GEL Pearson-Type Statistics Under Weak Identification

Patrik Guggenberger
Department of Economics
U.C.L.A.

Joaquim J.S. Ramalho
Department of Economics
University of Evora

Richard J. Smith
cemmap
I.F.S. and U.C.L.
and
Faculty of Economics
University of Cambridge

First Draft: October 2007
Very Preliminary and Incomplete

Abstract

Pearson-type test statistics based on implied probabilities obtained from generalized empirical likelihood methods are proposed for testing simple hypotheses involving the unknown parameter vector in moment condition time series models. Like the test statistics suggested in Guggenberger and Smith (2007), these statistics utilise smoothed versions of the moment indicators, a necessary alteration to ensure the statistics are asymptotically pivotal no matter the strength or weakness of identification. Like the statistics in Guggenberger and Smith (2007), Kleibergen (2005) and Otsu (2006), the empirical null rejection probabilities of the corresponding tests are not affected greatly by the strength or weakness of identification. More precisely, we show that the statistics are asymptotically chi square under both classical asymptotic theory and weak instrument asymptotics of Stock and Wright (2000). We also modify the statistics suggested in Guggenberger and Smith (2007) for a general form of kernel smoothing function.

JEL Classification: C13, C30
Keywords: Asymptotically Pivotal Statistics; Generalized Empirical Likelihood; Implied Probabilities; Nonlinear Moment Conditions.

1 Introduction

In many situations empirical researchers are confronted with instrumental variables only weakly correlated with the endogenous variables. A voluminous literature initiated in Phillips (1989) and Nelson and Startz (1990) demonstrates that, in such weakly identified contexts, the correspondence of the classical normal and chi-square limiting approximations to the finite-sample behaviour of estimators and test statistics may be poor.[1] A number of recent papers have examined the performance of generalized empirical likelihood (GEL) [Newey and Smith (2004), henceforth NS, Smith (1997, 2001)] and related methods for nonlinear moment condition models when identification may be weak as discussed in Stock and Wright (2000), henceforth SW. In particular, Guggenberger and Smith (2005), hereafter GSa, proposed GEL test statistics, which have chi-square asymptotic null distributions independent of the strength or weakness of identification. In related and independent work, Kleibergen (2005) proposes a GMM-based statistic for the time series context which is similar in structure to the continuous updating estimator (CUE) version of the Lagrange multiplier statistic in GSa. Guggenberger and Smith (2007), GSb, generalize the statistics in GSa from the i.i.d. and martingale difference sequence (m.d.s.) set up to the time series case. Their statistics are based on counterparts of the moment indicator functions smoothed using a kernel function and employing a bandwidth parameter [Kitamura and Stutzer (1997) and Smith (1997, 2001)], a procedure suggested for the weakly identified framework by Guggenberger (2003, Introduction of the first chapter). Otsu (2006) suggests a procedure based on the criterion function of the GEL estimator evaluated at transformed smoothed moment indicators. GSb provides computationally simple modifications to Otsu's (2006) statistic and two hybrid statistics that bridge their GEL-type and Kleibergen's (2005) GMM-type procedures.

The contributions of this paper are as follows. First, the Pearson-type GEL statistics

[1] See *inter alia* Dufour (1997), Staiger and Stock (1997), Stock and Wright (2000), Kleibergen (2002, 2005), Caner (2003), Moreira (2003), Andrews and Marmer (2004), Chao and Swanson (2005), Dufour and Taamouti (2005), Guggenberger and Smith (2005), Moreira et al. (2005a, 2005b), Andrews et al. (2006), Otsu (2006) and Andrews and Stock (2007). For a recent discussion of that literature, see Dufour (2003).

suggested in Ramalho and Smith (2004) are modified for time series observations and the weakly identified context. Secondly, the statistics developed in GSb are adapted for a general form of kernel function as are the Pearson-type GEL statistics of Ramalho and Smith (2004). The statistics proposed here, and those in GSb, Kleibergen (2005) and Otsu (2006), are asymptotically pivotal unlike classical Wald and likelihood ratio statistics no matter the strength of identification and are asymptotically chi-square under the null hypothesis. Thus, the empirical rejection probabilities (ERPs) under the null hypothesis of tests formed from these statistics do not vary much with the strength or weakness of identification. The tests deal with both simple hypotheses on the full parameter vector and composite hypotheses on a subvector of parameters, in the latter case assuming the parameters not under test are strongly identified with these parameters replaced by consistent estimators. Also see, e.g., Kleibergen (2004, 2005), Otsu (2006) and GSb.

The paper is organised as follows. Section 2 reviews GEL estimation and associated constructs. The test statistics are introduced in section 3 and their limiting behaviour stated. The Appendix contains the technical assumptions and proofs or results contained in the text.

The symbols " $\rightarrow_d$ " and " $\rightarrow_p$ " denote convergence in distribution and convergence in probability, respectively. Convergence "almost surely" is written as "a.s." and "with probability approaching 1" is replaced by "w.p.a.1". The space $C^i(S)$ contains all functions that are i–times continuously differentiable on the set S. Furthermore, $vec(M)$ stands for the column vectorization of the $k \times p$ matrix M, i.e. if $M = (m_1, ..., m_p)$ then $vec(M) = (m_1', ..., m_p')'$, "$M'$" denotes the transpose matrix of M, $(M)_{i,j}$ the element in the i-th row and j-th column, "$M > 0$" means that M is positive definite, $\lambda_{\min}(M)$ and $\lambda_{\max}(M)$ are the minimum and maximum eigenvalues of M, respectively, and $||M|| = \sqrt{\lambda_{\max}(M'M)}$. By I_p we denote the p-dimensional identity matrix.

2 Generalized Empirical Likelihood

This section outlines GEL estimation for time series observations; see Smith (1997, 2001) and GSb.[2] The set-up and notation is similar to that considered in GSa and GSb which also consider models specified by a finite number of unconditional moment restrictions.

2.1 Model and Notation

Let $\{z_i : i = 1, ..., n\}$ denote $\mathcal{R}^l$-valued time series data, where $n \in \mathcal{N}$ denotes the sample size. Also let $g_n : \mathcal{Z} \times \Theta \to \mathcal{R}^k$, where $\mathcal{Z} \subset \mathcal{R}^l$ and $\Theta \subset \mathcal{R}^p$ denote the sample and parameter spaces respectively. The model has a true parameter θ_0 for which the moment condition

$$Eg_n(z_i, \theta_0) = 0 \tag{2.1}$$

is satisfied. For ease of exposition, we adopt the notation $g_i(\theta)$ for $g_n(z_i, \theta)$ where no confusion should arise. The function g is allowed to depend on the sample size n to model weak identification, see Assumption **ID** below. As an illustration of this, GSb describes an i.i.d. linear instrumental variable model similar in spirit to Stock and Wright (2000).

We consider the simple null hypothesis

$$H_0 : \theta = \theta_0 \text{ versus } H_1 : \theta \neq \theta_0. \tag{2.2}$$

Define the recentered and rescaled sample average

$$\Psi_n(\theta) := n^{1/2}(\widehat{g}(\theta) - E\widehat{g}(\theta)),$$

where $\widehat{g}(\theta) := n^{-1} \sum_{i=1}^{n} g_i(\theta)$ and let

$$\Delta(\theta) := \lim_{n\to\infty} E\Psi_n(\theta)\Psi_n(\theta)' \in \mathcal{R}^{k\times k}$$

denote the long–run variance matrix of $g_i(\theta)$ which is proportional to the spectral density at zero frequency.

[2] For i.i.d. observations, NS show that GEL estimators possess attractive higher order properties as compared with alternatives to those based on generalized method of moments (GMM); see Hansen (1982), Newey (1985) and Newey and West (1987). In particular, for many models, the asymptotic bias of empirical likelihood (EL) does not grow with the number of moment restrictions, while that of GMM estimators grows without bound. Moreover, EL is higher order efficient among asymptotically bias adjusted estimators.

2.2 Moment Indicators

In a strongly identified context, Kitamura and Stutzer (1997) and Smith (1997, 2001) describe GEL and related methods appropriate for the general time series setup considered here. Guggenberger (2003) suggested adapting their approach for weakly identified models with time series data. See also Smith (2000, 2005) and Otsu (2006). An alternative procedure could use a blocking method as in Kitamura (1997).

Define the smoothed counterparts of the moment indicators $g_i(\theta)$; *viz.*

$$g_{in}(\theta) := S_n^{-1} \sum_{j=i-n}^{i-1} k(j/S_n) g_{i-j}(\theta),$$

where S_n is a bandwidth parameter ($S_n \to \infty$ as $n \to \infty$) and $k(\cdot)$ a kernel. The kernel weights $S_n^{-1}k(j/S_n)$ give rise to weights similar in nature to those used in heteroskedastic and autocorrelation consistent (HAC) variance matrix estimation; see *inter alia* Andrews (1991) and Newey and West (1987).

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

Define

$$\widehat{g}_n(\theta) := n^{-1} \sum_{i=1}^{n} g_{in}(\theta) \text{ and } \widehat{\Delta}(\theta) := S_n \sum_{i=1}^{n} g_{in}(\theta) g_{in}(\theta)'/n. \tag{2.3}$$

Under assumptions given in Lemma A.2 below, the estimator $\widehat{\Delta}(\theta_0)$ is shown to be consistent for $k_2\Delta(\theta_0)$.[3] As detailed in GSb, the consistency of $\widehat{\Delta}(\theta)$ is crucial for the testing procedures suggested below.[4]

2.3 GEL Estimation

The statistics below are based on GEL estimation; see Smith (1997, 2001), Newey and Smith (2004), and GSa for a more comprehensive discussion.

[3] The estimator $\widehat{\Omega}(\theta) := \sum_{i=1}^{n} g_i(\theta) g_i(\theta)'/n$ based on the unsmoothed indicators $g_i(\theta)$, $(i = 1, ..., n)$, is consistent in the i.i.d. or m.d.s. setup considered in GSa but is, however, inconsistent in the general time series context considered here. See GSa and GSb for further discussion.

[4] If the moment indicators $g_i(\theta_0)$, $(i = 1, 2, ...)$, are i.i.d. or constitute a m.d.s., the kernel $k(\cdot)$ is set as the Dirac delta function, i.e., $k(a) = 1$ if $a = 0$ and 0 otherwise.

Let ρ be a real-valued function $Q \to \mathcal{R}$, where Q is an open interval of the real line that contains 0 and

$$\widehat{\Lambda}_n(\theta) := \{\lambda \in \mathcal{R}^k : \lambda' g_{in}(\theta) \in Q, (i = 1, ..., n)\}. \tag{2.4}$$

If defined, let $\rho_j(v) := (\partial^j \rho / \partial v^j)(v)$ and $\rho_j := \rho_j(0)$ for integers $j \geq 0$.

The GEL criterion is defined by

$$\widehat{P}_\rho(\theta, \lambda) := 2 \sum_{i=1}^{n} (\rho(k\lambda' g_{in}(\theta)) - \rho_0)/n. \tag{2.5}$$

The normalisation $k := k_1/k_2$ leaves the GEL estimator for θ unaffected but makes the scale of the estimator of the auxiliary parameters λ comparable for different choices of kernel function $k(\cdot)$. The GEL estimator is the solution to a saddle point problem; *viz.*

$$\widehat{\theta}_\rho := \arg\min_{\theta \in \Theta} \sup_{\lambda \in \widehat{\Lambda}_n(\theta)} \widehat{P}_\rho(\theta, \lambda). \tag{2.6}$$

Assumption ρ: **(i)** ρ is concave on Q; **(ii)** ρ is C^2 in a neighborhood of 0 and $\rho_1 = \rho_2 = -1$.

Examples of GEL estimators include the CUE, $\rho(v) = -(1+v)^2/2$, [Hansen, Heaton and Yaron (1996)], empirical likelihood (EL), $\rho(v) = \ln(1-v)$, [Imbens (1997) and Qin and Lawless (1994)], and exponential tilting (ET), $\rho(v) = -\exp v$, [Kitamura and Stutzer (1997) and Imbens, Spady and Johnson (1998)].

2.4 Implied Probabilities

Implied or empirical probabilities for the observations which incorporate the moment restrictions (2.1) may be associated with each GMM and GEL estimator. For EL implied probabilities were given by Owen (1988), for ET by Kitamura and Stutzer (1997), for quadratic ρ by Back and Brown (1993), and for the general case by Brown and Newey (1992) and Brown and Newey (2003). Also see NS and Smith (1997, 2001) which also provides implied probabilities for the time series context. Since these probabilities form the basis for a class of statistics developed below we briefly describe them here.

For any function ρ satisfying Assumption $\boldsymbol{\rho}$, let, if it exists, $\lambda(\theta) = \arg\max_{\lambda\in\widehat{\Lambda}_n(\theta)} \widehat{P}_\rho(\theta,\lambda)$. For given $\theta \in \Theta$, implied probabilities are then defined as

$$\widehat{\pi}_i(\theta) := \frac{\rho_1(k\lambda(\theta)'g_{in}(\theta))}{\sum_{j=1}^n \rho_1(k\lambda(\theta)'g_{jn}(\theta))},\ (i = 1, ..., n). \tag{2.7}$$

See Smith (1997, 2001).

The empirical probabilities $\widehat{\pi}_i(\theta)$, $(i = 1, ..., n)$, sum to one by construction and are positive when the arguments $k\lambda(\theta)'g_{in}(\theta)$ are small uniformly $(i = 1, ..., n)$. Moreover, they impose the sample moment condition $\sum_{i=1}^n \widehat{\pi}_i(\theta)g_{in}(\theta) = 0$, $(i = 1, ..., n)$, when the first-order conditions for $\lambda(\theta)$ hold, mirroring the population moment condition (2.1).

In an i.i.d. strongly identified setting, for any function $a(z,\theta)$ and efficient GMM or GEL estimator $\widehat{\theta}$, $\sum_{i=1}^n \widehat{\pi}_i(\widehat{\theta})a(z_i,\widehat{\theta})$ is an efficient estimator of the expectation $E[a(z,\theta)]$; see Brown and Newey (1998). Smith (2001) provides a similar estimator for stationary data in a strongly identified context based on $\widehat{\pi}_i(\widehat{\theta})$, $(i = 1, ..., n)$, defined in (2.7), which is efficient within a class of GMM estimators. Of particular interest here, however, is the limiting average cumulative distribution function $\mu(z) = \lim_{n\to\infty} \sum_{i=1}^n \mathcal{P}\{z_i \leq z\}/n$, $z \in \mathcal{Z}$. Of course, in a stationary environment, $\mu(z) = \mathcal{P}\{z_i \leq z\}$. Since the probability $\mathcal{P}\{z_i \leq z\}$ may also be written in expectation form $E[1(z_i \leq z)]$, where $1(\cdot)$ denotes an indicator function, $1(A) = 1$ if A is true and 0 otherwise, it is convenient for our analysis to regard the conventional empirical distribution function (EDF) as a moment estimator for $\mu(z)$; *viz.*

$$\mu_n(z) := \sum_{i=1}^n 1(z_i \leq z)/n. \tag{2.8}$$

The corresponding GEL estimator for given θ obtained from the implied probabilities $\widehat{\pi}_i(\theta)$, $(i = 1, ..., n)$, is given by

$$\widehat{\mu}_n(z;\theta) := \sum_{i=1}^n \widehat{\pi}_i(\theta)1(z_i \leq z); \tag{2.9}$$

cf. Smith (2001).

3 Test Statistics

The main concern in this section is the adaptation of the goodness of fit class of statistics obtained in Ramalho and Smith (2004) to the weakly identified context to test (2.2). These Pearson-type statistics utilise the implied probabilities $\widehat{\pi}_i(\widehat{\theta})$, $(i = 1, ..., n)$, of (2.7). We also adapt the statistics considered in GSb for general kernel functions $k(\cdot)$ considered here and derive related statistics. Like GSb, all of these statistics are asymptotically equivalent and are asymptotically pivotal quantities with limiting chi square null distributions. Therefore these statistics lead to tests whose ERPs under the null should not be affected much by the strength or weakness of identification. Other asymptotically pivotal statistics appropriate for the general time series set-up considered here are Kleibergen's (2005a) GMM-based and Otsu's (2006) GEL-based statistics.

3.1 Full Vector Statistics

Let $\theta = (\alpha', \beta')'$, where $\alpha \in \mathcal{A}$, $\mathcal{A} \subset \mathcal{R}^{p_A}$, $\beta \in \mathcal{B}$, $\mathcal{B} \subset \mathcal{R}^{p_B}$, $\Theta = \mathcal{A} \times \mathcal{B}$, and $p_A + p_B = p$. The case $p_B = 0$ is allowed. The following assumption is adapted from Assumption C in Stock and Wright (2000) in which α_0 and β_0 are modelled as weakly and strongly identified parameter vectors, respectively. Let $\mathcal{N} \subset \mathcal{B}$ denote an open neighborhood β_0.

Assumption ID$_{\theta_0}$: The true parameter $\theta_0 = (\alpha_0', \beta_0')'$ is in the interior of the compact set $\Theta = \mathcal{A} \times \mathcal{B}$ and **(i)** $E\widehat{g}(\theta) = n^{-1/2} m_{1n}(\theta) + m_2(\beta)$, where $m_{1n}, m_1 : \Theta \to \mathcal{R}^k$ and (if $p_B > 0$) $m_2 : \mathcal{B} \to \mathcal{R}^k$ are continuous functions such that $m_{1n}(\theta) \to m_1(\theta)$ uniformly on Θ, $m_1(\theta_0) = 0$ and $m_2(\beta) = 0$ if and only if $\beta = \beta_0$; **(ii)** $m_2 \in C^1(\mathcal{N})$; **(iii)** let $M_2(\beta) := (\partial m_2/\partial\beta)(\beta) \in \mathcal{R}^{k \times p_B}$. $M_2(\beta_0)$ has full column rank p_B.

3.1.1 Over-Identifying Moments Statistics

Ramalho and Smith (2004) introduced a number of Pearson-type statistics appropriate for testing over-identified moment restrictions appropriate for i.i.d. observations in a strongly identified setting. Essentially, these statistics are consequent versions of the standard Pearson statistics using the normalised contrasts $n\widehat{\pi}_i(\theta) - 1$, $(i = 1, ..., n)$, obtained from comparing the GEL estimator $\widehat{\mu}_n(\cdot)$ (2.9) and EDF $\mu_n(\cdot)$ (A.10) for $\mu(\cdot)$. Adapted for the implied probabilities $\widehat{\pi}_i(\widehat{\theta})$, $(i = 1, ..., n)$, of (2.7),, these statistics may be expressed as

$$P_\rho^a(\theta) := S_n^{-1} \sum_{i=1}^n (n\widehat{\pi}_i(\theta) - 1)^2 / (k_1^2/k_2), P_\rho^b(\theta) := S_n^{-1} \sum_{i=1}^n \frac{(n\widehat{\pi}_i(\theta) - 1)^2}{n\widehat{\pi}_i(\theta)} / (k_1^2/k_2). \tag{3.1}$$

Ramalho and Smith (2004) also suggest an alternative class of Pearson-type tests for the over-identifying moment conditions (2.1) which are similar in spirit to those discussed by Andrews (1988a, 1988b). These statistics involves a finite collection of s (say) mutually exclusive subsets $\mathcal{Z}_j$, $(j = 1, ..., s)$, of $\mathcal{Z}$, where $s \geq m$ and $\mu(\mathcal{Z}_j) > 0$, $(j = 1, ..., s)$. Note that their union may not equal $\mathcal{Z}$, i.e., $\cup_{j=1}^s \mathcal{Z}_j \subset \mathcal{Z}$. Unlike Andrews (1988a, 1988b), a non-stochastic collection $\mathcal{Z}_j$, $(j = 1, ..., s)$, only is considered here for reasons of expositional simplicity.[5]

Define

$$\mu_n(\mathcal{Z}_j) := \sum_{i=1}^n 1\left(z_i \in \mathcal{Z}_j\right)/n, \widehat{\mu}_n(\mathcal{Z}_j; \theta) := \sum_{i=1}^n \widehat{\pi}_i(\theta) 1\left(z_i \in \mathcal{Z}_j\right), (j = 1, ..., s). \tag{3.2}$$

Let $\widehat{\mu}_n^s(\theta) := (\widehat{\mu}_n(\mathcal{Z}_1; \theta), ..., \widehat{\mu}_n(\mathcal{Z}_s; \theta))'$ and $\mu_n^s := (\mu_n(\mathcal{Z}_1), ..., \mu_n(\mathcal{Z}_s))'$. Also let

$$\widehat{B}_{ns}(\theta) := \sum_{i=1}^n (1(z_i \in \mathcal{Z}_1), ..., 1(z_i \in \mathcal{Z}_s))' g_{in}(\theta)/n$$

The test statistic $P_\rho(\theta)$ defined next is based on the normalised contrast $\widehat{\mu}_n^s(\theta) - \mu_n^s$ from (3.2); *viz.*

$$P_\rho(\theta) := (n/S_n^2)(\widehat{\mu}_n^s(\theta) - \mu_n^s)' \widehat{B}_s(\theta)' (\widehat{B}_s(\theta)\widehat{B}_s(\theta)')^{-1} \widehat{\Delta}(\theta) (\widehat{B}_s(\theta)\widehat{B}_s(\theta)')^{-1} \widehat{B}_s(\theta) (\widehat{\mu}_n^s(\theta) - \mu_n^s)/(k_1^2/k_2). \tag{3.3}$$

If $s = m$, statistic $P_\rho(\theta)$ takes the simpler form $(n/S_n^2)(\widehat{\mu}_n^s(\theta) - \mu_n^s)' \widehat{B}_s(\theta)^{-1} \widehat{\Delta}(\theta) \widehat{B}_s(\theta)'^{-1} (\widehat{\mu}_n^s(\theta) - \mu_n^s)/(k_1^2/k_2)$.

[5] A random partition as in Andrews (1988b) could be permitted which weakly converges to one with the properties ascribed below for $\mathcal{Z}_j$, $(j = 1, ..., s)$. See Andrews (1988b, Assumption RC1, p.1425, and Section 3.1, pp.1427-1431).

Theorem 3.1 below demonstrates the asymptotic equivalence of the Pearson-type statistics $P^a_\rho(\theta)$, $P^b_\rho(\theta)$, $P_\rho(\theta)$ and the GEL criterion function statistic suggested by GSb for the Bartlett kernel. For a general form of kernel function $k(\cdot)$, the GEL criterion function statistic is defined as[6]

$$GELR_\rho(\theta) := S_n^{-1} n \widehat{P}_\rho(\theta, \lambda(\theta))/(k_1^2/k_2). \tag{3.4}$$

It is well known that the statistic $GELR_\rho(\theta)$ has a nonparametric likelihood ratio interpretation; see, e.g., GSa.

Technical assumptions $\mathbf{M}_{\theta_0}$ are presented in the Appendix.

Theorem 3.1 *Suppose* $\mathbf{ID}_{\theta_0}$*,* $\boldsymbol{\rho}$*, and* $\mathbf{M}_{\theta_0}$**(i)-(iii)** *and* **(viii)** *hold. Then, if* $S_n \to \infty$ *as* $n \to \infty$ *and* $S_n = o(n^{1/2})$*,*

$$GELR_\rho(\theta_0), P^a_\rho(\theta_0), P^b_\rho(\theta_0), P_\rho(\theta_0) \to_d \chi^2(k).$$

The proof of Theorem 3.1 also suggests the statistics $n\widehat{g}_n(\theta_0)'\widehat{\Delta}(\theta_0)^{-1}\widehat{g}_n(\theta_0)/(k_1^2/k_2)$ and $(n/S_n^2)\lambda(\theta_0)'\widehat{\Delta}(\theta_0)\lambda(\theta_0)/k_2$ as asymptotically equivalent to $GELR_\rho(\theta_0)$, $P^a_\rho(\theta_0)$, $P^b_\rho(\theta_0)$ and $P_\rho(\theta_0)$.

Assumption $\mathbf{M}_{\theta_0}$**(viii)** ensures that $\widehat{B}_{ns}(\theta_0)$ is consistent for $k_1 B_s(\theta_0)$ where $B_s(\theta) := (b(\mathcal{Z}_1;\theta), ..., b(\mathcal{Z}_s;\theta))$ and $b(\mathcal{Z}_j;\theta) := E[1(z_i \in \mathcal{Z}_j) g_i(\theta)]$, $(j = 1, ..., s)$.

3.1.2 Parametric Restrictions Statistics

Ramalho and Smith (2004) introduced a further set of Pearson-type statistics appropriate for testing forms of parametric restrictions which include the special case $H_0 : \theta = \theta_0$ of (2.2).

For $\theta \in \Theta$, let

$$D_\rho(\theta) := \sum_{i=1}^{n} \rho_1(k\lambda(\theta)' g_{in}(\theta)) G_{in}(\theta)/n \in \mathcal{R}^{k \times p}, \tag{3.5}$$

where, if it is defined, $G_{in}(\theta) := (\partial g_{in}/\partial\theta)(\theta) \in \mathcal{R}^{k \times p}$. If the minimum of the objective function $\widehat{P}(\theta, \lambda(\theta))$ is obtained in the interior of Θ, the score vector with respect to θ must

[6] GSb and Otsu (2006) independently introduced the generalization to the time series context of the $GELR_\rho$ statistic provided in GSa for the i.i.d. setting.

equal 0 at the GEL estimator $\widehat{\theta}$ which, by the envelope theorem, implies $\lambda(\widehat{\theta})' D_\rho(\widehat{\theta}) = 0'$; see, e.g., GSa, Newey and Smith (2004) and Smith (2001).

The relevant Pearson-type statistic adapted from Ramalho and Smith (2004) is

$$\begin{aligned} P^r_\rho(\theta) \;:\; &= (n/S_n^2)(\widehat{\mu}^s_n(\theta) - \mu^s_n)' \widehat{B}_s(\theta)'(\widehat{B}_s(\theta)\widehat{B}_s(\theta)')^{-1} D_\rho(\theta) \\ &\times (D_\rho(\theta)'\widehat{\Delta}(\theta)^{-1} D_\rho(\theta))^{-1} D_\rho(\theta)(\widehat{B}_s(\theta)\widehat{B}_s(\theta)')^{-1}\widehat{B}_s(\theta)(\widehat{\mu}^s_n(\theta) - \mu^s_n)/(k_1^2/k_2). \end{aligned}$$

If $s = m$, statistic $P^r_\rho(\theta)$ takes the simpler form $(n/S_n^2)(\widehat{\mu}^s_n(\theta) - \mu^s_n)'\widehat{B}_s(\theta)^{-1} D_\rho(\theta)(D_\rho(\theta)'\widehat{\Delta}(\theta)^{-1} D_\rho(\theta))^{-1} D_\rho$ $\mu^s_n)/(k_1^2/k_2)$.

The proof of Theorem 3.2 below establishes the asymptotic equivalence of the expression $(\widehat{B}_s(\theta_0)\widehat{B}_s(\theta_0)')^{-1}\widehat{B}_s(\theta_0)(n/S_n^2)^{1/2}(\widehat{\mu}^s_n(\theta_0) - \mu^s_n)$ in $P^r_\rho(\theta_0)$ (3.6) with $-(k_1^2/k_2)\widehat{\Delta}(\theta_0)^{-1} n^{1/2}\widehat{g}(\theta_0)$. This latter term appears in the Lagrange multiplier form of statistic for testing (2.2) obtained in GSb and modified for the general kernel function $k(\cdot)$; *viz.*

$$LM_\rho(\theta) := n\widehat{g}_n(\theta)'\widehat{\Delta}(\theta)^{-1} D_\rho(\theta)(D_\rho(\theta)'\widehat{\Delta}(\theta)^{-1} D_\rho(\theta))^{-1} D_\rho(\theta)'\widehat{\Delta}(\theta)^{-1}\widehat{g}_n(\theta)/(k_1^2/k_2). \tag{3.6}$$

The score form of statistic is

$$S_\rho(\theta) := (n/S_n^2)\lambda(\theta)' D_\rho(\theta)(D_\rho(\theta)'\widehat{\Delta}(\theta)^{-1} D_\rho(\theta))^{-1} D_\rho(\theta)'\lambda(\theta)/k_2. \tag{3.7}$$

The next theorem details the asymptotic distribution of these test statistics evaluated at θ_0.

Theorem 3.2 *Suppose Assumptions* **ID**, $\boldsymbol{\rho}$ *and* $\mathbf{M}_{\theta_0}$**(i)-(viii)** *hold. Then, if* $S_n \to \infty$ *as* $n \to \infty$ *and* $S_n = o(n^{1/2})$,

$$P^r_\rho(\theta_0), S_\rho(\theta_0), LM_\rho(\theta_0) \to_d \chi^2(p).$$

Remarks: 1) Confidence regions or hypothesis tests for θ_0 based on the smoothed statistics are simply implemented based on Theorems 3.1 and 3.2. Unlike classical test statistics, the smoothed statistics are asymptotically pivotal statistics under Assumption **ID**. Therefore, ERPs under the null of tests based on these statistics should not vary much with the strength or weakness of identification in finite samples. The critical step in demonstrating that the statistics $P^r_\rho(\theta_0)$, $S_\rho(\theta_0)$ and $LM_\rho(\theta_0)$ are asymptotically pivotal, is the asymptotic independence of $D_\rho(\theta_0)$ and $n^{1/2}\widehat{g}_n(\theta_0)$. Also see Smith (2001) which demonstrates this property for the strongly identified case. Note also that the asymptotic null distribution of the test statistics is invariant to the choice of ρ.

2) Since the degrees of freedom k of their limiting distribution corresponds to the number of moment conditions rather than the number of parameters p, the statistics $GELR_\rho(\theta_0)$, $P^a_\rho(\theta_0)$, $P^b_\rho(\theta_0)$ and $P_\rho(\theta_0)$ can be expected to display poorer power properties than $P^r_\rho(\theta_0)$, $S_\rho(\theta_0)$ and $LM_\rho(\theta_0)$. The power of tests based on these latter statistics should be relatively invariant to the degree of over-identification.

3) As discussed in GSb, Assumption $\mathbf{M}_{\theta_0}$ is compatible with many time series models. Like Assumption 1 in Kleibergen (2005a), Assumption $\mathbf{M}_{\theta_0}$**(vii)** underpinning Theorem 3.2 states a CLT for $(vecG_i'(\theta_0), g_i'(\theta_0))'$ with possibly singular covariance matrix.

3.2 Subvector Statistics

We are now interested in testing

$$H_0 : \alpha = \alpha_0 \text{ versus } H_1 : \alpha \neq \alpha_0, \tag{3.8}$$

where $\alpha_0 \in \mathcal{R}^{p_A}$ and $\theta_0 = (\alpha_0', \beta_0')'$. Let $\theta = (\alpha_1', \alpha_2', \beta')'$, where $\alpha_j \in \mathcal{A}_j$, $\mathcal{A}_j \subset \mathcal{R}^{p_{A_j}}$, $(j = 1, 2)$, $p_{A_1} + p_{A_2} = p_A$ and $\beta \in \mathcal{B}$, $\mathcal{B} \subset \mathcal{R}^{p_B}$. We assume that the true parameter $\theta_0 = (\alpha_{01}', \alpha_{02}', \beta_0')'$ is in the interior of the compact space Θ, where $\Theta = \mathcal{A}_1 \times \mathcal{A}_2 \times \mathcal{B}$. We now modify Assumption $\mathbf{ID}_{\theta_0}$. Let $\mathcal{N} \subset \mathcal{A}_2 \times \mathcal{B}$ be an open neighborhood of (α_{02}, β_0).[7]

Assumption $\mathbf{ID}_{\alpha_0}$: (i) $E\widehat{g}(\theta) = n^{-1/2}m_{1n}(\theta) + m_2(\alpha_2, \beta)$, where $m_{1n}, m_1 : \Theta \to \mathcal{R}^k$ and (if $p_{A_2} + p_B > 0$) $m_2 : \mathcal{A}_2 \times \mathcal{B} \to \mathcal{R}^k$ are continuous functions such that $m_{1n}(\theta) \to m_1(\theta)$ uniformly on Θ, $m_1(\theta_0) = 0$ and $m_2(\alpha_2, \beta) = 0$ if and only if $(\alpha_2, \beta) = (\alpha_{02}, \beta_0)$; **(ii)** $m_2 \in C^1(\mathcal{N})$; **(iii)** Let $M_2(\cdot) := (\partial m_2/\partial(\alpha_2', \beta')')(\cdot) \in \mathcal{R}^{k\times(p_{A_2}+p_B)}$. $M_2(\alpha_{02}, \beta_0)$ has

[7] In this subsection, $m_2(\cdot)$ and $M_2(\cdot)$ (defined already in $\mathbf{ID}_{\theta_0}$ above as functions of β) now denote functions of α_2 and β.

full column rank $p_{A_2} + p_B$.

Assumption $\mathbf{ID}_{\alpha_0}$ implies that α_{01} is weakly and (α_{02}, β_0) is strongly identified. To adapt the full-vector test statistics to the subvector case, the basic idea is to replace β by an estimator $\widehat{\beta}(\alpha)$. Define the GEL estimator $\widehat{\beta}(\alpha)$ for β_0

$$\widehat{\beta}(\alpha) := \arg\min_{\beta \in B} \sup_{\lambda \in \widehat{\Lambda}_n(\alpha', \beta')'} \widehat{P}((\alpha', \beta')', \lambda). \tag{3.9}$$

Our assumptions below imply consistency $\widehat{\beta} := \widehat{\beta}(\alpha_0) \rightarrow_p \beta_0$ and efficiency under the null hypothesis, also see Smith (2001). Let

$$\widehat{\theta}_0 := (\alpha_0', \widehat{\beta}(\alpha_0)')' \text{ and } \theta_\beta := (\alpha_0', \beta')'.$$

We now introduce the subvector statistics.

3.2.1 Over-Identifying Moments Statistics

Recall the definitions of $P_\rho^a(\theta)$, $P_\rho^b(\theta)$ and $P_\rho(\theta)$ in (3.1) and (3.3). Evaluated at α_0, the subvector versions of these statistics are given by

$$P_\rho^{a,sub}(\alpha_0) := P_\rho^a(\widehat{\theta}_0), P_\rho^{b,sub}(\alpha_0) := P_\rho^b(\widehat{\theta}_0). \tag{3.10}$$

and

$$P_\rho^{sub}(\alpha_0) := P_\rho(\widehat{\theta}_0). \tag{3.11}$$

Likewise, the corresponding redefinition of $GELR_\rho(\theta)$ in (3.4) is

$$GELR_\rho^{sub}(\alpha_0) := GELR_\rho(\widehat{\theta}_0). \tag{3.12}$$

Under Assumption $\mathbf{M}_{\alpha_0}$ given in the Appendix we have the following theorem.

Theorem 3.3 *Assume* $1 \leq p_A < p$. *Suppose Assumptions* $\mathbf{ID}_{\alpha_0}$, $\mathbf{M}_{\alpha_0}$**(i)-(iv)** *and* $\boldsymbol{\rho}$ *hold. Then,*

$$P_\rho^{a,sub}(\alpha_0), P_\rho^{b,sub}(\alpha_0), P_\rho^{sub}(\alpha_0), GELR_\rho^{sub}(\alpha_0) \rightarrow_d \chi^2(k - p_B).$$

3.2.2 Parametric Restrictions Statistics

We now generalize the statistic P_ρ^r to the subvector case and S_ρ and LM_ρ in GSb likewise for the general kernel function $k(\cdot)$.

We need additional notation. For a full column rank matrix $A \in \mathcal{R}^{k\times p}$ and $0 < K \in \mathcal{R}^{k\times k}$, let $P_A(K) := A(A'K^{-1}A)^{-1}A'K^{-1}$ and $M_A(K) := I_k - P_A(K)$. We abbreviate this notation to P_A and M_A if $K = I_k$. If $p = 0$, set $M_A = I_k$. Let

$$D_\rho(\alpha_0) := \sum_{i=1}^n \rho_1(k\lambda(\widehat{\theta}_0)'g_{in}(\widehat{\theta}_0))G_{inA}(\widehat{\theta}_0)/n \in \mathcal{R}^{k\times p_A},$$

where $G_{inA}(\theta)$ is defined by $G_{in}(\theta) := (G_{inA}(\theta), G_{inB}(\theta))$ for $G_{inA}(\theta) \in \mathcal{R}^{k\times p_A}$ and $G_{inB}(\theta) \in \mathcal{R}^{k\times p_B}$, see below eq. (3.5). The definition of $D_\rho(\alpha_0)$ coincides with the one of $D_\rho(\theta_0)$ when α_0 is the full vector θ_0. If $p_B > 0$ let

$$\widehat{M}(\alpha_0) := \widehat{\Delta}(\widehat{\theta}_0)^{-1}M_{\widehat{G}_B(\widehat{\theta}_0)}(\widehat{\Delta}(\widehat{\theta}_0)/k_2) \tag{3.13}$$

and otherwise let $\widehat{M}(\alpha_0) := \widehat{\Delta}(\widehat{\theta}_0)^{-1}$, where

$$\widehat{G}(\theta) := n^{-1}\textstyle\sum_{i=1}^n G_i(\theta) \in \mathcal{R}^{k\times p}, \widehat{G}(\theta) := (\widehat{G}_A(\theta), \widehat{G}_B(\theta))$$

for $\widehat{G}_A(\theta) \in \mathcal{R}^{k\times p_A}$ and $\widehat{G}_B(\theta) \in \mathcal{R}^{k\times p_B}$.

The modified form of the Pearson-type statistics P_ρ^r (**??**) appropriate for testing (3.8) is

$$\begin{aligned} P_\rho^{r,sub}(\alpha_0) \ &: = (n/S_n^2)(\widehat{\mu}_n^s(\widehat{\theta}_0) - \mu_n^s)'\widehat{B}_s(\widehat{\theta}_0)'(\widehat{B}_s(\widehat{\theta}_0)\widehat{B}_s(\widehat{\theta}_0)')^{-1}D_\rho(\alpha_0) \\ &\times(D_\rho(\alpha_0)'\widehat{M}(\alpha_0)D_\rho(\alpha_0))^{-1}D_\rho(\alpha_0)(\widehat{B}_s(\widehat{\theta}_0)\widehat{B}_s(\widehat{\theta}_0)')^{-1}\widehat{B}_s(\widehat{\theta}_0)(\widehat{\mu}_n^s(\widehat{\theta}_0) - \mu_n^s)/(k_1^2/k_2). \end{aligned}$$

The subvector test statistic $S_\rho^{sub}(\alpha_0)$ is constructed as a quadratic form in the vector of FOC $\lambda(\widehat{\theta}_0)'D_\rho(\widehat{\theta}_0)$ with weighting matrix given by $\widehat{M}(\alpha_0)$; *viz.*

$$S_\rho^{sub}(\alpha_0) := (n/S_n^2)\lambda(\widehat{\theta}_0)'D_\rho(\alpha_0)(D_\rho(\alpha_0)'\widehat{M}(\alpha_0)D_\rho(\alpha_0))^{-1}D_\rho(\alpha_0)'\lambda(\widehat{\theta}_0)/k_2.$$

The statistic $LM_\rho^{sub}(\alpha_0)$ replaces $n^{1/2}S_n^{-1}\lambda(\widehat{\theta}_0)$ in $S_\rho^{sub}(\alpha_0)$ by the asymptotically equivalent expression $-\widehat{\Delta}(\widehat{\theta}_0)^{-1}n^{1/2}\widehat{g}_n(\widehat{\theta}_0)$. Therefore,

$$LM_\rho^{sub}(\alpha_0) := n\widehat{g}_n(\widehat{\theta}_0)'\widehat{\Delta}(\widehat{\theta}_0)^{-1}D_\rho(\alpha_0)(D_\rho(\alpha_0)'\widehat{M}(\alpha_0)D_\rho(\alpha_0))^{-1}D_\rho(\alpha_0)'\widehat{\Delta}(\widehat{\theta}_0)^{-1}\widehat{g}_n(\widehat{\theta}_0)/(k_1^2/k_2).$$

Under Assumption $\mathbf{M}_{\alpha_0}$ given in the Appendix we have the following theorem.

Theorem 3.4 *Assume $1 \leq p_A < p$. Suppose Assumptions* $\mathbf{ID}_{\alpha_0}$, $\mathbf{M}_{\alpha_0}$**(i)-(viii)** *and* $\boldsymbol{\rho}$ *hold. Then,*

$$P_\rho^{r,sub}(\alpha_0), S_\rho^{sub}(\alpha_0), LM_\rho^{sub}(\alpha_0) \rightarrow_d \chi^2(p_A).$$

Under the assumption used here, that the parameters not under test are strongly identified, there are various other alternatives for subvector inference besides $GELR_\rho^{sub}(\alpha_0)$, $S_\rho^{sub}(\alpha_0)$, and $LM_\rho^{sub}(\alpha_0)$. See, for example, the tests by Kleibergen (2004, 2005a) and Otsu (2006). An interesting recent contribution by Kleibergen (2005b) introduces boundedly pivotal tests for the linear IV model *without* additional identification assumptions. Alternatively, confidence intervals can be constructed by a projection argument, see Dufour (1997). However, this approach is conservative and in general computationally cumbersome. In a recent paper, Dufour and Taamouti (2005) show that the Anderson and Rubin (1949) statistic is an exception, in that a closed form solution is available. Another alternative is Guggenberger and Wolf (2004) who suggest a subsampling approach. Unlike some of the above procedures, subsampling leads to subvector tests whose null rejection probability converges to the desired nominal level *without* additional identification assumptions for each *fixed* degree of identification. Guggenberger and Wolf's (2004) Monte Carlos suggest that for subvector inference subsampling seems to do better in terms of power than Kleibergen (2004, 2005a) and Dufour and Taamouti (2005). In their simulation study, the former procedure tends to underreject when the components not under test are only weakly identified and the latter seems to underreject across all the scenarios. On the other hand, they find that for full–vector inference, subsampling is outperformed by the procedures in GS and Kleibergen (2005a). Andrews and Guggenberger's (2005b,c) size correction methods for subsampled tests could also be applied to subvector tests.

4 Simulation Evidence

To be completed.

Appendix

Additional notation is defined and then the assumptions for the results are stated. Throughout this Appendix, the argument θ_0 is suppressed for expositional simplicity when there is no possibility of confusion.

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

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

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

$$\mathcal{K}_2 = \{k^*(\cdot) \in \mathcal{K}_1 : K^*(\lambda) \geq 0 \text{ for all } \lambda \in \mathcal{R}\}, \qquad \text{(A.2)}$$

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

Define

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

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

Assumption $\mathbf{M}_{S_n,k(\cdot)}$: (a) $S_n \to \infty$ and $S_n/n^2 \to 0$; **(b)** $k(\cdot) : \mathcal{R} \to [-k_{\max}, k_{\max}]$, $k_{\max} < \infty$, $k(0) \neq 0$, $k_1 \neq 0$, and is continuous at 0 and almost everywhere; **(c)** $\int_{(-\infty,\infty)} \bar{k}(x)dx < \infty$; **(d)** $|K(\lambda)| \geq 0$ for all $\lambda \in \mathcal{R}$.
Assumptions (b) and (c) ensure $k_2 > 0$. Assumptions (b) and (c) also guarantee that the induced kernel

$$k^*(a) = \frac{1}{k_2} \int_{-\infty}^{\infty} k(b-a) k(b) db, \qquad \text{(A.3)}$$

is a member of the p.s.d. class of kernels $\mathcal{K}_2$ (A.2) used in HAC covariance matrix estimation [Andrews (1991, p.822)].[8] See Lemma C.3 in Appendix C of Smith (2001).

[8]Neither the square integrability condition $\int_{-\infty}^{\infty} k^*(x)^2 dx < \infty$ in Andrews (1991, (2.6), p.821) nor

Consistency of $\widehat{\Delta}(\theta_0)/k_2$ in (2.3) for the long–run variance matrix $\Delta(\theta_0)$ is essential for our analysis; see GSb for further details. For this, we assume consistency of the classical kernel HAC estimator defined in terms of the induced kernel $k^*(\cdot)$ (A.3); see *inter alia* Andrews (1991). We then show that this HAC estimator differs from $\widehat{\Delta}(\theta_0)/k_2$ by a $o_p(1)$ term only. This follows Lemmata 2.1 and A.3 in Smith (2001). Likewise, other long-run variance expressions are also consistently estimated by their corresponding counterparts, e.g., $\Delta_A(\theta_0)$, defined in $\mathbf{M}_{\theta_0}\mathbf{(vii)}$ below, and $\widehat{\Delta}_A(\theta_0)/k_2$, where $\widehat{\Delta}_A(\theta_0) := S_n \sum_{i=1}^{n}(vecG_{inA}(\theta_0))g_{in}'(\theta_0)/n$.

Partition $G_i(\theta) := (\partial g_i/\partial\theta)(\theta)$ as $(G_{iA}(\theta), G_{iB}(\theta))$, where $G_{iA}(\theta) \in \mathcal{R}^{k\times p_A}$ and $G_{iB}(\theta) \in \mathcal{R}^{k\times p_B}$.

The classical HAC estimator based on $k^*(\cdot)$ (A.3) of the long–run variance between sequences of mean zero random vectors $r = (r_i)_{i=1,\ldots,n}$ and $s = (s_i)_{i=1,\ldots,n}$ is defined as

$$\begin{aligned}\widetilde{J}_n(r,s) :&= \sum_{j=-n+1}^{n-1} k^*(j/S_n)\widetilde{\Gamma}_j(r,s), \text{ where} \\ \widetilde{\Gamma}_j(r,s) :&= \begin{cases} \sum_{i=j+1}^{n} r_i s_{i-j}'/n & \text{for } j \geq 0, \\ \sum_{i=-j+1}^{n} r_{i+j} s_i'/n & \text{for } j < 0, \end{cases};\end{aligned}$$

see Andrews (1991, eq. (3.2)).

A.1 Assumptions

The various assumptions required for our analysis are listed below.

A.1.1 Full Vector Tests

Assumptions $\mathbf{M}_{\theta_0}\mathbf{(i)}$-$\mathbf{(iii)}$ are required for the asymptotic distribution of $GELR_\rho$, P_ρ^a and P_ρ^b. The statistic P_ρ also needs $\mathbf{M}_{\theta_0}\mathbf{(viii)}$. For the statistics P_ρ^r, LM_ρ and S_ρ we also need $\mathbf{M}_\theta\mathbf{(iv)}$-$\mathbf{(vii)}$ and additionally $\mathbf{M}_{\theta_0}\mathbf{(viii)}$ for P_ρ^r. Denote by $\mathcal{Z}$ the set of integer numbers.

the stronger absolute integrability condition $\int_{-\infty}^{\infty} |k^*(x)|\, dx < \infty$ in Andrews and Monahan (1992, (2.5), p.955) is sufficient for the consistency results claimed in those papers; see Jansson (2002). The condition $\int_{[0,\infty)} \bar{k}^*(x)dx < \infty$ ensures that particular summations used in those papers converge appropriately; see Lemma 1 of Jansson (2002).

Let $B_{si}(\theta) := (1(z_i \in \mathcal{Z}_1), ..., 1(z_i \in \mathcal{Z}_s))' g_i(\theta)'$, $(i = 1, ..., n)$, and $\widehat{B}_s(\theta) := n^{-1} \sum_{i=1}^n B_{si}(\theta)$.

Assumption $\mathbf{M}_{\theta_0}$: Suppose **(i)** $\max_{1 \le i \le n} ||g_i(\theta_0)|| = o_p(S_n^{-1} n^{1/2})$; **(ii)** for $S_n \to \infty$ and $S_n = o(n^{1/2})$ we have $\widetilde{J}_n((g_i(\theta_0)), (g_i(\theta_0))) \to_p \Delta(\theta_0) > 0$; $\sup_{i,j \ge 1} E||g_i(\theta_0) g_j'(\theta_0)|| < \infty$; for any sequence $m \to \infty$ and $m = o(n^{1/2})$ $\sup_{k \in \mathcal{Z}} E||\frac{1}{nm} \sum_{j=1}^n \sum_{i=k}^{k+m} g_{j+i}(\theta_0) g_j'(\theta_0)|| = o(1)$; $S_n n^{-1} \sum_{i=1}^n ||g_{in}(\theta_0) g_{in}(\theta_0)'|| = O_p(1)$; **(iii)** $\Psi_n(\theta_0) \to_d \Psi(\theta_0)$, where $\Psi(\theta_0) \equiv N(0, \Delta(\theta_0))$;

$$\textbf{(iv)}\ M_{1n}(\theta_0) : = (\partial m_{1n}/\partial\theta)|_{\theta=\theta_0} \to M_1(\theta_0) := (\partial m_1/\partial\theta)|_{\theta=\theta_0} \in \mathcal{R}^{k \times p}, \quad \text{(A.4)}$$

$$E\widehat{G}(\theta_0) = n^{-1/2} M_{1n}(\theta_0) + (0, M_2(\beta_0)) \to (0, M_2(\beta_0)); \quad \text{(A.5)}$$

(v) $\widetilde{J}_n((vecG_{iA}), (g_i)) \to_p \Delta_A$ (Δ_A is defined in (vii)); $\sup_{i,j \ge 1} E||vecG_{iA} g_j'|| < \infty$; for any sequence $m \to \infty$ and $m = o(n^{1/2})$ $\sup_{k \in Z} E||\frac{1}{nm} \sum_{j=1}^n \sum_{i=k}^{k+m} vecG_{j+iA} g_j'|| = o(1)$; $\widehat{G}_B \to_p E\widehat{G}_B$; **(vi)** $\max_{1 \le i \le n} ||G_{iA}|| = o_p(S_n^{-1} n^{1/2})$; $S_n n^{-1} \sum_{i=1}^n ||vecG_{inA} g_{in}'|| = O_p(1)$; $\max_{1 \le i \le n} ||G_{iB}|| = o_p(S_n^{-1} n)$; $S_n n^{-3/2} \sum_{i=1}^n ||vecG_{inB} g_{in}'|| = o_p(1)$; **(vii)** $n^{-1/2} \sum_{i=1}^n ((vec(G_{iA} - EG_{iA}))', g_i')' \to_d N(0, V)$, where

$$V := \lim_{n \to \infty} var(n^{-1/2} \sum_{i=1}^n (vecG_{iA}', g_i'))' \in \mathcal{R}^{k(p_A+1) \times k(p_A+1)}$$

has full column rank. Decompose V into

$$V = \begin{pmatrix} \Delta_{AA} & \Delta_A \\ \Delta_A' & \Delta \end{pmatrix}, \text{ where } \Delta_{AA} \in \mathcal{R}^{p_A k \times p_A k};$$

(viii) $B_s = EB_{si}$ exists and is f.r.r.; $\widehat{B}_s \to_p B_s$; $\max_{1 \le i \le n} ||B_{si}|| = o_p(S_n^{-1} n)$.

GSb provides a discussion of Assumption $\mathbf{M}_{\theta_0}$. The assumption $\max_{1 \le i \le n} ||B_{si}|| = o_p(S_n^{-1} n)$ in $\mathbf{M}_{\theta_0}$**(viii)** guarantees $\widehat{B}_{sn} - k_1 \widehat{B}_s = o_p(1)$; cf. Lemma A.1. Besides technical assumptions, $\mathbf{M}_{\theta_0}$ essentially states that the HAC estimator $\widetilde{J}_n$ is consistent and that a CLT holds for $((vec(G_{iA} - EG_{iA}))', g_i')'$, cf. Assumption 1 in Kleibergen (2005).

A.1.2 Subvector Tests

For the subvector tests we give high level assumptions. More primitive assumptions along the lines of Assumption $\mathbf{M}_{\theta_0}$ could be stated at the cost of additional space.

Let $\widehat{G}_{A_j}(\theta) := n^{-1}\sum_{i=1}^{n}(\partial g_i/\partial\alpha_j)(\theta)$ and likewise $\widehat{G}_{nA_j}(\theta) := n^{-1}\sum_{i=1}^{n}(\partial g_{in}/\partial\alpha_j)(\theta)$.

Assumption $\mathbf{M}_{\alpha_0}$: For any consistent estimators $\widetilde{\beta},\overline{\beta}\rightarrow_p \beta_0$ we have **(i)** $\max_{1\leq i\leq n}\sup_{\beta\in\mathcal{B}}||g_i(\theta_\beta)|| = o_p(S_n^{-1}n^{1/2})$; $S_n^{-1}n^{-1}\sum_{i=1}^{n} g_i(\widehat{\theta}_0) = o_p(1)$; **(ii)** for $S_n\rightarrow\infty$, $S_n = o(n^{1/2})$ we have $\widehat{\Delta}(\theta_{\widetilde{\beta}})\rightarrow_p k_2\Delta(\theta_0) > 0$; $\lambda_{\max}(\widehat{\Delta}(\widehat{\theta}_0))$ is bounded w.p.a.1; $S_n n^{-1}\sum_{i=1}^{n}||g_{in}(\theta_{\widetilde{\beta}})g_{in}(\theta_{\overline{\beta}})'|| = O_p(1)$; **(iii)** $\widehat{G}_B(\theta_{\widetilde{\beta}})$ exists; $\widehat{G}_B(\theta_{\widetilde{\beta}})\rightarrow_p E\widehat{G}_B(\theta_{\widetilde{\beta}}) = n^{-1/2}(\partial m_{1n}/\partial\beta)(\theta_{\widetilde{\beta}})+(\partial m_2/\partial\beta)(\alpha_{02},\widetilde{\beta})\rightarrow (\partial m_2/\partial\beta)(\alpha_{02},\beta_0)$; $n^{-1}S_n^{-1}\sum_{i=1}^{n} G_{iB}(\theta_{\widetilde{\beta}}) = o_p(1)$; $\max_{1\leq i\leq n}||G_{iB}(\theta_{\widetilde{\beta}})|| = o_p(S_n^{-1}n)$; **(iv)** $\widehat{g}(\widehat{\theta}_0)\rightarrow_p E\widehat{g}(\widehat{\theta}_0)$; $\Psi_n(\theta_0)\rightarrow_d \Psi(\theta_0)$, where $\Psi(\theta_0)\equiv N(0,\Delta(\theta_0))$; **(v)** $(\partial vec\widehat{G}_{A_1}/\partial\beta)(\theta)$ exists on a neighborhood of θ_0 and $(\partial vec\widehat{G}_{A_1}/\partial\beta)(\theta_{\widetilde{\beta}})\rightarrow_p 0$; **(vi)** $\max_{1\leq i\leq n}||G_{iA_1}(\theta_{\widetilde{\beta}})|| = o_p(S_n^{-1}n^{1/2})$; $n^{-1/2}\sum_{i=1}^{n}((vec(G_{iA_1}(\theta_0)-EG_{iA_1}(\theta_0)))', (g_i(\theta_0)-Eg_i(\theta_0))')'\rightarrow_d N(0,V^\alpha)$, where V^α is the appropriate submatrix of V defined in $\mathbf{M}_{\theta_0}$**(vii)**; $V^\alpha > 0$; **(vii)** $S_n n^{-1}\sum_{i=1}^{n} vec(G_{inA_1}(\theta_{\widetilde{\beta}}))g_{in}(\theta_{\widetilde{\beta}})'\rightarrow_p k_2\Delta_{A_1}$ defined in (A.6) below; $S_n n^{-1}\sum_{i=1}^{n}||vecG_{inA_1}(\theta_{\widetilde{\beta}})g_{in}(\theta_{\widetilde{\beta}})'|| = O_p(1)$; $S_n n^{-3/2}\sum_{i=1}^{n}||vecG_{inA_2}(\theta_{\widetilde{\beta}})g_{in}(\theta_{\widetilde{\beta}})'|| = o_p(1)$; similar to **(iii)**, $\widehat{G}_{A_2}(\theta_{\widetilde{\beta}})$ exists and $\widehat{G}_{A_2}(\theta_{\widetilde{\beta}})\rightarrow_p (\partial m_2/\partial\alpha_2)\ (\alpha_{02},\beta_0)$; $\max_{1\leq i\leq n}||G_{iA_2}(\theta_{\widetilde{\beta}})|| = o_p(S_n^{-1}n)$; **(viii)** $\widehat{B}_s(\theta_{\widetilde{\beta}})$ exists; $\widehat{B}_s(\theta_{\widetilde{\beta}})\rightarrow_p B_s(\theta_0)$; $\max_{1\leq i\leq n}||B_{si}(\theta_{\widetilde{\beta}})|| = o_p(S_n^{-1}n)$.

In $\mathbf{M}_{\alpha_0}$**(vi)** write

$$V^\alpha = \begin{pmatrix} \Delta_{A_1A_1} & \Delta_{A_1} \\ \Delta_{A_1}' & \Delta \end{pmatrix},\ \text{where } \Delta_{A_1A_1}\in R^{p_{A_1}k\times p_{A_1}k}. \tag{A.6}$$

A.2 Proofs

The next lemmata are helpful in the proof of the main results. Where there is no possibility of confusion, the argument θ_0 is omitted to simplify the notation.

Let

$$\widehat{G}_n(\theta) := n^{-1}\sum_{i=1}^{n} G_{in}(\theta).$$

Write $\widehat{G}_n(\theta) := (\widehat{G}_{nA}(\theta),\widehat{G}_{nB}(\theta))$, where $\widehat{G}_{nA}(\theta)\in\mathcal{R}^{k\times p_A}$ and $\widehat{G}_{nB}(\theta)\in\mathcal{R}^{k\times p_B}$.

The assumptions made in Lemma A.1 below are implied by $\mathbf{M}_{\theta_0}$**(i)**, **(iii)**, **(vi)**, and **(vii)**, e.g. $\widehat{G}_A(\theta_0) = O_p(n^{-1/2})$ follows from $\mathbf{M}_{\theta_0}$**(vii)** and eq. (A.4).

Lemma A.1 *Suppose* $S_n \to \infty$, $S_n = o(n^{1/2})$ *and* $\int_{-\infty}^{\infty} |a|\, k(a)da = O(1)$.

$$\text{If } \max_{1\leq i\leq n} ||g_i|| = o_p(S_n^{-1}n^{1/2}), \widehat{g} = O_p(n^{-1/2}) \text{ then } n^{1/2}(\widehat{g}_n - k_1\widehat{g}) = o_p(1).$$
$$\text{If } \max_{1\leq i\leq n} ||G_{iA}|| = o_p(S_n^{-1}n^{1/2}), \widehat{G}_A = O_p(n^{-1/2}) \text{ then } n^{1/2}(\widehat{G}_{nA} - k_1\widehat{G}_A) = o_p(1).$$

Proof: For the first result

$$\widehat{g}_n = n^{-1}\sum_{i=1}^{n} g_{in} = S_n^{-1}\sum_{j=-n+1}^{n-1} k\left(\frac{j}{S_n}\right) n^{-1}\sum_{i=\max[1,1-j]}^{\min[n,n-j]} g_i.$$

The difference between $\sum_{i=\max[1,1-j]}^{\min[n,n-j]} g_i/n$ and $\widehat{g}$ consists of no more than $|j|$ terms. Since $\max_{1\leq i\leq n} ||g_i|| = o_p(S_n^{-1}n^{1/2})$

$$\left\| n^{-1}\sum_{i=\max[1,1-j]}^{\min[n,n-j]} g_i - \widehat{g}\right\| \leq \left|\frac{j}{n}\right| \max_{1\leq i\leq |j|} \|g_i\| = \left|\frac{j}{n}\right| o_p(S_n^{-1}n^{1/2}).$$

Substituting into (A.7),

$$\widehat{g}_n = S_n^{-1}\sum_{j=-n+1}^{n-1} k\left(\frac{j}{S_n}\right)\left(\widehat{g} + \left|\frac{j}{n}\right| o_p(S_n^{-1}n^{1/2})\right).$$

Now

$$\begin{aligned}
\lim_{n\to\infty} S_n^{-1}\sum_{j=-n+1}^{n-1}\left|\frac{j}{S_n}\right| k\left(\frac{j}{S_n}\right) &= \lim_{n\to\infty}\int_{(-n+1)/S_n}^{(n-1)/S_n} |a|\, k(a)da \\
&= \int_{-\infty}^{\infty} |a|\, k(a)da = O(1)
\end{aligned}$$

by hypothesis. Hence,

$$\begin{aligned}
\widehat{g}_n &= S_n^{-1}\sum_{j=-n+1}^{n-1} k\left(\frac{j}{S_n}\right)\widehat{g} + o_p(n^{-1/2}) \qquad (A.7)\\
&= (k_1 + o(1))\widehat{g} + o_p(n^{-1/2}).
\end{aligned}$$

Therefore, since $\widehat{g} = O_p(n^{-1/2})$ which render the remainder terms $o_p(n^{-1/2})$,

$$n^{1/2}\widehat{g}_n = k_1 n^{1/2}\widehat{g} + o_p(1).$$

The proof of the second relation is derived in an identical fashion. □

It is now shown that under $\mathbf{M}_{\theta_0}$, $\widehat{\Delta}/2$ and $\widehat{\Delta}_A/2$ are consistent for Δ and Δ_A, cf. Lemma 2 in GSb The first part of the following lemma is similar to Lemma A.3 in Smith (2004). Note that the assumptions in the Lemma are part of $\mathbf{M}_{\theta_0}$**(ii)** and **(v)**.

Lemma A.2 *For $S_n \to \infty$ assume $S_n = o(n^{1/2})$. If $\sup_{i,j\geq 1} E||g_i g_j'|| < \infty$ and $\sup_{k\in Z} E||\frac{1}{nS_n}\sum_{j=1}^{n}\sum_{i=k}^{k+S_n}$ $o(1)$ then*

$$\widehat{\Delta} - k_2 \widetilde{J}_n((g_i),(g_i)) = o_p(1).$$

If $\sup_{i,j\geq 1} E||vecG_{iA}g_j'|| < \infty$ and $\sup_{k\in Z} E||\frac{1}{nS_n}\sum_{j=1}^{n}\sum_{i=k}^{k+S_n} vecG_{j+iA}g_j'|| = o(1)$ then

$$\widehat{\Delta}_A - k_2 \widetilde{J}_n((vecG_{iA}),(g_i)) = o_p(1). \tag{A.8}$$

Proof: Similarly to the proof of Lemma A.3 in Smith (2004)

$$\widehat{\Delta} = \sum_{j=-n+1}^{n-1}\left[\frac{1}{S_n}\sum_{i=\max[1-n,1-n+j]}^{\min[n-1,n-1+j]} k\left(\frac{i-j}{S_n}\right)k\left(\frac{i}{S_n}\right)\right] n^{-1}\sum_{k=\max[1,1-j,1-i]}^{\min[n,n-j,n-i]} g_{k+j}g_k'.$$

The difference between $\sum_{k=\max[1,1-j,1-i]}^{\min[n,n-j,n-i]} g_{k+j}g_k'/n$ and $\widetilde{\Gamma}_j((g_i),(g_i))$ consists of no more than $|i|$ terms. By M,

$$\begin{aligned}\mathcal{P}\left\{\left|\frac{1}{n}\sum_{k=1}^{|i|}(g_{k+j}g_k' - Eg_{k+j}g_k'\right| \geq \varepsilon\right\} &\leq E\left[\left\|\sum_{k=1}^{|i|}(g_{k+j}g_k' - Eg_{k+j}g_k')\right\|\right]/n\varepsilon \\ &= |i|\,O(n^{-1})\end{aligned}$$

uniformly j and the $O(n^{-1})$ term is independent of i. Therefore,

$$\left\|\frac{1}{n}\sum_{k=\max[1,1-j,1-i]}^{\min[n,n-j,n-i]} g_{k+j}g_k' - \widetilde{\Gamma}_j((g_i),(g_i))\right\| \leq \frac{|i|}{n}\max_{1\leq j,k\leq n}\|Eg_jg_k'\| + |i|\,O_p(n^{-1})$$

uniformly j. Hence, using Lemmata C.1 and C.2 in Smith (2004), as $\Delta < \infty$ by M_{θ_0},

$$\begin{aligned}\widehat{\Delta}/k_2 &= \left(\sum_{j=-n+1}^{n-1}\frac{1}{S_n}\sum_{i=\max[1-n,1-n+j]}^{\min[n-1,n-1+j]} k\left(\frac{i-j}{S_n}\right)k\left(\frac{i}{S_n}\right)\right)\widetilde{\Gamma}_j((g_i),(g_i))/k_2 + o_p(1) \\ &= \sum_{j=-n+1}^{n-1}(k^*\left(\frac{j}{S_n}\right) + o(1))\widetilde{\Gamma}_j((g_i),(g_i)) + o_p(1) \\ &= \sum_{j=-n+1}^{n-1} k^*\left(\frac{j}{S_T}\right)\widetilde{\Gamma}_j((g_i),(g_i)) + o_p(1),\end{aligned}$$

where the remainder terms are uniform in j, since $\lim_{n\to\infty} E[\sum_{j=-n+1}^{n-1}\widetilde{\Gamma}_j((g_i),(g_i))] = \Delta$ and $\lim_{n\to\infty} var[\sum_{j=-n+1}^{n-1}\widetilde{\Gamma}_j((g_i),(g_i))] = O(1)$ by standard results on the inconsistency of the periodogram. Therefore, the result follows. The proof of the second claim is completely analogous and therefore omitted. □

Given the results in Lemma A.1 and consistency of $\widehat{\Delta}/2$ and $\widehat{\Delta}_A/2$, the proof of Theorems 3.1, 3.2, 3.3 and 3.4 follow those of Theorems 1 and 2 in GSb.

We require two lemmata that mirror Lemmata 3 and 4 in GSb which in turn modify Lemmata A1 and A2 in NS. These Lemmata are necessary since we allow for a more general form of kernel $k(\cdot)$ than GSb. Let $c_n := S_n n^{-1/2} \max_{1\leq i\leq n} ||g_{in}(\theta_0)||$. Let $\Lambda_n := \{\lambda \in \mathcal{R}^k : ||\lambda|| \leq S_n n^{-1/2} c_n^{-1/2}\}$ if $c_n \neq 0$ and $\Lambda_n = \mathcal{R}^k$ otherwise.

Let $\lambda_0 := \lambda(\theta_0)$ and $g_i := g_i(\theta_0)$, $g_{in} := g_{in}(\theta_0)$, $(i = 1, ..., n)$. Similarly, define $\widehat{g} := \widehat{g}(\theta_0)$ and $\widehat{g}_n := \widehat{g}_n(\theta_0)$.

Lemma A.3 *Assume* $\max_{1\leq i\leq n} ||g_i|| = o_p(S_n^{-1} n^{1/2})$. *Then* $\sup_{\lambda\in\Lambda_n, 1\leq i\leq n} |\lambda' g_{in}| \to_p 0$ *and* $\Lambda_n \subset \widehat{\Lambda}_n(\theta_0)$ *w.p.a.1.*

Proof: W.l.o.g. assume $c_n \neq 0$. Since $||g_{in}|| \leq S_n^{-1} \sum_{j=i-n}^{i-1} |k(j/S_n)| \, \|g_{i-j}\|$, by $\mathbf{M}_{S_n,k(\cdot)}$(iii)

$$\begin{aligned} \max_{1\leq i\leq n} ||g_{in}|| &\leq \max_{1\leq i\leq n} \|g_i\| \, S_n^{-1} \sum_{j=i-n}^{i-1} |k(j/S_n)| \\ &\leq \max_{1\leq i\leq n} \|g_i\| \lim_{n\to\infty} \int_{(-n+1)/S_n}^{(n-1)/S_n} |k(a)| \, da = o_p(S_n^{-1} n^{1/2}). \end{aligned}$$

Therefore, both parts of the statement follow immediately from the proof of Lemma 3 in GSb. □

Lemma A.4 *Suppose* $\max_{1\leq i\leq n} ||g_i|| = o_p(S_n^{-1} n^{1/2})$, $\lambda_{\min}(\widehat{\Delta}(\theta_0)) \geq \varepsilon$ *w.p.a.1 for some* $\varepsilon > 0$, $\widehat{g}_n = O_p(n^{-1/2})$ *and Assumption* $\boldsymbol{\rho}$ *holds. Then* $\lambda_0 \in \widehat{\Lambda}_n(\theta_0)$ *satisfying* $\widehat{P}_\rho(\theta_0, \lambda_0) = \sup_{\lambda\in\widehat{\Lambda}_n(\theta_0)} \widehat{P}_\rho(\theta_0, \lambda)$ *exists w.p.a.1,* $\lambda_0 = O_p(S_n n^{-1/2})$ *and* $\sup_{\lambda\in\widehat{\Lambda}_n(\theta_0)} \widehat{P}_\rho(\theta_0, \lambda) = O_p(S_n n^{-1})$.

Proof: The proof is identical to that for Lemma 4 in GSb. □

Lemma A.5 *Assume* $\max_{1\leq i\leq n}||g_i|| = o_p(S_n^{-1}n^{1/2})$*,* $\lambda_{\min}(\widehat{\Delta}(\theta_0)) \geq \varepsilon$ *w.p.a.1 for some* $\varepsilon > 0$*,* $\widehat{g}_n = O_p(n^{-1/2})$ *and Assumption* $\boldsymbol{\rho}$ *holds. If Assumptions 2.1, 2.2 and 2.3 are satisfied, then* $n\widehat{\pi}_i = 1 + o_p(1)$ *and*

$$(n/S_n^2)^{1/2}\left(\widehat{\pi}_i - \frac{1}{n}\right) = \frac{1}{n}kg_{in}'(n/S_n^2)^{1/2}\lambda_0(1 + o_p(1)) + O_p(n^{-3/2}),$$

uniformly $(i = 1, ..., n)$*.*

Proof: From the proof of Lemma A.4 $\lambda_0 \in int(\Lambda_n)$. Hence, the hypotheses of Lemma A.3 are satisfied and $\max_{1\leq i\leq n}|\lambda_0' g_{in}| \to_p 0$. A first order order Taylor expansion yields

$$\rho_1(k\lambda_0' g_{in}) = -1 + \rho_2(k\widetilde{\lambda}' g_{in})k\lambda_0' g_{in},$$

where $\widetilde{\lambda}$ is on the line segment joining λ_0 and 0. Now, $\max_{1\leq i\leq n}\left|\rho_2(k\widetilde{\lambda}' g_{in}) + 1\right| \xrightarrow{p} 0$ since $\max_{1\leq i\leq n}\left|\widetilde{\lambda}' g_{in}\right| \xrightarrow{p} 0$ by Lemma A.3 and so $\rho_2(\widetilde{\lambda}' g_{in})\lambda_0' g_{in} = -\lambda_0' g_{in}(1 + o_p(1)) = o_p(1)$, uniformly $(i = 1, ..., n)$. Therefore,

$$\rho_1(k\lambda_0' g_{in}) = -1 - k\lambda_0' g_{in}(1 + o_p(1)), \tag{A.9}$$

uniformly $(i = 1, ..., n)$. Similarly,

$$\begin{aligned}
\frac{1}{\sum_{j=1}^n \rho_1(k\lambda_0' g_{jn})} &= -\frac{1}{n} - \frac{1}{n(\sum_{j=1}^n \rho_1(k\lambda_0' g_{jn})/n)^2}\left(\sum_{j=1}^n \rho_2(k\widetilde{\lambda}' g_{jn})g_{jn}'/n\right)k\lambda_0 \\
&= -\frac{1}{n} - \frac{1}{n}\left(\sum_{j=1}^n \rho_2(k\widetilde{\lambda}' g_{jn})g_{jn}'/n\right)k\lambda_0 \\
&= -\frac{1}{n} + \frac{1}{n}(1 + o_p(1))k\lambda_0'\widehat{g}_n \\
&= -\frac{1}{n}(1 + O_p(S_n n^{-1})),
\end{aligned}$$

where $\widehat{g}_n = \widehat{g}_n(\theta_0)$. In (A.10), the second equality follows from (A.9) since $\rho_1(k\lambda_0' g_{in}) = -1 + o_p(1)$, uniformly $(i = 1, ..., n)$, the third equality from $\rho_2(\widetilde{\lambda}' g_{in})\lambda_0' g_{in} = -\lambda_0' g_{in}(1 + o_p(1))$, uniformly $(i = 1, ..., n)$, and the final equality holds since $\widehat{g}_n = O_p(n^{-1/2})$ by hypothesis and $\lambda_0 = O_p(S_n n^{-1/2})$ by Lemma A.4. Combining (A.9) and (A.10)

$$\widehat{\pi}_i = \frac{1}{n}(1 + k\lambda_0' g_{in}(1 + o_p(1)))(1 + O_p(S_n n^{-1})).$$

Therefore, from Lemma A.3,

$$\begin{aligned} n\widehat{\pi}_i - 1 &= k\lambda_0' g_{in}(1 + o_p(1)) + O_p(S_n n^{-1}) \\ &= o_p(1), \end{aligned}$$

uniformly $(i = 1, ..., n)$, yielding the first part of the lemma. For the second part

$$(n/S_n^2)^{1/2}\left(\widehat{\pi}_i - \frac{1}{n}\right) = \frac{1}{n}kg_{in}'(n/S_n^2)^{1/2}\lambda_0(1 + o_p(1)) + O_p(n^{-3/2}),$$

uniformly $(i = 1, ..., n)$. □

Define the observation distribution function estimator and the empirical distribution function (EDF)

$$\hat{\mu}_n(z) = \sum_{i=1}^{n} \hat{\pi}_i 1(z_i \leq z), \mu_n(z) = \sum_{i=1}^{n} 1(z_i \leq z)/n. \tag{A.10}$$

Lemma A.6 *Assume* $\max_{1\leq i\leq n} ||g_i|| = o_p(S_n^{-1}n^{1/2})$, $\max_{1\leq i\leq n} ||1(z_i \leq z)g_i|| = o_p(S_n^{-1}n)$, $\sum_{i=1}^{n} 1(z_i \leq z)g_i/n \to_p b(z)$, $\Psi_n \to_d \Psi = N(0, \Delta)$, $\lambda_{\min}(\widehat{\Delta}(\theta_0)) \geq \varepsilon$ *w.p.a.1 for some* $\varepsilon > 0$ *and Assumption* ρ *holds. If Assumptions 2.1, 2.2 and 2.3 are satisfied, then*

$$(n/S_n^2)^{1/2}(\widehat{\mu}_n - \mu_n) \Rightarrow (k_1^2/k_2)\widehat{\Lambda}$$

where $\widehat{\Lambda}$ *is a Gaussian stochastic process on* $\mathcal{R}^l$ *with mean zero and covariance function* $E[\hat{\Lambda}(z_1)\hat{\Lambda}(z_2)] = b(z_1)'\Delta^{-1}b(z_2)$.

Proof: A Taylor expansion of the FOC for λ_0 gives

$$0 = -\widehat{g}_n - k\widehat{\Delta}_{\widetilde{\lambda}} S_n^{-1}\lambda_0$$

where $\widehat{\Delta}_{\widetilde{\lambda}} = -S_n \sum_{i=1}^{n} \rho_2(\widetilde{\lambda}' g_{in}) g_{in} g_{in}'/n$ with $\widetilde{\lambda}$ is a mean value $\widetilde{\lambda}$ between 0 and λ_0 (that may be different for each row). Since $\lambda_0 = O_p(S_n n^{-1/2})$, Lemma A.3 and Assumption ρ imply that $\max_{1\leq i\leq n} |\rho_2(\widetilde{\lambda}' g_{in}) + 1| \to_p 0$. Note that Assumption $\mathbf{M}_{\theta_0}$**(ii)** and Lemma A.2 imply $\widehat{\Delta}_{\widetilde{\lambda}} \to_p k_2\Delta > 0$. Thus $\widehat{\Delta}_{\widetilde{\lambda}}$ is invertible w.p.a.1 and $(\widehat{\Delta}_{\widetilde{\lambda}})^{-1} \to_p \Delta^{-1}/k_2$. Therefore

$$\begin{aligned} (n/S_n^2)^{1/2}\lambda_0 &= -(\widehat{\Delta}_{\widetilde{\lambda}})^{-1} n^{1/2}\widehat{g}_n \\ &= -((k_1)^{-1}\Delta^{-1} + o_p(1))n^{1/2}\widehat{g}_n \\ &= -\Delta^{-1}n^{1/2}\widehat{g} + o_p(1) \end{aligned}$$

w.p.a.1., where the third equality follows from Lemma A.1 since $n^{1/2}\widehat{g}_n = k_1 n^{1/2}\widehat{g} + o_p(1)$. Therefore

$$\begin{aligned}
(n/S_n^2)^{1/2}[\widehat{\mu}_n(z) - \mu_n(z)] &= (n/S_n^2)^{1/2}\sum_{i=1}^{n}\left(\widehat{\pi}_i - \frac{1}{n}\right)1(z_i \leq z) \\
&= \frac{1}{n}\sum_{i=1}^{n} 1(z_i \leq z)[kg_{in}'(n/S_n^2)^{1/2}\lambda_0(1 + o_p(1)) + O_p(n^{-1/2})] \\
&= k\left(\frac{1}{n}\sum_{i=1}^{n} 1(z_i \leq z)g_{in}'\right)(n/S_n^2)^{1/2}\lambda_0(1 + o_p(1)) + O_p(n^{-1/2}) \\
&= (k_1^2/k_2)[b(z) + o_p(1)]'(n/S_n^2)^{1/2}\lambda_0 + o_p(1) \\
&= -(k_1^2/k_2)[b(z) + o_p(1)]'\Delta^{-1}n^{1/2}\widehat{g} + o_p(1) \\
&\Rightarrow (k_1^2/k_2)\hat{\Lambda}(z),
\end{aligned}$$

where Lemma A.5 yields the second equality, the third and fourth equalities by hypothesis and $n^{-1}\sum_{i=1}^{n} 1(z_i \leq z)(g_{in} - k_1 g_i) = o_p(1)$ similarly to Lemma A.1, and the final equality from eq. (A.11). The result then follows from Assumption $\mathbf{M}_{\theta_0}$**(iii)** since $n^{1/2}\widehat{g} \rightarrow_d N(0, \Delta)$. □

A.2.1 Full Vector Tests

Proof of Theorem 3.1: The proof closely follows that of Theorem 1(i) in GSb.

Substituting for λ_0 into a second order Taylor expansion for $\widehat{P}(\theta_0, \lambda_0)$ at $(\theta_0, 0)$ with mean value λ^*

$$(n/S_n)\widehat{P}_\rho(\theta_0, \lambda_0) = 2n\widehat{g}_n'\widehat{\Delta}_{\widetilde{\lambda}}^{-1}\widehat{g}_n - n\widehat{g}_n'\widehat{\Delta}_{\widetilde{\lambda}}^{-1}\widehat{\Delta}_{\lambda^*}\widehat{\Delta}_{\widetilde{\lambda}}^{-1}\widehat{g}_n \tag{A.11}$$

w.p.a.1. where $\widehat{\Delta}_{\lambda^*}$ is defined similarly to $\widehat{\Delta}_{\widetilde{\lambda}}$ in the proof of Lemma A.6 above. Hence, $\widehat{\Delta}_{\lambda^*} \rightarrow_p k_2\Delta$. Therefore, from Lemmata A.1 and A.6, w.p.a.1

$$\begin{aligned}
(n/S_n)\widehat{P}_\rho(\theta_0, \lambda_0) &= n\widehat{g}_n'\Delta^{-1}\widehat{g}_n/k_2 \\
&= (k_1^2/k_2)n\widehat{g}'\Delta^{-1}\widehat{g} \\
&= (k_1^2/k_2)(n/S_n^2)\lambda_0'\Delta\lambda_0 \\
&\rightarrow_d (k_1^2/k_2)\chi^2(m).
\end{aligned}$$

From Lemma A.5

$$\begin{aligned} S_n^{-1}\sum_{i=1}^n (n\widehat{\pi}_i - 1)^2 &= \sum_{i=1}^n (k\lambda_0' g_{in}(1+o_p(1)) + O_p(S_n n^{-1}))^2/S_n \\ &= k^2(n/S_n^2)\lambda_0'(S_n \sum_{i=1}^n g_{in}g_{in}'/n)\lambda_0(1+o_p(1)) \\ &\quad + O_p(S_n n^{-1})(1+o_p(1))k(n/S_n^2)^{1/2}\lambda_0' n^{1/2}\widehat{g}_n + O_p(S_n n^{-2}) \\ &= (k_1^2/k_2)(n/S_n^2)\lambda_0'\Delta\lambda_0(1+o_p(1)) + O_p(S_n n^{-1}) + O_p(S_n n^{-2}) \\ &= (k_1^2/k_2)(n/S_n^2)\lambda_0'\Delta\lambda_0(1+o_p(1)) + o_p(1). \end{aligned}$$

The third equality holds since $\widehat{g}_n = O_p(n^{-1/2})$ by the hypotheses of the theorem and Lemma A.4.

For P_ρ, by a similar reasoning to that in Lemma A.6,

$$\begin{aligned} (n/S_n^2)^{1/2}(\widehat{\mu}_n^s - \mu_n^s) &= (n/S_n^2)^{1/2}\sum_{i=1}^n \left(\widehat{\pi}_i - \frac{1}{n}\right)\begin{pmatrix} 1(z_i \in \mathcal{Z}_1) \\ \dots \\ 1(z_i \in \mathcal{Z}_s) \end{pmatrix} \qquad \text{(A.12)} \\ &= k\left(\frac{1}{n}\sum_{i=1}^n \begin{pmatrix} 1(z_i \in \mathcal{Z}_1) \\ \dots \\ 1(z_i \in \mathcal{Z}_s) \end{pmatrix} g_{in}'\right)(n/S_n^2)^{1/2}\lambda_0(1+o_p(1)) + O_p(n^{-1/2}) \\ &= k(k_1 B_s + o_p(1))'(n/S_n^2)^{1/2}\lambda_0 + o_p(1) \\ &= -kB_s'\Delta^{-1}n^{1/2}\widehat{g}_n + o_p(1) \\ &= -(k_1^2/k_2)B_s'\Delta^{-1}n^{1/2}\widehat{g} + o_p(1) \\ &\to_d (k_1^2/k_2)N(0, B_s'\Delta^{-1}B_s). \end{aligned}$$

Since B_s is full row rank, $B_s'(B_sB_s')^{-1}\Delta(B_sB_s')^{-1}B_s$ is a g-inverse for $B_s'\Delta^{-1}B_s$ and the result follows. □

Proof of Theorem 3.2: The proof closely follows that of Theorem 1(ii) in GSb.

Define $D := D_\rho(\theta_0)\Lambda$ where the $p \times p$ diagonal matrix $\Lambda := diag(n^{1/2}, ..., n^{1/2}, 1, ..., 1)$ has first p_A diagonal elements equal to $n^{1/2}$ and the remainder equal to unity. Then, it follows that

$$LM_\rho = n\widehat{g}_n'\widehat{\Delta}^{-1}D(D'\widehat{\Delta}^{-1}D)^{-1}D'\widehat{\Delta}^{-1}\widehat{g}_n/(k_1^2/k_2). \qquad \text{(A.13)}$$

From (A.11) and $n^{1/2}\widehat{g}_n = O_p(1)$ that

$$(n/S_n^2)^{1/2}\lambda_0 = -(k_1)^{-1}\Delta^{-1}n^{1/2}\widehat{g}_n + o_p(1). \tag{A.14}$$

Therefore, only the statistic LM_ρ (A.13) is analysed, the argument for S_ρ following immediately from eq. (A.14).

As in GSb, D is asymptotically independent of $n^{1/2}\widehat{g}_n$. To see this, we follow GSb. Now $\rho_1(k\lambda_0' g_{in}) = -1 + o_p(1)$, uniformly $(i = 1, ..., n)$, by Lemma A.3. By (3.5), (A.14) and the definition of Λ (modulo $o_p(1)$)

$$D = -(n^{-1/2}\sum_{i=1}^{n} G_{inA} - (S_n n^{-1}\sum_{i=1}^{n} G_{inA}g_{in}'/k_2)\Delta^{-1}n^{1/2}\widehat{g}_n, k_2 M_2(\beta_0)), \tag{A.15}$$

using (A.4) and Assumptions $\mathbf{M}_{\theta_0}\mathbf{(v)}$ and $\mathbf{(vi)}$. By Assumption $\mathbf{M}_\theta\mathbf{(v)}$ and eq. (A.8) $\widehat{\Delta}_A = S_n n^{-1}\sum_{i=1}^n vec(G_{inA})g_{in}'/k_2 \to_p \Delta_A$. Define

$$\begin{aligned} w_1 &:= vec(0, -M_2(\beta_0), 0) \in \mathcal{R}^{kp_A + kp_B + k}, \\ M &:= \begin{pmatrix} -I_{kp_A} & \Delta_A\Delta^{-1} \\ 0 & 0 \\ 0 & I_k \end{pmatrix}, \ v := n^{-1/2}\sum_{i=1}^{n}\begin{pmatrix} vecG_{inA} \\ g_{in} \end{pmatrix}. \end{aligned}$$

where M and v are $(kp_A + kp_B + k) \times (kp_A + k)$ and $(kp_A + k) \times 1$ matrices respectively. Thus

$$vec(D, n^{1/2}\widehat{g}_n) = k_2 w_1 + Mv + o_p(1). \tag{A.16}$$

By Assumption $\mathbf{ID}_{\theta_0}$, $\mathbf{M}_{\theta_0}\mathbf{(vii)}$, Lemma A.1 and (A.4) it follows that $v \to_d k_2 N(w_2, V)$, where $w_2 := ((vecM_{1A})', 0)'$ and M_{1A} are the first p_A columns of M_1. Therefore,

$$vec(D, n^{1/2}\widehat{g}_n) \to_d k_2 N(w_1 + Mw_2, \begin{pmatrix} \Psi & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Delta \end{pmatrix}), \tag{A.17}$$

where $\Psi := \Delta_{AA} - \Delta_A\Delta^{-1}\Delta_A'$ has full column rank, proving that D and $n^{1/2}\widehat{g}_n$ are asymptotically independent.

Let $\overline{D}$ and $\overline{g}$ denote the limiting normal random matrices corresponding to D and $n^{1/2}\widehat{g}_n$, respectively, see (A.17). The function $h : \mathcal{R}^{k\times p} \to \mathcal{R}^{p\times k}$ defined by $h(d) :=$

$(d'\Delta^{-1}d)^{-1/2}d'$ for $d \in \mathcal{R}^{k\times p}$ is continuous on a set $C \subset \mathcal{R}^{k\times p}$ with $\mathcal{P}\{\overline{D} \in C\} = 1$; see GSb. Hence, by the Continuous Mapping Theorem and $\mathbf{M}_{\theta_0}(\mathbf{v})$ it follows that

$$(D'(\widehat{\Delta})^{-1}D)^{-1/2}D'(\widehat{\Delta})^{-1}n^{1/2}\widehat{g}_n \to_d (\overline{D}'\Delta^{-1}\overline{D})^{-1/2}\overline{D}'\Delta^{-1}\overline{g}. \tag{A.18}$$

Therefore, given the independence of $\overline{D}$ and $\overline{g}$, the latter random variable is distributed as ζ, where $\zeta \sim k_1 N(0, I_p)$.

By a similar argument to that for LM_ρ,

$$P_\rho^r = (n/S_n^2)(\widehat{\mu}_n^s - \mu_n^s)'\widehat{B}_s'(\widehat{B}_s\widehat{B}_s')^{-1}D(D'(\widehat{\Delta})^{-1}D)^{-1}D'(\widehat{B}_s\widehat{B}_s')^{-1}\widehat{B}_s(\widehat{\mu}_n^s - \mu_n^s).$$

Since $\widehat{B}_s \to_p B_s$ by hypothesis, from (A.12),

$$\begin{aligned}(D'(\widehat{\Delta})^{-1}D)^{-1/2}D'(\widehat{B}_s\widehat{B}_s')^{-1}\widehat{B}_s(n/S_n^2)^{1/2}(\widehat{\mu}_n^s - \mu_n^s) &\to {}_d -k(\overline{D}'\Delta^{-1}\overline{D})^{-1/2}\overline{D}'\Delta^{-1}\overline{g} \\ &\sim (k_1^2/k_2)N(0, I_p).\end{aligned}$$

□

A.2.2 Subvector Tests

Proof of Theorem 3.3: The proof closely follows that of Theorem 2(i) in GSb.

By an identical argument to that in the proof of Theorem 2(i) of GSb $\widehat{\beta} \to_p \beta_0$.

Likewise, similarly to Lemma A.4, $\widehat{g}_n(\widehat{\theta}_0) = O_p(n^{-1/2})$, gives $\widehat{\lambda} := \lambda(\widehat{\theta}_0)$ exists and a FOC of $\widehat{P}_\rho(\widehat{\theta}_0, \lambda)$ w.r.t. λ holds w.p.a.1. Moreover, cf. eqs. (A.11) and (A.14),

$$GELR_\rho^{sub}(\alpha_0) = n^{1/2}\widehat{g}_n(\widehat{\theta}_0)'\Delta(\theta_0)^{-1}n^{1/2}\widehat{g}_n(\widehat{\theta}_0)/k_1^2 + o_p(1). \tag{A.19}$$

By the implicit function and envelope theorems, $0 = n^{-1}\sum_{i=1}^n \rho_1(k\widehat{\lambda}'g_{in}(\widehat{\theta}_0))\,(\partial g_{in}/\partial\beta)'(\widehat{\theta}_0)S_n^{-1}\widehat{\lambda}$ holds. Togther with a mean–value expansion of the FOC for $(\widehat{\beta}, \widehat{\lambda})$ about $(\beta_0, 0)$ we obtain

$$\begin{pmatrix} 0 \\ -\widehat{g}_n(\theta_0) \end{pmatrix} + M \begin{pmatrix} \widehat{\beta} - \beta_0 \\ S_n^{-1}\widehat{\lambda} \end{pmatrix} = 0, \text{ where} \tag{A.20}$$

$$M := n^{-1}\sum_{i=1}^n \begin{pmatrix} 0 & \rho_1(k\widehat{\lambda}'g_{in}(\widehat{\theta}_0))(\partial g_{in}/\partial\beta)'(\widehat{\theta}_0) \\ \rho_1(k\overline{\lambda}'g_{in}(\widehat{\theta}_0))(\partial g_{in}/\partial\beta)(\theta_{\overline{\beta}}) & kS_n\rho_2(k\overline{\lambda}'g_{in}(\widehat{\theta}_0))g_{in}(\theta_{\overline{\beta}})g_{in}(\widehat{\theta}_0)' \end{pmatrix}$$

and $(\overline{\beta}', \overline{\lambda}')$ are mean values on the line segment joining $(\widehat{\beta}, \widehat{\lambda})$ and $(\beta_0', 0')$; cf. Newey and Smith (2004, Proof of Theorem 3.2, p.240). From the last two conditions in $\mathbf{M}_{\alpha_0}\mathbf{(iii)}$ and similarly to (A.7), $(\widehat{G}_{nB} - k_1\widehat{G}_B)(\theta_{\widetilde{\beta}}) = o_p(1)$ for any argument $\theta_{\widetilde{\beta}}$ as in $\mathbf{M}_{\alpha_0}$. Thus, also by $\mathbf{M}_{\alpha_0}\mathbf{(iii)}$, $\widehat{G}_B(\theta_{\widetilde{\beta}}) \to_p M_{2\beta}(\alpha_{02}, \beta_0)$, where $M_{2\beta}(\cdot) := (\partial m_2/\partial\beta)(\cdot) \in \mathcal{R}^{k\times p_B}$. Therefore, by $\mathbf{M}_{\alpha_0}$(ii), $M \to_p \overline{M}$, where

$$\begin{aligned}\overline{M} &: = -k_1 \begin{pmatrix} 0 & M_{2\beta}' \\ M_{2\beta} & \Delta \end{pmatrix}, \ \overline{M}^{-1} = -(k_1)^{-1} \begin{pmatrix} -\Sigma & H \\ H' & P \end{pmatrix}, \\ \Sigma &: = (M_{2\beta}'\Delta^{-1}M_{2\beta})^{-1}, \ H := \Sigma M_{2\beta}'\Delta^{-1}, P := \Delta^{-1} - \Delta^{-1}M_{2\beta}\Sigma M_{2\beta}'\Delta^{-1}\end{aligned}$$

and the respective arguments (α_{02}, β_0) in $M_{2\beta}$ and θ_0 in Δ have been omitted. By (A.20) w.p.a.1

$$(n^{1/2}(\widehat{\beta} - \beta_0)', (n/S_n^2)^{1/2}\widehat{\lambda}')' = M^{-1}(0', n^{1/2}\widehat{g}_n(\theta_0)')' = k_1\overline{M}^{-1}(0', n^{1/2}\widehat{g}(\theta_0)')' + o_p(1), \tag{A.21}$$

where the second equality holds by Lemma A.1 using $\mathbf{M}_{\alpha_0}\mathbf{(i)}$ and $\mathbf{(iv)}$. From an expansion of $\widehat{g}(\widehat{\theta}_0)$ in β around β_0 and (A.21), up to $o_p(1)$ terms,

$$n^{1/2}\widehat{g}_n(\widehat{\theta}_0) = n^{1/2}k_1\widehat{g}(\widehat{\theta}_0) = n^{1/2}k_1[\widehat{g}(\theta_0) + \widehat{G}_B(\overline{\theta})(\widehat{\beta} - \beta_0)] = k_1(I_k - M_{2\beta}H)n^{1/2}\widehat{g}(\theta_0) \tag{A.22}$$

for some appropriate mean–value $\overline{\theta}$, where the first equality can be established by an analogous expansion for $n^{1/2}\widehat{g}_n(\widehat{\theta}_0)$, Lemma A.1, and $n^{1/2}(\widehat{\beta} - \beta_0) = O_p(1)$. Note that $M_{M_{2\beta}}(\Delta) = I_k - M_{2\beta}H$ and $\Delta^{-1/2}M_{M_{2\beta}}(\Delta)\Delta^{1/2} = M_{\Delta^{-1/2}M_{2\beta}}$. Then, by (A.19) and $\mathbf{M}_{\alpha_0}\mathbf{(iv)}$, $GELR_\rho^{sub}(\alpha_0) \to_d \xi' M_{\Delta^{-1/2}M_{2\beta}}\xi$ for $\xi \sim N(0, I_k)$ and since $\Delta^{-1/2}M_{2\beta}$ is of rank p_B we obtain $GELR_\rho^{sub}(\alpha_0) \to_d \chi^2(k - p_B)$ as claimed.

Similarly to the proof of Lemma A.5, $\sup_{\lambda\in\Lambda_n, 1\leq i\leq n} \left|\rho_2(k\lambda' g_{in}(\widehat{\theta}_0)) + 1\right| \xrightarrow{p} 0$ where, now, $c_n = (n/S_n^2)^{-1/2} \max_{1\leq i\leq n} \sup_{\beta\in\mathcal{B}} ||g_{in}(\theta_\beta)||$. Therefore, again by a similar argument to that in the proof of Lemma A.5,

$$\begin{aligned} n\widehat{\pi}_i(\widehat{\theta}_0) - 1 &= k\widehat{\lambda}' g_{in}(\widehat{\theta}_0)(1 + o_p(1)) + O_p(S_n n^{-1}) \\ &= o_p(1), \end{aligned}$$

uniformly $(i = 1, ..., n)$, and

$$(n/S_n^2)^{1/2}\left(\widehat{\pi}_i(\widehat{\theta}_0) - \frac{1}{n}\right) = \frac{1}{n}kg_{in}(\widehat{\theta}_0)'(n/S_n^2)^{1/2}\widehat{\lambda}(1 + o_p(1)) + O_p(n^{-3/2}),$$

uniformly ($i = 1, ..., n$). Therefore, following the proof of Theorem 3.1) for P_ρ^a, from eqs. (A.21) and (A.22), Assumption $\mathbf{M}_{\alpha_0}\mathbf{(ii)}$,

$$\begin{aligned}
S_n^{-1}\sum_{i=1}^n (n\widehat{\pi}_i(\widehat{\theta}_0)-1)^2 &= \sum_{i=1}^n (k\widehat{\lambda}' g_{in}(\widehat{\theta}_0)(1+o_p(1)) + O_p(S_n n^{-1}))^2/S_n \\
&= k^2(n/S_n^2)\widehat{\lambda}'(S_n\sum_{i=1}^n g_{in}(\widehat{\theta}_0)g_{in}(\widehat{\theta}_0)'/n)\widehat{\lambda}(1+o_p(1)) \\
&\quad + O_p(S_n n^{-1})(1+o_p(1))k(n/S_n^2)^{1/2}\widehat{\lambda}' n^{1/2}\widehat{g}_n(\widehat{\theta}_0) + O_p(S_n n^{-2}) \\
&= (k_2)^{-1} n^{1/2}\widehat{g}_n(\widehat{\theta}_0)' P\Delta P n^{1/2}\widehat{g}_n(\widehat{\theta}_0)(1+o_p(1)) + O_p(S_n n^{-1}) + O_p(S_n n^{-2}) \\
&= (k_1^2/k_2) n^{1/2}\widehat{g}(\theta_0)' P\Delta P n^{1/2}\widehat{g}(\theta_0)(1+o_p(1)) + o_p(1) \\
&= (k_1^2/k_2)\xi' M_{\Delta^{-1/2}M_{2\beta}}\xi(1+o_p(1)) + o_p(1).
\end{aligned}$$

For $P_\rho(\alpha_0)$,

$$\begin{aligned}
(n/S_n^2)^{1/2}(\widehat{\mu}_n^s(\widehat{\theta}_0) - \mu_n^s) &= n^{1/2}\sum_{i=1}^n \left(\widehat{\pi}_i(\widehat{\theta}_0) - \frac{1}{n}\right)\begin{pmatrix} 1(z_i \in \mathcal{Z}_1) \\ \cdots \\ 1(z_i \in \mathcal{Z}_s)\end{pmatrix} \\
&= k\left(\frac{1}{n}\sum_{i=1}^n \begin{pmatrix} 1(z_i \in \mathcal{Z}_1) \\ \cdots \\ 1(z_i \in \mathcal{Z}_s)\end{pmatrix} g_{in}(\widehat{\theta}_0)\right)(n/S_n^2)^{1/2}\widehat{\lambda}(1+o_p(1)) + O_p(n^{-1/2}) \\
&= -k(B_s + o_p(1))' P n^{1/2}\widehat{g}_n(\theta_0) + o_p(1) \\
&= -(k_1^2/k_2) B_s' P n^{1/2}\widehat{g}(\theta_0) + o_p(1) \\
&\to_d (k_1^2/k_2) N(0, B_s' P\Delta P B_s).
\end{aligned}$$

□

Proof of Theorem 3.4: The proof closely follows that of Theorem 2(ii) in GSb.

Renormalize $D := D_\rho(\alpha_0)\Lambda$, where $\Lambda := diag(n^{1/2}, ..., n^{1/2}, 1, ..., 1)$ has p_{A_1} elements equal to $n^{1/2}$ and p_{A_2} elements equal to 1.

It follows immediately from the proof of Theorem 2(ii) in GSb that

$$n^{1/2} vec\widehat{G}_{A_1}(\widehat{\theta}_0) = n^{1/2} vec\widehat{G}_{A_1}(\theta_0) + o_p(1)$$

and, thus, that $vec\widehat{G}_{A_1}(\widehat{\theta}_0) = O_p(n^{-1/2})$. Then, by an analysis as in Lemma A.1 and the first part of $\mathbf{M}_{\alpha_0}\mathbf{(vi)}$

$$n^{1/2} vec\widehat{G}_{nA_1}(\widehat{\theta}_0) = n^{1/2} k_1 vec\widehat{G}_{A_1}(\theta_0) + o_p(1). \tag{A.23}$$

By $\mathbf{M}_{\alpha_0}\mathbf{(vii)}$, (A.22), and (A.23),

$$vec(D, n^{1/2}\widehat{g}_n(\widehat{\theta}_0)) = k_2 m + k_1 M v + o_p(1), \tag{A.24}$$

where $M \in \mathcal{R}^{(kp_{A_1}+kp_{A_2}+k)\times(kp_{A_1}+k)}$ and

$$M := \begin{pmatrix} -I_{kp_{A_1}} & \Delta_{A_1}\Delta^{-1} \\ 0 & 0 \\ 0 & I_k \end{pmatrix} \begin{pmatrix} I_{kp_{A_1}} & 0 \\ 0 & I_k - M_{2\beta}H \end{pmatrix},$$
$$v := n^{-1/2}\sum_{i=1}^{n} \begin{pmatrix} vecG_{iA_1}(\theta_0) \\ g_i(\theta_0) \end{pmatrix}, \ m := vec(0, -(\partial m_2/\partial\alpha_2), 0),$$

where the arguments (α_{02}, β_0) in $M_{2\beta}$ and $(\partial m_2/\partial\alpha_2)$ and θ_0 in Δ_{A_1} and Δ are suppressed. Note here that the last two conditions in $\mathbf{M}_{\alpha_0}\mathbf{(vii)}$ and analysis as in Lemma A.1 imply $\widehat{G}_{nA_2}(\widehat{\theta}_0) - k_1\widehat{G}_{A_2}(\widehat{\theta}_0) = o_p(1)$. By M_{α_0}(vi), v is asymptotically normal with full rank covariance matrix V^α and thus the asymptotic covariance matrix of $vec(D, n^{1/2}\widehat{g}_n(\widehat{\theta}_0))$ is given by $k_1^2 M V^\alpha M'$. For independence of D and $n^{1/2}\widehat{g}_n(\widehat{\theta}_0)$ It is straightforward to demonstrate that the upper right $k(p_{A_1}+p_{A_2})\times k$-submatrix of $MV^\alpha M'$ is null 0 guaranteeing the independence of D and $n^{1/2}\widehat{g}_n(\widehat{\theta}_0)$. Denote by $\overline{D}$ and $\overline{g}$ the limiting normal distributions of D and $n^{1/2}\widehat{g}_n(\widehat{\theta}_0)$, implied by (A.24). Note that $k_2\widehat{M}(\alpha_0) \to_p \Delta^{-1}M_{M_{2\beta}}(\Delta) := M^*$ for the matrix in (3.13). As in Theorem 3.2) above and GSa and GSb, $h : \mathcal{R}^{k\times p_A} \to \mathcal{R}^{p_A\times k}$, where $h(d) := (d'M^*d)^{-1/2}d'$ for $d \in \mathcal{R}^{k\times p_A}$, is continuous on a set $C \subset \mathcal{R}^{k\times p_A}$ with $\Pr(\overline{D} \in C) = 1$. By the Continuous Mapping Theorem and (A.22)

$$(D'M^*D)^{-1/2}D'\Delta^{-1}n^{1/2}\widehat{g}_n(\widehat{\theta}_0) \to_d (\overline{D}'M^*\overline{D})^{-1/2}\overline{D}'\Delta^{-1}\overline{g} \sim k_1 N(0, I_{p_A}).$$

Because $\widehat{\Delta}(\widehat{\theta}_0) \to_p k_2\Delta$ the claim follows. By a modification of (A.14), the result for $LM_\rho^{sub}(\alpha_0)$ implies the result for $S_\rho^{sub}(\alpha_0)$.

The proof for $P_\rho^{r,sub}(\alpha_0)$ is virtually identical to that for P_ρ^r. $\square$

References

Anderson, T.W. and H. Rubin (1949): "Estimators of the parameters of a single equation in a complete set of stochastic equations", *The Annals of Mathematical Statistics* 21, 570–582.

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

Andrews, D. W. K. and P. Guggenberger (2005a): "The Limit of finite-sample Size and a Problem with Subsampling," unpublished manuscript, Cowles Foundation, Yale University.

——— (2005b): "Hybrid and Size-corrected Subsample Methods," unpublished manuscript, Cowles Foundation, Yale University.

——— (2005c): "Applications of Hybrid and Size-Corrected Subsample Methods," unpublished manuscript, Cowles Foundation, Yale University.

Andrews, D.W.K. and V. Marmer (2004): "Exactly Distribution-Free Inference in Instrumental Variables Regression with Possibly Weak Instruments", unpublished manuscript.

Andrews, D.W.K., M. Moreira, and J.H. Stock (2006): "Optimal invariant similar tests for instrumental variables regression", *Econometrica* 74, 715-752.

Andrews, D.W.K., and J.H. Stock (2007): "Inference with Weak Instruments". Chapter 6 in *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society*, Volume 3, eds. R.W. Blundell, W.K. Newey and T. Persson, Econometric Society Mongraphs, ESM 43, 122-173. Cambridge University Press: Cambridge.

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

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

——— (1998): "Efficient Semiparametric Estimation of Expectations", *Econometrica* 66, 453-464.

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

Caner, M. (2003): "Exponential Tilting with Weak Instruments: Estimation and Testing", unpublished manuscript.

Chao, J.C. and N.R. Swanson (2005): "Consistent Estimation With a Large Number of Weak Instruments", *Econometrica* 73, 1673–1692.

Dufour, J. (1997): "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models", *Econometrica* 65, 1365–1387.

——— (2003): "Identification, Weak Instruments and Statistical Inference in Econometrics. Presidential Address to the Canadian Economics Association", *Canadian Journal of Economics* 36, 767–808.

Dufour, J. and M. Taamouti (2005): "Projection–based statistical inference in linear structural models with possibly weak instruments", *Econometrica* 73, 1351–1365.

Guggenberger, P. (2003): "Econometric Essays on Generalized Empirical Likelihood, Long–memory Time Series, and Volatility", Ph.D. thesis, Yale University.

Guggenberger, P. and R.J. Smith (2005): "Generalized Empirical Likelihood Estimators and Tests under Partial, Weak and Strong Identification", *Econometric Theory* 21, 667–709.

Guggenberger, P. and R.J. Smith (2007): "Generalized Empirical Likelihood Tests in Time Series Models With Potential Identification Failure". Forthcoming *Journal of Econometrics*.

Guggenberger, P. and M. Wolf (2004): "Subsampling tests of parameter hypotheses and overidentifying restrictions with possible failure of identification", unpublished manuscript.

Hamilton, J.D. (1994): "Time Series Analysis", Princeton University Press.

Hannan, E.J. (1957): "The variance of the mean of a stationary process", *Journal of the Royal Statistical Society. Series B*, 19, 282–285.

——— (1970): "Multiple Time Series", New York: Wiley.

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

Hansen, L.P., J. Heaton and A. Yaron (1996): "Finite–sample properties of some alternative GMM estimators", *Journal of Business & Economic Statistics* 14, 262–280.

Imbens, G. (1997): "One–step estimators for over–identified Generalized Method of Moments models", *Review of Economic Studies* 64, 359–383.

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

Jansson, M. (2002): "Consistent Covariance Matrix Estimation for Linear Processes", *Econometric Theory* 18, 1449–1459.

Kiefer, N. and T. Vogelsang (2002a): "Heteroskedasticity–Autocorrelation Robust Testing Using Bandwidth Equal To Sample Size", *Econometric Theory* 18, 1350–1366.

——— (2002b): "Heteroskedasticity–Autocorrelation Robust Standard Errors Using the Bartlett Kernel Without Truncation", *Econometrica*, 70, 2093–2095.

——— (2005): "A New Asymptotic Theory for Heteroskedasticity–Autocorrelation Robust Tests", *Econometric Theory* 21, 1130–1164.

Kiefer, N., T. Vogelsang, and H. Bunzel (2000): "Simple Robust Testing of Regression Hypothesis", *Econometrica* 68, 695–714.

Kitamura, Y. (1997): "Empirical likelihood methods with weakly dependent processes", *Annals of Statistics* 25, 2084–2102.

Kitamura, Y. and M. Stutzer (1997): "An information–theoretic alternative to Generalized Method of Moments estimation", *Econometrica* 65, 861–874.

Kleibergen, F. (2002): "Pivotal statistics for testing structural parameters in instrumental variables regression", *Econometrica* 70, 1781–1805.

——— (2004): "Testing subsets of structural parameters in the instrumental variables regression model", *Review of Economics and Statistics* 86, 418–423.

——— (2005): "Testing parameters in GMM without assuming that they are identified", *Econometrica* 73, 1103–1123.

——— (2006): "Subset statistics in the linear IV regression model", unpublished manuscript, Brown University.

Moreira, M.J. (2003): "A Conditional Likelihood Ratio Test for Structural Models", *Econometrica* 71, 1027–1048.

Moreira, M.J., J. Porter, and G.A. Suarez (2005a): "Bootstrap and Higher–Order Expansion Validity When Instruments May Be Weak", unpublished manuscript.

——— (2005b): "Bootstrap Validity for the Score Test When Instruments May Be Weak", unpublished manuscript.

Nelson, C.R. and R. Startz (1990) Some further results on the exact small sample properties of the instrumental variable estimator. *Econometrica* 58, 967-976.

Newey, W.K. and R.J. Smith (2004): "Higher order properties of GMM and Generalized Empirical Likelihood estimators", *Econometrica* 72, 219–255.

Newey, W.K. and K.D. West (1994): "Automatic Lag Selection in Covariance Matrix Estimation", *Review of Economic Studies* 61, 631–653.

Otsu, T. (2006): "Generalized Empirical Likelihood Inference for Nonlinear and Time Series Models under Weak Identification", *Econometric Theory* 22, 513–527.

Parzen, E. (1956): "On consistent estimates of the spectral density of a stationary time series", *Proc. Nat. Acad. Sci.,* 42, 154–157.

——— (1957): "On consistent estimates of the spectrum of a stationary time series", *The Annals of Mathematical Statistics* 28, 329–348.

Phillips, P.C.B. (1989): "Partially Identified Econometric Models", *Econometric Theory* 5, 181-240.

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

Ramalho, J.J.S., and R.J. Smith (2004): "Goodness of Fit Tests for Moment Condition Models", unpublished manuscript, University of Cambridge.

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

——— (2000): "Empirical Likelihood Estimation and Inference", Chapter 4 in *Applications of Differential Geometry to Econometrics*, eds. M. Salmon and P. Marriott, 119–150. Cambridge University Press: Cambridge.

——— (2001): "GEL Criteria for Moment Condition Models", mimeo, University of Bristol. Revised version CWP 19/04, c*e*mmap, I.F.S. and U.C.L. http://cemmap.ifs.org.uk/wps/cw

——— (2005): "Automatic Positive Semi–Definite HAC Covariance Matrix and GMM Estimation", *Econometric Theory* 21, 158–170.

Staiger, D. and J.H. Stock (1997): "Instrumental Variables Regression With Weak Instruments", *Econometrica* 65, 557–586.

Stock, J.H. and J.H. Wright (2000): "GMM with Weak Identification", *Econometrica* 68, 1055–1096.

Wooldridge, J.M. and H. White (1988): "Some invariance principles and central limit theorems for dependent heterogeneous processes", *Econometric Theory* 4, 210–230.