theta}) \\ 0 \\ k_1r(\hat{\theta})\end{pmatrix} &= k_1\begin{pmatrix}\Sigma & Q & 0 \\ Q' & 0 & R' \\ 0 & R & 0\end{pmatrix}T^{1/2}\begin{pmatrix}(\tilde{\varphi}-\hat{\varphi})/S_T \\ (\tilde{\theta}-\hat{\theta}) \\ \tilde{\mu}/S_T\end{pmatrix} \\ &\quad +o_p(1)\end{aligned} \tag{B.22}$$

as $\sum_{t=1}^{T}\rho_1(k\hat{\lambda}'g_{tT}(\hat{\beta}))G_{tT}(\hat{\beta})'\hat{\lambda}=0$. Hence,

$$\begin{aligned}-k_1T^{1/2}\begin{pmatrix}(\tilde{\varphi}-\hat{\varphi})/S_T \\ (\tilde{\theta}-\hat{\theta}) \\ \tilde{\mu}/S_T\end{pmatrix} &= -\begin{pmatrix}\Sigma & Q & 0 \\ Q' & 0 & R' \\ 0 & R & 0\end{pmatrix}^{-1}T^{1/2}\sum_{t=1}^{T}\rho_1(k\hat{\lambda}'g_{tT}(\hat{\beta}))\begin{pmatrix} q_{tT}(\hat{\theta}) \\ 0 \\ k_1r(\hat{\theta})\end{pmatrix} \\ &\quad +o_p(1) \\ &= -\begin{pmatrix}\Sigma & Q & 0 \\ Q' & 0 & R' \\ 0 & R & 0\end{pmatrix}^{-1}S_{\psi,\mu}T^{1/2}\sum_{t=1}^{T}\rho_1(k\hat{\lambda}'g_{tT}(\hat{\beta}))\begin{pmatrix} h_{tT}(\hat{\theta}) \\ k_1r(\hat{\theta}\end{pmatrix} \\ &\quad +o_p(1)\end{aligned} \tag{B.23}$$

as $\sum_{t=1}^{T}\rho_1(k\hat{\lambda}'g_{tT}(\hat{\beta}))g_{tT}(\hat{\beta})=0$. Secondly, the GEL LM-like statistic may alternatively be expressed as

$$\begin{aligned}\mathcal{LM}_a &= T\begin{pmatrix}(\tilde{\varphi}-\hat{\varphi})/S_T \\ (\tilde{\theta}-\hat{\theta}) \\ \tilde{\mu}/S_T\end{pmatrix}'S_{\psi,\mu}\left(S_{\psi,\mu}'\begin{pmatrix}\Sigma & Q & 0 \\ Q' & 0 & R' \\ 0 & R & 0\end{pmatrix}^{-1}S_{\psi,\mu}\right)^{-1}S_{\psi,\mu}'\begin{pmatrix}(\tilde{\eta}-\hat{\eta})/S_T \\ (\tilde{\theta}-\hat{\theta}) \\ \tilde{\mu}/S_T\end{pmatrix} \\ &\quad +o_p(1).\end{aligned} \tag{B.24}$$

Therefore, substituting (B.23) into (B.24),

$$\mathcal{LM}_a=(k_1)^{-2}T\sum_{t=1}^{T}\rho_1(k\hat{\lambda}'g_{tT}(\hat{\beta}))\begin{pmatrix} q_{tT}(\hat{\theta}) \\ 0 \\ k_1r(\hat{\theta})\end{pmatrix}'\begin{pmatrix}\Sigma & Q & 0 \\ Q' & 0 & R' \\ 0 & R & 0\end{pmatrix}^{-1} \tag{B.25}$$

$$\times \sum_{t=1}^{T} \rho_1(k\hat{\lambda}' g_{tT}(\hat{\beta})) \begin{pmatrix} q_{tT}(\hat{\theta}) \\ 0 \\ k_1 r(\hat{\theta}) \end{pmatrix} + o_p(1).$$

Thirdly, as

$$\begin{pmatrix} \Sigma & Q & 0 \\ Q' & 0 & R' \\ 0 & R & 0 \end{pmatrix}^{-1} = \begin{pmatrix} \Sigma^{-1} - \Sigma^{-1}QKQ'\Sigma^{-1} & \Sigma^{-1}QK & -\Sigma^{-1}QMR'(RMR')^{-1} \\ KQ'\Sigma^{-1} & -K & MR'(RMR')^{-1} \\ -(RMR')^{-1}RMQ'\Sigma^{-1} & (RMR')^{-1}RM & (RMR')^{-1} - I_r \end{pmatrix},$$

from (5.5), $\mathcal{S}_a$ may also be expressed as

$$\mathcal{S}_a = (k_1)^{-2} T^{-1} \sum_{t=1}^{T} \rho_1(k\hat{\lambda}' g_{tT}(\hat{\beta})) \begin{pmatrix} h_{tT}(\hat{\theta}) \\ k_1 r(\hat{\theta}) \end{pmatrix}' S'_{\psi,\mu} \begin{pmatrix} \tilde{\Sigma} & \tilde{Q} & 0 \\ \tilde{Q}' & 0 & \tilde{R}' \\ 0 & \tilde{R} & 0 \end{pmatrix}^{-1} \qquad \text{(B.26)}$$

$$\times S_{\psi,\mu} \sum_{t=1}^{T} \rho_1(k\hat{\lambda}' g_{tT}(\hat{\beta})) \begin{pmatrix} h_{tT}(\hat{\theta}) \\ k_1 r(\hat{\theta}) \end{pmatrix}.$$

Therefore, from (B.23), (B.25) and (B.26), $\mathcal{LM}_a = \mathcal{S}_a + o_p(1)$.

Substituting (B.22) into (B.25) yields

$$\mathcal{LM}_a = T \sum_{t=1}^{T} \begin{pmatrix} (\tilde{\varphi} - \hat{\varphi})/S_T \\ (\tilde{\theta} - \hat{\theta}) \\ \tilde{\mu}/S_T \end{pmatrix}' \begin{pmatrix} \Sigma & Q & 0 \\ Q' & 0 & R' \\ 0 & R & 0 \end{pmatrix} \sum_{t=1}^{T} \begin{pmatrix} (\tilde{\varphi} - \hat{\varphi})/S_T \\ (\tilde{\theta} - \hat{\theta}) \\ \tilde{\mu}/S_T \end{pmatrix} + o_p(1).$$

Eq. (B.21) obtains apart from asymptotically negligible terms upon recalling $Q'(T/S_T^2)^{1/2}\tilde{\varphi} + R'(T/S_T^2)^{1/2}\tilde{\mu} = o_p(1)$ and $G'(T/S_T^2)^{1/2}\hat{\lambda} = o_p(1)$. ■

Appendix C: Technical Lemmata

The following result is an adaptation of Kronecker's Lemma.

Lemma C.1 *Let Assumption 2.2 hold. Then*

$$\lim_{T\to\infty} \frac{1}{S_T} \sum_{t=1-T}^{T-1} \frac{|t|}{T} \left| k\left(\frac{t}{S_T}\right) \right| = 0.$$

Proof: We may re-express

$$\lim_{T\to\infty} \frac{1}{S_T} \sum_{t=1-T}^{T-1} \frac{|t|}{T} \left| k\left(\frac{t}{S_T}\right) \right| = \lim_{T\to\infty} \int_{|a| \le (T-1)/S_T} \frac{|a|}{(T/S_T)} |k_T(a)| \, da.$$

Because $k_T(\cdot) \to k(\cdot)$ a.e., the result will follow by the dominated convergence theorem as $|k(\cdot)| \leq \bar{k}(\cdot)$ a.e. if

$$\lim_{T\to\infty} \int_{|a|\leq (T-1)/S_T} \frac{|a|}{(T/S_T)} \bar{k}(a)da = 0.$$

From Assumption 2.2 (c), given any $\epsilon > 0$, there exists a $\bar{T}$ such that

$$\int_{|a|\geq \bar{T}/S_{\bar{T}}} \bar{k}(a)da < \epsilon.$$

Therefore, for all $T >\bar{T}$,

$$\begin{aligned}
\int_{|a|\leq (T-1)/S_T} \frac{|a|}{(T/S_T)} \bar{k}(a)da &= \int_{|a|\leq (\bar{T}-1)/S_{\bar{T}}} \frac{|a|}{(T/S_T)} \bar{k}(a)da + \int_{\bar{T}/S_{\bar{T}}\leq |a|\leq (T-1)/S_T} \frac{|a|}{(T/S_T)} \bar{k}(a)da \\
&< \frac{1}{(T/S_T)} \int_{|a|\leq (\bar{T}-1)/S_{\bar{T}}} |a|\, \bar{k}(a)da + \epsilon.
\end{aligned}$$

For fixed $\bar{T}$, as $S_T/T \to 0$ by Assumption 2.2 (a),

$$\lim_{T\to\infty} \frac{1}{(T/S_T)} \int_{|a|\leq (\bar{T}-1)/S_{\bar{T}}} |a|\, \bar{k}(a)da = 0.$$

Since ϵ is arbitrary,

$$\lim_{T\to\infty} \int_{|a|\leq (T-1)/S_{\bar{T}}} \frac{|a|}{(T/S_T)} \bar{k}(a)da = 0$$

which concludes the proof. ■

Lemma C.2 *If Assumption 2.2 is satisfied, then*

$$\sum_{t=\max[1-T,1-T+s]}^{\min[T-1,T-1+s]} k\left(\frac{t-s}{S_T}\right) k\left(\frac{t}{S_T}\right) / \sum_{t=1-T}^{T-1} k\left(\frac{t}{S_T}\right)^2 = k^*\left(\frac{s}{S_T}\right) + o(1)$$

uniformly s.

Proof: Consider the difference

$$\begin{aligned}
&\sum_{t=-\infty}^{\infty} k\left(\frac{t-s}{S_T}\right) k\left(\frac{t}{S_T}\right) - \sum_{t=\max[1-T,1-T+s]}^{\min[T-1,T-1+s]} k\left(\frac{t-s}{S_T}\right) k\left(\frac{t}{S_T}\right) \\
= &\sum_{t=\min[T,T+s]}^{\infty} k\left(\frac{t-s}{S_T}\right) k\left(\frac{t}{S_T}\right) + \sum_{t=\infty}^{\max[-T,-T+s]} k\left(\frac{t-s}{S_T}\right) k\left(\frac{t}{S_T}\right)
\end{aligned}$$

Firstly, if $s \geq 0$, $\min[T, T+s] = T$. Then

$$\begin{aligned}\left|\sum_{t=T}^{\infty} k\left(\frac{t-s}{S_T}\right) k\left(\frac{t}{S_T}\right)\right| &\leq \sum_{t=T}^{\infty}\left|k\left(\frac{t-s}{S_T}\right) k\left(\frac{t}{S_T}\right)\right| \\ &\leq \sup_a \bar{k}(a) \sum_{t=T}^{\infty}\left|k\left(\frac{t}{S_T}\right)\right|.\end{aligned}$$

Secondly, if $s \leq 0$, $\min[T, T+s] = T+s$. Then

$$\begin{aligned}\left|\sum_{t=T+s}^{\infty} k\left(\frac{t-s}{S_T}\right) k\left(\frac{t}{S_T}\right)\right| &\leq \sup_a \bar{k}(a) \sum_{t=T+s}^{\infty}\left|k\left(\frac{t-s}{S_T}\right)\right| \\ &= \sup_a \bar{k}(a) \sum_{t=T}^{\infty}\left|k\left(\frac{t}{S_T}\right)\right|.\end{aligned}$$

Therefore,

$$\left|\sum_{t=\min[T,T+s]}^{\infty} k\left(\frac{t-s}{S_T}\right) k\left(\frac{t}{S_T}\right)\right| \leq \sup_a \bar{k}(a) \sum_{t=T}^{\infty}\left|k\left(\frac{t}{S_T}\right)\right|.$$

Similarly,

$$\left|\sum_{t=-\infty}^{\max[-T,-T+s]} k\left(\frac{t-s}{S_T}\right) k\left(\frac{t}{S_T}\right)\right| \leq \sup_a \bar{k}(a) \sum_{t=-\infty}^{-T}\left|k\left(\frac{t}{S_T}\right)\right|.$$

Using the change of variable $t = [S_T b]$, by Assumption 2.2 (c),

$$\begin{aligned}\lim_{T\to\infty} \frac{1}{S_T} \sum_{t=T}^{\infty}\left|k\left(\frac{t}{S_T}\right)\right| &= \lim_{T\to\infty} \int_{T/S_T}^{\infty} |k_T(b)|\, db \\ &\leq \lim_{T\to\infty} \int_{T/S_T}^{\infty} \bar{k}(b) db = o(1).\end{aligned}$$

Likewise,

$$\lim_{T\to\infty} \frac{1}{S_T} \sum_{t=-\infty}^{-T}\left|k\left(\frac{t}{S_T}\right)\right| = o(1).$$

Therefore,

$$\frac{1}{S_T} \sum_{t=\max[1-T,1-T+s]}^{\min[T-1,T-1+s]} k\left(\frac{t-s}{S_T}\right) k\left(\frac{t}{S_T}\right) = \frac{1}{S_T} \sum_{t=-\infty}^{\infty} k\left(\frac{t-s}{S_T}\right) k\left(\frac{t}{S_T}\right) + o(1)$$

uniformly s. A similar argument establishes

$$\frac{1}{S_T} \sum_{t=1-T}^{T-1} k\left(\frac{t-s}{S_T}\right) k\left(\frac{t}{S_T}\right) = \frac{1}{S_T} \sum_{t=-\infty}^{\infty} k\left(\frac{t-s}{S_T}\right) k\left(\frac{t}{S_T}\right) + o(1)$$

uniformly s. Therefore,

$$\frac{1}{S_T} \sum_{t=\max[1-T,1-T+s]}^{\min[T-1,T-1+s]} k\left(\frac{t-s}{S_T}\right) k\left(\frac{t}{S_T}\right) = \frac{1}{S_T} \sum_{t=1-T}^{T-1} k\left(\frac{t-s}{S_T}\right) k\left(\frac{t}{S_T}\right) + o(1)$$

uniformly s.

Using the change of variables $s = [S_T a]$ and $t = [S_T b]$,

$$\begin{aligned}
\frac{1}{S_T}\sum_{t=-\infty}^{\infty} k\left(\frac{t-s}{S_T}\right)k\left(\frac{t}{S_T}\right) &= \lim_{T\to\infty}\frac{1}{S_T}\sum_{t=1-T}^{T-1} k\left(\frac{t-s}{S_T}\right)k\left(\frac{t}{S_T}\right) \\
&= \lim_{T\to\infty}\int_{(1-T)/S_T}^{(T-1)/S_T} k_T(b-a)k_T(b)db \\
&\quad +\frac{1}{S_T}(k(0)k\left(\frac{s}{S_T}\right)+k\left(\frac{-s}{S_T}\right)k(0)) \\
&= \int_{-\infty}^{\infty} k(b-a)k(b)db + o(1)
\end{aligned}$$

uniformly s as $|k(a)| \leq \sup_a \bar{k}(a)$. Therefore,

$$\begin{aligned}
\sum_{t=\max[1-T,1-T+s]}^{\min[T-1,T-1+s]} k\left(\frac{t-s}{S_T}\right)k\left(\frac{t}{S_T}\right)/\sum_{t=1-T}^{T-1} k\left(\frac{t}{S_T}\right)^2 &= \frac{1}{S_T}\sum_{t=1-T}^{T-1} k\left(\frac{t-s}{S_T}\right)k\left(\frac{t}{S_T}\right) \\
&\quad \div \frac{1}{S_T}\sum_{t=1-T}^{T-1} k\left(\frac{t}{S_T}\right)^2 + o(1) \\
&= k^*\left(\frac{s}{S_T}\right) + o(1)
\end{aligned}$$

uniformly s as, using the change of variable $t = [S_T b]$, by the dominated convergence theorem,

$$\begin{aligned}
\lim_{T\to\infty}\frac{1}{S_T}\sum_{t=1-T}^{T-1} k\left(\frac{t}{S_T}\right)^2 &= \lim_{T\to\infty}\int_{(1-T)/S_T}^{(T-1)/S_T} k_T(b)^2 da + \frac{1}{S_T}k(0)^2 \\
&= \int_{-\infty}^{\infty} k(b)^2 db + o(1) \\
&= k_2 + o(1) > 0.
\end{aligned}$$

■

Lemma C.3 *Let Assumptions 2.2 (b) and (c) hold. Then* $k^*(\cdot) \in \mathcal{K}_2$.

Proof: Firstly, $k^*(\cdot) : \mathcal{R} \to [-1, 1]$ by CS and $k^*(0) = 1$.

Secondly, $k^*(\cdot)$ is symmetric as $k^*(a) = \int_{-\infty}^{\infty} k(b-a)k(b)db/k_2 = \int_{-\infty}^{\infty} k(c)k(c-(-a))dc/k_2 = k^*(-a)$ using the change of variable $c = b - a$.

Thirdly, we show that $\int_{[0,\infty)} \bar{k}^*(a)da < \infty$. Initially, however, note that $\int_{(-\infty,\infty)} k(c-b)k(c)dc = \int_{(-\infty,0)} k(c-b)k(c)dc + \int_{[-b,\infty)} k(d+b)k(d)dd$. Now,

$$\begin{aligned}\sup_{b\geq a}\left|\int_{(-\infty,0)} k(c-b)k(c)dc\right| &\leq \int_{(-\infty,0)} \sup_{b\geq a}|k(c-b)|\,|k(c)|\,dc \\ &\leq \bar{k}(-a)\int_{(-\infty,0)} |k(c)|\,dc.\end{aligned}$$

Therefore,

$$\int_{[0,\infty)} \sup_{b\geq a}\left|\int_{(-\infty,0)} k(c-b)k(c)dc\right| \leq \int_{[0,\infty)} \bar{k}(-a)\left(\int_{(-\infty,0)} |k(c)|\,dc\right) da \leq \left(\int_{(-\infty,0]} \bar{k}(a)da\right)^2.$$

Next, note $\int_{[-b,\infty)} k(d+b)k(d)dd = \left(\int_{[-b,-a)} + \int_{[-a,\infty)}\right) k(d+b)k(d)dd$. Firstly,

$$\begin{aligned}\int_{[0,\infty)} \sup_{b\geq a}\left|\int_{[-a,\infty)} k(d+b)k(d)dd\right| da &\leq \int_{-\infty}^{\infty} |k(d)| \left(\int_{-\infty}^{\infty} \bar{k}(a)da\right) dd \\ &\leq \left(\int_{-\infty}^{\infty} \bar{k}(a)da\right)^2\end{aligned}$$

as

$$\begin{aligned}\sup_{b\geq a}\left|\int_{[-a,\infty)} k(d+b)k(d)dd\right| &\leq \int_{[-a,\infty)} \bar{k}(d+a)\,|k(d)|\,dd \\ &\leq \int_{-\infty}^{\infty} \bar{k}(d+a)\,|k(d)|\,dd.\end{aligned}$$

Also,

$$\left|\int_{[-b,-a)} k(d+b)k(d)dd\right| \leq \bar{k}(-a)\int_{[-b,-a)} |k(d+b)|\,dd \leq \bar{k}(-a)\int_{[0,\infty)} \bar{k}(c)dc$$

yielding

$$\int_{[0,\infty)} \sup_{b\geq a}\left|\int_{[-b,-a)} k(d+b)k(d)dd\right| da \leq \left(\int_{(-\infty,0]} \bar{k}(a)da\right)\left(\int_{[0,\infty)} \bar{k}(a)da\right).$$

Therefore $\int_{[0,\infty)} \bar{k}^*(a)da < \infty$.

Fourthly, by Assumption 2.2 (b), $k^*(\cdot)$ is continuous at 0 and almost everywhere. Therefore, $k^*(\cdot) \in \mathcal{K}_1$.

Finally, $K^*(\lambda) = (2\pi)^{-1}\int \exp(-ia\lambda)k^*(a)da = 2\pi\,|K(\lambda)|^2/(k_2)$. Therefore, as $|K(\lambda)| \geq 0$ for all $\lambda \in \mathcal{R}$ by Assumption 2.2 (d), $K^*(\lambda) \geq 0$ and, moreover, $k^*(\cdot) \in \mathcal{K}_2$. ■

Appendix D: Second Order Derivatives

Differentiating twice with respect to λ and β,

$$\frac{\partial \hat{P}(\lambda,\beta)}{\partial\lambda\partial\beta'} = \sum_{t=1}^{T}\left(k^2\rho_2(k\lambda' g_{tT}(\beta))g_{tT}(\beta)(\lambda' G_{tT}(\beta)) + k\rho_1(k\lambda' g_{tT}(\beta))G_{tT}(\beta)\right)/T. \quad \text{(D.1)}$$

Moreover,

$$\frac{\partial \hat{P}(\lambda,\beta)}{\partial\lambda\partial\lambda'} = k^2\sum_{t=1}^{T}\rho_2(k\lambda' g_{tT}(\beta))g_{tT}(\beta)g_{tT}(\beta)'/T, \quad \text{(D.2)}$$

which is n.d. if $\sum_{t=1}^{T} g_{tT}(\beta)g_{tT}(\beta)'$ is p.d. as $\rho_2(k\lambda' g_{tT}(\beta)) < 0$ by the concavity of $\rho(\cdot)$ on its domain $\mathcal{V}$. Hence, $\hat{\lambda}(\beta)$ defines a unique minimum of $\hat{P}(\lambda,\beta)$ and is continuously differentiable by the implicit function theorem. Hence,

$$\frac{\partial\hat{\lambda}(\beta)}{\partial\beta'} = -\left(\frac{\partial \hat{P}(\lambda,\beta)}{\partial\lambda\partial\lambda'}\right)^{-1}\frac{\partial \hat{P}(\lambda,\beta)}{\partial\lambda\partial\beta'}.$$

Also

$$\frac{\partial \hat{P}(\lambda,\beta)}{\partial\beta\partial\beta'} = \sum_{t=1}^{T}\left(k^2\rho_2(k\lambda' g_{tT}(\beta))(G_{tT}(\beta)'\lambda)(\lambda' G_{tT}(\beta)) + k\rho_1(k\lambda' g_{tT}(\beta))\sum_{k=1}^{m}\frac{\partial g_{tT,k}(\beta)}{\partial\beta\partial\beta'}\lambda_k\right)/T,$$

where $g_{tT,k}(\beta)$ denotes the kth element of $g_{tT}(\beta)$, $(k = 1, ..., m)$.

References

Aitchison, J. (1962): "Large-Sample Restricted Parametric Tests", *Journal of the Royal Statistical Society* (*Series B*), 24, 234-250.

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

Andrews, D.W.K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation", *Econometrica*, 59, 817-858.

Back, K., and D.P. Brown (1993): "Implied Probabilities in GMM Estimators", *Econometrica*, 61, 971-975.

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

Brown, B.W., and W.K. Newey (2002): "Generalized Method of Moments, Efficient Bootstrapping, and Improved Inference", *Journal of Business and Economic Statistics*, 20, 507-517.

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

Chesher, A., and R.J. Smith (1997): "Likelihood Ratio Specification Tests", *Econometrica*, 65, 627-646.

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

Gouriéroux, C., and A. Monfort (1989): "A General Framework for Testing a Null Hypothesis in a "Mixed Form"", *Econometric Theory*, 5, 63-82.

Gradshteyn, I.S., and I.M. Ryzhik (1980): *Tables of Integrals, Series and Products*, New York: Academic Press.

Grenander,U., and M. Rosenblatt (1984): *Statistical Analysis of Stationary Time Series.* Chelsea: New York.

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

Kitamura, Y. (1997): "Empirical ikelihood with Weakly Dependent Processes", *Annals of Statistics*, 25, 2084-2102.

Kitamura, Y., and M. Stutzer (1995): "An Information-Theoretic Alternative to Generalized Method of Moments Estimation", *Econometrica*, 65, 861-874.

Newey, W.K. (1985a): "Maximum Likelihood Specification Testing and Conditional Moment Tests," *Econometrica*, 53, 1047-1070.

Newey, W.K. (1985b): "Generalized Method of Moments Specification Testing", *Journal of Econometrics*, 29, 229-256.

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

Newey, W. K. and R. J. Smith (2004): "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators", *Econometrica*, 72, 219-255.

Newey, W.K., and K.D. West (1987): "A Simple, Positive Semi-Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", *Econometrica*, 55, 703-708.

Neyman, J. (1959): "Optimal Asymptotic Tests of Composite Statistical Hypotheses", in Grenander, U., ed., *Probability and Statistics*, New York: Wiley.

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

Rao, C.R., and S.K. Mitra (1971): *Generalized Inverse of Matrices and its Applications*, New York: Wiley.

Sargan, J.D. (1980): "Some Tests of Dynamic Specification for a Single Equation", *Econometrica*, 48, 879-897.

Seber, G.A.F. (1964): "The Linear Hypothesis and Large Sample Theory", *Annals of Mathematical Statistics* 35, 773-779.

Smith, R.J. (1987): "Alternative Asymptotically Optimal Tests and Their Application to Dynamic Specification", *Review of Economic Studies*, 54, 665-680.

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

Smith, R.J. (2000): "Empirical Likelihood Estimation and Inference," in Salmon, M., and P. Marriott, eds., *Applications of Differential Geometry to Econometrics*, Cambridge: Cambridge University Press.

Smith, R.J. (2005): "Automatic Positive Semi-Definite HAC Covariance Matrix and GMM Estimation." Forthcoming *Econometric Theory.*

Szroeter, J. (1983): "Generalized Wald Methods for Testing Nonlinear Implicit and Overidentifying Restrictions", *Econometrica* 51, 335-353.

Tauchen, G. (1985): "Diagnostic Testing and Evaluation of Maximum Likelihood Models", *Journal of Econometrics* 30, 415-443.

White, H. (1984): *Asymptotic Theory for Econometricians.* New York: Academic Press.