Generalized Empirical Likelihood Criteria for Generalized Method of Moments Estimation and Inference*

Richard J. Smith
Department of Economics
University of Bristol

This Revision: October 1998

Abstract

Since Hansen's (1982) seminal paper, the generalized method of moments (GMM) has become an increasingly important method for estimation and inference in econometrics. This paper examines alternative generalized empirical likelihood approaches. Essentially, these methods embed sample versions of the moment conditions used in GMM in a non-parametric quasi-likelihood function by use of additional parameters associated with these moment conditions. Specification and misspecification tests may be defined which are similar in nature to the classical tests and are first order equivalent to the corresponding GMM statistics.

*Financial support for this research by the ESRC (Grant No. R000237386) is gratefully acknowledged.

1 Introduction

Since Hansen's (1982) seminal paper, the *generalized method of moments* (GMM) has become an increasingly important basis for estimation and inference in econometrics. Estimation using GMM is semi-parametric in the sense that it does not fully specify the form of the probability density from which the sample was drawn. Therefore, a particular advantage of GMM is that it imposes less stringent assumptions than, for example, the method of maximum likelihood (ML). Although consequently more robust, GMM is generally less efficient than ML. Given assumed population moment conditions, the GMM estimation method minimises a quadratic form in the sample counterparts of these moment conditions constructed using a positive definite metric. The resultant GMM estimator is asymptotically efficient if this metric is the inverse of a positive semi-definite consistent estimator for the asymptotic variance matrix of the sample moments.

Although the efficient GMM estimator is asymptotically optimal, there is now a considerable body of Monte-Carlo evidence which indicates that it may be severely biased in small samples; see, for example, the July 1996 Special Issue of *Journal of Business and Economic Statistics*. This bias seems mainly to arise from the metric used. For example, Altonji and Segal (1996) suggest that GMM estimators using an identity matrix metric may perform better in finite samples than the efficient form. Other studies, Brown and Newey (1997), Burnside and Eichenbaum (1996), Hall and Horowitz (1994) and Horowitz (1998) comment on the poor finite sample characteristics of the efficient GMM estimator and associated statistics.

The aim of this paper is to synthesise and extend the recent statistics and econometrics literature which has proposed alternative criteria for the estimation of moment condition models; see *inter alia* Back and Brown (1993), Qin and Lawless (1994), Imbens (1997), Kitamura and Stutzer (1997) and Imbens, Spady and Johnson (1998). These criteria are quasi-likelihood in construction and may be viewed as variants on *empirical likelihood* (EL); see DiCiccio, Hall and Romano (1989) and Owen (1988, 1990, 1991). They give rise to estimators which are first order asymptotically equivalent to the efficient GMM estimator and, importantly, they are one-step estimators which do not require explicit calculation or estimation of the efficient metric. Hence, the correspondence between the asymptotic analysis and the finite sample behaviour of the resultant estimators may be improved over and above that for efficient GMM estimators. Some initial simulation evidence indicates that there might be reasons for hope in this direction; see, for example, Qin and Lawless (1994), Imbens (1997) and Imbens, Spady and Johnson (1998). Moreover, the construction of classical-type test statistics for over-identifying moment conditions, additional moment conditions and parametric restrictions is straightforward as will be seen below. The majority of this earlier literature has been concerned with independent and identically distributed data; see Qin and Lawless (1994), Imbens (1997) and Imbens, Spady and Johnson (1998). However, Kitamura and Stutzer (1997) and Smith (1998) examine particular criteria in a time-series context; see also Imbens (1997, Section 6).

The approach adopted in this paper is a development of the method of Chesher and Smith

(1997) for deriving likelihood ratio tests for moment conditions in a parametric model context which is adapted for time-series moment condition models. To deal with the time-series nature of the data, a smoothed version of the moment indicators rather than the moment indicators themselves as in standard GMM estimation forms the basis of the suggested estimation procedure; cf. Kitamura and Stutzer (1997). Moreover, an emphasis of this paper concerns issues of misspecification. In particular, this paper discusses misspecification and specification test statistics based on generalized empirical likelihood 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 Newey (1985a, 1985b). The tests presented here mimic in an obvious way classical tests; *viz.* likelihood ratio, score and Wald statistics.

Section 2 briefly reviews GMM estimation and introduces a generalized empirical likelihood criterion for time-series data which is formed by augmenting empirical measures associated with the empirical distribution function by a carrier function which incorporates a weighted smoothed sample version of the moment indicators; cf. Chesher and Smith (1997). The parameter vector and these weights are the respective objects of interest for estimation and inference, the former from an economic-theoretic standpoint and the latter for tests of misspecification and specification. The maximum generalized empirical likelihood estimation procedure is then described and the limiting distribution of the estimators is obtained. An estimator for the stationary distribution function of the time-series process is given which efficiently incorporates the moment information and which, therefore, dominates the empirical distribution function in section 3. Section 4 is concerned with deriving classical-type tests for over-identifying moment restrictions and additional moments; cf. Hansen (1982) and Newey (1985a, 1985b). Section 5 presents classical-type tests for parametric restrictions expressed in mixed implicit and constraint equation form encountered in simultaneous equations models and which is sufficiently general to include constraint equation [Aitchison (1962), Sargan (1980)], freedom equation [Seber (1964)] and mixed form restrictions [Gourieroux and Monfort (1989)] as special cases; cf. Szroeter (1983). The application of semi-parametric quasi-likelihood criteria to testing non-nested hypotheses is explored in Section 6. The paper is concluded by Section 7. Sketches of various proofs are given in the Appendix.

2 Generalized Empirical Likelihood Estimation

2.1 Generalized Method of Moments Estimation

Many applications of the generalized method of moments (GMM) are concerned with the modelling of stationary time series. Given the observations $\mathbf{x}_t$, $t = 1, ..., T$, on the strictly stationary process $\{\mathbf{x}_t\}_{t=-\infty}^{\infty}$, consider GMM estimation of the p-element parameter vector $\boldsymbol{\theta}$ based on the m-vector moment indicators $\mathbf{g}(\mathbf{x}_t; \boldsymbol{\theta})$, $t = 1, ..., T$, where $m \geq p$.[1] It is assumed that the functional $\mathbf{g}(.; \boldsymbol{\theta})$ is known up to $\boldsymbol{\theta}$ and that there exists a unique value $\boldsymbol{\theta}_0$ of $\boldsymbol{\theta} \in \Theta$,

[1] In practice, the moment indicators $\mathbf{g}(\mathbf{x}_t; \boldsymbol{\theta})$ will typically also involve lagged values $\mathbf{x}_{t-s}$, $s = 1, 2, ...$, and current and lagged values of exogenous variables. However, for economy of notation, but without loss of generality, such conditional dependence is suppressed in the following discussion.

with the parameter space Θ compact, such that:

$$E_{\mu}\{\mathbf{g}(\mathbf{x}_t;\boldsymbol{\theta}_0)\} = \mathbf{0}, \tag{2.1}$$

where $E_{\mu}\{.\}$ denotes expectation taken with respect to the true but unknown distribution (probability measure) μ of $\mathbf{x}_t$, $t = 0, \pm 1, ...$ It is further assumed that the normalised sample counterpart $T^{-1/2}\sum_{t=1}^{T}\mathbf{g}(\mathbf{x}_t;\boldsymbol{\theta})$ of the moment condition (2.1) obeys the central limit theorem $T^{-1/2}\sum_{t=1}^{T}\mathbf{g}(\mathbf{x}_t;\boldsymbol{\theta}_0) \rightarrow^{L} N(\mathbf{0},\mathbf{V})$ and the asymptotic variance matrix

$$\mathbf{V} = \lim_{T\rightarrow\infty} var_{\mu}\{T^{-1/2}\sum_{t=1}^{T}\mathbf{g}(\mathbf{x}_t;\theta_0)\}. \tag{2.2}$$

is assumed positive definite, where "$\rightarrow^{L}$" denotes convergence in distribution.[2]

The efficient GMM estimator $\hat{\boldsymbol{\theta}}_T$ is based on the quadratic form $T^{-1}\sum_{t=1}^{T}\mathbf{g}(\mathbf{x}_t;\boldsymbol{\theta})'\hat{\mathbf{V}}_T^{-1}\sum_{t=1}^{T}\mathbf{g}(\mathbf{x}_t;\boldsymbol{\theta})$, where $\hat{\mathbf{V}}_T$ denotes a positive semi-definite consistent estimator for $\mathbf{V}$, and is defined by

$$\hat{\boldsymbol{\theta}}_T \equiv \arg\min_{\boldsymbol{\theta}\in\Theta} T^{-1}\sum_{t=1}^{T}\mathbf{g}(\mathbf{x}_t;\boldsymbol{\theta})'\hat{\mathbf{V}}_T^{-1}\sum_{t=1}^{T}\mathbf{g}(\mathbf{x}_t;\boldsymbol{\theta}). \tag{2.3}$$

See *inter alia* Andrews (1991) and Newey and West (1987) for positive semi-definite consistent estimation of asymptotic variance matrices. Under suitable conditions and (2.1)

$$T^{1/2}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \rightarrow^{L} N[\mathbf{0}, (\mathbf{G}'\mathbf{V}^{-1}\mathbf{G})^{-1}], \tag{2.4}$$

where $\mathbf{G} = E_{\mu}\{\nabla'_{\boldsymbol{\theta}}\mathbf{g}(\mathbf{x}_t;\boldsymbol{\theta}_0)\}$ is assumed full column rank and $\nabla_{\boldsymbol{\theta}} \equiv \partial/\partial\boldsymbol{\theta}$; see Hansen (1982).

2.2 Generalized Empirical Likelihood Criteria

The approach taken in this paper is an adaptation of Chesher and Smith (1997) which is concerned with generating likelihood ratio (LR) test statistics for implied moment conditions in a fully parametric likelihood context. Recall that many tests for misspecification in classical settings may be constructed as tests for implied moment conditions; see, for example, Chesher (1984), Newey (1985a), Chesher and Irish (1987), Gourieroux, Monfort, Renault and Trognon (1987) and Smith (1987a, 1989).

Chesher and Smith (1997) augment the null parametric density for the observations multiplicatively by a carrier function of a weighted version of the moment indicator vector which underpins the implied moment conditions. These weights and the parameters characterising the null density constitute the objects of inferential interest in Chesher and Smith's (1997) framework. However, in contradistinction to Chesher and Smith (1997), in the GMM context, there is no explicit knowledge of the underlying density function for $\{\mathbf{x}_t\}_{t=-\infty}^{\infty}$, the only parametric information being contained in the moment conditions (2.1). This difficulty is circumvented by replacing the parametric density function in Chesher and Smith's (1997)

[2] In the following, "$\rightarrow^{P}$" denotes convergence in probability.

procedure by the non-parametric substitutes $d\mu_t = T^{-1}$, $t = 1, ..., T$. The empirical measures $d\mu_t$, $t = 1, ..., T$, are used to construct the empirical distribution function (EDF) $\mu_T(\mathbf{x}) = \sum_{t=1}^{T} 1(\mathbf{x}_t \leq \mathbf{x}) d\mu_t$, where $1(\mathbf{x}_t \leq \mathbf{x}) = 1$ if $\mathbf{x}_t \leq \mathbf{x}$ and 0 otherwise.[3]

Because of the time-series context, a further difficulty concerns the potential serial correlation of the moment indicators $\{\mathbf{g}(\mathbf{x}_t; \boldsymbol{\theta}_0)\}_{t=-\infty}^{\infty}$. Consider the following reformulation of the moment conditions (2.1)

$$\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta}) \equiv \sum_{s=-[(T-1)/2]}^{[(T-1)/2]} \omega(s; S_T) \mathbf{g}(\mathbf{x}_{t-s}; \boldsymbol{\theta}) \tag{2.5}$$

$t = 1, ..., T$, $T = 1, 2, ...$, where the weights $\omega(s; S_T) \equiv S_T^{-1} k(s/S_T)$, $k(.)$ is a kernel and S_T a band-width parameter.[4] Under certain conditions on $k(.)$ and S_T, the smoothed linear function $\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})$ of the moment indicators $\{\mathbf{g}(\mathbf{x}_t; \boldsymbol{\theta})\}_{t=1}^{T}$ renders the implicit metric imposed by the generalized empirical likelihood criterion appropriate for efficient estimation. The weights $\{\omega(s; S_T)\}$ are similar in spirit to those used in heteroskedastic and autocorrelation consistent (HAC) variance matrix estimation; see *inter alia* Andrews (1991) and Newey and West (1987). Precise conditions governing the choice of the kernel $k(.)$ and the band-width parameter S_T are detailed below. Effectively, the definition of $\{\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})\}_{t=1}^{T}$ in (2.5) from the moment indicators $\{\mathbf{g}(\mathbf{x}_t; \boldsymbol{\theta}_0)\}_{t=1}^{T}$, $T = 1, 2, ...$, induces a restatement of the moment conditions (2.1) as $E_\mu\{\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta}_0)\} = \mathbf{0}$, $t = 1, 2, ...$, $T = 1, 2, ...$ In particular, the choice of weights $\{\omega(.; S_T)\}$ requires $E_\mu\{\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta}_0)\} = \mathbf{0}$ if and only if (2.1) rendering (2.1) and the moment condition $E_\mu\{\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta}_0)\} = \mathbf{0}$, $t = 1, 2, ...$, $T = 1, 2, ...$, based on (2.5) as essentially equivalent.

Therefore, the generalized empirical likelihood criterion is based on the augmented quasi-density

$$r_t(\boldsymbol{\varphi}) = T^{-1} h[\boldsymbol{\phi}' \mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})], \tag{2.6}$$

where $\boldsymbol{\varphi} = (\boldsymbol{\theta}', \boldsymbol{\phi}')'$ with the m-element parameter vector $\boldsymbol{\phi}$ associated with the smoothed moment indicator $\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})$, $t = 1, ..., T$; cf. Chesher and Smith (1997). The quasi-density function $r_t(\boldsymbol{\varphi})$, $t = 1, ..., T$, of (2.6) need not necessarily be a density. However, in some circumstances, the quantities $r_t(\boldsymbol{\varphi})$, $t = 1, ..., T$, may themselves be interpreted as empirical measures; that is, $r_t(\boldsymbol{\varphi}) > 0$, $t = 1, ..., T$, and $\sum_{t=1}^{T} r_t(\boldsymbol{\varphi}) = 1$, cf. the EDF. As will become evident below, it is not necessary to ascribe probabilistic content to the quasi-density (2.6) itself but rather associated quantities which appear in the first-order log-derivatives of $r_t(\boldsymbol{\varphi})$, $t = 1, ..., T$.

Consequently, from (2.6), the generalized empirical likelihood (GEL) function is defined as $\prod_{t=1}^{T} r_t(\boldsymbol{\varphi})$ with its log counterpart given by

$$\mathcal{R}(\boldsymbol{\varphi}) \equiv -T \ln T + \sum_{t=1}^{T} \ln h[\boldsymbol{\phi}' \mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})], \tag{2.7}$$

[3]Under quite general conditions, the EDF $\mu_T(.)$, is a consistent estimator for the underlying distribution function $\mu(.)$ of $\{\mathbf{x}_t\}_{t=-\infty}^{\infty}$. Moreover, the empirical measures $d\mu_t$, $t = 1, ..., T$, are nonparametric ML estimators which maximise the non-parametric log-likelihood $\sum_{t=1}^{T} \ln d\mu_t$ subject to the constraints $0 < d\mu_t < 1$, $t = 1, ..., T$, and $\sum_{t=1}^{T} d\mu_t = 1$.

[4]As in Andrews (1991), the class of kernels $k(.)$ is restricted to the scale parameter type in this paper.

and associated score vector given by

$$\mathbf{S}(\boldsymbol{\varphi}) = \begin{pmatrix} \nabla_{\theta}\mathcal{R}(\boldsymbol{\varphi}) \\ \nabla_{\phi}\mathcal{R}(\boldsymbol{\varphi}) \end{pmatrix} = \sum_{t=1}^{T} \pi_t(\boldsymbol{\varphi}) \begin{pmatrix} \nabla_{\theta}\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})'\boldsymbol{\phi} \\ \mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta}) \end{pmatrix}, \tag{2.8}$$

where

$$\pi_t(\boldsymbol{\varphi}) = \nabla \ln h[\boldsymbol{\phi}'\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})], \tag{2.9}$$

$t = 1, ..., T$.

Under the conditions given below on $\ln h(.)$, the ratios

$$\pi_t(\boldsymbol{\varphi}) / \sum_{t=1}^{T} \pi_t(\boldsymbol{\varphi}), \tag{2.10}$$

$t = 1, ..., T$, may be thought of as implied probabilities as in Back and Brown (1993) being positive and summing to unity. Thus, these ratios may be regarded as providing empirical measure counterparts to the expectation operator in (2.1); cf. the use of the empirical measures $d\mu_t = T^{-1}$, $t = 1, ..., T$, comprising the EDF used in the calculation of sample mean-like quantities.[5]

2.3 Maximum Generalized Empirical Likelihood Estimation

To create a well-defined optimisation problem, the log-carrier function $\ln h(.)$ in (2.7) is assumed to be strictly monotonic, convex and continuously differentiable to the second order with $h(.)$ positive-valued.[6] Without loss of generality, the normalisations $h(0) = 1$ and $h'(0) = 1$ are imposed; cf. Chesher and Smith (1997). The resultant log-GEL $\mathcal{R}(\boldsymbol{\varphi})$ of (2.7) is treated as a saddle function; cf. Chesher and Smith (1997).[7] The above specification for the carrier function $h(.)$ includes the Cressie-Read family of power divergence criteria discussed in Imbens, Spady and Johnson (1998, p.336) which also specialises to the criteria associated with empirical likelihood and empirical information or exponential tilting.

Firstly, the log-GEL $\mathcal{R}(\boldsymbol{\varphi})$ for given $\boldsymbol{\theta}$ is minimised by $\hat{\boldsymbol{\phi}}_T(\boldsymbol{\theta})$; that is

$$\hat{\boldsymbol{\phi}}_T(\boldsymbol{\theta}) \equiv \arg\min_{\phi} \mathcal{R}(\boldsymbol{\varphi}),$$

and $\hat{\boldsymbol{\phi}}_T(\boldsymbol{\theta})$ satisfies the GEL equations, cf. (2.8),

$$\sum_{t=1}^{T} \pi_t(\boldsymbol{\theta}, \hat{\boldsymbol{\phi}}_T(\boldsymbol{\theta}))\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta}) = \mathbf{0}. \tag{2.11}$$

[5]Brown and Newey (1997) independently proposed a similar definition to (2.10) albeit for a different purpose.

[6]Strictly speaking, it is only necessary that these properties hold in a neighbourhood of zero. For further details, see Kitamura and Stutzer (1997, p.866).

[7]In some circumstances, the parameter vector $\boldsymbol{\phi}$ may be interpreted as a vector of Lagrange multipliers associated with a sample version of the moment conditions (2.1), *viz.* $\sum_{t=1}^{T} \pi_t(\boldsymbol{\varphi})\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta}) = \mathbf{0}$; cf. (2.11).

The second order derivative matrix of the log-GEL $\mathcal{R}(\boldsymbol{\varphi})$ with respect to $\boldsymbol{\phi}$ for given $\boldsymbol{\theta}$ is given in the Appendix and is positive definite if $\sum_{t=1}^{T}\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})'$ is positive definite by the strict convexity of $\ln h(.)$. Hence, the solution $\hat{\boldsymbol{\phi}}_T(\boldsymbol{\theta})$ is a unique minimiser of $\mathcal{R}(\boldsymbol{\varphi})$ for given $\boldsymbol{\theta}$ and is continuously differentiable in $\boldsymbol{\theta}$ by the implicit function theorem; see, for example, Bartle (1976, Theorem 41.9, p.384). When $\boldsymbol{\phi} = \mathbf{0}$, $\mathcal{R}(\boldsymbol{\varphi}) = -T\ln T$ which is that obtained with the EDF empirical measures $d\mu_t = T^{-1}$, $t = 1, ..., T$. Therefore, the choice of the minimum value $\hat{\boldsymbol{\phi}}_T(\boldsymbol{\theta})$ maximises the discrepancy between the log-GEL $\mathcal{R}(\boldsymbol{\varphi})$ and the value corresponding to the EDF empirical measures. In this sense, the profile or concentrated log-GEL $\mathcal{R}(\boldsymbol{\theta}, \hat{\boldsymbol{\phi}}_T(\boldsymbol{\theta}))$ may be viewed as creating a least favourable family of criteria for the estimation of $\boldsymbol{\theta}_0$; cf. DiCiccio and Romano (1990).

Secondly, the maximised GEL estimator (MGELE) $\hat{\boldsymbol{\theta}}_T$ is the maximiser of the profile log-GEL $\mathcal{R}(\boldsymbol{\theta}, \hat{\boldsymbol{\phi}}_T(\boldsymbol{\theta}))$; *viz.*

$$\hat{\boldsymbol{\theta}}_T \equiv \arg\max_{\boldsymbol{\theta}\in\Theta} \mathcal{R}(\boldsymbol{\theta}, \hat{\boldsymbol{\phi}}_T(\boldsymbol{\theta})) = \arg\max_{\boldsymbol{\theta}\in\Theta}\min_{\boldsymbol{\phi}} \mathcal{R}(\boldsymbol{\theta}, \boldsymbol{\phi}), \tag{2.12}$$

and satisfies the GEL equations

$$\sum_{t=1}^{T} \pi_t(\hat{\boldsymbol{\varphi}}_T)\nabla_{\boldsymbol{\theta}}\mathbf{g}_{tT}^{\omega}(\hat{\boldsymbol{\theta}}_T)'\hat{\boldsymbol{\phi}}_T = \mathbf{0}, \tag{2.13}$$

where $\hat{\boldsymbol{\varphi}}_T = (\hat{\boldsymbol{\theta}}_T', \hat{\boldsymbol{\phi}}_T')'$ and $\hat{\boldsymbol{\phi}}_T \equiv \hat{\boldsymbol{\phi}}_T(\hat{\boldsymbol{\theta}}_T)$.[8] Hence, the solution $\hat{\boldsymbol{\varphi}}_T$ defines a saddle point of the log-GEL $\mathcal{R}(\boldsymbol{\varphi})$. As is evident from (2.11) and (2.13), the introduction of the m-vector of auxiliary parameters $\boldsymbol{\phi}$ has rendered the first order conditions determining the MGELE $\hat{\boldsymbol{\theta}}_T$ and $\hat{\boldsymbol{\phi}}_T$ as corresponding to a just-identified GMM problem. In the just-identified case $m = p$, $\hat{\boldsymbol{\phi}}_T = \mathbf{0}$ from (2.13) and, thus, (2.11) reduces to the familiar GMM first order conditions for a just-identified problem. A consistent estimator for $\boldsymbol{\theta}_0$ to initiate an iterative process to locate $\hat{\boldsymbol{\varphi}}_T$ is a GMM estimator, optimal or otherwise.

In the Appendix, it is shown that $\boldsymbol{\varphi} = (\boldsymbol{\theta}_0', \mathbf{0}')'$ defines a unique saddle point of the expectation of the quasi-density $r_t(\boldsymbol{\varphi})$ under (2.1). Hence, the MGELE $\hat{\boldsymbol{\theta}}_T$ and $\hat{\boldsymbol{\phi}}_T$ are consistent estimators for $\boldsymbol{\theta}_0$ and $\mathbf{0}$ respectively as is suggested by the expectation of (2.8) under (2.1) being zero at $\boldsymbol{\varphi} = (\boldsymbol{\theta}_0', \mathbf{0}')'$.

Lemma 2.1 *(Consistency of the MGELE $\hat{\boldsymbol{\theta}}_T$ and $\hat{\boldsymbol{\phi}}_T$.) Under (2.1), the MGELE $\hat{\boldsymbol{\theta}}_T$ of (2.12) is a consistent estimator for $\boldsymbol{\theta}_0$ and $\hat{\boldsymbol{\phi}}_T \to^P \mathbf{0}$.*

Defining $\tau \equiv \int k(x)dx$ and $\alpha \equiv \int k(x)^2 dx$, the following theorem demonstrates the first order equivalence of the MGELE $\hat{\boldsymbol{\theta}}_T$ with the optimal GMM estimator of (2.3).

Theorem 2.1 *(Limit Distribution of the MGELE $\hat{\boldsymbol{\theta}}_T$ and $\hat{\boldsymbol{\phi}}_T$.) Under (2.1),*

$$T^{1/2}\begin{pmatrix} (\alpha/\tau)\nabla^2 \ln h(0) S_T^{-1}\hat{\boldsymbol{\phi}}_T \\ \hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0 \end{pmatrix} \to^L N\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{G}(\mathbf{G}'\mathbf{V}^{-1}\mathbf{G})^{-1}\mathbf{G}'\mathbf{V}^{-1} & \mathbf{0}' \\ \mathbf{0} & (\mathbf{G}'\mathbf{V}^{-1}\mathbf{G})^{-1} \end{pmatrix}\right).$$

[8] The same notation $\hat{\boldsymbol{\theta}}_T$ is used for the MGELE in (2.12) as for the optimal GMM estimator of (2.3) because, as will be seen in Theorem 2.1 below, the MGELE, although not numerically identical, shares the same first order asymptotic properties as the optimal GMM estimator of (2.3).

From Theorem 2.1, we see that $\hat{\boldsymbol{\phi}}_T$ is asymptotically uncorrelated with the MGELE $\hat{\boldsymbol{\theta}}_T$ which ensures that the moment conditions (2.1) are used efficiently in its construction.[9]

2.4 The Choice of Kernel $k(.)$ and Band-Width Parameter S_T

In the course of the proof of Theorem 2.1, it is required that

$$\mathbf{V}_T = \sum_{s=-2[(T-1)/2]}^{2[(T-1)/2]} k^* \left(\frac{s}{S_T} \right) \mathbf{C}(s) \to^P \mathbf{V}, \tag{2.14}$$

where the sample autocovariances $\mathbf{C}(s) = T^{-1} \sum_{t=s+1}^{T} \mathbf{g}(\mathbf{x}_t; \boldsymbol{\theta}_0) \mathbf{g}(\mathbf{x}_{t-s}; \boldsymbol{\theta}_0)'$, $\mathbf{C}(-s) = \mathbf{C}(s)'$, $s = 0, ..., 2[(T-1)/2]$, and the induced kernel $k^*(.)$ obeys

$$k^*(x) = \alpha^{-1} \int k(x+y) k(y) dy. \tag{2.15}$$

Clearly, therefore, for (2.14), the induced kernel $k^*(.)$ and S_T should satisfy the conditions of Andrews (1991, Theorem 1(a), p.827); in particular, $k^*(.)$ should belong to the positive semi-definite class $\mathcal{K}_2$ defined in Andrews (1991, p.822) and $S_T = o(T^{1/2})$. The optimal rate for S_T will depend on the particular kernel $k(.)$ (and, thus, $k^*(.)$) chosen; see Andrews (1991, Section 5, pp.830-832). Now, $k^*(.) : \mathcal{R} \to [-1, 1]$, $k^*(0) = 1$ and is symmetric, $k^*(x) = k^*(-x)$ for all $x \in \mathcal{R}$. Moreover, the boundedness condition $\int k^*(x)^2 dx < \infty$ is satisfied if $\int \int k(x+y)^2 dx dy < \infty$. Hence, $k^*(.) \in \mathcal{K}_1$; see Andrews (1991, equation (2.6), p.821). The spectral window generator for $k^*(.)$ is $K^*(\lambda) \equiv (2\pi)^{-1} \int \exp(-ix\lambda) k^*(x) dx = [2\pi\sigma^2]^{-1} |K(\lambda)|^2$ where $K(\lambda) \equiv \int \exp(-ix\lambda) k(x) dx$ is the spectral window generator for the kernel $k(.)$. Therefore, $K^*(\lambda) \geq 0$ and $k^*(.) \in \mathcal{K}_2$ as long as $k(\lambda) = 0$ almost everywhere is excluded. This relationship between the spectral window generators $K(.)$ and $K^*(.)$ for the kernels $k(.)$ and $k^*(.)$ respectively may be used to deduce suitable choices for the kernel $k(.)$ in (2.5) as the following examples attest.

2.4.1 Example 1: Truncated and Bartlett Kernels

The truncated kernel is defined as $k(x) = 1$, $|x| \leq 1$, and 0, $|x| > 1$, and 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 the spectral window generator for the Bartlett kernel is $K^*(\lambda) = (2\pi)^{-1}[(\sin \lambda/2)/(\lambda/2)]^2$. The choice of the band-width parameter $S_T = O(T^{1/3})$ is optimal; see Andrews (1991, (5.3), p.830). Cf. Kitamura and Stutzer (1997).

[9] Alternatively, $\nabla^2 \ln h(0)$ may be consistently estimated by $T^{-1} \sum_{t=1}^{T} \nabla^2 \ln h[\hat{\boldsymbol{\phi}}_T' \mathbf{g}_{tT}^{\omega}(\hat{\boldsymbol{\theta}}_T)]$ or $\sum_{t=1}^{T} \pi_t(\hat{\boldsymbol{\varphi}}_T) \nabla^2 \ln h[\hat{\boldsymbol{\phi}}_T' \mathbf{g}_{tT}^{\omega}(\hat{\boldsymbol{\theta}}_T)] / \sum_{t=1}^{T} \pi_t(\hat{\boldsymbol{\varphi}}_T)$ and substituted into Theorem 2.1 with no alteration to the result.

2.4.2 Example 2: Bartlett and Parzen Kernels

Again because the spectral window generator for the Parzen kernel is $K^*(\lambda) = (8\pi/3)^{-1}[(\sin\lambda/4)/(\lambda/4)]^2$, it is immediate that the induced kernel corresponding to the Bartlett kernel $k(x) = 1 - |x|$, $|x| \leq 1$, and 0, $|x| > 1$, 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 choice of the band-width parameter $S_T = O(T^{1/5})$ is optimal; see Andrews (1991, (5.3), p.830).

For other examples, see Andrews (1991). Hence, the kernel $k(.)$ gives rise to the (infeasible) positive-semidefinite consistent estimator $\mathbf{V}_T$ for $\mathbf{V}$. A feasible positive semi-definite consistent estimator for $\mathbf{V}$ of (2.2) may therefore be defined as

$$\hat{\mathbf{V}}_T = \sum_{s=-2[(T-1)/2]}^{2[(T-1)/2]} k^*\left(\frac{s}{S_T}\right)\hat{\mathbf{C}}(s), \tag{2.16}$$

where $\hat{\mathbf{C}}(s) = T^{-1}\sum_{t=\max(1,s+1)}^{\min(T,T-s)} \mathbf{g}(\mathbf{x}_t;\hat{\boldsymbol{\theta}}_T)\mathbf{g}(\mathbf{x}_{t-s};\hat{\boldsymbol{\theta}}_T)'$, $\hat{\mathbf{C}}(-s) = \hat{\mathbf{C}}(s)'$ and the kernel $k^*(.)$ is defined by (2.15) above. An alternative positive semi-definite consistent estimator to $\hat{\mathbf{V}}_T$ is output as $(\alpha\nabla^2\ln h(0)TS_T^{-1})^{-1}$ or $(\alpha S_T^{-1}\sum_{t=1}^T \nabla^2\ln h[\hat{\boldsymbol{\phi}}_T'\mathbf{g}_{tT}^{\omega}(\hat{\boldsymbol{\theta}}_T)])^{-1}$ times the $(\boldsymbol{\phi},\boldsymbol{\phi})$ component of the Hessian of the log-GEL $\mathcal{R}(\boldsymbol{\varphi})$ of (2.7) evaluated at $\hat{\boldsymbol{\varphi}}_T$, namely $\sum_{t=1}^T \nabla^2\ln h[\hat{\boldsymbol{\phi}}_T'\mathbf{g}_{tT}^{\omega}(\hat{\boldsymbol{\theta}}_T)]\mathbf{g}_{tT}^{\omega}(\hat{\boldsymbol{\theta}}_T$ see the Appendix.

In a similar manner, we may obtain consistent estimators for the expected derivative matrix $\mathbf{G}$ as $T^{-1}\sum_{t=1}^T \nabla_{\boldsymbol{\theta}}'\mathbf{g}_{tT}^{\omega}(\hat{\boldsymbol{\theta}}_T)$ or

$$\hat{\mathbf{G}}_T = T^{-1}\sum_{t=1}^T \nabla_{\boldsymbol{\theta}}'\mathbf{g}(\mathbf{x}_t;\hat{\boldsymbol{\theta}}_T). \tag{2.17}$$

An alternative consistent estimator to $\hat{\mathbf{G}}_T$ is output as T^{-1} or $(\sum_{t=1}^T \nabla\ln h[\hat{\boldsymbol{\phi}}_T'\mathbf{g}_{tT}^{\omega}(\hat{\boldsymbol{\theta}}_T)])^{-1}$ times the $(\boldsymbol{\phi},\boldsymbol{\theta})$ component of the Hessian of the log-GEL $\mathcal{R}(\boldsymbol{\varphi})$ of (2.7) evaluated at $\hat{\boldsymbol{\varphi}}_T$; see the Appendix.

3 Efficient Cumulative Distribution Function Estimation

The information contained in the moment conditions $E_\mu\{\mathbf{g}(\mathbf{x}_t;\boldsymbol{\theta}_0)\} = \mathbf{0}$, $t = ..., -1, 0, 1, ...$, of (2.1) may be exploited to provide a more efficient estimator of the stationary distribution μ of the process $\{\mathbf{x}_t\}_{t=-\infty}^{\infty}$ than the EDF $\mu_T(\mathbf{x}) = T^{-1}\sum_{t=1}^T 1(\mathbf{x}_t \leq \mathbf{x})$. Consider the GEL cumulative distribution function (GELCDF) estimator based on the estimated ratios or implied probabilities $\pi_t(\hat{\boldsymbol{\varphi}}_T)/\sum_{t=1}^T \pi_t(\hat{\boldsymbol{\varphi}}_T)$, $t = 1, ..., T$, from (2.10):

$$\hat{\mu}_T(\mathbf{x}) = \sum_{t=1}^T\left(\sum_{r=-[(T-1)/2]}^{[(T-1)/2]} \omega(s;S_T)1(\mathbf{x}_{t-r} \leq \mathbf{x})\right)\frac{\pi_t(\hat{\boldsymbol{\varphi}}_T)}{\sum_{t=1}^T \pi_t(\hat{\boldsymbol{\varphi}}_T)}; \tag{3.18}$$

cf. Section 2.2.

Theorem 3.1 *(Limit Distribution of the GELCDF estimator $\hat{\mu}_T(\mathbf{x})$.) Under (2.1), the GELCDF estimator $\hat{\mu}_T(\mathbf{x})$ of (3.18) has limiting distribution given by*

$$T^{1/2}\left(\hat{\mu}_T(\mathbf{x}) - \mu(\mathbf{x})\right) \to^L N(0, \omega^2),$$

where $\omega^2 \equiv \sigma^2 - \tau^2 \mathbf{B}'\left(\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{G}(\mathbf{G}'\mathbf{V}^{-1}\mathbf{G})^{-1}\mathbf{G}'\mathbf{V}^{-1}\right)\mathbf{B}$, $\mathbf{B} \equiv \sum_{s=-\infty}^{\infty} E_\mu\{1(\mathbf{x}_t \leq \mathbf{x})\mathbf{g}(\mathbf{x}_{t-s}; \boldsymbol{\theta}_0)\}$ *and* $\sigma^2 \equiv \sum_{s=-\infty}^{\infty}\left(E_\mu\{1(\mathbf{x}_t \leq \mathbf{x})1(\mathbf{x}_{t-s} \leq \mathbf{x})\} - \mu(\mathbf{x})^2\right)$.

The EDF $\mu_T(\mathbf{x})$ has a limiting distribution described by

$$T^{1/2}\left(\mu_T(\mathbf{x}) - \mu(\mathbf{x})\right) \to^L N(0, \sigma^2).$$

Hence, the GELCDF estimator $\hat{\mu}_T(\mathbf{x})$ is more efficient than the EDF $\mu_T(\mathbf{x})$. Moreover, it may be straightforwardly seen that the GELCDF estimator $\hat{\mu}_T(\mathbf{x})$ is not dominated by any other estimator which incorporates the information in the moment conditions (2.1).

The various components in the variance ω^2 may be consistently estimated by

$$\hat{\sigma}^2 = \sum_{s=-2[(T-1)/2]}^{2[(T-1)/2]} k^*\left(\frac{s}{S_T}\right)\hat{\sigma}^2(s),$$

where $\hat{\sigma}^2(s) = T^{-1}\sum_{t=s+1}^{T}\left(1(\mathbf{x}_t \leq \mathbf{x})1(\mathbf{x}_{t-s} \leq \mathbf{x}) - \mu_T(\mathbf{x})^2\right)$, $\hat{\sigma}^2(-s) = \hat{\sigma}^2(s)$, $s = 0, ..., 2[(T-1)/2]$,

$$\hat{\mathbf{B}} = \sum_{s=-2[(T-1)/2]}^{2[(T-1)/2]} k^*\left(\frac{s}{S_T}\right)\hat{\mathbf{B}}(s),$$

where $\hat{\mathbf{B}}(s) = T^{-1}\sum_{t=\max(1,s+1)}^{\min(T,T-s)} 1(\mathbf{x}_t \leq \mathbf{x})\mathbf{g}(\mathbf{x}_{t-s}; \hat{\boldsymbol{\theta}}_T)$, and $\hat{\mathbf{V}}_T$ and $\hat{\mathbf{G}}_T$ are given in (2.16) and (2.17) above.

4 Generalized Empirical Likelihood Misspecification Tests

4.1 Tests for Over-identifying Moment Conditions

In order to gauge the validity of the moment conditions $E_\mu\{\mathbf{g}(\mathbf{x}_t; \boldsymbol{\theta}_0)\} = \mathbf{0}$, $t = 0, \pm 1, ...$, of (2.1), Hansen (1982) suggested using the optimised GMM criterion; that is, the statistic

$$\mathcal{G} = T^{-1}\sum_{t=1}^{T}\mathbf{g}(\mathbf{x}_t; \hat{\boldsymbol{\theta}}_T)'\hat{\mathbf{V}}_T^{-1}\sum_{t=1}^{T}\mathbf{g}(\mathbf{x}_t; \hat{\boldsymbol{\theta}}_T)' \tag{4.1}$$

which may be shown under suitable conditions and the validity of (2.1) to possess a limiting chi-squared distribution with $(m-p)$ degrees of freedom.

In the context of the log-GEL criterion (2.7), the validity of the moment conditions (2.1) as corresponding to the parametric restrictions $\boldsymbol{\phi} = \mathbf{0}$. This viewpoint allows us to define straightforwardly classical-type tests for the validity of the moment conditions (2.1) based directly on consideration of the parametric hypothesis $\boldsymbol{\phi} = \mathbf{0}$ with respect to the log-GEL criterion $\mathcal{R}(\boldsymbol{\varphi})$ (2.7). However, we emphasise that the hypothesis $\boldsymbol{\phi} = \mathbf{0}$ is equivalent to the validity of the moment conditions (2.1).

Firstly, we consider a likelihood ratio (LR) test based on the log-GEL (2.7). Now, $\mathcal{R}(\boldsymbol{\theta}, \mathbf{0}) = -T \ln T$ which corresponds to the parametric hypothesis $\boldsymbol{\phi} = \mathbf{0}$. When $\boldsymbol{\phi} \neq \mathbf{0}$, the log-GEL criterion $\mathcal{R}(\boldsymbol{\varphi})$ is evaluated at the saddle point estimator $\hat{\boldsymbol{\varphi}}_T$; *viz.* $\mathcal{R}(\hat{\boldsymbol{\varphi}}_T)$. From the classical viewpoint then, the difference of log-GEL criteria may be used to define a generalized empirical LR (GELR) statistic for testing $\boldsymbol{\phi} = \mathbf{0}$ or, rather, (2.1); *viz.*[10]

$$\mathcal{GELR}_\phi = 2(\alpha/\tau^2)\nabla^2 \ln h(0) S_T^{-1} \left(\mathcal{R}(\boldsymbol{\theta}_0, \mathbf{0}) - \mathcal{R}(\hat{\boldsymbol{\varphi}}_T)\right). \tag{4.2}$$

Theorem 4.1 *(Limit Distribution of the GELR Statistic for Over-Identifying Moment Conditions.) Under (2.1), the GELR statistic $\mathcal{GELR}_\phi$ of (4.2) has a limiting distribution described by*

$$\mathcal{GELR}_\phi \to^L \chi^2(m-p).$$

Secondly, by analogy with a classical Wald test for $\phi = \mathbf{0}$, a GEL Wald (GELW) statistic for the over-identifying moment conditions (2.1) is defined in terms of the estimator $\hat{\boldsymbol{\phi}}_T$; *viz.*

$$\mathcal{GELW}_\phi = [(\alpha/\tau)\nabla^2 \ln h(0) T^{1/2} S_T^{-1}]^2 \hat{\boldsymbol{\phi}}_T' \hat{\mathbf{V}}_T \hat{\boldsymbol{\phi}}_T, \tag{4.3}$$

where $\hat{\mathbf{V}}_T$ is defined in (2.16) and is a generalized inverse for the estimated asymptotic variance matrix of $(\alpha/\tau)\nabla^2 \ln h(0) T^{1/2} S_T^{-1} \hat{\boldsymbol{\phi}}_T$; see Theorem 2.1.

Proposition 4.1 *(Limit Distribution of the GELW Statistic for Over-Identifying Moment Conditions.) Under (2.1), the GELW statistic $\mathcal{GELW}_\phi$ of (4.3) has a limiting distribution described by*

$$\mathcal{GELW}_\phi \to^L \chi^2(m-p).$$

Defining a classical score test for $\boldsymbol{\phi} = \mathbf{0}$ based on the log-GEL $\mathcal{R}(\boldsymbol{\varphi})$ raises a particular difficulty; *viz.* when $\boldsymbol{\phi} = \mathbf{0}$ the parameter vector $\boldsymbol{\theta}$ is no longer identified. This class of problem has been considered by Davies (1977, 1987) in the classical and other contexts. For a recent treatment of this problem, see Andrews and Ploberger (1994). To circumvent this difficulty, consider a score statistic based on the first order derivatives of the SPEQLL $\mathcal{R}(\boldsymbol{\varphi})$ evaluated at $\hat{\boldsymbol{\theta}}_T$ and $\mathbf{0}$, the MGELE being regarded as a least favourable choice of estimator for $\boldsymbol{\theta}_0$. Denote the score vector (2.8) evaluated at $\hat{\boldsymbol{\theta}}_T$ and $\mathbf{0}$ by $\hat{\mathbf{S}}_T$. Therefore

$$\hat{\mathbf{S}}_T = \sum_{t=1}^T \begin{pmatrix} \mathbf{g}_{tT}^\omega(\hat{\boldsymbol{\theta}}_T) \\ \mathbf{0} \end{pmatrix},$$

as $\pi_t(\hat{\boldsymbol{\theta}}_T, \mathbf{0}) = 1$, $t = 1, ..., T$. Hence, the GEL score (GELS) statistic is given by

$$\mathcal{GELS}_\phi = \tau^{-2} T^{-1} \sum_{t=1}^T \mathbf{g}_{tT}^\omega(\hat{\boldsymbol{\theta}}_T)' \hat{\mathbf{V}}_T^{-1} \sum_{t=1}^T \mathbf{g}_{tT}^\omega(\hat{\boldsymbol{\theta}}_T), \tag{4.4}$$

where $\hat{\mathbf{V}}_T$ is defined in (2.16); cf. the optimal GMM statistic $\mathcal{G}$ of (4.1).

[10]Alternatively, $\nabla^2 \ln h(0)$ may be consistently estimated by $T^{-1}\sum_{t=1}^T \nabla^2 \ln h[\hat{\boldsymbol{\phi}}_T' \mathbf{g}_{tT}^\omega(\hat{\boldsymbol{\theta}}_T)]$ or $\sum_{t=1}^T \pi_t(\hat{\boldsymbol{\varphi}}_T)\nabla^2 \ln h[\hat{\boldsymbol{\phi}}_T' \mathbf{g}_{tT}^\omega(\hat{\boldsymbol{\theta}}_T)] / \sum_{t=1}^T \pi_t(\hat{\boldsymbol{\varphi}}_T)$ and substituted into the following results with no change in conclusion.

Proposition 4.2 *(Limit Distribution of the GELS Statistic for Over-Identifying Moment Conditions.) Under (2.1), the GELS statistic $\mathcal{GELS}_\phi$ of (4.4) has a limiting distribution described by*

$$\mathcal{GELS}_\phi \to^L \chi^2(m-p).$$

Because the MGELE $\hat{\boldsymbol{\theta}}_T$ of (2.12) is first order equivalent to the optimal GMM estimator of (2.3), the MGELE asymptotically obeys the optimal GMM estimator's first order conditions; see Proof of Proposition 3.2. Hence, an equivalent score-type statistic may be based on

$$T^{-1}\sum_{t=1}^{T}\mathbf{g}(\mathbf{x}_t;\hat{\boldsymbol{\theta}}_T)'\hat{\mathbf{V}}_T^{-1}\sum_{t=1}^{T}\mathbf{g}(\mathbf{x}_t;\hat{\boldsymbol{\theta}}_T);$$

viz. the GMM statistic $\mathcal{G}$ of (4.1) using $\hat{\boldsymbol{\theta}}_T$ of (2.12). Moreover, all three classical-type statistics (4.2), (4.3) and (4.4) and the GMM statistic $\mathcal{G}$ are first order equivalent as $(\alpha/\tau)\nabla^2 \ln h(0)T^{1/2}S_T^{-1}\hat{\boldsymbol{\phi}}_T = -\mathbf{V}^{-1}T^{-1/2}\sum_{t=1}^{T}\mathbf{g}(\mathbf{x}_t;\hat{\boldsymbol{\theta}}_T)+o_P(1)$; that is, $\mathcal{GELW}_\phi = \mathcal{G}+o_P(1)$. Therefore, the classical-type GEL statistics offer a potentially useful alternative class of tests for the over-identifying moment conditions (2.1). Although not discussed here, other first order equivalent tests based on the $\mathcal{C}(\alpha)$ principle may also be defined in a parallel fashion; see *inter alia* Neyman (1959) and Smith (1987b).

4.2 Tests for Additional Moment Conditions

It may be of interest to examine whether an additional s-vector of moments also has zero mean and, thus, might be usefully incorporated to improve inferences on the vector $\boldsymbol{\theta}_0$; *viz.*

$$E_\mu\{\mathbf{q}(\mathbf{x}_t;\boldsymbol{\theta}_0)\} = \mathbf{0}, \tag{4.5}$$

$t = 0, \pm 1, \ldots.$ The approach due to Newey (1985a, 1985b) would set up a conditional moment test based on a quadratic form in the estimated sample moments $T^{-1/2}\sum_{t=1}^{T}\mathbf{q}(\mathbf{x}_t;\hat{\boldsymbol{\theta}}_T)$ using as metric the inverse of the asymptotic variance matrix of $T^{-1/2}\sum_{t=1}^{T}\mathbf{q}(\mathbf{x}_t;\hat{\boldsymbol{\theta}}_T)$, where $\hat{\boldsymbol{\theta}}_T$ is the optimal GMM estimator defined in (2.3).

Following Section 2.2, we may define a log-GEL criterion similar to (2.7) appropriate for the incorporation of (4.5); *viz.*

$$\mathcal{R}^*(\boldsymbol{\varphi},\boldsymbol{\psi}) = -T\ln T + \sum_{t=1}^{T}\ln h[\boldsymbol{\phi}'\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta}) + \boldsymbol{\psi}'\mathbf{q}_{tT}^{\omega}(\boldsymbol{\theta})], \tag{4.6}$$

where $\mathbf{q}_{tT}^{\omega}(\boldsymbol{\theta}) = \sum_{s=-[(T-1)/2]}^{[(T-1)/2]}\omega(s/S_T)\mathbf{q}(\mathbf{x}_{t-s};\boldsymbol{\theta})$, $t = 1, \ldots, T$.

The saddle point estimators $(\tilde{\boldsymbol{\phi}}_T(\theta), \tilde{\boldsymbol{\psi}}_T(\theta))$ minimise the log-GEL $\mathcal{R}^*(\boldsymbol{\varphi},\boldsymbol{\psi})$ of (4.6) for given $\boldsymbol{\theta}$:

$$(\tilde{\boldsymbol{\phi}}_T(\boldsymbol{\theta}), \tilde{\boldsymbol{\psi}}_T(\boldsymbol{\theta})) = \arg\min_{\phi,\psi}\mathcal{R}^*(\boldsymbol{\varphi},\boldsymbol{\psi}),$$

and the MGELE $\tilde{\boldsymbol{\theta}}_T$ under the moment conditions (2.1) and (4.5) maximises $\mathcal{R}^*(\boldsymbol{\theta}, \tilde{\boldsymbol{\phi}}_T(\boldsymbol{\theta}), \tilde{\boldsymbol{\psi}}_T(\boldsymbol{\theta}))$:

$$\tilde{\boldsymbol{\theta}}_T = \arg\max_{\theta\in\Theta}\mathcal{R}^*(\boldsymbol{\theta}, \tilde{\boldsymbol{\phi}}_T(\boldsymbol{\theta}), \tilde{\boldsymbol{\psi}}_T(\boldsymbol{\theta})) = \arg\max_{\theta\in\Theta}\min_{\phi,\psi}\mathcal{R}^*(\boldsymbol{\varphi},\boldsymbol{\psi}).$$

Define $\tilde{\boldsymbol{\phi}}_T \equiv \tilde{\boldsymbol{\phi}}_T(\tilde{\boldsymbol{\theta}}_T)$ and $\tilde{\boldsymbol{\psi}}_T \equiv \tilde{\boldsymbol{\psi}}_T(\tilde{\boldsymbol{\theta}}_T)$.

As above, we may construct GEL counterparts to the classical LR, Wald and score statistics for the additional moments (4.5) by considering GEL-based tests for the parametric hypothesis $\boldsymbol{\psi} = \mathbf{0}$ with respect to the log-GEL $\mathcal{R}^*(\boldsymbol{\varphi}, \boldsymbol{\psi})$ of (4.6).

Firstly, consider the difference of GELR statistics under (2.1) (4.5) and (2.1) respectively, which, as $\mathcal{R}(\boldsymbol{\varphi}) = \mathcal{R}^*(\boldsymbol{\varphi}, \mathbf{0})$, corresponds to a GELR test for $\boldsymbol{\psi} = \mathbf{0}$; *viz.*

$$\begin{aligned} \mathcal{GELR} &= \mathcal{GELR}_{\phi,\psi} - \mathcal{GELR}_{\phi} \\ &= 2(\alpha/\tau^2)\nabla^2 \ln h(0) S_T^{-1} \left(\mathcal{R}^*(\hat{\boldsymbol{\varphi}}_T, \mathbf{0}) - \mathcal{R}^*(\tilde{\boldsymbol{\varphi}}_T, \tilde{\boldsymbol{\psi}}_T) \right). \end{aligned} \tag{4.7}$$

Therefore

Theorem 4.2 *(Limiting Distribution of the GELR Statistic for Additional Moment Restrictions.) Under (2.1) and (4.5), the GELR statistic $\mathcal{GELR}$ of (4.7) has a limiting distribution described by*

$$\mathcal{GELR} \to^L \chi^2(s).$$

If one were interested in the full vector of moment conditions (2.1) (4.5), one obtains the statistic $\mathcal{GELR}_{\phi,\psi} = 2\alpha^2\nabla^2 \ln h(0) S_T^{-1} \left(\mathcal{R}^*(\boldsymbol{\theta}_0, \mathbf{0}, \mathbf{0}) - \mathcal{R}^*(\tilde{\boldsymbol{\varphi}}_T, \tilde{\boldsymbol{\psi}}_T) \right)$, cf. (4.2), which has a limiting chi-squared distibution with $(m+s)-p$ degrees of freedom; cf. Theorem 4.1 and Hansen (1982).

Secondly, the GEL Wald statistic for $\boldsymbol{\psi} = \mathbf{0}$ or (4.5) is defined in a standard way. Define the $(r+s, s)$ matrix $\mathbf{S}_\psi$ to select the elements of $\boldsymbol{\psi}$ from the vector $(\boldsymbol{\phi}', \boldsymbol{\psi}')'$; that is, $\mathbf{S}_\psi'(\boldsymbol{\phi}', \boldsymbol{\psi}')' = \boldsymbol{\psi}$. The asymptotic variance matrix of the vector $T^{-1/2}\sum_{t=1}^T (\mathbf{g}(\mathbf{x}_t; \boldsymbol{\theta}_0)', \mathbf{q}(\mathbf{x}_t; \boldsymbol{\theta}_0)')'$ is denoted by

$$\mathbf{V}_* = \lim_{T\to\infty} var\left\{ T^{-1/2} \sum_{t=1}^T \begin{pmatrix} \mathbf{g}(\mathbf{x}_t; \boldsymbol{\theta}_0)' & \mathbf{q}(\mathbf{x}_t; \boldsymbol{\theta}_0)' \end{pmatrix}' \right\},$$

which is assumed positive definite. A positive semi-definite consistent estimator $\tilde{\mathbf{V}}_{*T}$ for $\mathbf{V}_*$ may be defined similarly to $\hat{\mathbf{V}}_T$ of (2.16) after appropriate substitution of the estimators $\tilde{\boldsymbol{\varphi}}_T$ and $\tilde{\psi}_T$. Similarly, $\tilde{\mathbf{G}}_{*T}$ is a consistent estimator for $\mathbf{G}_* = E_\mu\{(\nabla_\theta \mathbf{g}(\mathbf{x}_t; \boldsymbol{\theta})', \nabla_\theta \mathbf{q}(\mathbf{x}_t; \boldsymbol{\theta})')'\}$ based on $\tilde{\boldsymbol{\varphi}}_T$ and $\tilde{\psi}_T$; cf. (2.17). The GEL Wald statistic for the additional moment conditions is then defined as[11]

$$\mathcal{GELW} = [(\alpha/\tau)\nabla^2 \ln h(0) T^{1/2} S_T^{-1}]^2 \tilde{\boldsymbol{\psi}}_T' \left(\mathbf{S}_\psi' (\tilde{\mathbf{V}}_{*T}^{-1} - \tilde{\mathbf{V}}_{*T}^{-1}\tilde{\mathbf{G}}_{*T}(\tilde{\mathbf{G}}_{*T}'\tilde{\mathbf{V}}_{*T}^{-1}\tilde{\mathbf{G}}_{*T})^{-1}\tilde{\mathbf{G}}_{*T}'\tilde{\mathbf{V}}_{*T}^{-1})\mathbf{S}_\psi \right)^{-1} \tilde{\boldsymbol{\psi}}_T. \tag{4.8}$$

Therefore

Proposition 4.3 *(Limiting Distribution of the GEL Wald Statistic for Additional Moment Restrictions.) Under (2.1) and (4.5), the GEL Wald statistic $\mathcal{GELW}$ of (4.8) has a limiting distibution described by*

$$\mathcal{GELW} \to^L \chi^2(s).$$

[11]The (s,s) matrix $\mathbf{S}_\psi'(\mathbf{V}_*^{-1} - \mathbf{V}_*^{-1}\mathbf{G}_*(\mathbf{G}_*'\mathbf{V}_*^{-1}\mathbf{G}_*)^{-1}\mathbf{G}_*'\mathbf{V}_*^{-1})\mathbf{S}_\psi$ is positive definite as the $(m+s, s+p)$ matrix $(\mathbf{S}_\psi, \mathbf{G}_*)$ is full column rank.

The score test for $\boldsymbol{\psi} = \mathbf{0}$ is constructed in a standard manner using the score from the log-GEL $\mathcal{R}^*(\boldsymbol{\varphi}, \boldsymbol{\psi})$ evaluated at the estimators $(\hat{\boldsymbol{\varphi}}_T, \mathbf{0})$; cf. (2.8). The score vector of the log-GEL $\mathcal{R}^*(\boldsymbol{\varphi}, \boldsymbol{\psi})$ is

$$\mathbf{S}_{*T} = \sum_{t=1}^{T} \pi_t^*(\boldsymbol{\varphi}, \boldsymbol{\psi}) \begin{pmatrix} \mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta}) \\ \mathbf{q}_{tT}^{\omega}(\boldsymbol{\theta}) \\ (\nabla_\theta \mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})', \nabla_\theta \mathbf{q}_{tT}^{\omega}(\boldsymbol{\theta})')\,(\boldsymbol{\phi}', \boldsymbol{\psi}')' \end{pmatrix}, \tag{4.9}$$

where

$$\pi_t^*(\boldsymbol{\varphi}, \boldsymbol{\psi}) = \nabla \ln h[\boldsymbol{\phi}' \mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta}) + \boldsymbol{\psi}' \mathbf{q}_{tT}^{\omega}(\boldsymbol{\theta})],$$

$t = 1, ..., T$; cf. $\pi_t(\boldsymbol{\varphi})$, $t = 1, ..., T$, of (2.9). It is easily seen that when evaluated at $(\hat{\boldsymbol{\varphi}}_T, \mathbf{0})$, the score has zeroes in its first and last blocks from the first order conditions determining the MGELE $\hat{\boldsymbol{\theta}}_T$ and $\hat{\boldsymbol{\phi}}_T$. Moreover, $\pi_t^*(\boldsymbol{\varphi}, \boldsymbol{\psi}) = \nabla \ln h[\boldsymbol{\phi}' \mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta}) + \boldsymbol{\psi}' \mathbf{q}_{tT}^{\omega}(\boldsymbol{\theta})]$ at $(\hat{\boldsymbol{\varphi}}_T, \mathbf{0})$ is identical to $\pi_t(\hat{\boldsymbol{\varphi}}_T)$, $t = 1, ..., T$. Hence, the GEL score statistic for $\boldsymbol{\psi} = \mathbf{0}$ or (4.5) is given by

$$\begin{aligned} \mathcal{GELS} = \tau^{-2} T^{-1} \sum_{t=1}^{T} \pi_t(\hat{\boldsymbol{\varphi}}_T) \mathbf{q}_{tT}^{\omega}(\hat{\boldsymbol{\theta}}_T)' \mathbf{S}_\psi' \left(\hat{\mathbf{V}}_{*T}^{-1} - \hat{\mathbf{V}}_{*T}^{-1} \hat{\mathbf{G}}_{*T} (\hat{\mathbf{G}}_{*T}' \hat{\mathbf{V}}_{*T}^{-1} \hat{\mathbf{G}}_{*T})^{-1} \hat{\mathbf{G}}_{*T}' \hat{\mathbf{V}}_{*T}^{-1} \right) \\ \times \mathbf{S}_\psi \sum_{t=1}^{T} \pi_t(\hat{\boldsymbol{\varphi}}_T) \mathbf{q}_{tT}^{\omega}(\hat{\boldsymbol{\theta}}_T), \end{aligned} \tag{4.10}$$

where $\hat{\mathbf{V}}_{*T}$ and $\hat{\mathbf{G}}_{*T}$ denote a positive semi-definite consistent estimator for $\mathbf{V}_*$ and consistent estimator for $\mathbf{G}_*$ based on the estimators $\hat{\boldsymbol{\theta}}_T$ and $\hat{\boldsymbol{\phi}}_T$ constructed similarly to $\hat{\mathbf{V}}_T$ of (2.16) and $\hat{\mathbf{G}}_T$ of (2.17).

Therefore

Proposition 4.4 *(Limiting Distribution of the GEL Score Statistic for Additional Moment Restrictions.) Under (2.1) and (4.5), the GEL score statistic $\mathcal{GELS}$ of (4.10) has a limiting chi-squared distibution with s degrees of freedom.*

As in Section 4.1, under (2.1) and (4.5), because $\pi_t(\hat{\boldsymbol{\varphi}}_T) - 1 = o_P(1)$,

$$T^{-1/2} \sum_{t=1}^{T} \pi_t(\hat{\boldsymbol{\varphi}}_T) \mathbf{q}_{tT}^{\omega}(\hat{\boldsymbol{\theta}}_T) = \tau T^{-1/2} \sum_{t=1}^{T} \mathbf{q}(\mathbf{x}_t; \hat{\boldsymbol{\theta}}_T) + o_P(1).$$

Hence, $\mathcal{GELS}$ is first order equivalent to the optimal GMM statistic for (4.5); cf. Newey (1985a, 1985b).

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

$$\mathcal{MC} = [(\alpha/\tau)\nabla^2 \ln h(0) T^{1/2} S_T^{-1}]^2 \begin{pmatrix} (\tilde{\boldsymbol{\phi}}_T - \hat{\boldsymbol{\phi}}_T)' & \tilde{\boldsymbol{\psi}}_T' \end{pmatrix} \hat{\mathbf{V}}_{*T} \begin{pmatrix} (\tilde{\boldsymbol{\phi}}_T - \hat{\boldsymbol{\phi}}_T)' & \tilde{\boldsymbol{\psi}}_T' \end{pmatrix}'. \tag{4.11}$$

The proofs of Theorem 4.2 and Propositions 4.3, 4.4 and 4.5 show that the GEL-based statistics, (4.7), (4.8), (4.10) and (4.11), are all first order equivalent. It also immediately follows from the expression of the GEL statistic $\mathcal{GELR}$ of (4.7) as the difference of the GELR statistics, $\mathcal{GELR}_{\phi,\psi}$ and $\mathcal{GELR}_{\phi}$, that equivalent statistics may be obtained as the difference of GEL Wald and score statistics; *viz.* $\mathcal{GELW}_{\phi,\psi} - \mathcal{GELW}_{\phi}$ and $\mathcal{GELS}_{\phi,\psi} - \mathcal{GELS}_{\phi}$ respectively. However, these latter statistics may not possess positive support; a common choice of estimator for $\mathbf{V}_*$ and $\mathbf{G}_*$ may ameliorate this difficulty. Furthermore, given the discussion in Section 4.1 concerning the equivalence of those GEL-based statistics with the GMM statistic $\mathcal{G}$ 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 additional moment restrictions. A final point to note is that, under the moment conditions (2.1) and (4.5), all of the statistics of this sub-section are asymptotically independent of those of Section 4.1, a property also displayed by the classical tests for a sequence of nested hypotheses; see *inter alia* Aitchison (1962) and Sargan (1980).

5 Generalized Empirical Likelihood Specification Tests

The limiting distribution results given in Section 4.2 may be suitably adapted to allow various classical-type (likelihood ratio, Lagrange multiplier and Wald) specification tests on the parameter vector $\boldsymbol{\theta}_0$ to be defined. Consider the following parametric null hypothesis expressed in mixed implicit and constraint equation form [Szroeter (1983)]

$$\mathbf{q}(\boldsymbol{\theta}_0, \boldsymbol{\alpha}_0) = \mathbf{0}, \mathbf{r}(\boldsymbol{\alpha}_0) = \mathbf{0}, \tag{5.1}$$

where $\mathbf{q}(.,.)$ and $\mathbf{r}(.)$ are respectively known s- and r-vectors of functions which are continuously differentiable in $\boldsymbol{\theta}$ and the q-vector of parameters $\boldsymbol{\alpha}$. The derivative matrices $\mathbf{Q}_{\boldsymbol{\theta}} = \nabla'_{\boldsymbol{\theta}}\mathbf{q}(\boldsymbol{\theta}_0, \boldsymbol{\alpha}_0)$, $\mathbf{R} = \nabla'_{\boldsymbol{\alpha}}\mathbf{r}(\boldsymbol{\alpha}_0)$ and $(\mathbf{Q}'_{\boldsymbol{\alpha}}, \mathbf{R}')$, where $\mathbf{Q}_{\boldsymbol{\alpha}} = \nabla'_{\boldsymbol{\alpha}}\mathbf{q}(\boldsymbol{\theta}_0, \boldsymbol{\alpha}_0)$, are full row rank. These rank conditions are sufficient to guarantee the local independence of the restrictions (5.1) and the local identifiability of $\boldsymbol{\alpha}_0$ with respect to $\boldsymbol{\theta}_0$. We have assumed that any restrictions expressed in freedom equation form, *viz.* $\boldsymbol{\delta}_0 = \delta(\boldsymbol{\alpha}_0)$, have been substituted out. The parametric restrictions (5.1) are sufficently general to include other types of constraints; *viz.* freedom equation [Seber (1964)], $\boldsymbol{\theta}_0 = \theta(\boldsymbol{\alpha}_0)$, mixed form [Gouriéroux and Monfort (1989)], $\mathbf{q}(\boldsymbol{\theta}_0, \boldsymbol{\alpha}_0) = \mathbf{0}$, and the familiar constraint equation [Aitchison (1962) and Sargan (1980)], $\mathbf{r}(\boldsymbol{\alpha}_0) = \mathbf{0}$, restrictions.

The parametric restrictions $\mathbf{q}(\boldsymbol{\theta}_0, \boldsymbol{\alpha}_0) = \mathbf{0}$, $\mathbf{r}(\boldsymbol{\alpha}_0) = \mathbf{0}$ of (5.1) may be integrated directly into the framework of sub-section 4.2 by defining the corresponding log-GEL criterion incorporating the parametric restrictions (5.1) as

$$\mathcal{R}^*(\theta, \phi, \psi) = -T \ln T + \sum_{t=1}^{T} \ln h[\boldsymbol{\phi}'\mathbf{g}^{\omega}_{tT}(\boldsymbol{\theta}) + \boldsymbol{\psi}'\mathbf{q}(\boldsymbol{\theta}, \boldsymbol{\alpha}) + \boldsymbol{\eta}'\mathbf{r}(\boldsymbol{\alpha})]. \tag{5.2}$$

As previously in sub-section 4.2, we denote the saddle point estimators based on $\mathcal{R}^*(\boldsymbol{\varphi}, \boldsymbol{\alpha}, \boldsymbol{\psi}, \boldsymbol{\eta})$

of (5.2) by $(\tilde{\boldsymbol{\varphi}}_T, \tilde{\boldsymbol{\alpha}}_T, \tilde{\boldsymbol{\psi}}_T, \tilde{\boldsymbol{\eta}}_T)$. The associated score vector is

$$\mathbf{S}_{*T} = \sum_{t=1}^{T} \pi_t^*(\boldsymbol{\varphi}, \boldsymbol{\alpha}, \boldsymbol{\psi}, \boldsymbol{\eta}) \begin{pmatrix} \mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta}) \\ \mathbf{q}(\boldsymbol{\theta}, \boldsymbol{\alpha}) \\ \mathbf{r}(\boldsymbol{\alpha}) \\ \nabla_{\boldsymbol{\theta}} \mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})' \boldsymbol{\phi} + \nabla_{\boldsymbol{\theta}} \mathbf{q}(\boldsymbol{\theta}, \boldsymbol{\alpha})' \boldsymbol{\psi} \\ \nabla_{\boldsymbol{\alpha}} \mathbf{q}(\boldsymbol{\theta}, \boldsymbol{\alpha})' \boldsymbol{\psi} + \nabla_{\boldsymbol{\alpha}} \mathbf{r}(\boldsymbol{\alpha})' \boldsymbol{\eta} \end{pmatrix},$$

where, now,

$$\pi_t^*(\boldsymbol{\varphi}, \boldsymbol{\alpha}, \boldsymbol{\psi}, \boldsymbol{\eta}) \equiv \nabla \ln h[\boldsymbol{\phi}' \mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta}) + \boldsymbol{\psi}' \mathbf{q}(\boldsymbol{\theta}, \boldsymbol{\alpha}) + \boldsymbol{\eta}' \mathbf{r}(\boldsymbol{\alpha})].$$

From the first order conditions, cf. (2.8), the MGELE $(\tilde{\boldsymbol{\theta}}_T, \tilde{\boldsymbol{\alpha}}_T)$ satisfies $\mathbf{q}(\tilde{\boldsymbol{\theta}}_T, \tilde{\boldsymbol{\alpha}}_T) = \mathbf{0}$ and $\mathbf{r}(\tilde{\boldsymbol{\alpha}}_T) = \mathbf{0}$. Therefore, the weights $\pi_t^*(\tilde{\boldsymbol{\varphi}}_T, \tilde{\boldsymbol{\alpha}}_T, \tilde{\boldsymbol{\psi}}_T, \tilde{\boldsymbol{\eta}}_T) = \pi_t(\tilde{\boldsymbol{\varphi}}_T)$, $t = 1, ..., T$, and, hence, $\mathcal{R}^*(\tilde{\boldsymbol{\varphi}}_T, \tilde{\boldsymbol{\alpha}}_T, \tilde{\boldsymbol{\psi}}_T, \tilde{\boldsymbol{\eta}}_T) = \mathcal{R}(\tilde{\boldsymbol{\varphi}}_T)$.

It is straightforward to demonstrate the consistency of the estimators $(\tilde{\boldsymbol{\varphi}}_T, \tilde{\boldsymbol{\alpha}}_T, \tilde{\boldsymbol{\psi}}_T, \tilde{\boldsymbol{\eta}}_T)$ for $(\boldsymbol{\theta}_0, \mathbf{0}, \boldsymbol{\alpha}_0, \mathbf{0}, \mathbf{0})$ under (2.1) and (5.1); cf. Lemma 2.1. Moreover, the limiting distribution of the MGELE $(\tilde{\boldsymbol{\theta}}_T, \tilde{\boldsymbol{\alpha}}_T)$ and $(\tilde{\boldsymbol{\phi}}_T, \tilde{\boldsymbol{\psi}}_T, \tilde{\boldsymbol{\eta}}_T)$ may be derived along similar lines to Theorem 2.1; *viz.*

Theorem 5.1 *(Limiting Distribution of the MGELE $(\tilde{\boldsymbol{\theta}}_T, \tilde{\boldsymbol{\alpha}}_T)$ and $(\tilde{\boldsymbol{\phi}}_T, \tilde{\boldsymbol{\psi}}_T, \tilde{\boldsymbol{\eta}}_T)$.) Under (2.1) and (5.1), the MGELE $(\tilde{\boldsymbol{\theta}}_T, \tilde{\boldsymbol{\alpha}}_T)$ and $(\tilde{\boldsymbol{\phi}}_T, \tilde{\boldsymbol{\psi}}_T, \tilde{\boldsymbol{\eta}}_T)$ have limiting distribution given by*

$$T^{1/2} \begin{pmatrix} (\alpha/\tau)\nabla^2 \ln h(0) S_T^{-1} \tilde{\boldsymbol{\phi}}_T \\ (\alpha/\tau)\nabla^2 \ln h(0) S_T^{-1} \tilde{\boldsymbol{\psi}}_T \\ (\alpha/\tau)\nabla^2 \ln h(0) S_T^{-1} \tilde{\boldsymbol{\eta}}_T \\ \tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0 \\ \tilde{\boldsymbol{\alpha}}_T - \boldsymbol{\alpha}_0 \end{pmatrix} \to^L N \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{G}\mathbf{A}\mathbf{G}'\mathbf{V}^{-1} & \mathbf{D} & \mathbf{F} & \mathbf{0} & \mathbf{0} \\ \mathbf{D}' & \mathbf{C} & \mathbf{H} & \mathbf{0} & \mathbf{0} \\ \mathbf{F}' & \mathbf{H}' & \mathbf{E} & \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{A} & \mathbf{B} \\ \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & \mathbf{B}' & \mathbf{K} \end{pmatrix} \right), \tag{5.3}$$

where

$$\mathbf{A} = \mathbf{N} - \mathbf{N}\mathbf{Q}_{\boldsymbol{\theta}}'\mathbf{C}\mathbf{Q}_{\boldsymbol{\theta}}\mathbf{N}, \mathbf{B} = -\mathbf{N}\mathbf{Q}_{\boldsymbol{\theta}}'\mathbf{S}\mathbf{Q}_{\boldsymbol{\alpha}}\mathbf{K}, \mathbf{C} = \mathbf{S} - \mathbf{S}\mathbf{Q}_{\boldsymbol{\alpha}}\mathbf{K}\mathbf{Q}_{\boldsymbol{\alpha}}'\mathbf{S},$$
$$\mathbf{D} = -\mathbf{V}^{-1}\mathbf{G}\mathbf{N}\mathbf{Q}_{\boldsymbol{\theta}}'\mathbf{C}, \mathbf{E} = (\mathbf{R}\mathbf{M}\mathbf{R}')^{-1}\mathbf{R}\mathbf{M}\mathbf{Q}_{\boldsymbol{\alpha}}'\mathbf{S}\mathbf{Q}_{\boldsymbol{\alpha}}\mathbf{M}\mathbf{R}'(\mathbf{R}\mathbf{M}\mathbf{R}')^{-1},$$
$$\mathbf{F} = \mathbf{V}^{-1}\mathbf{G}\mathbf{N}\mathbf{Q}_{\boldsymbol{\theta}}'\mathbf{S}\mathbf{Q}_{\boldsymbol{\alpha}}\mathbf{M}\mathbf{R}'(\mathbf{R}\mathbf{M}\mathbf{R}')^{-1}, \mathbf{H} = -\mathbf{C}\mathbf{Q}_{\boldsymbol{\alpha}}\mathbf{M}\mathbf{R}'(\mathbf{R}\mathbf{M}\mathbf{R}')^{-1},$$

and $\mathbf{N} = (\mathbf{G}'\mathbf{V}^{-1}\mathbf{G})^{-1}$, $\mathbf{S} = (\mathbf{Q}_{\boldsymbol{\theta}}\mathbf{N}\mathbf{Q}_{\boldsymbol{\theta}}')^{-1}$, $\mathbf{M} = (\mathbf{Q}_{\boldsymbol{\alpha}}'\mathbf{S}\mathbf{Q}_{\boldsymbol{\alpha}} + \mathbf{R}'\mathbf{R})^{-1}$, $\mathbf{K} = \mathbf{M} - \mathbf{M}\mathbf{R}'(\mathbf{R}\mathbf{M}\mathbf{R}')^{-1}\mathbf{R}\mathbf{M}$.

The corresponding GELR statistic for (5.1) is

$$\mathcal{GELR} = 2(\alpha/\tau^2)\nabla^2 \ln h(0) S_T^{-1} \left(\mathcal{R}(\hat{\boldsymbol{\varphi}}_T) - \mathcal{R}(\tilde{\boldsymbol{\varphi}}_T)\right), \tag{5.4}$$

whereas from (5.3), because from the first order conditions $\tilde{\boldsymbol{\eta}}_T$ is linearly dependent on $\tilde{\boldsymbol{\psi}}_T$, the GELL Wald statistic is given by

$$\mathcal{GELW} = [(\alpha/\tau)\nabla^2 \ln h(0) T^{1/2} S_T^{-1}]^2 \tilde{\boldsymbol{\psi}}_T' \tilde{\mathbf{Q}}_{\boldsymbol{\theta}T} (\tilde{\mathbf{G}}_T' \tilde{\mathbf{V}}_T^{-1} \tilde{\mathbf{G}}_T)^{-1} \tilde{\mathbf{Q}}_{\boldsymbol{\theta}T}' \tilde{\boldsymbol{\psi}}_T, \tag{5.5}$$

where $\tilde{\mathbf{Q}}_{\boldsymbol{\theta}T} = \nabla_{\boldsymbol{\theta}}' \mathbf{q}(\tilde{\boldsymbol{\theta}}_T, \tilde{\boldsymbol{\alpha}}_T)$ and $\tilde{\mathbf{V}}_T$ and $\tilde{\mathbf{G}}_T$ are consistent estimators for $\mathbf{V}$ and $\mathbf{G}$ respectively using the MGELE $(\tilde{\boldsymbol{\theta}}_T, \tilde{\boldsymbol{\alpha}}_T)$ and $(\tilde{\boldsymbol{\phi}}_T, \tilde{\boldsymbol{\psi}}_T, \tilde{\boldsymbol{\eta}}_T)$; cf. (2.16) and (2.17).

A similar problem is encountered in the derivation of the GEL score statistic as arose in sub-section 4.1; namely, $\boldsymbol{\alpha}_0$ is unidentified under (2.1). Hence, the score $\mathbf{S}_{*T}$ is evaluated at $(\hat{\boldsymbol{\theta}}_T, \tilde{\boldsymbol{\alpha}}_T)$ and $(\hat{\boldsymbol{\phi}}_T, \mathbf{0}, \mathbf{0})$. The only non-zero block of the estimated score vector is $T^{1/2}\mathbf{q}(\hat{\boldsymbol{\theta}}_T, \tilde{\boldsymbol{\alpha}}_T)$, the other blocks being zero. Hence, the GEL score statistic for (5.1) is given by

$$\mathcal{SPEQS} = T\mathbf{q}(\hat{\boldsymbol{\theta}}_T, \tilde{\boldsymbol{\alpha}}_T)' \left(\hat{\mathbf{Q}}_{\boldsymbol{\theta}T}(\hat{\mathbf{G}}_T'\hat{\mathbf{V}}_T^{-1}\hat{\mathbf{G}}_T)^{-1}\hat{\mathbf{Q}}_{\boldsymbol{\theta}T}'\right)^{-1} \mathbf{q}(\hat{\boldsymbol{\theta}}_T, \tilde{\boldsymbol{\alpha}}_T) \tag{5.6}$$

where $\hat{\mathbf{V}}_T$ and $\hat{\mathbf{G}}_T$ are defined in (2.16) and (2.17) respectively and $\hat{\mathbf{Q}}_{\boldsymbol{\theta}T} = \nabla_{\boldsymbol{\theta}}'\mathbf{q}(\hat{\boldsymbol{\theta}}_T, \tilde{\boldsymbol{\alpha}}_T)$.[12] The forms of the GEL Wald (5.5) and score (5.6) statistics are, respectively, those of Lagrange multiplier and Wald statistics in the GMM framework.

Proposition 5.1 *(Limiting Distribution of GEL Statistics for Mixed Implicit and Constraint Equation Parametric Restrictions.) Under (2.1) and (5.1), the GEL statistics $\mathcal{GELR}$, $\mathcal{GELW}$ and $\mathcal{GELS}$ statistics of (5.4), (5.5) and (5.6) respectively each have limiting distributions described by*

$$\mathcal{GELR}, \mathcal{GELW}, \mathcal{GELS} \to^L \chi^2(s + r - q).$$

As the MGELE $\hat{\boldsymbol{\theta}}_T$ is first order equivalent to the optimal GMM estimator, not only are all three statistics $\mathcal{GELR}$, $\mathcal{GELW}$ and $\mathcal{GELS}$ asymptotically equivalent [see Proof of Proposition 5.1] but they are equivalent to GMM tests for $\mathbf{q}(\boldsymbol{\theta}_0, \boldsymbol{\alpha}_0) = \mathbf{0}$, $\mathbf{r}(\boldsymbol{\alpha}_0) = \mathbf{0}$ of (5.1); cf. (5.6).

The Proof of Proposition 4.5 also shows that the following minimum chi-squared statistics are first order equivalent to $\mathcal{GELR}$, $\mathcal{GELW}$ and $\mathcal{GELS}$:

$$\mathcal{GELMC}_\phi = [(\alpha/\tau)\nabla^2 \ln h(0) T^{1/2} S_T^{-1}]^2 (\tilde{\boldsymbol{\phi}}_T - \hat{\boldsymbol{\phi}}_T)'\hat{\mathbf{V}}_T(\tilde{\boldsymbol{\phi}}_T - \hat{\boldsymbol{\phi}}_T),$$

$$\mathcal{GELMC}_{\boldsymbol{\theta}} = T(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T)'\hat{\mathbf{G}}_T'\hat{\mathbf{V}}_T^{-1}\hat{\mathbf{G}}_T(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T),$$

as is the statistic $[(\alpha/\tau)\nabla^2 \ln h(0) T^{1/2} S_T^{-1}]^2 \left(\tilde{\boldsymbol{\phi}}_T'\tilde{\mathbf{V}}_T\tilde{\boldsymbol{\phi}}_T - \hat{\boldsymbol{\phi}}_T'\hat{\mathbf{V}}_T\hat{\boldsymbol{\phi}}_T\right)$. Moreover, differences of GEL Wald and score statistics for $\boldsymbol{\phi} = \mathbf{0}, \boldsymbol{\psi} = \mathbf{0}, \boldsymbol{\eta} = \mathbf{0}$ and $\boldsymbol{\phi} = \mathbf{0}$ are also equivalent to the above statistics as is seen from the structure of the statistic $\mathcal{GELR}$ of (5.4) but, of course, may be prone to the difficulty of not possessing positive support as described in sub-section 4.2.

6 Generalized Empirical Likelihood Non-Nested Tests

Denote the model embodied in the moment conditions $E_\mu\{\mathbf{g}(\mathbf{x}_t; \boldsymbol{\theta}_0)\} = \mathbf{0}$, $t = 0, \pm 1, ...$, of (2.1) by H_g. Consider an alternative model H_q based on the assumed m_q moment conditions

$$E_\mu\{\mathbf{q}(\mathbf{x}_t; \boldsymbol{\beta}_0)\} = \mathbf{0}, \tag{6.1}$$

[12] In fact, any $T^{1/2}$-consistent estimator for α_0 may be used in place of $\tilde{\alpha}_T$ in (5.6) with no alteration to the following result.

$t = 0, \pm 1, \ldots$. The log-GEL criterion under H_q of (6.1) is denoted as

$$\mathcal{R}_q(\boldsymbol{\zeta}) = -T \ln T + \sum_{t=1}^{T} \ln h[\boldsymbol{\delta}' \mathbf{q}_{tT}^{\omega}(\boldsymbol{\beta})], \tag{6.2}$$

where $\boldsymbol{\zeta} = (\boldsymbol{\beta}', \boldsymbol{\delta}')'$ and

$$\mathbf{q}_{tT}^{\omega}(\boldsymbol{\beta}) = \sum_{s=-[(T-1)/2]}^{[(T-1)/2]} \omega(s; S_T) \mathbf{q}(\mathbf{x}_{t-s}; \boldsymbol{\beta}),$$

$t = 0, \pm 1, \ldots$. For convenience of exposition, the carrier function $h(.)$, the weights $\{\omega(s; S_T)\}$ and bandwidth parameter S_T are chosen identically to those under H_g; see sub-sections 2.2 and 2.3.

Correspondingly, the saddle point estimator for $\boldsymbol{\delta}$ given $\boldsymbol{\beta}$ is defined by

$$\hat{\boldsymbol{\delta}}_T(\boldsymbol{\beta}) = \arg\min_{\boldsymbol{\delta}} \mathcal{R}_q(\boldsymbol{\zeta})$$

and the MGELE $\hat{\boldsymbol{\beta}}_T$ for $\boldsymbol{\beta}$ by

$$\hat{\boldsymbol{\beta}}_T = \arg\max_{\boldsymbol{\beta} \in B} \mathcal{R}_h(\boldsymbol{\beta}, \hat{\boldsymbol{\delta}}_T(\boldsymbol{\beta})) = \arg\max_{\boldsymbol{\beta} \in B} \min_{\boldsymbol{\delta}} \mathcal{R}_q(\boldsymbol{\zeta}). \tag{6.3}$$

Under H_g of (2.1), the pseudo-true value (PTV) of $\hat{\boldsymbol{\delta}}_T \equiv \hat{\boldsymbol{\delta}}_T(\hat{\boldsymbol{\beta}}_T)$ is defined by

$$\boldsymbol{\delta}_*(\boldsymbol{\beta}) = \arg\min_{\boldsymbol{\delta}} \lim_{T \to \infty} E_\mu\{\ln h[\boldsymbol{\delta}' \mathbf{q}_{tT}^{\omega}(\boldsymbol{\beta})]\}.$$

The PTV for the MGELE $\hat{\boldsymbol{\beta}}_T$ is given by

$$\boldsymbol{\beta}_* = \arg\max_{\boldsymbol{\beta} \in B} \min_{\boldsymbol{\delta}} \lim_{T \to \infty} E_\mu\{\ln h[\boldsymbol{\delta}' \mathbf{q}_{tT}^{\omega}(\boldsymbol{\beta})]\};$$

in the discussion below, $\boldsymbol{\delta}_*$ denotes $\boldsymbol{\delta}_*(\boldsymbol{\beta}_*)$. Therefore, by a similar argument to the Proof of Lemma 2.1, under H_g of (2.1),

$$\hat{\boldsymbol{\beta}}_T \to^P \boldsymbol{\beta}_*, \hat{\boldsymbol{\delta}}_T \to^P \boldsymbol{\delta}_*.$$

In order to avoid the difficulty of observational equivalence between the log-GEL of H_g and H_h, it is assumed that $\lim_{T \to \infty} E_\mu\{\ln h[\boldsymbol{\delta}_*' \mathbf{q}_{tT}^{\omega}(\boldsymbol{\beta}_*)]\} < 0$. Hence, $\boldsymbol{\delta}_* \neq \mathbf{0}$.[13]

6.1 Cox-Type Non-Nested Tests

The usual approach to constructing Cox-type (1961, 1962) tests of H_g of (2.1) against H_q of (6.1) involves the contrast of consistent estimators under H_g of the probability limit of the requisite criterion function under H_q evaluated at the corresponding estimator PTV

[13]As in the Proof of Lemma 2.1, the minimising PTV $\boldsymbol{\delta}_*(\boldsymbol{\beta})$ is unique for given $\boldsymbol{\beta}$.

viz. $\lim_{T\to\infty} E_\mu\{\ln h[\boldsymbol{\delta}_*'\mathbf{q}_{tT}^\omega(\boldsymbol{\beta}_*)]\}$; see Smith (1992, equation (2.3), p.973). Clearly, the log-GEL (6.3) evaluated at $\hat{\boldsymbol{\zeta}}_T = (\hat{\boldsymbol{\beta}}_T', \hat{\boldsymbol{\delta}}_T')'$, $T^{-1}\mathcal{R}_q(\hat{\boldsymbol{\zeta}}_T)$, less $T^{-1}\mathcal{R}_q(\hat{\boldsymbol{\beta}}_T, \mathbf{0})$ provides one such consistent estimator. To provide another consistent estimator to contrast with $T^{-1}\mathcal{R}_q(\hat{\boldsymbol{\zeta}}_T) - T^{-1}\mathcal{R}_q(\hat{\boldsymbol{\beta}}_T, \mathbf{0})$, recall that the implied probabilities $\pi_t(\hat{\boldsymbol{\varphi}}_T)/\sum_{t=1}^T \pi_t(\hat{\boldsymbol{\varphi}}_T) = T^{-1}(1 + o_P(1))$, $t = 1, ..., T$, under (2.1). Hence, consider the alternative log-GEL criterion

$$\mathcal{R}_q^*(\boldsymbol{\zeta}) = -T\ln T + T\sum_{t=1}^T \frac{\pi_t(\hat{\boldsymbol{\varphi}}_T)}{\sum_{t=1}^T \pi_t(\hat{\boldsymbol{\varphi}}_T)} \ln h[\boldsymbol{\delta}'\mathbf{q}_{tT}^\omega(\boldsymbol{\beta})], \tag{6.4}$$

and denote the corresponding saddle point estimator by $\tilde{\boldsymbol{\zeta}}_T$. Therefore, as $\hat{\boldsymbol{\varphi}}_T$ is consistent for $(\theta_0, \mathbf{0})$ under H_g, $\tilde{\boldsymbol{\zeta}}_T$ is consistent for $\boldsymbol{\zeta}_* = (\boldsymbol{\beta}_*', \boldsymbol{\delta}_*')'$. Hence, an alternative consistent estimator for $\lim_{T\to\infty} E_\mu\{\ln h[\boldsymbol{\delta}_*'\mathbf{q}_{tT}^\omega(\boldsymbol{\beta}_*)]\}$ is obtained by evaluation of (6.4) at $\tilde{\boldsymbol{\zeta}}_T$;

$$T^{-1}\mathcal{R}_q^*(\tilde{\boldsymbol{\zeta}}_T) = -\ln T + \sum_{t=1}^T \frac{\pi_t(\hat{\boldsymbol{\varphi}}_T)}{\sum_{t=1}^T \pi_t(\hat{\boldsymbol{\varphi}}_T)} \ln h[\tilde{\boldsymbol{\delta}}_T'\mathbf{q}_{tT}^\omega(\tilde{\boldsymbol{\beta}}_T)].$$

Consequently, a GEL Cox-type statistic for H_g against H_q is based on the contrast between the optimised log-GEL criteria, $\mathcal{R}_q(\hat{\boldsymbol{\zeta}}_T)$ of (6.2) and $\mathcal{R}_q^*(\tilde{\boldsymbol{\zeta}}_T)$ of (6.4); *viz.*

$$\mathcal{C}(H_g|H_h) = (\alpha/\tau)\nabla^2 \ln h(0) T^{-1/2} S_T^{-1}\left(\mathcal{R}_q(\hat{\boldsymbol{\zeta}}_T) - \mathcal{R}_q^*(\tilde{\boldsymbol{\zeta}}_T)\right) \tag{6.5}$$

$$= (\alpha/\tau)\nabla^2 \ln h(0) T^{1/2} S_T^{-1}\sum_{t=1}^T\left(T^{-1}\ln h[\hat{\boldsymbol{\delta}}_T'\mathbf{q}_{tT}^\omega(\hat{\boldsymbol{\beta}}_T)] - \frac{\pi_t(\hat{\boldsymbol{\varphi}}_T)}{\sum_{t=1}^T \pi_t(\hat{\boldsymbol{\varphi}}_T)}\ln h[\tilde{\boldsymbol{\delta}}_T'\mathbf{q}_{tT}^\omega(\tilde{\boldsymbol{\beta}}_T)]\right).$$

Theorem 6.1 *(Limiting Distribution of the GEL Cox-type Statistic.) Under (2.1), the Cox statistic $\mathcal{C}(H_g|H_h)$ of (6.5) has limiting distribution described by*

$$\mathcal{C}(H_g|H_h) \to^L N(0, \sigma^2),$$

where

$$\sigma^2 = \boldsymbol{\xi}_*'\left(\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{G}(\mathbf{G}'\mathbf{V}^{-1}\mathbf{G})^{-1}\mathbf{G}'\mathbf{V}^{-1}\right)\boldsymbol{\xi}_*,$$
$$\boldsymbol{\xi}_* = \lim_{T\to\infty} E_\mu\{\ln h[\boldsymbol{\delta}_*'\mathbf{q}_{tT}^\omega(\boldsymbol{\beta}_*)]\mathbf{g}_{tT}^\omega(\boldsymbol{\theta}_0)\},$$

and σ^2 is assumed non-zero.

The asymptotic variance σ^2 may be consistently estimated under H_g by

$$\hat{\sigma}_T^2 = \hat{\boldsymbol{\xi}}_T'\left(\hat{\mathbf{V}}_T^{-1} - \hat{\mathbf{V}}_T^{-1}\hat{\mathbf{G}}_T(\hat{\mathbf{G}}_T'\hat{\mathbf{V}}_T^{-1}\hat{\mathbf{G}}_T)^{-1}\hat{\mathbf{G}}_T'\hat{\mathbf{V}}_T^{-1}\right)\hat{\boldsymbol{\xi}}_T,$$

where

$$\hat{\boldsymbol{\xi}}_T = T^{-1}\sum_{t=1}^T \ln h[\hat{\boldsymbol{\delta}}_T'\mathbf{q}_{tT}^\omega(\hat{\boldsymbol{\beta}}_T)]\mathbf{g}_{tT}^\omega(\hat{\boldsymbol{\theta}}_T) \tag{6.6}$$

and $\hat{\mathbf{V}}_T$ and $\hat{\mathbf{G}}_T$ are described in (2.16) and (2.17) respectively.

The Proof of Theorem 6.1 suggests an alternative first order equivalent linearised Cox-type statistic, which may be regarded as a Wald-type form:

$$\mathcal{LC}(H_g|H_h) = \hat{\boldsymbol{\xi}}_T'(\alpha/\tau)\nabla^2 \ln h(0)T^{1/2}S_T^{-1}\hat{\boldsymbol{\phi}}_T, \tag{6.7}$$

where $\hat{\boldsymbol{\xi}}_T$ is defined in (6.6); cf. Smith (1992, Section 2.1, pp.974-6).

Corollary 6.1 *(Limiting Distribution of the Linearised GEL Cox-type Statistic.) Under (2.1), the linearised Cox statistic $\mathcal{LC}(H_g|H_h)$ of (6.7) for H_g against H_q has limiting distribution described by*

$$\mathcal{LC}(H_g|H_h) \to^L N(0, \sigma^2),$$

where σ^2, defined in Theorem 6.1, is assumed non-zero.

Because $[(\alpha/\tau)\nabla^2 \ln h(0)T^{1/2}S_T^{-1}]\hat{\boldsymbol{\phi}}_T = -\mathbf{V}^{-1}T^{-1/2}\sum_{t=1}^T \mathbf{g}(\mathbf{x}_t; \hat{\boldsymbol{\theta}}_T) + o_P(1)$, the form of the linearised statistic $\mathcal{LC}(H_g|H_h)$ emphasises that non-nested tests of competing hypotheses expressed in moment form are (asymptotically) particular linear combinations of the estimated sample moment vector under the null hypothesis as this vector represents the sole information feasible and available for inference purposes; cf. Smith (1992). Hence, an alternative score form of the Cox-type statistic for H_g against H_q with identical first order asymptotic properties to those above is

$$\tau^{-1}\hat{\boldsymbol{\xi}}_T'\hat{\mathbf{V}}_T^{-1}T^{-1/2}\sum_{t=1}^T \mathbf{g}_{tT}^{\omega}(\hat{\boldsymbol{\theta}}_T).$$

A simplified form of GEL Cox-type statistic evaluates the criterion $\mathcal{R}_q^*(\boldsymbol{\zeta})$ at $\hat{\boldsymbol{\zeta}}_T$ and is also equivalent to the above Cox-type statistics, $\mathcal{C}(H_g|H_h)$ of (6.5) and $\mathcal{LC}(H_g|H_h)$ of (6.7); *viz.*

$$\mathcal{SC}(H_g|H_h) = [(\alpha/\tau)\nabla^2 \ln h(0)T^{-1/2}S_T^{-1}]\left(\mathcal{R}_q(\hat{\boldsymbol{\zeta}}_T) - \mathcal{R}_q^*(\hat{\boldsymbol{\zeta}}_T)\right) \tag{6.8}$$

$$= [(\alpha/\tau)\nabla^2 \ln h(0)T^{1/2}S_T^{-1}]\sum_{t=1}^T \left(T^{-1} - \frac{\pi_t(\hat{\boldsymbol{\varphi}}_T)}{\sum_{t=1}^T \pi_t(\hat{\boldsymbol{\varphi}}_T)}\right)\ln h[\hat{\boldsymbol{\delta}}_T'\mathbf{q}_{tT}^{\omega}(\hat{\boldsymbol{\beta}}_T)].$$

Proposition 6.1 *(Limiting Distribution of the Simplified GEL Cox-type Statistic.) Under (2.1), the simplified Cox statistic $\mathcal{SC}(H_g|H_h)$of (6.8) for H_g against H_q has limiting distribution described by*

$$\mathcal{SC}(H_g|H_h) \to^L N(0, \sigma^2),$$

where σ^2, defined in Theorem 6.1, is assumed non-zero.

6.2 Encompassing Non-Nested Tests

To be completed

7 Summary

By utilising and adapting the approach of Chesher and Smith (1997) which descibes likelihood ratio tests for implied moment conditions in fully parametric models, this paper has introduced a class of generalized empirical likelihood procedures for time-series models as alternative criteria to generalized method of moments estimation and inference. This class of criteria includes empirical likelihood [Imbens (1993), Qin and Lawless (1995), Smith (1995)], empirical information or exponential tilting [Imbens, Johnson and Spady (1995), Kitamura and Stutzer (1995)] and the Cressie-Read power divergence criteria as special cases. The asymptotic properties of the associated estimators and statistics are identical to those derived from the generalized method of moments. Being likelihood-like in construction, similar statistics to the classical likelihood ratio, Wald and score statistics may be defined. Non-nested Cox-type tests may also be obtained.

Given that the asymptotic distribution may poorly approximate the finite sample distribution of the efficient generalized method of moments estimator, a future avenue for research is to explore which choices of carrier function can ameliorate this problem. Furthermore, given the likelihood nature of the suggested estimators and statistics, one may be able to adapt the existing theory of Edgeworth expansions to improve the quality of asymptotic approximation or to detect circumstances in which the asymptotic approximations are likely to be poor. This research agenda is currently being investigated by the author.

Appendix

Second Derivatives

$$\nabla_{\phi\phi}\mathcal{R}(\boldsymbol{\varphi})=\sum_{t=1}^{T}\nabla^2\ln h[\boldsymbol{\phi}'\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})]\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})',$$

which is positive definite if $\sum_{t=1}^{T}\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})'$ is positive definite as $\nabla^2\ln h[\boldsymbol{\phi}'\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})]>0$ by the strict convexity of $\ln h(.)$. Hence, $\boldsymbol{\phi}=\boldsymbol{\phi}(\boldsymbol{\theta})$ defines a unique minimum of $\mathcal{R}(\boldsymbol{\varphi})$ for given $\boldsymbol{\theta}$ and is continuously differentiable by the implicit function theorem. Moreover,

$$\nabla_{\phi\boldsymbol{\theta}}\mathcal{R}(\boldsymbol{\varphi})=\sum_{t=1}^{T}\left(\nabla^2\ln h[\boldsymbol{\phi}'\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})]\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})\boldsymbol{\phi}'+\nabla\ln h[\boldsymbol{\phi}'\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})]\mathbf{I}_r\right)\nabla_{\boldsymbol{\theta}}'\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta}).$$

Hence,

$$\nabla_{\boldsymbol{\theta}}'\boldsymbol{\phi}(\boldsymbol{\theta})=-\left(\nabla_{\phi\phi}\mathcal{R}(\boldsymbol{\theta},\boldsymbol{\phi}(\boldsymbol{\theta}))\right)^{-1}\nabla_{\phi\boldsymbol{\theta}}\mathcal{R}(\boldsymbol{\theta},\boldsymbol{\phi}(\boldsymbol{\theta})).$$

$$\nabla_{\boldsymbol{\theta\theta}}\mathcal{R}(\boldsymbol{\varphi})=\sum_{t=1}^{T}\left(\nabla^2\ln h[\boldsymbol{\phi}'\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})]\nabla_{\boldsymbol{\theta}}\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})'\boldsymbol{\phi}\boldsymbol{\phi}'\nabla_{\boldsymbol{\theta}}'\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})+\nabla\ln h[\boldsymbol{\phi}'\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})]\sum_{i=1}^{r}\nabla_{\boldsymbol{\theta\theta}}\mathbf{g}_{tT,i}^{\omega}(\boldsymbol{\theta})\phi_i\right).$$

Therefore,

$$\nabla_{\boldsymbol{\theta\theta}}\mathcal{R}(\boldsymbol{\theta},\boldsymbol{\phi}(\boldsymbol{\theta}))=\nabla_{\boldsymbol{\theta\theta}}\mathcal{R}(\boldsymbol{\theta},\boldsymbol{\phi}(\boldsymbol{\theta}))-\nabla_{\boldsymbol{\theta}\phi}\mathcal{R}(\boldsymbol{\theta},\boldsymbol{\phi}(\boldsymbol{\theta}))\left(\nabla_{\phi\phi}\mathcal{R}(\boldsymbol{\theta},\boldsymbol{\phi}(\boldsymbol{\theta}))\right)^{-1}\nabla_{\phi\boldsymbol{\theta}}\mathcal{R}(\boldsymbol{\theta},\boldsymbol{\phi}(\boldsymbol{\theta})).$$

At $\boldsymbol{\phi}(\boldsymbol{\theta})=\mathbf{0}$, $\nabla_{\boldsymbol{\theta\theta}}\mathcal{R}(\boldsymbol{\theta},\mathbf{0})=-\nabla_{\boldsymbol{\theta}\phi}\mathcal{R}(\boldsymbol{\theta},\mathbf{0}))\left(\nabla_{\phi\phi}\mathcal{R}(\boldsymbol{\theta},\mathbf{0}))\right)^{-1}\nabla_{\phi\boldsymbol{\theta}}\mathcal{R}(\boldsymbol{\theta},\mathbf{0}))$ which is negative definite if $\sum_{t=1}^{T}\nabla_{\boldsymbol{\theta}}'\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})$ is full column rank indicating that a local maximum will exist.

Proofs of Results

Proof of Lemma 2.1: Define $\mathcal{M}_T(\boldsymbol{\varphi})=E_\mu\{\ln h[\boldsymbol{\phi}'\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})]\}$, where from (2.1) $E_\mu\{\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})\}=\mathbf{0}\Leftrightarrow\boldsymbol{\theta}=\boldsymbol{\theta}_0$. Now, $E_\mu\{\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})'\}$ is positive definite if $E_\mu\{\mathbf{g}(\mathbf{x}_t;\boldsymbol{\theta})\mathbf{g}(\mathbf{x}_t;\boldsymbol{\theta})'\}$ is positive definite. Assume that $E_\mu\{\nabla^2\ln h[\boldsymbol{\phi}'\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})]\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})'\}$ exists and define

$$\boldsymbol{\phi}_T(\boldsymbol{\theta})=\arg\min_{\boldsymbol{\phi}}\mathcal{M}_T(\boldsymbol{\varphi}).$$

Hence, $\boldsymbol{\phi}_T(\boldsymbol{\theta})$ satisfies

$$E_\mu\{\nabla\ln h[\boldsymbol{\phi}_T(\boldsymbol{\theta})'\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})]\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})\}=\mathbf{0}.$$

It follows from the Lemma and the Proof of Theorem 1 in the Appendix of Chesher and Smith (1995) that $\boldsymbol{\phi}_T(\boldsymbol{\theta})$ is unique. Hence, as $\boldsymbol{\theta}_0$ uniquely satisfies the moment conditions (2.1), $\boldsymbol{\phi}_T(\boldsymbol{\theta}_0)=\mathbf{0}$ and $\boldsymbol{\phi}_T(\boldsymbol{\theta})\neq\mathbf{0}$, $\boldsymbol{\theta}\neq\boldsymbol{\theta}_0$. Now, $0=\mathcal{M}_T(\boldsymbol{\theta},\boldsymbol{\phi}_T(\boldsymbol{\theta}_0))\geq\mathcal{M}_T(\boldsymbol{\theta},\boldsymbol{\phi}(\boldsymbol{\theta}))$, $\boldsymbol{\theta}\neq\boldsymbol{\theta}_0$, and, as $\boldsymbol{\phi}_T(\boldsymbol{\theta})$ is a unique minimiser, $0=\mathcal{M}_T(\boldsymbol{\theta},\boldsymbol{\phi}_T(\boldsymbol{\theta}_0))>\mathcal{M}_T(\boldsymbol{\theta},\boldsymbol{\phi}(\boldsymbol{\theta}))$, $\boldsymbol{\theta}\neq\boldsymbol{\theta}_0$. Therefore, $\boldsymbol{\theta}_0$ is the unique maximiser of $\mathcal{M}_T(\boldsymbol{\theta},\boldsymbol{\phi}(\boldsymbol{\theta}))$; *viz.*

$$\boldsymbol{\theta}_0=\arg\max_{\boldsymbol{\theta}\in\Theta}\mathcal{M}_T(\boldsymbol{\theta},\boldsymbol{\phi}(\boldsymbol{\theta})),$$

and $\mathcal{M}_T(\boldsymbol{\theta}_0,\boldsymbol{\phi}_T(\boldsymbol{\theta}_0))=0$.

Define the neighbourhood of $\boldsymbol{\theta}_0$, $\mathcal{N}_\delta(\boldsymbol{\theta}_0)$, and a sequence of open neighbourhoods, $\mathcal{N}_{\delta_j}(\boldsymbol{\theta}_j)$, $\boldsymbol{\theta}_j\in\Theta$, $j=1,...,J$, such that $\cup_{j=1}^{J}\mathcal{N}_{\delta_j}(\boldsymbol{\theta}_j)$ covers $\Theta-\mathcal{N}_\delta(\boldsymbol{\theta}_0)$. From continuity and the Dominated Convergence Theorem

$$E_\mu\{\sup_{\boldsymbol{\theta}'\in\mathcal{N}_{\delta_j}(\boldsymbol{\theta}_j)}\ln h[\boldsymbol{\phi}_T(\boldsymbol{\theta})'\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})]\}=-2\nu_j<0,j=1,...,J.$$

A point-wise double array WLLN gives:

$$\Pr\{T^{-1}\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}'\in\mathcal{N}_{\delta_j}(\boldsymbol{\theta}_j)}\ln h[\boldsymbol{\phi}_T(\boldsymbol{\theta})'\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})]>-\nu_j\}<\epsilon/2J, j=1,...,J.$$

Hence,

$$\Pr\{\sup_{\boldsymbol{\theta}'\in\Theta-\mathcal{N}_{\delta}(\boldsymbol{\theta}_0)}T^{-1}\sum_{t=1}^{T}\ln h[\boldsymbol{\phi}_T(\boldsymbol{\theta})'\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})]>-\nu\}<\epsilon/2,$$

where $\nu=\min_j\nu_j$. The saddle-point minimisation property of $\hat{\boldsymbol{\phi}}_T(.)$ implies

$$\Pr\{\sup_{\boldsymbol{\theta}'\in\Theta-\mathcal{N}_{\delta}(\boldsymbol{\theta}_0)}T^{-1}\sum_{t=1}^{T}\ln h[\hat{\boldsymbol{\phi}}_T(\boldsymbol{\theta})'\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta})]>-\nu\}<\epsilon/2. \tag{A.1}$$

Noting by convexity

$$\ln h[\hat{\boldsymbol{\phi}}_T(\boldsymbol{\theta}_0)'T^{-1}\sum_{t=1}^{T}\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta}_0)]\leq T^{-1}\sum_{t=1}^{T}\ln h[\hat{\boldsymbol{\phi}}_T(\boldsymbol{\theta}_0)'\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta}_0)]\leq 0,$$

and as $T^{-1}\sum_{t=1}^{T}\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta}_0)=T^{-1}\sum_{t=1}^{T}\mathbf{g}(\mathbf{x}_t;\boldsymbol{\theta}_0)+o_P(1)\to^P\mathbf{0}$, we have $\hat{\boldsymbol{\phi}}_T(\boldsymbol{\theta}_0)\to^P\mathbf{0}$ and $T^{-1}\sum_{t=1}^{T}\ln h[\hat{\boldsymbol{\phi}}_T(\boldsymbol{\theta}_0)'\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta}_0)]\to^P 0$. Therefore

$$\Pr\{T^{-1}\sum_{t=1}^{T}\ln h[\hat{\boldsymbol{\phi}}_T(\boldsymbol{\theta}_0)'\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta}_0)]<-\nu/2\}<\epsilon/2. \tag{A.2}$$

Combining (A.1) and (A.2), we have

$$\Pr\{\hat{\boldsymbol{\theta}}_T\in\mathcal{N}_{\delta}(\boldsymbol{\theta}_0)\}>1-\epsilon.$$

Proof of Theorem 2.1: Consider the first order conditions determining $\hat{\theta}_T$ and $\hat{\phi}_T$

$$\sum_{t=1}^{T}\pi_t(\hat{\boldsymbol{\varphi}}_T)\begin{pmatrix}\mathbf{g}_{tT}^{\omega}(\hat{\boldsymbol{\theta}}_T)\\ \nabla_{\theta}\mathbf{g}_{tT}^{\omega}(\hat{\boldsymbol{\theta}}_T)'\hat{\boldsymbol{\phi}}_T\end{pmatrix}=\begin{pmatrix}\mathbf{0}\\ \mathbf{0}\end{pmatrix}. \tag{A.3}$$

Now, defining $\tau\equiv\int k(x)dx$,

$$\begin{aligned}T^{-1}\sum_{t=1}^{T}\nabla_{\theta}\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta}_0)'&=\tau T^{-1}\sum_{t=1}^{T}\nabla_{\theta}\mathbf{g}(\mathbf{x}_t;\boldsymbol{\theta}_0)'+o_P(1)\\ &=\tau\mathbf{G}+o_P(1).\end{aligned}$$

Therefore, a first order Taylor series expansion of (A.3) about θ_0 and $\mathbf{0}$ yields

$$\begin{pmatrix}o_P(1)\\ o_P(1)\end{pmatrix}=\begin{pmatrix}T^{-1/2}\sum_{t=1}^{T}\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta}_0)\\ \mathbf{0}\end{pmatrix}+\begin{pmatrix}\nabla^2\ln h(0)\mathbf{V} & \tau\mathbf{G}\\ \tau\mathbf{G}' & \mathbf{0}\end{pmatrix}T^{1/2}\begin{pmatrix}\alpha S_T^{-1}\hat{\boldsymbol{\phi}}_T\\ \hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0\end{pmatrix}.$$

Moreover, by a similar argument to that above,

$$T^{-1/2}\sum_{t=1}^{T}\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta}_0)=\tau T^{-1/2}\sum_{t=1}^{T}\mathbf{g}(\mathbf{x}_t;\boldsymbol{\theta}_0)+o_P(1).$$

Therefore, as $\mathbf{G}'\alpha T^{1/2}S_T^{-1}\hat{\boldsymbol{\phi}}_T=o_P(1)$,

$$T^{1/2}(\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0)=-(\mathbf{G}'\mathbf{V}^{-1}\mathbf{G})^{-1}\mathbf{G}'\mathbf{V}^{-1}T^{-1/2}\sum_{t=1}^{T}\mathbf{g}(\mathbf{x}_t;\boldsymbol{\theta}_0)+o_P(1). \tag{A.4}$$

Hence,

$$(\alpha/\tau)\nabla^2 \ln h(0) T^{1/2} S_T^{-1} \hat{\boldsymbol{\phi}}_T = -\left(\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{G}(\mathbf{G}'\mathbf{V}^{-1}\mathbf{G})^{-1}\mathbf{G}'\mathbf{V}^{-1}\right) T^{-1/2} \sum_{t=1}^{T} \mathbf{g}(\mathbf{x}_t; \boldsymbol{\theta}_0) + o_P(1). \quad \text{(A.5)}$$

Proof of Theorem 3.1: Consider a first order Taylor series expansion of $T^{1/2}(\hat{\mu}_T(\mathbf{x}) - \mu(\mathbf{x}))$ about θ_0 and $\mathbf{0}$:

$$T^{1/2}(\hat{\mu}_T(\mathbf{x}) - \mu(\mathbf{x})) = T^{1/2}(\mu_T(\mathbf{x}) - \mu(\mathbf{x})) - \tau \mathbf{B}' \alpha \nabla^2 \ln h(0) T^{1/2} S_T^{-1} \hat{\boldsymbol{\phi}}_T + o_P(1).$$

We have made use of the approximations

$$T^{-1/2} \sum_{t=1}^{T} \sum_{s=-[(T-1)/2]}^{[(T-1)/2]} \omega(s; S_T) 1(\mathbf{x}_{t-r} \leq \mathbf{x}) = \tau T^{-1/2} \sum_{t=1}^{T} 1(\mathbf{x}_t \leq \mathbf{x}) + o_P(1),$$

and

$$\left(\left(\alpha T S_T^{-1}\right)^{-1} \sum_{t=1}^{T} \sum_{s=-[(T-1)/2]}^{[(T-1)/2]} \omega(s; S_T) 1(\mathbf{x}_{t-r} \leq \mathbf{x}) \mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta}_0)' \right) = \tau \mathbf{B}' + o_P(1);$$

cf. the Proof of Theorem 2.1. As

$$T^{-1/2} \sum_{t=1}^{T} \begin{pmatrix} 1(\mathbf{x}_t \leq \mathbf{x}) - \mu(\mathbf{x}) \\ \mathbf{g}(\mathbf{x}_t; \boldsymbol{\theta}_0) \end{pmatrix} \Rightarrow N\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \sigma^2 & \mathbf{B}' \\ \mathbf{B} & \mathbf{V} \end{pmatrix} \right).$$

Therefore, from (A.5),

$$T^{1/2}(\hat{\mu}_T(\mathbf{x}) - \mu(\mathbf{x})) = \begin{pmatrix} 1, & -\tau \mathbf{B}' \left(\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{G}(\mathbf{G}'\mathbf{V}^{-1}\mathbf{G})^{-1}\mathbf{G}'\mathbf{V}^{-1}\right) \end{pmatrix}$$

$$\times T^{-1/2} \sum_{t=1}^{T} \begin{pmatrix} 1(\mathbf{x}_t \leq \mathbf{x}) - \mu(\mathbf{x}) \\ \mathbf{g}(\mathbf{x}_t; \boldsymbol{\theta}_0) \end{pmatrix} + o_P(1),$$

and, hence, the result follows.

Proof of Theorem 4.1: A second order Taylor series expansion of $\mathcal{R}(\theta_0, \mathbf{0}) = -T \ln T$ about $\hat{\boldsymbol{\varphi}}_T$ gives

$$(\alpha/\tau^2) S_T^{-1} (\mathcal{R}(\boldsymbol{\theta}_0, \mathbf{0}) - \mathcal{R}(\hat{\boldsymbol{\varphi}}_T))$$

$$= (1/2) T \begin{pmatrix} (\alpha/\tau) S_T^{-1} \hat{\boldsymbol{\phi}}_T' & (\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \end{pmatrix} \begin{pmatrix} \nabla^2 \ln h(0) \mathbf{V} & \tau \mathbf{G} \\ \tau \mathbf{G}' & \mathbf{0} \end{pmatrix} \begin{pmatrix} (\alpha/\tau) S_T^{-1} \hat{\boldsymbol{\phi}}_T \\ \hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0 \end{pmatrix} + o_P(1). \quad \text{(A.6)}$$

Noting from the Proof of Theorem 2.1 that $\mathbf{G}' \alpha \nabla^2 \ln h(0) T^{1/2} S_T^{-1} \eta(T) \hat{\boldsymbol{\phi}}_T = o_P(1)$, after substitution from (A.5) into (A.6), we obtain

$$2(\alpha/\tau^2) \nabla^2 \ln h(0) S_T^{-1} (\mathcal{R}(\boldsymbol{\theta}_0, \mathbf{0}) - \mathcal{R}(\hat{\boldsymbol{\varphi}}_T)) = [(\alpha/\tau) \nabla^2 \ln h(0) T^{1/2} S_T^{-1}]^2 \hat{\boldsymbol{\phi}}_T' \mathbf{V} \hat{\boldsymbol{\phi}}_T + o_P(1)$$

$$= T^{-1/2} \sum_{t=1}^{T} \mathbf{g}(\mathbf{x}_t; \boldsymbol{\theta}_0)' [\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{G}(\mathbf{G}'\mathbf{V}^{-1}\mathbf{G})^{-1}\mathbf{G}'\mathbf{V}^{-1}] T^{-1/2} \sum_{t=1}^{T} \mathbf{g}(\mathbf{x}_t; \boldsymbol{\theta}_0) + o_P(1). \quad \text{(A.7)}$$

Proof of Proposition 4.1: Follows directly from the expansion (A.7).

Proof of Proposition 4.2: From (A.5), $(\alpha/\tau) \nabla^2 \ln h(0) T^{1/2} S_T^{-1} \hat{\boldsymbol{\phi}}_T = -\mathbf{V}^{-1} T^{-1/2} \sum_{t=1}^{T} \mathbf{g}(\mathbf{x}_t; \hat{\boldsymbol{\theta}}_T) + o_P(1)$. Moreover, similarly to the Proof of Theorem 2.2, $T^{-1/2} \sum_{t=1}^{T} \mathbf{g}_{tT}^{\omega}(\hat{\boldsymbol{\theta}}_T) = \tau T^{-1/2} \sum_{t=1}^{T} \mathbf{g}(\mathbf{x}_t; \hat{\boldsymbol{\theta}}_T) + o_P(1)$.

Proof of Theorem 4.2: A Taylor series expansion for $\mathcal{R}^*(\tilde{\boldsymbol{\varphi}}_T, \tilde{\boldsymbol{\psi}}_T)$ similar to that in (A.7) for $\mathcal{R}(\hat{\boldsymbol{\varphi}}_T)$ results in

$$2(\alpha/\tau^2)\nabla^2 \ln h(0) S_T^{-1}\left(\mathcal{R}(\theta_0, \mathbf{0}) - \mathcal{R}^*(\tilde{\theta}_T, \tilde{\phi}_T, \tilde{\psi}_T)\right) = [(\alpha/\tau)\nabla^2 \ln h(0) T^{1/2} S_T^{-1}]^2 \begin{pmatrix} \tilde{\boldsymbol{\phi}}_T' & \tilde{\boldsymbol{\psi}}_T' \end{pmatrix} \mathbf{V}_* \begin{pmatrix} \tilde{\boldsymbol{\phi}}_T \\ \tilde{\boldsymbol{\psi}}_T \end{pmatrix} + o_P(1)$$

$$= T^{-1}\sum_{t=1}^{T} \begin{pmatrix} \mathbf{g}(\mathbf{x}_t;\boldsymbol{\theta}_0)' & \mathbf{q}(\mathbf{x}_t;\boldsymbol{\theta}_0)' \end{pmatrix} \mathbf{A}_* \sum_{t=1}^{T} \begin{pmatrix} \mathbf{g}(\mathbf{x}_t;\boldsymbol{\theta}_0) \\ \mathbf{q}(\mathbf{x}_t;\boldsymbol{\theta}_0) \end{pmatrix} + o_P(1), \qquad \text{(A.8)}$$

where $\mathbf{A}_* = \mathbf{V}_*^{-1} - \mathbf{V}_*^{-1}\mathbf{G}_*(\mathbf{G}_*'\mathbf{V}_*^{-1}\mathbf{G}_*)^{-1}\mathbf{G}_*'\mathbf{V}_*^{-1}$, $\mathbf{G}_* = \begin{pmatrix} \mathbf{G}' & \mathbf{Q}' \end{pmatrix}'$ and $\mathbf{Q} = E_\mu\{\nabla_{\boldsymbol{\theta}}'\mathbf{q}(\mathbf{x};\boldsymbol{\theta}_0)\}$. Hence

$$2(\alpha/\tau^2)\nabla^2 \ln h(0) S_T^{-1}\left(\mathcal{R}^*(\hat{\boldsymbol{\varphi}}_T, \mathbf{0}) - \mathcal{R}^*(\tilde{\boldsymbol{\varphi}}_T, \tilde{\boldsymbol{\psi}}_T)\right)$$

$$= T^{-1}\sum_{t=1}^{T} \begin{pmatrix} \mathbf{g}(\mathbf{x}_t;\boldsymbol{\theta}_0)' & \mathbf{q}(\mathbf{x}_t;\boldsymbol{\theta}_0)' \end{pmatrix} [\mathbf{A}_* - \mathbf{A}] \sum_{t=1}^{T} \begin{pmatrix} \mathbf{g}(\mathbf{x}_t;\boldsymbol{\theta}_0) \\ \mathbf{q}(\mathbf{x}_t;\boldsymbol{\theta}_0) \end{pmatrix} + o_P(1), \qquad \text{(A.9)}$$

where $\mathbf{A} = \mathbf{S}_\phi[\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{G}(\mathbf{G}'\mathbf{V}^{-1}\mathbf{G})^{-1}\mathbf{G}'\mathbf{V}^{-1}]\mathbf{S}_\phi'$ and $\mathbf{S}_\phi$ is a selection matrix such that $\mathbf{S}_\phi' \begin{pmatrix} \mathbf{g}(\mathbf{x}_t;\boldsymbol{\theta}_0)' & \mathbf{q}(\mathbf{x}_t;\boldsymbol{\theta}_0)' \end{pmatrix}' = \mathbf{g}(\mathbf{x}_t;\boldsymbol{\theta}_0)$. Note that $\mathbf{V} = \mathbf{S}_\phi'\mathbf{V}_*\mathbf{S}_\phi$ and $\mathbf{S}_\phi'\mathbf{G}_* = \mathbf{G}$. Now

$$\mathbf{V}_*[\mathbf{A}_* - \mathbf{A}]\mathbf{V}_*[\mathbf{A}_* - \mathbf{A}]\mathbf{V}_* = \mathbf{V}_*[\mathbf{A}_* - \mathbf{A}]\mathbf{V}_*$$

as $\mathbf{A}_*\mathbf{V}_*\mathbf{A}_* = \mathbf{A}_*$ and $\mathbf{A}\mathbf{V}_*\mathbf{A}_* = \mathbf{A}$. Therefore, Proposition 3.2 follows from Rao and Mitra (1971, Theorem 9.2.1, p.171) with the degrees of freedom given by

$$tr\{\mathbf{V}_*[\mathbf{A}_* - \mathbf{A}]\} = tr\{\mathbf{V}_*\mathbf{A}_*\} - tr\{\mathbf{V}_*\mathbf{A}\}$$

$$= (m + s - p) - (m - p) = s.$$

Proof of Proposition 4.3: Firstly, we have from (A.5) that

$$(\alpha/\tau)\nabla^2 \ln h(0) T^{1/2} S_T^{-1} \begin{pmatrix} \tilde{\boldsymbol{\phi}}_T' & \tilde{\boldsymbol{\psi}}_T' \end{pmatrix}' = \mathbf{A}_* T^{-1/2}\sum_{t=1}^{T} \begin{pmatrix} \mathbf{g}(\mathbf{x}_t;\boldsymbol{\theta}_0)' & \mathbf{q}(\mathbf{x}_t;\boldsymbol{\theta}_0)' \end{pmatrix} + o_P(1). \qquad \text{(A.10)}$$

Combining (A.5), (A.8) yields the intermediate result that the GEL minimum chi-squared statistic $\mathcal{MC}$ of (4.11)

$$[(\alpha/\tau)\nabla^2 \ln h(0) T^{1/2} S_T^{-1}]^2 \begin{pmatrix} (\tilde{\boldsymbol{\phi}}_T - \hat{\boldsymbol{\phi}}_T)' & \tilde{\boldsymbol{\psi}}_T' \end{pmatrix} \hat{\mathbf{V}}_{*T} \begin{pmatrix} (\tilde{\boldsymbol{\phi}}_T - \hat{\boldsymbol{\phi}}_T)' & \tilde{\boldsymbol{\psi}}_T' \end{pmatrix}'$$

$$= T^{-1}\sum_{t=1}^{T} \begin{pmatrix} \mathbf{g}(\mathbf{x}_t;\boldsymbol{\theta}_0)' & \mathbf{q}(\mathbf{x}_t;\boldsymbol{\theta}_0)' \end{pmatrix}' (\mathbf{A}_* - \mathbf{A}) \sum_{t=1}^{T} \begin{pmatrix} \mathbf{g}(\mathbf{x}_t;\boldsymbol{\theta}_0)' & \mathbf{q}(\mathbf{x}_t;\boldsymbol{\theta}_0)' \end{pmatrix} + o_P(1), \qquad \text{(A.11)}$$

which, therefore, from (A.7), is asymptotically equivalent to the difference of GELR statistics (4.7).

Now,

$$\begin{pmatrix} \mathbf{V}_* & \mathbf{G}_* \\ \mathbf{G}_*' & \mathbf{0} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{V}_*^{-1} - \mathbf{V}_*^{-1}\mathbf{G}_*(\mathbf{G}_*'\mathbf{V}_*^{-1}\mathbf{G}_*)^{-1}\mathbf{G}_*'\mathbf{V}_*^{-1} & \mathbf{V}_*^{-1}\mathbf{G}_*(\mathbf{G}_*'\mathbf{V}_*^{-1}\mathbf{G}_*)^{-1} \\ (\mathbf{G}_*'\mathbf{V}_*^{-1}\mathbf{G}_*)^{-1}\mathbf{G}_*'\mathbf{V}_*^{-1} & -(\mathbf{G}_*'\mathbf{V}_*^{-1}\mathbf{G}_*)^{-1} \end{pmatrix}.$$

Define the $(r + s + p, s)$ selection matrix $\mathbf{S}_{*\psi}$ such that $\mathbf{S}_{*\psi}'(\boldsymbol{\phi}', \boldsymbol{\psi}', \boldsymbol{\theta}')' = \boldsymbol{\psi}$. Therefore, the SPEQL Wald statistic $\mathcal{GELW}$ of (4.8) may be expressed as

$$\mathcal{GELW} = T \begin{pmatrix} (\alpha/\tau)\nabla^2 \ln h(0) S_T^{-1}(\tilde{\boldsymbol{\phi}}_T - \hat{\boldsymbol{\phi}}_T) \\ (\alpha/\tau)\nabla^2 \ln h(0) S_T^{-1}\tilde{\boldsymbol{\psi}}_T \\ (\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T) \end{pmatrix}' \mathbf{S}_{*\psi} \left(\mathbf{S}_{*\psi}' \begin{pmatrix} \mathbf{V}_* & \mathbf{G}_* \\ \mathbf{G}_*' & \mathbf{0} \end{pmatrix}^{-1} \mathbf{S}_{*\psi}\right)^{-1}$$

$$\times \mathbf{S}_{*\psi}' \begin{pmatrix} (\alpha/\tau)\nabla^2 \ln h(0) S_T^{-1}(\tilde{\boldsymbol{\phi}}_T - \hat{\boldsymbol{\phi}}_T) \\ (\alpha/\tau)\nabla^2 \ln h(0) S_T^{-1}\tilde{\boldsymbol{\psi}}_T \\ (\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T) \end{pmatrix} + o_P(1).$$

A Taylor series expansion for the score at $(\tilde{\boldsymbol{\varphi}}_T, \tilde{\boldsymbol{\psi}}_T)$ around the score at $(\hat{\boldsymbol{\varphi}}_T, \mathbf{0})$, $\hat{\mathbf{S}}_{*T}$, cf. (4.9), yields

$$T^{-1/2}\hat{\mathbf{S}}_{*T} = -\begin{pmatrix} \mathbf{V}_* & \mathbf{G}_* \\ \mathbf{G}_*' & \mathbf{0} \end{pmatrix} T^{1/2} \begin{pmatrix} (\alpha/\tau)\nabla^2 \ln h(0) S_T^{-1}(\tilde{\boldsymbol{\phi}}_T - \hat{\boldsymbol{\phi}}_T) \\ (\alpha/\tau)\nabla^2 \ln h(0) S_T^{-1}\tilde{\boldsymbol{\psi}}_T \\ (\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T) \end{pmatrix} + o_P(1), \tag{A.12}$$

noting that $\mathbf{G}_*'(\alpha/\tau)\nabla^2 \ln h(0) T^{1/2} S_T^{-1} \begin{pmatrix} \tilde{\boldsymbol{\phi}}_T{}' & \tilde{\boldsymbol{\psi}}_T' \end{pmatrix}' = o_P(1)$ and $\mathbf{G}'(\alpha/\tau)\nabla^2 \ln h(0) T^{1/2} S_T^{-1}\hat{\boldsymbol{\phi}}_T = o_P(1)$. Hence,

$$T^{1/2} \begin{pmatrix} (\alpha/\tau)\nabla^2 \ln h(0) S_T^{-1}(\tilde{\boldsymbol{\phi}}_T - \hat{\boldsymbol{\phi}}_T) \\ (\alpha/\tau)\nabla^2 \ln h(0) S_T^{-1}\tilde{\boldsymbol{\psi}}_T \\ (\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T) \end{pmatrix} = -\begin{pmatrix} \mathbf{V}_* & \mathbf{G}_* \\ \mathbf{G}_*' & \mathbf{0} \end{pmatrix}^{-1} T^{-1/2}\hat{\mathbf{S}}_{*T} + o_P(1),$$

Therefore, noting that $\hat{\mathbf{S}}_{*T} = \mathbf{S}_{*\psi} \sum_{t=1}^T \pi_t(\hat{\boldsymbol{\varphi}}_T)\mathbf{q}_{tT}^{\omega}(\hat{\boldsymbol{\theta}}_T)$,

$$\begin{aligned} \mathcal{GELW} &= T^{-1}\hat{\mathbf{S}}_{*T}' \begin{pmatrix} \mathbf{V}_* & \mathbf{G}_* \\ \mathbf{G}_*' & \mathbf{0} \end{pmatrix}^{-1} \hat{\mathbf{S}}_{*T} + o_P(1) \\ &= \mathcal{GELS} + o_P(1). \end{aligned} \tag{A.13}$$

Substituting into (A.13) for $T^{-1/2}\hat{\mathbf{S}}_{*T}$ from (A.12) and recalling $\mathbf{G}_*'(\alpha/\tau)\nabla^2 \ln h(0) T^{1/2} S_T^{-1} \begin{pmatrix} \tilde{\boldsymbol{\phi}}_T{}' & \tilde{\boldsymbol{\psi}}_T' \end{pmatrix}' = o_P(1)$ and $\mathbf{G}'(\alpha/\tau)\nabla^2 \ln h(0) T^{1/2} S_T^{-1}\hat{\boldsymbol{\phi}}_T = o_P(1)$ yields (A.11) apart from asymptotically negligible terms.

Proof of Proposition 4.4: This result is immediate from (A.13).

Proof of Theorem 5.1: Consider the first order conditions determining $(\tilde{\boldsymbol{\varphi}}_T, \tilde{\boldsymbol{\psi}}_T)$:

$$\sum_{t=1}^T \pi_t(\tilde{\boldsymbol{\varphi}}_T) \begin{pmatrix} \mathbf{g}_{tT}^{\omega}(\tilde{\boldsymbol{\theta}}_T) \\ \mathbf{q}(\tilde{\boldsymbol{\theta}}_T, \tilde{\boldsymbol{\alpha}}_T) \\ \mathbf{r}(\tilde{\boldsymbol{\alpha}}_T) \\ \nabla_{\boldsymbol{\theta}}\mathbf{g}_{tT}^{\omega}(\tilde{\boldsymbol{\theta}}_T)'\tilde{\boldsymbol{\phi}}_T + \nabla_{\boldsymbol{\theta}}\mathbf{q}(\tilde{\boldsymbol{\theta}}_T, \tilde{\boldsymbol{\alpha}}_T)'\tilde{\boldsymbol{\psi}}_T \\ \nabla_{\boldsymbol{\alpha}}\mathbf{q}(\tilde{\boldsymbol{\theta}}_T, \tilde{\boldsymbol{\alpha}}_T)'\tilde{\boldsymbol{\psi}}_T + \nabla_{\boldsymbol{\alpha}}\mathbf{r}(\tilde{\boldsymbol{\alpha}}_T)'\tilde{\boldsymbol{\eta}}_T \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}. \tag{A.14}$$

Because $\mathbf{q}(\boldsymbol{\theta}_0, \boldsymbol{\alpha}_0) = \mathbf{0}$ and $\mathbf{r}(\boldsymbol{\alpha}_0) = \mathbf{0}$, a first order Taylor series expansion of (A.14) about $(\boldsymbol{\theta}_0, \mathbf{0}, \boldsymbol{\alpha}_0, \mathbf{0}, \mathbf{0})$ yields

$$\mathbf{0} = T^{-1/2}\sum_{t=1}^T \mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta}_0) + \mathbf{V}\alpha\nabla^2 \ln h(0) T^{1/2} S_T^{-1}\tilde{\boldsymbol{\phi}}_T + \tau\mathbf{G}T^{1/2}(\tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) + o_P(1), \tag{A.15}$$

$$\mathbf{0} = \mathbf{Q}_{\boldsymbol{\theta}} T^{1/2}(\tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) + \mathbf{Q}_{\boldsymbol{\alpha}} T^{1/2}(\tilde{\boldsymbol{\alpha}}_T - \boldsymbol{\alpha}_0) + o_P(1), \tag{A.16}$$

$$\mathbf{0} = \mathbf{R}T^{1/2}(\tilde{\boldsymbol{\alpha}}_T - \boldsymbol{\alpha}_0) + o_P(1), \tag{A.17}$$

$$\mathbf{0} = \mathbf{G}'T^{1/2}S_T^{-1}\tilde{\boldsymbol{\phi}}_T + \mathbf{Q}_{\boldsymbol{\theta}}' T^{1/2}S_T^{-1}\tilde{\boldsymbol{\psi}}_T + o_P(1), \tag{A.18}$$

$$\mathbf{0} = \mathbf{Q}_{\boldsymbol{\alpha}}' T^{1/2}S_T^{-1}\tilde{\boldsymbol{\psi}}_T + \mathbf{R}'T^{1/2}S_T^{-1}\tilde{\boldsymbol{\eta}}_T + o_P(1). \tag{A.19}$$

From (A.19)

$$T^{1/2}S_T^{-1}\tilde{\boldsymbol{\eta}}_T = -(\mathbf{RMR}')^{-1}\mathbf{RMQ}_{\boldsymbol{\alpha}}' T^{1/2}S_T^{-1}\tilde{\boldsymbol{\psi}}_T + o_P(1), \tag{A.20}$$

and, thus,

$$\mathbf{KQ}_{\boldsymbol{\alpha}}' T^{1/2}S_T^{-1}\tilde{\boldsymbol{\psi}}_T = o_P(1).$$

Hence, pre-multiplying (A.18) by $\mathbf{KQ}_{\boldsymbol{\alpha}}'\mathbf{SQ}_{\boldsymbol{\theta}}\mathbf{N}$ yields

$$\mathbf{KQ}_{\boldsymbol{\alpha}}'\mathbf{SQ}_{\boldsymbol{\theta}}\mathbf{NG}'T^{1/2}S_T^{-1}\tilde{\boldsymbol{\phi}}_T = o_P(1).$$

Hence, pre-multiplying (A.15) by $\mathbf{K}\mathbf{Q}_{\alpha}'\mathbf{S}\mathbf{Q}_{\theta}\mathbf{N}\mathbf{G}'\mathbf{V}^{-1}$ gives

$$T^{1/2}(\tilde{\boldsymbol{\alpha}}_T - \boldsymbol{\alpha}_0) = \tau^{-1}\mathbf{K}\mathbf{Q}_{\alpha}'\mathbf{S}\mathbf{Q}_{\theta}\mathbf{N}\mathbf{G}'\mathbf{V}^{-1}T^{-1/2}\sum_{t=1}^{T}\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta}_0) + o_P(1), \tag{A.21}$$

as $\mathbf{K}\mathbf{Q}_{\alpha}'\mathbf{S}\mathbf{Q}_{\alpha} = \mathbf{I} - \mathbf{M}\mathbf{R}'(\mathbf{R}\mathbf{M}\mathbf{R}')^{-1}\mathbf{R}$ and using (A.17).

From (A.18)

$$T^{1/2}S_T^{-1}\tilde{\boldsymbol{\psi}}_T = -\mathbf{S}\mathbf{Q}_{\theta}\mathbf{N}\mathbf{G}'T^{1/2}S_T^{-1}\tilde{\boldsymbol{\phi}}_T + o_P(1), \tag{A.22}$$

and, thus,

$$(\mathbf{I}_p - \mathbf{Q}_{\theta}'\mathbf{S}\mathbf{Q}_{\theta}\mathbf{N})\mathbf{G}'T^{1/2}S_T^{-1}\tilde{\boldsymbol{\phi}}_T = o_P(1).$$

Hence, pre-multiplying (A.15) by $(\mathbf{I}_p - \mathbf{Q}_{\theta}'\mathbf{S}\mathbf{Q}_{\theta}\mathbf{N})\mathbf{G}'\mathbf{V}^{-1}$ yields

$$\mathbf{0} = (\mathbf{I}_p - \mathbf{Q}_{\theta}'\mathbf{S}\mathbf{Q}_{\theta}\mathbf{N})\mathbf{G}'\mathbf{V}^{-1}T^{-1/2}\sum_{t=1}^{T}\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta}_0) + \mathbf{N}^{-1}T^{1/2}(\tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) + \mathbf{Q}_{\theta}'\mathbf{S}\mathbf{Q}_{\alpha}T^{1/2}(\tilde{\boldsymbol{\alpha}}_T - \boldsymbol{\alpha}_0) + o_P(1),$$

from (A.16). Therefore, from (A.21)

$$T^{1/2}(\tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) = -\tau^{-1}\mathbf{N}\left[\mathbf{I}_p - \mathbf{Q}_{\theta}'(\mathbf{S} - \mathbf{S}\mathbf{Q}_{\alpha}\mathbf{K}\mathbf{Q}_{\alpha}'\mathbf{S})\mathbf{Q}_{\theta}\mathbf{N}\right]\mathbf{G}'\mathbf{V}^{-1}T^{-1/2}\sum_{t=1}^{T}\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta}_0)$$
$$+o_P(1). \tag{A.23}$$

Hence, from (A.15) and (A.23)

$$\alpha\nabla^2\ln h(0)T^{1/2}S_T^{-1}\tilde{\boldsymbol{\phi}}_T = -\mathbf{V}^{-1}\left(\mathbf{I}_r - \mathbf{G}\left[\mathbf{N} - \mathbf{N}\mathbf{Q}_{\theta}'(\mathbf{S} - \mathbf{S}\mathbf{Q}_{\alpha}\mathbf{K}\mathbf{Q}_{\alpha}'\mathbf{S})\mathbf{Q}_{\theta}\mathbf{N}\right]\mathbf{G}'\mathbf{V}^{-1}\right)T^{-1/2}\sum_{t=1}^{T}\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta}_0) + o_P(1),$$
$$+o_P(1), \tag{A.24}$$

from (A.22)

$$\alpha\nabla^2\ln h(0)T^{1/2}S_T^{-1}\tilde{\boldsymbol{\psi}}_T = (\mathbf{S} - \mathbf{S}\mathbf{Q}_{\alpha}\mathbf{K}\mathbf{Q}_{\alpha}'\mathbf{S})\mathbf{Q}_{\theta}\mathbf{N}\mathbf{G}'\mathbf{V}^{-1}T^{-1/2}\sum_{t=1}^{T}\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta}_0) + o_P(1), \tag{A.25}$$

and from (A.20)

$$\alpha\nabla^2\ln h(0)T^{1/2}S_T^{-1}\tilde{\boldsymbol{\eta}}_T = -(\mathbf{R}\mathbf{M}\mathbf{R}')^{-1}\mathbf{R}\mathbf{M}\mathbf{Q}_{\alpha}'\mathbf{S}\mathbf{Q}_{\theta}\mathbf{N}\mathbf{G}'\mathbf{V}^{-1}T^{-1/2}\sum_{t=1}^{T}\mathbf{g}_{tT}^{\omega}(\boldsymbol{\theta}_0) + o_P(1).$$

Proof of Proposition 5.1: A Taylor series expansion for $\mathcal{R}(\tilde{\boldsymbol{\varphi}}_T)$ about $(\boldsymbol{\theta}_0, \mathbf{0})$ results in

$$2(\alpha/\tau^2)\nabla^2\ln h(0)S_T^{-1}\left(\mathcal{R}(\theta_0,\mathbf{0}) - \mathcal{R}(\tilde{\boldsymbol{\varphi}}_T)\right) = [(\alpha/\tau)\nabla^2\ln h(0)T^{1/2}S_T^{-1}]\tilde{\boldsymbol{\phi}}_T'\mathbf{V}\tilde{\boldsymbol{\phi}}_T + o_P(1). \tag{A.26}$$

Combining (A.7) and (A.26):

$$2(\alpha/\tau^2)\nabla^2\ln h(0)S_T^{-1}\left(\mathcal{R}(\tilde{\boldsymbol{\varphi}}_T) - \mathcal{R}(\hat{\boldsymbol{\varphi}}_T)\right) = [(\alpha/\tau)\nabla^2\ln h(0)T^{1/2}S_T^{-1}]^2\left(\tilde{\boldsymbol{\phi}}_T'\mathbf{V}\tilde{\boldsymbol{\phi}}_T - \hat{\boldsymbol{\phi}}_T'\mathbf{V}\hat{\boldsymbol{\phi}}_T\right) + o_P(1)$$
$$= T^{-1/2}\sum_{t=1}^{T}\mathbf{g}(\mathbf{x}_t;\boldsymbol{\theta}_0)'\mathbf{V}^{-1}\mathbf{G}\mathbf{N}\mathbf{Q}_{\theta}\mathbf{C}\mathbf{Q}_{\theta}\mathbf{N}\mathbf{G}'\mathbf{V}^{-1}T^{-1/2}\sum_{t=1}^{T}\mathbf{g}(\mathbf{x}_t;\boldsymbol{\theta}_0) + o_P(1)$$
$$= [(\alpha/\tau)\nabla^2\ln h(0)T^{1/2}S_T^{-1}]^2\tilde{\boldsymbol{\psi}}_T'\mathbf{S}^{-1}\tilde{\boldsymbol{\psi}}_T + o_P(1);$$

the second equality following from (A.5) and (A.24) and the third from (A.25) as $\mathbf{C}\mathbf{S}^{-1}\mathbf{C} = \mathbf{C}$.

A Taylor series expansion about $(\boldsymbol{\theta}_0, \boldsymbol{\alpha}_0)$ yields from (A.4) and (A.21)

$$T^{1/2}\mathbf{q}(\hat{\boldsymbol{\theta}}_T, \tilde{\boldsymbol{\alpha}}_T) = -(\mathbf{I}_s - \mathbf{Q}_\alpha \mathbf{K}\mathbf{Q}'_\alpha \mathbf{S})\mathbf{Q}_\theta \mathbf{N}\mathbf{G}'\mathbf{V}^{-1}T^{-1/2}\sum_{t=1}^{T}\mathbf{g}(\mathbf{x}_t; \boldsymbol{\theta}_0) + o_P(1)$$

$$= -\mathbf{S}^{-1}[(\alpha/\tau)\nabla^2 \ln h(0)T^{1/2}S_T^{-1}]\tilde{\boldsymbol{\psi}}_T + o_P(1),$$

by (A.25).

We also note that, from (A.4), (A.23) and (A.25),

$$T^{1/2}(\tilde{\boldsymbol{\theta}}_T - \hat{\boldsymbol{\theta}}_T) = \mathbf{N}\mathbf{Q}'_\theta[(\alpha/\tau)\nabla^2 \ln h(0)T^{1/2}S_T^{-1}]\tilde{\boldsymbol{\psi}}_T + o_P(1),$$

and, from (A.5), (A.24) and (A.25),

$$T^{1/2}S_T^{-1}(\tilde{\boldsymbol{\phi}}_T - \hat{\boldsymbol{\phi}}_T) = \mathbf{V}^{-1}\mathbf{G}\mathbf{N}\mathbf{Q}'_\theta T^{1/2}S_T^{-1}\tilde{\boldsymbol{\psi}}_T + o_P(1).$$

Proof of Theorem 6.1: From the first order conditions determining $(\hat{\beta}_T, \hat{\delta}_T)$, a first order Taylor series expansion about $(\boldsymbol{\theta}_0, \mathbf{0})$ and $\hat{\boldsymbol{\zeta}}_T$ for the optimised criterion $\mathcal{R}_h^*(\tilde{\boldsymbol{\zeta}}_T)$ yields

$$(\alpha/\tau)\nabla^2 \ln h(0)T^{-1/2}S_T^{-1}\left(\mathcal{R}_h(\hat{\boldsymbol{\zeta}}_T) - \mathcal{R}_h^*(\tilde{\boldsymbol{\zeta}}_T)\right)$$

$$= \left(T^{-1}\sum_{t=1}^{T}\ln h[\hat{\boldsymbol{\delta}}'_T \mathbf{q}_{tT}^\omega(\hat{\boldsymbol{\beta}}_T)]\mathbf{g}_{tT}^\omega(\boldsymbol{\theta}_0)'\right)[(\alpha/\tau)\nabla^2 \ln h(0)T^{1/2}S_T^{-1}]\hat{\boldsymbol{\phi}}_T + o_P(1)$$

$$= \boldsymbol{\xi}'_*[(\alpha/\tau)\nabla^2 \ln h(0)T^{1/2}S_T^{-1}]\hat{\boldsymbol{\phi}}_T + o_P(1).$$

Proof of Corollary 6.1: Immediate from that of Theorem 6.1.

Proof of Proposition 6.1: A first order Taylor series expansion of the statistic $[(\alpha/\tau)\nabla^2 \ln h(0)T^{-1/2}S_T^{-1}]$ $\left(\mathcal{R}_h(\hat{\boldsymbol{\zeta}}_T) - \mathcal{R}_h^*(\hat{\boldsymbol{\zeta}}_T)\right)$ about $(\boldsymbol{\theta}_0, \mathbf{0})$ yields an identical expansion to that in the Proof of Theorem 6.1.

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 (1994): 'Small Sample Bias in GMM Estimation of Covariance Structures', *mimeo*, Center for Urban Affairs and Policy Research, Northwestern University.

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

Andrews, D.W.K., and W. Ploberger (1994): 'Optimal Tests When a Nuisance Parameter is Present Only Under the Alternative, *Econometrica*, 62, 1383-1414.

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

Brown, B., and W.K. Newey (1992): 'Bootstrapping for GMM', *mimeo*, MIT.

Burnside, C., and M. Eichenbaum (1994): 'Small Sample Properties of Generalized Method of Moments Based Wald Tests', *Technical Working Paper* 155, NBER.

Chesher, A., and R.J. Smith (1994): 'Likelihood Ratio Specification Tests'. University of Bristol Discussion Paper No. 94/370 and CREST-INSEE Document de Travail No. 9411.

Cox, D.R. (1961): 'Tests of Separate Families of Hypotheses', *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics*, 1, 105-123.

Cox, D.R. (1962): 'Further Results on Tests of Separate Families of Hypotheses', *Journal of the Royal Statistical Society*, Series B, 24, 406-424.

Davies, R.B. (1977): 'Hypothesis Testing When a Parameter is Present Only Under the Alternative', *Biometrika*, 64, 247-254.

Davies, R.B. (1987): 'Hypothesis Testing When a Parameter is Present Only Under the Alternative', *Biometrika*, 74, 33-43.

DiCiccio, T.J., Hall, P., and J.P. Romano (1989): 'Comparison of Parametric and Empirical Likelihood Functions', *Biometrika*, 76, 465-476.

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

Hall, P., and J. Horowitz (1994): 'Bootstrap Critical Values for Tests based on Generalized Method of Moments Estimators', *mimeo*, University of Iowa.

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

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

Hansen, L.P., and J.A. Scheinkman (1995): 'Back to the Future: Generating Moment Implications for Continuous-Time Markov Processes', *Econometrica* 63, 767-804.

Imbens, G.W. (1993): 'A New Approach to Generalized Method of Moments Estimation', Working Paper No.1633, Harvard Institute of Economic Research, Harvard University.

Imbens, G.W., Johnson, P., and R.H. Spady (1995): 'Information-Theoretic Approaches to Inference in Moment Condition Models', *mimeo*, Harvard University.

Kitamura, Y., and M. Stutzer (1995): 'An Information-Theoretic Alternative to Generalized Method of Moments Estimation', *mimeo*, University of Minnesota.

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 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.

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

Owen, A.B. (1990): 'Empirical Likelihood Confidence Regions', *Annals of Statistics*, 18, 90-120.

Owen, A.B. (1991): 'Empirical Likelihood for Linear Models', *Annals of Statistics*, 19, 1725-1747.

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. (1987a): 'Testing the Normality Assumption in Multivariate Simultaneous Limited Dependent Variable Models', *Journal of Econometrics* 34, 105-123.

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

Smith, R.J. (1989): 'On the Use of Distributional Mis-Specification Checks in Limited Dependent Variable Models', *Economic Journal (Supplement)* 99, 178-192.

Smith, R.J. (1992): 'Non-Nested Tests for Competing Models Estimated by Generalized Method of Moments', *Econometrica*, 60, 973-980.

Smith, R.J. (1995): 'Empirical Likelihood Estimation and Inference', *mimeo*, University of Bristol.

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

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