Tests of Additional Conditional Moment Restrictions*

Paulo M.D.C. Parente
Instituto Universitário de Lisboa (ISCTE-IUL)
Business Research Unit (BRU-IUL)

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

This draft: March 2016

Abstract

The primary focus of this article is the provision of tests for the validity of a set of conditional moment constraints additional to those defining the maintained hypothesis that are relevant for either independent cross-sectional data or short panel data with independent cross-sections contexts. The point of departure and principal contribution of the paper is the explicit and full incorporation of the conditional moment information defining the maintained hypothesis in the design of the test statistics. Thus, the approach mirrors that of the classical parametric likelihood setting by defining *restricted* tests in contradistinction to *unrestricted* tests that partially or completely fail to incorporate the maintained information in their formulation. The framework is quite general allowing the parameters defining the additional and maintained conditional moment restrictions to differ and permitting the conditioning variates to differ likewise. GMM and generalized empirical likelihood test statistics are suggested. The asymptotic properties of the statistics are described under both null hypothesis and a suitable sequence of local alternatives. An extensive set of simulation experiments explores the practical efficacy of the various test statistics in terms of empirical size and size-adjusted power confirming the superiority of restricted over unrestricted tests. A number of restricted tests possess both sufficiently satisfactory empirical size and power characteristics to allow their recommendation for econometric practice.

JEL Classification: C12, C14, C30

Key-words: GMM; Generalized Empirical Likelihood; Series Approximations; Restricted Tests; Unrestricted Tests; Local Power.

*We would like to thank an Editor, Associate Editor and referees for their helpful and constructive comments on a previous version of the paper. The authors are also grateful for comments on versions of this paper by participants at Microdata Methods and Practice: A Cemmap Celebration, the Cowles Econometrics Seminar, the CEMFI-UC3M Econometrics Seminar, and seminars at Brown University and the Universities of Cambridge, Hull and Surrey.

1 Introduction

The primary focus of this article is the provision of tests for the validity of a set of conditional moment constraints in addition to those defining the maintained hypothesis relevant for independent cross-sectional data or short panel data with independent cross-sections when a finite dimensional parameter vector is the object of inferential interest. Examples include moment conditional homoskedasticity and instrument validity.[1] The main point of departure and principal contribution of the paper is the explicit incorporation of the maintained conditional moment information in the formulation of the test statistics. Thus, our approach mirrors that of the classical parametric likelihood setting by defining *restricted* tests for these additional conditional moments in contradistinction to *unrestricted* tests that partially or completely fail to incorporate the maintained moment condition information in their design with a similar consequent advantage that the former dominate the latter tests in terms of asymptotic local power; cf. Aitchison (1962). The framework is quite general allowing the parameters defining the additional and maintained conditional moment restrictions to differ and permitting the conditioning variates to differ likewise. The paper also contributes a number of new theoretical results required to address the null and local alternative asymptotic distributions of the test statistics.

The approach taken in the paper exploits an equivalence between conditional moment constraints and a countably infinite number of unconditional restrictions noted elsewhere; see Chamberlain (1987). Test statistics are consequently defined in terms of an appropriate set of additional infinite unconditional moment conditions. These tests adapt and generalise those of Donald, Imbens and Newey (2003) which approximates conditional moments by an appropriate finite set of unconditional moments. Tests for a finite number of unconditional moment restrictions, cf. *inter alia* Newey (1985), Eichenbaum, Hansen and Singleton (1988) and Ruud (2000) for GMM, Hansen (1982), and Smith (1997, 2011) for generalized empirical likelihood (GEL), see also Kitamura and Stutzer (1997), Imbens, Spady and Johnson (1998) and Newey and Smith (2004), are well-known to be inconsistent against all alternatives implied by conditional moment conditions; see, e.g., Bierens (1990). GMM and GEL test statistics defined in Donald, Imbens and Newey (2003) circumvent this difficulty by allowing the number of unconditional moments to grow with sample size at an appropriate rate.[2] Likewise here both maintained and null hypothesis conditional moment constraints are approximated by corresponding sets of unconditional moment restrictions with the former a subset of the latter, both of whose dimensions grow with sample size at appropriate rates. Restricted GMM- and GEL-based test statistics for additional conditional moment restrictions, after location and scale standardization, are asymptotically equivalent and converge

[1]Instrument validity tests are the concern of the application in section 6 to a parametric specification of an Engel curve relationship discussed elsewhere in, e.g., Muellbauer (1976), Banks, Blundell and Lewbel (1997) and, more recently, Blundell and Horowitz (2007). See fn. 15 below.

[2]Consistent tests of goodness of fit in regression models have received substantial attention in the literature. See, e.g., Eubank and Spiegelman (1990) for the nonlinear regression context. See also *inter alia* De Jong and Bierens (1994), Hong and White (1995) and Jayasuriya (1996).

in distribution to a standard normal variate under the null hypothesis. Intuitively this result reflects the implicit infinite number of unconditional moments under test since standardised chi-square distributed statistics are asymptotically standard normally distributed when the statistic degrees of freedom diverges to infinity. A similar result obtains for unrestricted statistics that partially or completely neglect the maintained conditional moment information although the limit standard normal variate differs.[3] Interestingly, unlike finite dimensional test statistics, efficient parameter estimation is no longer required for test implementation. Under a suitable sequence of local alternatives, restricted and unrestricted test statistics are asymptotically non-central standard normally distributed. The non-centrality parameter of the restricted statistics exceeds those of unrestricted statistics thereby demonstrating the deficiency of these latter tests mirroring the results for restricted tests in the classical parametric likelihood, Aitchison (1962), and unconditional moment condition, Newey (1985), settings. The asymptotic local power results also indicate that one-sided tests of the additional conditional moment restrictions are apposite.

The paper is organized as follows. Section 2 provides some initial definitions, details the test problem and describes moment conditional homoskedasticity and instrument validity examples that are used throughout the paper. GMM and GEL restricted test statistics are then specified in section 3; an initial discussion presents the equivalence between conditional moment restrictions and an appropriately defined infinite set of unconditional moment constraints together with the assumptions that underpin the analysis in the paper. Section 4 provides the limiting distributions of these and unrestricted statistics under the null hypothesis of the additional conditional moment validity; the large sample independence of the restricted test statistics and GMM and GEL test statistics for the maintained hypothesis is shown which thus permits the overall test size of a sequential test of the maintained and then additional conditional moment restrictions to be controlled. Section 5 considers the local asymptotic behaviour of the restricted and unrestricted test statistics demonstrating the one-sided nature of the tests and the relative deficiency of the latter tests. Section 6 presents a set of simulation results on the size and power of the test statistics based on an application to a parametric specification of an Engel curve relationship. Section 7 concludes. Proofs of the results in the text and certain subsidiary lemmata are given in Appendix A.

The paper uses the generic subscript notation "m" and "a" to denote quantities associated with the maintained hypothesis and additional moment constraints. Conditional moment indicator vectors are denoted by $u(\cdot, \beta)$ of generic dimension J, with parameter vector β of dimension p and associated parameter space $\mathcal{B}$; instrument vectors are denoted as s with dimension d. The abbreviations a.s., f.r.r., n.s. and p.d. indicate "almost surely", "full row rank", "nonsingular" and "positive definite" respectively. $[\cdot]$ is the integer part of $\cdot$. Statistics are "asymptotically equivalent" if they differ by an

[3] Alternative unrestricted tests could also be based *inter alia* on the approaches of Bierens (1982, 1990), Wooldridge (1992), Yatchew (1992), Härdle and Mammen (1993), Fan and Li (1996), Zheng (1996,1998), Lavergne and Vuong (2000), Ellison and Ellison (2000) and Domínguez and Lobato (2004). The continuum of moment conditions method suggested in Carrasco and Florens (2000) offers another possible approach; see also Hsu and Kuan (2011).

$o_p(1)$ term.

2 Some Preliminaries

2.1 Definitions

The *maintained* hypothesis is defined in terms of the moment indicator vector $u_m(z, \beta_m)$ which is a J_m-vector of known functions of the d_z-vector of data observables z and the p_m-vector of parameters β_m. In many cases $u_m(z, \beta_m)$ may be interpreted as an error vector. It is assumed that there exists an observable d_m-vector of instruments s_m such that

$$E[u_m(z, \beta_{m0})|s_m] = 0 \text{ a.s. } s_m \tag{2.1}$$

for some unknown value $\beta_{m0} \in \mathcal{B}_m$ of the parameter vector β_m where $\mathcal{B}_m$ denotes the corresponding parameter space.

The central interest of the paper is the provision of tests of the additional conditional moment restrictions

$$E[u_a(z, \beta_{a0})|s_a] = 0 \text{ a.s. } s_a \tag{2.2}$$

for some $\beta_{a0} \in \mathcal{B}_a$. Here the moment indicator vector $u_a(z, \beta_a)$ denotes a J_a-vector of known functions of z and the unknown p_a-vector of parameters β_a with $\mathcal{B}_a$ the corresponding parameter space and s_a an observable d_a-vector of instruments. Together the parameter vectors β_{m0} and β_{a0} constitute the objects of inferential interest. Note that β_a may or may not be coincident with the maintained hypothesis parameter vector β_m. Likewise, the notation s_a for the instrument vector defining the additional conditional moment constraints (2.2) explicitly permits circumstances in which the maintained instruments s_m may or may not be strictly included in the additional instruments s_a or vice-versa.[4]

2.2 Test Problem

The maintained hypothesis is given by the conditional moment constraint $E[u_m(z, \beta_{m0})|s_m] = 0$ (2.1) and is assumed to hold throughout. The null hypothesis H_0 of interest is consequently defined in terms of the validity of the additional conditional moment constraints (2.2), i.e.,

$$H_0 : E[u_a(z, \beta_{a0})|s_a] = 0 \text{ a.s. } s_a \text{ and } E[u_m(z, \beta_{m0})|s_m] = 0 \text{ a.s. } s_m \tag{2.3}$$

with the corresponding alternative hypothesis H_1 given by

$$H_1 : E[u_a(z, \beta_a)|s_a] \neq 0 \text{ all } \beta_a \in \mathcal{B}_a,\ s_a \in \mathcal{S}_a, \text{ and } E[u_m(z, \beta_{m0})|s_m] = 0 \text{ a.s. } s_m \tag{2.4}$$

for some $\mathcal{S}_a$ with non-zero probability content.

[4]Nonparametric components are excluded from the moment indicator vector definitions. The theoretical analysis of the paper could in principle be extended to deal with such models; see, e.g., Chen and Pouzo (2009, 2012).

2.3 Examples

Example 2.1 (Conditional Homoskedasticity): This example concerns the conditional homoskedasticity of the maintained conditional moment indicator vector $u_m(z, \beta_m)$; hence the maintained hypothesis and additional instrument vectors are identical, i.e., $s_m = s_a$. The additional conditional moment indicator is defined by

$$u_a(z, \beta_a) = vech(u_m(z, \beta_m)u_m(z, \beta_m)' - \Sigma)$$

where $vech(\cdot)$ denotes the vectorised upper triangle of $\cdot$. Thus $J_a = J_m(J_m+1)/2$ and $\beta_a = (\beta_m', vech(\Sigma)')'$ includes the maintained parameter vector β_m. Let $\Sigma_0(s_m) = E[u_m(z, \beta_{m0})u_m(z, \beta_{m0})'|s_m]$ and $\Sigma_0 = E[u_m(z, \beta_{m0})u_m(z, \beta_{m0})']$. Therefore the null hypothesis may be expressed as

$$H_0 : \Sigma_0(s_m) = \Sigma_0 \text{ and } E[u_m(z, \beta_{m0})|s_m] = 0 \text{ a.s. } s_m,$$

with alternative hypothesis $H_1 : \Sigma_0(s_m) \neq \Sigma$ all p.d. Σ, $s_m \in \mathcal{S}_m$, where $\mathcal{S}_m$ has non-zero probability mass, and $E[u_m(z, \beta_{m0})|s_m] = 0$ a.s. s_m.

Remark 2.1: The standard instrumental variable (IV) linear regression model defines $u_m(z, \beta_m) = y - \beta_m x$, with $J_m = 1$ and thus $J_a = 1$. With maintained unconditional moment indicator vector $s_m u_m(z, \beta_m) = s_m(y - \beta_m x)$, continuous updating estimation (CUE) of β_m, Hansen, Heaton and Yaron (1996), uses the inverse of the sample moment matrix $\sum_{i=1}^n s_{mi}s_{mi}'(y_i - \beta_m x_i)^2/n$ as metric whereas, under conditional homoskedasticity, the LIML metric, i.e., the inverse of $\sigma_n^2(\beta_m)\sum_{i=1}^n s_{mi}s_{mi}'/n$, where $\sigma_n^2(\beta_m) = \sum_{i=1}^n (y_i - \beta_m x_i)^2/n$, is apposite.

Example 2.2 (Instrument Validity): In this example both maintained and additional conditional moment indicators coincide, i.e., $u_m(z, \beta_m) = u_a(z, \beta_a)$ with $\beta_m = \beta_a$ and, thus, $J_a = J_m$. The issue here is the validity of the additional instrument vector s_a. The null hypothesis is therefore defined by

$$H_0 : E[u_m(z, \beta_{m0})|s_a] = 0 \text{ all } s_a, E[u_m(z, \beta_{m0})|s_m] = 0 \text{ all } s_m,$$

with alternative hypothesis $H_1 : E[u_m(z, \beta_m)|s_a] \neq 0$ all $\beta_m \in \mathcal{B}_m$, $s_a \in \mathcal{S}_a$, where $\mathcal{S}_a$ has non-zero probability content, and $E[u_m(z, \beta_0)|s_m] = 0$ all s_m.

Remark 2.2: Blundell and Horowitz (2007) defines a form of exogeneity hypothesis for nonparametric regression in which s_a coincides with the covariate vector x but does not include the maintained instrument vector s_m. The structural regression function may therefore be consistently estimated by nonparametric least squares (LS) thus avoiding the difficulties associated with nonparametric IV estimation. Since x but not s_m is included in s_a, this hypothesis might be regarded as a *marginal* form of exogeneity hypothesis (ME). Alternatively the inclusion of both s_m and x in s_a constitutes a *joint* form of exogeneity (JE) and might be of interest if the regression function included both x and elements of s_m.[5] Note that JE is more stringent than ME. For linear regression, see Remark 2.1 above, if $E[y - \beta_{m0}x|s_a] = 0$, i.e., $E[y|s_a] = \beta_{m0}x$, LS estimation of β_{m0} is consistent but inefficient in the presence of conditional heteroskedasticity, Cragg (1983), with IV estimation incorporating the additional $E[y - \beta_{m0}x|s_a] = 0$ and maintained $E[y - \beta_{m0}x|s_m] = 0$ conditional moments efficient.

3 GMM and GEL Test Statistics

3.1 Approximating Conditional Moment Restrictions

Conditional moment constraints of the form (2.1) and (2.2) are equivalent to a countable number of unconditional moment restrictions under certain regularity conditions; see Chamberlain (1987). The following assumption, Assumption 1, p.58, of Donald, Imbens and Newey (2003), henceforth DIN, provides precise conditions. The discussion is initially framed for a generic vector of instruments s and moment indicator vector $u(z, \beta)$.

For each positive integer K, let $q^K(s) = (q_{1K}(s), ..., q_{KK}(s))'$ denote a K-vector of approximating functions.

Assumption 3.1 *For all K, $E[q^K(s)'q^K(s)]$ is finite and for any $a(s)$ with $E[a(s)^2] < \infty$ there are K-vectors γ_K such that as $K \to \infty$,*

$$E[(a(s) - q^K(s)'\gamma_K)^2] \to 0.$$

Possible approximating functions which satisfy Assumption 3.1 are splines, power series and Fourier series. See *inter alia* DIN, Newey (1997) and Powell (1981) for further discussion.

The next result, DIN Lemma 2.1, p.58, establishes a formal equivalence between conditional moment restrictions of the type (2.1) and (2.2) and a sequence of unconditional moment restrictions.

Lemma 3.1 *Suppose that Assumption 3.1 is satisfied and $E[u(z, \beta_0)'u(z, \beta_0)]$ is finite. If $E[u(z, \beta_0)|s] = 0$, then $E[u(z, \beta_0) \otimes q^K(s)] = 0$ for all K. Furthermore, if $E[u(z, \beta_0)|s] \neq 0$, then $E[u(z, \beta_0) \otimes q^K(s)] \neq 0$ for all K large enough.*

DIN defines the unconditional moment indicator vector as $u(z, \beta) \otimes q^K(s)$. By considering the moment conditions $E[u(z, \beta_0) \otimes q^K(s)] = 0$, if K approaches infinity at an appropriate rate, dependent on the sample size n and the estimation method, EL, IV, GMM or GEL, DIN demonstrates that under certain conditions these estimators are consistent and achieve the semi-parametric efficiency lower bound. To do so, however, requires the imposition of a normalization condition on the approximating functions, DIN Assumption 2, p.59, which now follows. Let $\mathcal{S}$ denote the support of the random vector s.

[5] JE has been a central concern in the literature on classical likelihood-based tests for (weak) exogeneity. See *inter alia* Durbin (1954), Wu (1973), Hausman (1978), Engle (1982), Engle et al. (1983) and Smith (1994).

Assumption 3.2 *For each K there is a constant scalar $\zeta(K)$ and matrix B_K such that $\tilde{q}^K(s) = B_K q^K(s)$ for all $s \in \mathcal{S}$, $\sup_{s\in\mathcal{S}} \left\|\tilde{q}^K(s)\right\| \leq \zeta(K)$, $E[\tilde{q}^K(s)\tilde{q}^K(s)']$ has smallest eigenvalue bounded away from zero uniformly in K and $\sqrt{K} \leq \zeta(K)$.*

Hence to formulate a test statistic appropriate for the null hypothesis (2.3) requires that its constituent conditional moment constraints, $E[u_m(z,\beta_{m0})|s_m] = 0$ (2.1) and $E[u_a(z,\beta_{a0})|s_a] = 0$ (2.2), are re-interpreted as suitably defined sequences of unconditional moment restrictions based on Assumptions 3.1 and 3.2. The maintained conditional moment restrictions (2.1) are re-expressed as the sequence of $J_m K$ unconditional moment restrictions

$$E[u_m(z,\beta_{m0}) \otimes q_m^K(s_m)] = 0, K \to \infty, \tag{3.1}$$

for approximating functions $q_m^K(s_m)$ satisfying Assumptions 3.1 and 3.2. Likewise let $q_a^{MK}(s_a)$ be a MK-vector of approximating functions that depends on s_a and that also satisfies Assumptions 3.1 and 3.2, where for ease of exposition M is a positive integer. Thus the additional conditional moment restrictions (2.2) are rewritten as the sequence of $J_a MK$ unconditional moment restrictions

$$E[u_a(z,\beta_{a0}) \otimes q_a^{MK}(s_a)] = 0, K \to \infty. \tag{3.2}$$

The null hypothesis (2.3) is then formally equivalent to the sequence of $(J_m + J_a M)K$ unconditional moments[6]

$$E[u_m(z,\beta_{m0}) \otimes q_m^K(s_m)] = 0, E[u_a(z,\beta_{a0}) \otimes q_a^{MK}(s_a)] = 0, K \to \infty. \tag{3.3}$$

Remark 3.1: Strictly speaking, the succeeding theoretical analysis requires the dimension of $q_a^{MK}(\cdot)$, the integer $d_{q_a}(K)$ say, should satisfy $\lim_{K\to\infty} d_{q_a}(K)/K = M$, M a positive constant, e.g., $d_{q_a}(K) = [MK]$, i.e., the same order as that of $q_m^K(\cdot)$. The multiplicative choice MK with M a positive integer is adopted for simplicity and for ease of implementation and exposition. Restricted test statistics for (2.3) defined in section 3.3 below are expressed as (or are asymptotically equivalent to) the difference of an unrestricted statistic and a statistic apposite for testing the maintained conditional moment restrictions (2.1); see section 4. Their respective large sample behaviours are determined by the relative number of approximating functions used to express the null and maintained hypotheses in unconditional form. If the dimension of $q_a^{MK}(\cdot)$ diverges at a rate different from that of $q_m^K(\cdot)$, the limit theory used in sections 4

[6] To illustrate the construction of $q_m^K(s_m)$ and $q_a^{MK}(s_a)$ for polynomial approximating functions suppose s_m and s_a have d_{am} elements in common. Let the approximating functions vector $q_m^K(s_m)$ for the maintained conditional moment restrictions (2.1) be a polynomial of order $k_m - 1$ which yields $K = k_m^{d_m}$. Thus k_m could be chosen as $[K^{1/d_m}] + 1$ for given K. Similarly let the components of the vector of approximating functions $q_a^{MK}(s_a)$ for the additional conditional moment restrictions (2.2) corresponding to the d_{am} elements in common between s_m and s_a be formed from a polynomial of order $k_a - 1$. Also suppose a polynomial of order k_a excluding the constant term is used for those components corresponding to the $d_a - d_{am}$ unique elements in s_a. Then the dimension of the vector of approximating functions $q_a^{MK}(s_a)$ is $k_a^{d_{am}}((k_a+1)^{d_a - d_{am}} - 1)$. Therefore the order of the dimension of $q_a^{MK}(s_a)$ is $k_a^{d_a}$. Examples: (a) ME: $d_{am} = 0$; thus $MK = (k_a+1)^{d_a} - 1$, e.g., $d_a = 1$, $MK = k_a$. (b) JE: $d_{am} = d_m$; thus $MK = k_a^{d_m}((k_a+1)^{d_a - d_m} - 1)$, e.g., $d_m = 1$, $d_a = 2$, $MK = k_a^2$. For the general case this suggests choosing $k_a = [(MK)^{1/d_a}] + 1$.

and 5 to establish the asymptotic behaviour of the unrestricted statistic under null and local alternative hypotheses no longer applies.

EXAMPLE 2.1 (CONDITIONAL HOMOSKEDASTICITY CONT.): Recall that $u_a(z,\beta_a) = vech(u_m(z,\beta_m)u_m(z,\beta_m)' - \Sigma)$ with $\beta_a = (\beta_m', vech(\Sigma)')'$. In this case $s_a = s_m$ and thus the additional approximating functions are defined as $q_a^{MK}(s_a) = q_m^K(s_m)$. Therefore $M = 1$. Hence, the null hypothesis $H_0 : \Sigma_0(s_m) = \Sigma_0$, $E[u_m(z,\beta_{m0})|s_m] = 0$ is re-expressed in unconditional form as

$$E[u_a(z,\beta_{a0}) \otimes q_m^K(s_m)] = 0, E[u_m(z,\beta_{m0}) \otimes q_m^K(s_m)] = 0, K \to \infty.$$

EXAMPLE 2.2 (INSTRUMENT VALIDITY CONT.): Recall that $u_a(z,\beta_a) = u_m(z,\beta_m)$ with $J_m = J_a$ and $\beta_a = \beta_m$. The vector of additional approximating functions is $q_a^{MK}(s_a)$ with dimension MK. Thus, the null hypothesis $H_0 : E[u_m(z,\beta_{m0})|s_a] = 0$, $E[u_m(z,\beta_{m0})|s_m] = 0$ is re-expressed in unconditional form as

$$E[u_m(z,\beta_{m0}) \otimes q_a^{MK}(s_a)] = 0, E[u_m(z,\beta_{m0}) \otimes q_m^K(s_m)] = 0, K \to \infty.$$

REMARK 3.2: For regression the special cases ME $s_a = x$ with $q_a^{MK}(s_a)$ functions of x only and JE $s_a = (s_m, x)$ with $q_a^{MK}(s_a)$ additional functions of s_m and x are of particular interest.

3.2 Basic Assumptions and Notation

Let β denote the distinct elements of β_m and β_a with β_0 and the composite parameter space $\mathcal{B}$ defined similarly with p the number of parameters comprising β. The vector s collects the distinct elements of the maintained and additional instrument vectors s_m and s_a. Also let $u(z,\beta)$ and $q^K(s)$ denote the non-redundant elements of $u_m(z,\beta_m)$ and $u_a(z,\beta_a)$ and $q_m^K(s_m)$ and $q_a^{MK}(s_a)$ respectively. It will be helpful to define a number of f.r.r. selection matrices S_m^u, S_a^u and S_m^q, S_a^q; *viz.*, $S_m^u u(z,\beta) = u_m(z,\beta_m)$, $S_a^u u(z,\beta) = u_a(z,\beta_a)$ and $S_m^q q^K(s) = q_m^K(s_m)$, $S_a^q q^K(s) = q_a^{MK}(s_a)$.[7] Correspondingly $S_m = S_m^u \otimes S_m^q$ and $S_a = S_a^u \otimes S_a^q$ are both f.r.r. selection matrices. Importantly for the theoretical analysis underpinning the results in the paper, the unconditional forms of moment indicator vectors corresponding to the maintained and null hypotheses, cf. (3.1) and (3.3), may be expressed as $S_m(u(z,\beta) \otimes q^K(s))$ and $S(u(z,\beta) \otimes q^K(s))$ respectively where $S = (S_m', S_a')'$. Necessarily S is n.s. otherwise either $u(z,\beta)$ or $q^K(s)$ would contain redundant elements.

EXAMPLE 2.1 (CONDITIONAL HOMOSKEDASTICITY CONT.): Here $u(z,\beta) = (u_m(z,\beta_m)', u_a(z,\beta_a)')'$ and $q^K(s) = q_m^K(s_m)$. Hence $S_m^q = S_a^q = I_K$ and $S_m^u = (I_{J_m}, 0_{(J_m \times J_a)})$, $S_a^u = (0_{(J_a \times J_m)}, I_{J_a})$. The unconditional form of the moment indicator vector corresponding to the null hypothesis $H_0 : \Sigma_0(s_m) =$

[7] The row and column dimensions of the selection matrices S_m^q and S_a^q depend on K but to avoid a burdensome notation this dependence is not made explicit.

Σ_0, $E[u(z,\beta_0)|s_m]=0$ is then

$$S(u(z,\beta)\otimes q^K(s))=\left(\begin{array}{c} u_m(z,\beta_m) \\ u_a(z,\beta_a) \end{array}\right)\otimes q_m^K(s_m), K\rightarrow\infty,$$

with that for the maintained hypothesis expressed as $S_m(u(z,\beta)\otimes q^K(s))=u_m(z,\beta_m)\otimes q_m^K(s_m)$, $K\rightarrow\infty$.

EXAMPLE 2.2 (INSTRUMENT VALIDITY CONT.): Here $u(z,\beta)=u_a(z,\beta_a)=u_m(z,\beta_m)$ with $J_m=J_a$ and $\beta=\beta_a=\beta_m$. Thus $S_m^u=S_a^u=I_{J_m}$ and $S_m^q=(I_K,0_{(K\times MK)})$, $S_a^q=(0_{(MK\times K)},I_{MK})$. The unconditional moment indicator vector $u_m(z,\beta_m)\otimes(q_m^K(s_m)',q_a^{MK}(s_a)')'$ corresponding to the null hypothesis $H_0:E[u_m(z,\beta_{m0})|s_a]=0$, $E[u_m(z,\beta_{m0})|s_m]=0$ may equivalently be re-arranged as

$$S(u(z,\beta)\otimes q^K(s))=\left(\begin{array}{c} u_m(z,\beta_m)\otimes q_m^K(s_m) \\ u_m(z,\beta_m)\otimes q_a^{MK}(s_a) \end{array}\right), K\rightarrow\infty,$$

with that for the maintained hypothesis given by $S_m(u(z,\beta)\otimes q^K(s))=u_m(z,\beta_m)\otimes q_m^K(s_m)$, $K\rightarrow\infty$, as above.

Standard conditions are imposed to derive the limiting distributions of the test statistics discussed below; *viz.*

Assumption 3.3 *(a) The data are i.i.d.; (b) there exists* $\beta_0\in int(\mathcal{B})$ *such that* $E[u_m(z,\beta_{m0})|s_m]=0$ *and* $E[u_a(z,\beta_{a0})|s_a]=0$*; (c)* $\sqrt{n}(\hat{\beta}-\beta_0)=O_p(1)$*; (d)* $E[\sup_{\beta\in\mathcal{B}}\|u(z,\beta)\|^2|s]$ *is bounded.*

Unlike DIN Assumption 6(b), p.67, it is unnecessary to impose $E[\sup_{\beta\in\mathcal{B}}\|u(z,\beta)\|^{\gamma}]<\infty$ for some $\gamma>2$ for GEL; see Guggenberger and Smith (2005).[8]

REMARK 3.3: Assumption 3.3(a) requires only a root-n consistent rather than an efficient estimator $\hat{\beta}$ of β_0. Global identification of β_0 and thus root-n consistency of GMM and GEL are not necessarily guaranteed if based on an arbitrary finite set of unconditional moments derived from the conditional moment restrictions; see, e.g., Domínguez and Lobato (2004) and Hsu and Kuan (2011). If $\beta_0\in\mathcal{B}$ uniquely satisfies $E[u(z,\beta)|s]=0$ a.s., $\beta\in\mathcal{B}$, Lemma 3.1 guarantees global identification of β_0 for sufficiently large K and root-n consistency of GMM and GEL follows with the imposition of the additional assumptions described in section 5, pp.64-67, of DIN if Assumptions 3.1 and 3.2 on the vector of approximating functions $q^K(s)$ are satisfied. See also Kitamura, Tripathi and Ahn (2004). Domínguez and Lobato (2004) and Hsu and Kuan (2011) also propose root-n consistent GMM-type methods based on particular classes of unconditional moment constraints which do not require global identification.

[8]Lemma A.1 in Appendix A may be substituted for Lemma A.10 in DIN rendering $\gamma=2$ sufficient for the succeeding DIN Lemmas and Theorems concerned with GEL.

Define $u_\beta(z,\beta) = \partial u(z,\beta)/\partial\beta'$, $D(s) = E[u_\beta(z,\beta)|s]$ and $u_{\beta\beta j}(z,\beta) = \partial^2 u_j(z,\beta)/\partial\beta\partial\beta'$, $j = 1,...,J$, where J denotes the dimension of $u(z,\beta)$.[9] Also let $\mathcal{N}$ denote a neighbourhood of β_0.

Assumption 3.4 *(a) $u(z,\beta)$ is twice continuously differentiable in $\mathcal{N}$, $E[\sup_{\beta\in\mathcal{N}} \|u_\beta(z,\beta)\|^2 |s]$ and $E[\|u_{\beta\beta j}(z,\beta_0)\|^2 |s]$, $(j = 1,...,J)$, are bounded;* **(b)** *$\Sigma(s) = E[u(z,\beta_0)u(z,\beta_0)'|s]$ has smallest eigenvalue bounded away from zero*; *(c) $E[\sup_{\beta\in\mathcal{N}} \|u(z,\beta)\|^4 |s]$ is bounded; (d) for all $\beta\in\mathcal{N}$, $\|u(z,\beta) - u(z,\beta_0)\| \leq \delta(z)\|\beta-\beta_0\|$ and $E[\delta(z)^2|s]$ is bounded; (e) $E[D(s)'D(s)]$ is nonsingular.*

3.3 Test Statistics

Let $g_{mi}(\beta_m) = S_m(u(z_i,\beta)\otimes q^K(s_i)) = u_m(z_i,\beta_m)\otimes q^K_m(s_{mi})$, $g_{ai}(\beta_a) = S_a(u(z_i,\beta)\otimes q^K(s_i)) = u_a(z_i,\beta)\otimes q^{MK}_a(s_{ai})$ and $g_i(\beta) = S(u(z_i,\beta)\otimes q^K(s_i))$, $(i = 1,...,n)$. Write $\hat{g}_m(\beta_m) = \sum_{i=1}^n g_{mi}(\beta_m)/n$ and $\hat{g}(\beta) = \sum_{i=1}^n g_i(\beta)/n$.

GMM statistics appropriate for tests of maintained and null hypotheses expressed unconditionally in (3.1) and (3.3) take the standard forms

$$\mathcal{T}^{g_m}_{GMM} = n\hat{g}_m(\hat{\beta}_m)'\hat{\Omega}_m^{-1}\hat{g}_m(\hat{\beta}_m) \tag{3.4}$$

and

$$\mathcal{T}^{g}_{GMM} = n\hat{g}(\hat{\beta})'\hat{\Omega}^{-1}\hat{g}(\hat{\beta}) \tag{3.5}$$

where $\hat{\beta}_m$ denotes the subvector of $\hat{\beta}$ corresponding to β_m, $\hat{\Omega}_m = \sum_{i=1}^n g_{mi}(\hat{\beta}_m)g_{mi}(\hat{\beta}_m)'/n$ and $\hat{\Omega} = \sum_{i=1}^n g_i(\hat{\beta})g_i(\hat{\beta})'/n$. Cf., for example, DIN, section 4, pp.63-64.

In the remainder of the paper tests that fully incorporate the information contained in the maintained hypothesis (2.1), or (3.1), in their formulation are referred to as *restricted* tests whereas those that partially or completely fail to do so are termed *unrestricted* tests.

A restricted GMM statistic appropriate for testing the null hypothesis (2.3) against the maintained hypothesis (2.4) may be based on the difference of GMM criterion function statistics (3.5) and (3.4) for the respective revised hypotheses (3.3) and (3.1); *viz.*

$$\mathcal{J}^r = \frac{\mathcal{T}^{g}_{GMM} - \mathcal{T}^{g_m}_{GMM} - (J_a MK - (p-p_m))}{\sqrt{2(J_a MK - (p-p_m))}}, \tag{3.6}$$

where $p - p_m$ is the number of additional parameters in β_a defining the additional conditional moment conditions (2.2) as compared with the maintained hypothesis (2.1) parameters β_m.

Remark 3.4: For fixed and finite K, under suitable conditions, GMM, Newey (1985), and GEL, Smith (2011), test statistics for the validity of additional moment restrictions, e.g., $\mathcal{T}^{g}_{GMM} - \mathcal{T}^{g_m}_{GMM}$,

[9] Nonsmooth moment indicators could be accommodated by appropriately modifying the theoretical analysis. See, e.g., Chen and Pouzo (2009, 2012) and Parente and Smith (2011).

are asymptotically chi-square distributed with $J_aMK - (p - p_m)$ degrees of freedom. The mean location $J_aMK - (p - p_m)$ and standard deviation scale $\sqrt{2(J_aMK - (p - p_m))}$ standardisations of $\mathcal{T}^{g}_{GMM} - \mathcal{T}^{g_m}_{GMM}$ in $\mathcal{J}^r$ (3.6) mimic those introduced to render chi-square random variates with large degrees of freedom approximately standard normally distributed.

A number of alternative test statistics to GMM-based procedures for a finite number of additional moment restrictions using GEL, Newey and Smith (2004) and Smith (1997, 2011), may be adapted for the framework considered here. As in DIN and Newey and Smith (2004) let $\rho(v)$ denote a function of a scalar v that is concave on its domain, an open interval $\mathcal{V}$ containing zero. Define the respective GEL criteria under null and alternative hypotheses as

$$\begin{aligned} \hat{P}^{g}_{\rho}(\beta, \lambda) &= \sum\nolimits_{i=1}^{n} [\rho(\lambda' g_i(\beta)) - \rho_0]/n, \\ \hat{P}^{g_m}_{\rho}(\beta_m, \lambda_m) &= \sum\nolimits_{i=1}^{n} [\rho(\lambda'_m g_{mi}(\beta_m)) - \rho_0]/n, \end{aligned}$$

where λ and $\lambda_m = S_m\lambda$ are the corresponding $(J_m + J_aM)K$- and J_mK-vectors of Lagrange multipliers associated with the unconditional moment constraints (3.1) and (3.3). Let $\rho_j(v) = \partial^j\rho(v)/\partial v^j$ and $\rho_j = \rho_j(0)$, $(j = 0, 1, 2, ...)$ where, without loss of generality, the normalisation $\rho_1 = \rho_2 = -1$ is imposed.[10]

Let $\hat{\Lambda}^{g_m}_n(\beta_m) = \{\lambda_m : \lambda'_m g_{mi}(\beta_m) \in \mathcal{V},\ i = 1, ..., n\}$ and $\hat{\Lambda}^{g}_n(\beta) = \{\lambda : \lambda' g_i(\beta) \in \mathcal{V},\ i = 1, ..., n\}$. Given β, the respective Lagrange multiplier estimators for λ_m and λ are defined by

$$\hat{\lambda}_m(\beta_m) = \arg\max_{\lambda_m \in \hat{\Lambda}^{g_m}_n(\beta_m)} \hat{P}^{g_m}_{\rho}(\beta_m, \lambda_m), \hat{\lambda}(\beta) = \arg\max_{\lambda \in \hat{\Lambda}^{g}_n(\beta)} \hat{P}^{g}_{\rho}(\beta, \lambda).$$

The corresponding respective Lagrange multiplier estimators for λ_m and λ are then defined as $\hat{\lambda}_m = \hat{\lambda}_m(\hat{\beta}_m)$ and $\hat{\lambda} = \hat{\lambda}(\hat{\beta})$, cf. Assumption 3.3(c),

Similarly to the restricted GMM statistic $\mathcal{J}^r$ (3.6), a restricted form of GEL likelihood ratio (LR) statistic for testing the null hypothesis (2.3) against the maintained hypothesis (2.4) may be based on the difference of GEL criterion function (3.7) statistics; *viz.*

$$\mathcal{LR}^r = \frac{2n(\hat{P}^{g}_{\rho}(\hat{\beta}, \hat{\lambda}) - \hat{P}^{g_m}_{\rho}(\hat{\beta}_m, \hat{\lambda}_m)) - (J_aMK - (p - p_m))}{\sqrt{2(J_aMK - (p - p_m))}}. \tag{3.7}$$

[10]EL is GEL with $\rho(v) = \log(1 - v)$, Imbens (1997), Qin and Lawless (1994) and Smith (2000). ET is also GEL with $\rho(v) = -\exp(v)$, Imbens, Spady and Johnson (1998), Kitamura and Stutzer (1997), as is CUE if $\rho(\cdot)$ is quadratic, Hansen, Heaton and Yaron (1996); see Theorem 2.1, p.223, of Newey and Smith (2004). More generally, members of the Cressie-Read (1984) power divergence family of discrepancies discussed by Imbens et al. are GEL with $\rho(v) = -(1+\gamma v)^{(\gamma+1)/\gamma}/(\gamma+1)$; see Newey and Smith (2004), Section 2.1, pp.223-224.

Restricted Lagrange multiplier, score and Wald-type statistics are defined respectively as[11]

$$\mathcal{LM}^r = \frac{n(\hat{\lambda} - S_m'\hat{\lambda}_m)'\hat{\Omega}(\hat{\lambda} - S_m'\hat{\lambda}_m) - (J_a MK - (p - p_m))}{\sqrt{2(J_a MK - (p - p_m))}}, \tag{3.8}$$

$$\mathcal{S}^r = \frac{\sum_{i=1}^n \rho_1(\hat{\lambda}_m' g_{mi}(\hat{\beta}_m))g_{ai}(\hat{\beta}_a)'S_a\hat{\Omega}^{-1}S_a' \sum_{i=1}^n \rho_1(\hat{\lambda}_m' g_{mi}(\hat{\beta}_m))g_{ai}(\hat{\beta}_a)/n - (J_a MK - (p - p_m))}{\sqrt{2(J_a MK - (p - p_m))}} \tag{3.9}$$

and

$$\mathcal{W}^r = \frac{n\hat{\lambda}'S_a'(S_a\hat{\Omega}^{-1}S_a')^{-1}S_a\hat{\lambda} - (J_a MK - (p - p_m))}{\sqrt{2(J_a MK - (p - p_m))}}. \tag{3.10}$$

An additional assumption on the GEL function $\rho(\cdot)$ is required for statistics based on GEL as in DIN, Assumption 6, p.67.

Assumption 3.5 $\rho(\cdot)$ *is a twice continuously differentiable concave function with Lipschitz second derivative in a neighborhood of* 0.

4 Asymptotic Null Distribution

The following theorem provides a statement of the limiting distribution of the restricted GMM statistic $\mathcal{J}^r$ (3.6) under the null hypothesis H_0 (2.3).

Theorem 4.1 *If Assumptions 3.1-3.4 hold and if* $K \to \infty$ *and* $\zeta(K)^2 K^2/n \to 0$*, then* $\mathcal{J}^r \xrightarrow{d} N(0,1)$.

The next result details the limiting properties of the restricted GEL-based statistics for the null hypothesis (2.3) and their relationship to that of the GMM statistic $\mathcal{J}^r$ (3.6).

Theorem 4.2 *Let Assumptions 3.1-3.5 hold and suppose in addition* $K \to \infty$ *and* $\zeta(K)^2 K^3/n \to 0$*. Then* $\mathcal{LR}^r$*,* $\mathcal{LM}^r$*,* $\mathcal{S}^r$ *and* $\mathcal{W}^r$ *converge in distribution to a standard normal random variate. Moreover all of these statistics are asymptotically equivalent to* $\mathcal{J}^r$*.*

Remark 4.1: The large sample analysis in section 5 of the local alternative behaviour of restricted and unrestricted statistics discussed below indicates that one-sided tests of the null hypothesis H_0 (2.3) are appropriate. E.g., the critical region $\{\mathcal{J}^r \geq z_\alpha\}$ for the *standardised* GMM statistic $\mathcal{J}^r$ (3.6) has asymptotic size α where $\mathcal{P}\{N(0,1) \geq z_\alpha\} = \alpha$. Alternatively, valid critical regions

[11] Alternative restricted score and Wald statistics robust to estimation effects may be defined; *viz.*

$$\bar{\mathcal{S}}^r = \frac{\sum_{i=1}^n \rho_1(\hat{\lambda}_m' g_{mi}(\hat{\beta}_m))g_i(\hat{\beta})'(\hat{\Omega}^{-1} - \hat{\Omega}^{-1}\hat{G}(\hat{G}'\hat{\Omega}^{-1}\hat{G})^{-1}\hat{G}'\hat{\Omega}^{-1}) \sum_{i=1}^n \rho_1(\hat{\lambda}_m' g_{mi}(\hat{\beta}_m))g_i(\hat{\beta})/n - (J_a MK - (p - p_m))}{\sqrt{2(J_a MK - (p - p_m))}}$$

$$\bar{\mathcal{W}}^r = \frac{n\hat{\lambda}_a'(S_a(\hat{\Omega}^{-1} - \hat{\Omega}^{-1}\hat{G}(\hat{G}'\hat{\Omega}^{-1}\hat{G})^{-1}\hat{G}'\hat{\Omega}^{-1})S_a')^{-1}\hat{\lambda}_a - (J_a MK - (p - p_m))}{\sqrt{2(J_a MK - (p - p_m))}}.$$

See Smith (1997, section II.2, pp.511-514) and Smith (2011, section 5, pp.1209-1213).

based on *non-standardised* statistics may also be defined. E.g., for $\mathcal{T}^{g}_{GMM} - \mathcal{T}^{g_m}_{GMM}$, the critical region $\{\mathcal{T}^{g}_{GMM} - \mathcal{T}^{g_m}_{GMM} \geq \chi^2_{J_a MK-(p-p_m)}(\alpha)\}$ where $\chi^2_k(\alpha)$ is the α-level critical value of the chi-square distribution with k degrees of freedom.[12] Note that $p - p_m$ is negligible in the large K, large n asymptotic analysis of Theorems 4.1 and 4.2.

Unrestricted statistics fail to take into account some or all of the information contained in the maintained hypothesis (2.1) in their formulation. The standard forms of unrestricted GEL-based statistic, cf. Aitchison (1962), do not incorporate the component of the restricted statistic corresponding to the maintained hypothesis (2.1), cf. $\mathcal{LR}^r$ (3.7), $\mathcal{LM}^r$ (3.8) and $\mathcal{S}^r$ (3.9); i.e.,

$$\mathcal{LR}^u = \frac{2n\hat{P}^g_n(\hat{\beta},\hat{\lambda}) - ((J_a MK + J_m K) - p)}{\sqrt{2((J_a MK + J_m K) - p)}}, \tag{4.1}$$

$$\mathcal{LM}^u = \frac{n\hat{\lambda}'\hat{\Omega}\hat{\lambda} - ((J_a MK + J_m K) - p)}{\sqrt{2((J_a MK + J_m K) - p)}} \tag{4.2}$$

with the score form based on $\mathcal{T}^{g}_{GMM}$ (3.5)

$$\mathcal{S}^u = \frac{n\hat{g}(\hat{\beta})'\hat{\Omega}^{-1}\hat{g}(\hat{\beta}) - ((J_a MK + J_m K) - p)}{\sqrt{2((J_a MK + J_m K) - p)}}. \tag{4.3}$$

By a similar analysis to that used to establish Theorems 4.1 and 4.2 the statistics $\mathcal{LR}^u$, $\mathcal{LM}^u$ and $\mathcal{S}^u$ converge in distribution to a standard normal random variate and are mutually asymptotically equivalent but not to the restricted statistics above.[13]

Remark 4.2: Other forms of unrestricted statistics may also be defined that incorporate the maintained information (2.1) to a lesser extent than restricted statistics, e.g., a GMM statistic solely based on the additional conditional moment restrictions (2.2); *viz.*

$$\mathcal{J}^a = \frac{\mathcal{T}^{g_a}_{GMM} - (J_a MK - p_a)}{\sqrt{2(J_a MK - p_a)}}, \tag{4.4}$$

[12] To see this let the statistic $S_n(k)$ be such that $S_n(k) \xrightarrow{d} \chi^2_{d(k)}$, $n \to \infty$, for fixed k where $d(k)$ is the associated degrees of freedom. Define

$$Z_n(k) = \frac{S_n(k) - d(k)}{\sqrt{2d(k)}} \text{ and } z_k(\alpha) = \frac{\chi^2_{d(k)}(\alpha) - d(k)}{\sqrt{2d(k)}}.$$

Assume that there exists a sequence $k_n \to \infty$ such that $Z_n(k_n) \xrightarrow{d} N(0,1)$, $n \to \infty$. Consider the critical region $\{S_n(k_n) \geq \chi^2_{d(k_n)}(\alpha)\}$. Since $\lim_{n\to\infty} \mathcal{P}_n\{Z_n(k_n) \geq z_\alpha\} = \alpha$,

$$\begin{aligned}\lim_{n\to\infty} \mathcal{P}_n\{S_n(k_n) &\geq \chi^2_{d(k_n)}(\alpha)\} = \lim_{n\to\infty} \mathcal{P}_n\{Z_n(k_n) \geq z_{k_n}(\alpha)\} \\ &= \lim_{n\to\infty} \mathcal{P}_n\{Z_n(k_n) \geq z_\alpha\} = \alpha.\end{aligned}$$

The second equality follows from $Z_n(k_n) \xrightarrow{d} N(0,1)$, the absolute continuity of the $N(0,1)$ distribution function and $\lim_{n\to\infty} z_{k_n}(\alpha) = z_\alpha$.

[13] These unrestricted statistics are apposite for a *joint* test of the additional (2.2) *and* maintained (2.1) conditional moment restrictions. The statistics $\mathcal{LR}^u$ and $\mathcal{S}^u$ are forms of GMM and GEL statistics suggested in DIN section 6, pp.67-71, adapted for testing the null hypothesis (2.3).

where $\mathcal{T}_{GMM}^{g_a} = n\hat{g}_a(\hat{\beta}_a)'\hat{\Omega}_a^{-1}\hat{g}_a(\hat{\beta}_a)$ with $\hat{\beta}_a$ the subvector of $\hat{\beta}$ corresponding to β_a, $\hat{g}_a(\beta_a) = \sum_{i=1}^n g_{ai}(\beta_a)/n$ and $\hat{\Omega}_a = \sum_{i=1}^n g_{ai}(\hat{\beta}_a)g_{ai}(\hat{\beta}_a)'/n$. GEL forms $\mathcal{LR}^a$, $\mathcal{LM}^a$ and $\mathcal{S}^a$ follow similarly; cf. (4.1), (4.2) and (4.3) respectively. The proofs of Theorems 4.1 and 4.2 may be adapted to demonstrate that these statistics each converge in distribution to a standard normal random variate and are mutually asymptotically equivalent but not to the restricted statistics or the unrestricted GEL class defined above.

This section concludes with an asymptotic independence result between the restricted GMM statistic $\mathcal{J}^r$ for testing (2.3) and the corresponding statistic for testing the maintained hypothesis (2.1); *viz.*

$$\mathcal{J}^m = \frac{\mathcal{T}_{GMM}^{g_m} - (J_m K - p_m)}{\sqrt{2(J_m K - p_m)}}; \tag{4.5}$$

viz.

Theorem 4.3 *If Assumptions 3.1-3.4 hold and if* $K \to \infty$ *and* $\zeta(K)^2 K^2/n \to 0$*, then (a)* $\mathcal{J}^m \xrightarrow{d} N(0,1)$ *and (b)* $\mathcal{J}^r$ *is asymptotically independent of* $\mathcal{J}^m$*.*

A similar result holds for the associated restricted GEL statistics $\mathcal{LR}^r$, $\mathcal{LM}^r$, $\mathcal{S}^r$ and $\mathcal{W}^r$ and their counterparts for testing (2.1) if the additional assumption $\zeta(K)^2K^3/n \to 0$ is imposed.

Remark 4.3: The practical import of Theorem 4.3 is that the overall asymptotic size of the test sequence for (2.1) and (2.2) may be controlled, e.g., (a) test (2.1) using $\mathcal{J}^m$; (b) given (2.1), test (2.2) using $\mathcal{J}^r$, with overall asymptotic test size $1-(1-\alpha_m)(1-\alpha_a)$, where α_m and α_a are the respective asymptotic sizes of the individual tests in (a) and (b).

Remark 4.4: The asymptotic independence of $\mathcal{J}^r$ and $\mathcal{J}^m$ mirrors that of classical and unconditional moment GMM and GEL tests for a sequence of parametric restrictions; see Newey (1985) and Smith (2011). Indeed the unrestricted statistic $\mathcal{J}^u$ is the sum of suitably rescaled restricted $\mathcal{J}^r$ and maintained hypothesis $\mathcal{J}^m$ statistics; cf. the decomposition of standard unrestricted classical or GMM and GEL statistics for parametric restrictions.

5 Asymptotic Local Power

This section considers the asymptotic distribution of the statistics of the previous sections under a suitable sequence of local alternatives. Critically, this discussion demonstrates the deficiency in terms of asymptotic local power of unrestricted tests which fail to fully incorporate the maintained conditional information (2.1) and thereby the superiority of restricted tests.

The set-up is similar to that in Eubank and Spielgeman (1990) and Hong and White (1995), see also

Tripathi and Kitamura (2003), utilising local alternatives to the null hypothesis (2.3) of the form

$$H_{1n} : E[u(z, \beta_{n,0})|s] = \frac{\sqrt[4]{J_a MK}}{\sqrt{n}} \xi(s), \tag{5.1}$$

where $\beta_{n,0} \in \mathcal{B}$ is a non-stochastic sequence such that $\beta_{n,0} \to \beta_0$. It is assumed that $E[\xi_m(s)|s_m] = 0$, where $\xi_m(s) = S_m^u \xi(s)$, thus ensuring that the maintained hypothesis $E[u_m(z, \beta_{m0})|s_m] = 0$ (2.1) is not violated.

REMARK 5.1: The sequence of local alternatives (5.1) is particularly apposite for the instrumental validity Example 2.2 in which $u(z, \beta) = u_m(z, \beta_m) = u_a(z, \beta_a)$ with $\beta = \beta_m = \beta_a$. If the maintained instruments s_m are a subvector of s_a, i.e., $s = s_a$, $E[\xi(s)|s_m] = 0$. Similarly, when s_m is not a subvector of s_a, the relevant sequence of local alternatives to $E[u(z, \beta_0)|s_m] = 0$ is the expectation of (5.1) conditional on s_a, i.e.,

$$E[u(z, \beta_{n,0})|s_a] = \frac{\sqrt[4]{J_a MK}}{\sqrt{n}} E[\xi(s)|s_a].$$

The asymptotic local alternative distributions of the statistics described above are obtained under the following assumption.

Assumption 5.1 *(a) $\beta_{n,0}$ is a non-stochastic sequence such that (5.1) holds and $\beta_{n,0} \to \beta_0$; (b) $\sqrt{n}(\hat{\beta} - \beta_{n,0}) = O_p(1)$; (c) for all $\beta \in \mathcal{N}$, $\Sigma(s; \beta) = E[u(z, \beta)u(z, \beta)'|s]$ and $\Sigma_m(s_m; \beta_m) = E[u_m(z, \beta_m)u_m(z, \beta_m)'|s_m]$ each have smallest eigenvalue bounded away from zero; (d) $\|\xi(s)\|$ is bounded; (e) $\Sigma(s; \beta)$, $\Sigma_m(s_m; \beta_m)$ and $D(s; \beta) = E[u_\beta(z, \beta)|s]$, $D_m(s_m; \beta_m) = E[u_{m\beta}(z, \beta_m)|s_m]$ are continuous functions on a compact closure of $\mathcal{N}$.*

The next result summarises the limiting distribution of the restricted statistics $\mathcal{J}^r$, $\mathcal{LR}^r$, $\mathcal{LM}^r$, $\mathcal{S}^r$ and $\mathcal{W}^r$ under the sequence of local alternatives (5.1). Let $\Sigma(s) = \Sigma(s; \beta_0)$.

Theorem 5.1 *Let Assumptions 3.1-3.4 and 5.1 hold, $K \to \infty$ and $\zeta(K)^2 K^2/n \to 0$. Then $\mathcal{J}^r$ converges in distribution to a $N(\mu^r/\sqrt{2}, 1)$ random variate, where*

$$\mu^r = E[\xi(s)'\Sigma(s)^{-1}\xi(s)].$$

If additionally Assumption 3.5 is satisfied and $\zeta(K)^2 K^3/n \to 0$, then $\mathcal{LR}^r$, $\mathcal{LM}^r$, $\mathcal{S}^r$ and $\mathcal{W}^r$ are asymptotically equivalent to $\mathcal{J}^r$.

REMARK 5.2: Since $\mu^r \geq 0$ tests of the null hypothesis H_0 (2.3) based on these statistics should be one-sided. Although not discussed here, a similar analysis to that underpinning DIN Lemma 6.5, p.71, demonstrates the consistency of tests based on the statistics $\mathcal{J}^r$, $\mathcal{LR}^r$, $\mathcal{LM}^r$, $\mathcal{S}^r$ and $\mathcal{W}^r$. Note that the non-centrality parameter μ^r of Theorem 5.1 does not depend on M and hence the asymptotic local power of restricted tests is therefore unaffected by the choice of M.

The following corollary to Theorem 5.1 details the limiting distribution of the standard forms of unrestricted statistics $\mathcal{LR}^u$ (4.1), $\mathcal{LM}^u$ (4.2) and $\mathcal{S}^u$ (4.3) under the same local alternative sequence (5.1).

Corollary 5.1 *Let Assumptions 3.1-3.4 and 5.1 hold and $\zeta(K)^2K^2/n \to 0$. Then $\mathcal{S}^u$ converges in distribution to a $N(\mu^u/\sqrt{2}, 1)$ random variate, where*

$$\mu^u = \sqrt{\frac{J_a M}{J_a M + J_m}}\mu^r.$$

If additionally Assumption 3.5 is satisfied and $\zeta(K)^2K^3/n \to 0$, then $\mathcal{LR}^u$, $\mathcal{LM}^u$ are asymptotically equivalent to $\mathcal{S}^u$.

REMARK 5.3: Since $\mu^r > \mu^u$ Corollary 5.1 demonstrates that for fixed M restricted tests dominate the standard unrestricted tests in terms of asymptotic local power. Other unrestricted tests that partially or completely fail to incorporate the maintained conditional moment information (2.1) in their formulation are likewise relatively deficient. For example, using a similar analysis to that for Theorem 5.1, the GMM statistic $\mathcal{J}^a$ (4.4) and associated GEL statistics $\mathcal{LR}^a$, $\mathcal{LM}^a$ and $\mathcal{S}^a$ may be shown to converge in distribution under the local alternatives sequence (5.1) to a $N(\mu^a/\sqrt{2}, 1)$ random variable where $\mu^a = E[\xi(s)' S_a^{u\prime}(S_a^u \Sigma(s) S_a^{u\prime})^{-1} S_a^u \xi(s)]$. Hence $\mu^r - \mu^a \geq 0$. Therefore tests based on these and other unrestricted statistics are asymptotically less powerful relative to restricted tests.

REMARK 5.4: Corollary 5.1 also shows that the difference in local asymptotic power between restricted and unrestricted tests declines with increasing M since the noncentrality parameter μ^u would differ little from μ^r with consequential similar discriminatory power for both standard unrestricted and restricted tests for local departures from the null hypothesis H_0 (2.3).

REMARK 5.5. Theorem 5.1 and Corollary 5.1 provide no guidance for the choice of M. The effect of M on power for given sample size n and K will depend on the specific alternative hypothesis and correspondingly the relevance of any additional unconditional moment functions included by increasing M. More precisely the efficacy in terms of power of including extra elements in $q_a^{MK}(s_a)$, i.e., increasing M, for given n and K, will depend on the correlation between these extra elements and the conditional expectation $E[u(z, \beta_0)|s]$. If this correlation is zero or weak then although not strictly speaking applicable here an asymptotic local power analysis for the unconditional moment context would indicate that power should be expected to be diminished since test chi-square degrees of freedom will increase with M but the noncentrality parameter will remain relatively unaltered. Cf. section 3, pp.238-244, and, in particular, the discussion following Proposition 6, p.242, in Newey (1985). If this correlation is strong there will be a trade-off between increases in both degrees of freedom and noncentrality parameter with power potentially enhanced. Simulation evidence reported next in section 6 suggests that for a given sample size n and fixed value of K the correspondence between empirical and nominal test size deteriorates with increasing M; a similar deterioration is also observed for size-corrected empirical power but it should be emphasised against specific sets of alternatives.

6 Simulation Evidence

This section reports the results from a simulation study to assess the performance of some of the tests for ME and JE forms of instrument validity in the linear regression model, see Example 2.2, based on the GMM and GEL statistics developed in previous sections. To provide a realistic setting, the investigation is based on an application to a dataset where the issue of instrument validity is of some interest and importance.

Overall these experiments revealed that nominal size is approximated relatively more closely by the empirical size of (a) the *non-standardised* tests, see Remark 4.1, and (b) tests based on efficient estimators, cf. Tripathi and Kitamura (2003), although Assumption 3.3(c) only requires $\sqrt{n}$-consistent estimation. Consequently only results for these forms of statistics are presented. The Wald test statistic $\mathcal{W}^r$ (3.10) and score test statistic $\mathcal{S}^r$ (3.9) are also excluded for similar reasons. Likewise only the results for restricted tests are reported as they dominate the unrestricted forms in terms of empirical power reflecting their theoretical superiority; see Corollary 5.1.[14]

All experiments concern a parametric specification for the Engel curve relationship between the expenditure share of leisure services y and the logarithm of total expenditures x and employ the same data as those in Blundell and Horowitz (2007). These data correspond to a subsample of the household-level observations from the British Family Expenditure Survey and consist of a sample of 1518 married couples with one or two children and an employed head of household. Since many parametric Engel curve specifications are often linear or quadratic in x, see, e.g., Muellbauer (1976) and Banks, Blundell and Lewbel (1997), the experimental basis here is the linear regression model

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + u. \tag{6.1}$$

The maintained instrument s_m is the annual income from wages and salaries of the head of household. Thus $\beta = \beta_m = \beta_a = (\beta_0, \beta_1, \beta_2)'$, $J_a = J_m = 1$ and $u(z, \beta) = u_m(z, \beta_m) = u_a(z, \beta_a)$ where $u(z, \beta) = y - \beta_0 - \beta_1 x - \beta_2 x^2$; see Example 2.2. Cf. section 5, p.1051, of Blundell and Horowitz (2007).

The regression design incorporates both ME and JE forms of additional conditional constraint restrictions (2.2); see Remark 2.2. Therefore the hypotheses of interest are as follows. First, the maintained

[14] The full set of simulation results is available from the authors upon request.

hypothesis (2.1) $E[u|s_m] = 0$. Secondly, the additional conditional moment constraints (2.2): ME $E[u|x] = 0$, i.e., $s_a = x$, and JE $E[u|s_m, x] = 0$, i.e., $s_a = (s_m, x)$.

6.1 Experimental Design

The parameter vector β is estimated using the full data set by efficient two step (2S) GMM, with weight matrix computed using two stage least squares with the single instrument s_m, see section 4, pp.63-65, of DIN, based on the maintained conditional moment restriction $E[u(z,\beta)|s_m] = 0$. The maintained 2SGMM vector of approximating functions is $q_m^K(s_m)$ with $K = 25$.[15] 2SGMM estimates are denoted as β_0^e, β_1^e and β_2^e with 2SGMM residual $u^e = y - \beta_0^e - \beta_1^e x - \beta_2^e x^2$.

The structure of the data generating process underpinning the design is similar to that in section 4, pp.1049-1051, of Blundell and Horowtiz (2007). To ensure that the maintained hypothesis $E[u(z,\beta)|s_m] = 0$ holds in the sample consider the residual from a nonparametric series regression of u^e on s_m for the full data set, i.e., $u_{s_m}^{e\perp} = u^e - q_m^{25}(s_m)' \left(Q_{25}(s_m)'Q_{25}(s_m)\right)^{-} Q_{25}(s_m)'u^e$, where $^-$ denotes a generalised inverse and $Q_{25}(s_m) = (q_m^{25}(s_{m1}), ..., q_m^{25}(s_{m1518}))'$ with the vector $q_m^{25}(s_m)$ defined below in section 6.1.1 for $n = 1518$. Hence $E[u_{s_m}^{e\perp}|s_m] = 0$ approximately; see, e.g., section 3, pp.6-8, of Newey (1994). To impose the JE hypothesis $E[u(z,\beta)|s_a] = 0$, where $s_a = (s_m, x)$, the error term $u_{s_m x}^{e\perp}$ is obtained as the residual from the nonparametric series regression of u^e on s_m and x, i.e., $u_{s_m x}^{e\perp} = u^e - q^{25}(s)' \left(Q_{25}(s)'Q_{25}(s)\right)^{-} Q_{25}(s)'u^e$, where $Q_{25}(s) = (q^{25}(s_1), ..., q^{25}(s_{1518}))'$ with $q^{25}(s) = (q_m^{25}(s_m)', q_a^{25}(s_a)')'$, and then generating the dependent variable as $y^{mc} = \beta_0^e + \beta_1^e x + \beta_2^e x^2 + u_{s_m x}^{e\perp}$. Then $E[u_{s_m x}^{e\perp}|s_m] = 0$ and $E[u_{s_m x}^{e\perp}|s_m, x] = 0$ approximately. Deviations from the JE null hypothesis are formulated as in $y^{mc} = \beta_0^e + \beta_1^e x + \beta_2^e x^2 + u_{s_m x}^e$ where $u_{s_m x}^e = s^{s_m x}(u_{s_m x}^{e\perp} + \rho(u_{s_m}^{e\perp} - u_{s_m x}^{e\perp}))/s_\rho^{s_m x}$ with $s^{s_m x}$ and $s_\rho^{s_m x}$ the standard deviations of $u_{s_m x}^{e\perp}$ and $u_{s_m x}^{e\perp} + \rho(u_{s_m}^{e\perp} - u_{s_m x}^{e\perp})$ respectively.

Experimental data are generated as random samples of size n from (s_{mi}, x_i, y_i^{mc}), $(i = 1, ..., 1518)$; simulation random samples are denoted by $z_i = (s_{mi}, x_i, y_i^{mc})$, $(i = 1, ..., n)$, below. Empirical test size is examined for sample sizes $n = 200$, 500, 1000 and 1500 with nominal sizes 0.01, 0.05 and 0.10. Sample sizes of $n = 200$ and 500 only are considered in those experiments concerned with empirical power. All experiments employ 5000 replications and were programmed using `MATLAB`.

[15] Efficient 2SGMM estimates are

$$\hat{y} = -\underset{(0.662)}{1.29} + \underset{(0.268)}{0.629}\, x - \underset{(0.0269)}{0.0609}\, x^2.$$

Estimated standard errors are in parentheses. Tests for ME $E[u|x] = 0$, i.e., $s_a = x$, and JE $E[u|s_m, x] = 0$, i.e., $s_a = (s_m, x)$, discussed in section 6.1.3 were conducted on the full data set using the value $K = 8$ indicated by the rule in section 6.2 below. All ME tests rejected the null hypothesis at nominal levels 0.01, 0.05 and 0.10 for $M = 1$ and at levels 0.05 and 0.10 when $M = 2$ providing further support for the results reported in section 5, p.1051, of Blundell and Horowitz (2007). At nominal level 0.01 for $M = 2$ tests based on the GEL LR-type, LM-type and Wald statistics failed to reject the ME null hypothesis whereas those based on the statistics $\mathcal{J}^{\mathrm{M}}$, $\mathcal{LR}_{\mathrm{CUE}}^{\mathrm{M}}$ and $\mathcal{LR}_{\mathrm{CUE(GEL)}}^{\mathrm{M}}$ evaluated at EL and ET estimators, $\widetilde{\mathcal{LM}}^{\mathrm{M}}$ and score statistics did reject at the 0.01 level. These latter tests are precisely those that displayed a close correspondence between empirical and nominal size in the experiments reported below. All tests for the JE null hypothesis $E[u|s_m, x] = 0$ rejected at nominal levels 0.01, 0.05 and 0.1 for both $M = 1$ and 2.

6.1.1 Approximating Functions

Legendre polynomials are used to form the approximating functions in the simulations because of their good collinearity properties, see Example 3.1 of Belloni et al. (2015), and are defined as

$$\begin{aligned} \mathrm{P}_0(v) &= 1, \mathrm{P}_1(v) = v, \\ \mathrm{P}_{r+1}(v) &= \frac{(2r+1)v\mathrm{P}_r(v) - r\mathrm{P}_{r-1}(v)}{r+1}, r = 1, 2, 3, ... \end{aligned}$$

where $v \in [-1, 1]$; see eq. 8.5.3, p.334, of Abramowitz and Stegun (1970).[16] Since neither s_m nor x has support $[-1, 1]$ the transformations $\tilde{s}_m = 2\Phi\left(\frac{s_m - \bar{s}_m}{s_{s_m}}\right) - 1$ and $\tilde{x} = 2\Phi\left(\frac{x - \bar{x}}{s_x}\right) - 1$ are employed where $\Phi(\cdot)$ is the $N(0, 1)$ cumulative distribution function; for a given replication of sample size n, $\bar{s}_m = \sum_{i=1}^n s_{mi}/n$, $s_{s_m} = \sum_{i=1}^n (s_{mi} - \bar{s}_m)^2/n$ and $\bar{x} = \sum_{i=1}^n x_i/n$, $s_x = \sum_{i=1}^n (x_i - \bar{x})^2/n$.[17]

The maintained conditional moment $E[u(z, \beta)|s_m]$, cf. (2.1), is approximated using the vector of functions $q_m^K(s_m)$ with elements $\mathrm{P}_j(\tilde{s}_m)$, $(j = 0, ..., K - 1)$. For ME $E[u(z, \beta)|x]$ is approximated using a polynomial of order MK in x, i.e., $q_a^{MK}(s_a)$ has elements $\mathrm{P}_k(\tilde{x})$, $(k = 1, ..., MK)$. The JE case $E[u(z, \beta)|s_a]$, $s_a = (s_m, x)$, uses the $[(MK)^{1/2}]^2$-vector of approximating functions $q_a^{MK}(s_a)$ with elements $\mathrm{P}_j(\tilde{s}_m)\mathrm{P}_k(\tilde{x})$, $(k = 0, ..., [(MK)^{1/2}] - 1, l = 1, ..., [(MK)^{1/2}])$ resulting in the null hypothesis vector of approximating functions $q^K(s) = (q_m^K(s_m)', q_a^{MK}(s_a)')'$. See fn. 6.

6.1.2 Estimators

Efficient estimation methods examined include 2SGMM (GMM) with weight matrix computed as above, continuous updating (CUE), empirical likelihood (EL) and exponential tilting (ET). The subscripts MA, ME and JE indicate estimation incorporating maintained, ME and JE restrictions respectively.

GMM, CUE and ET are computed using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm of `MATLAB`. EL is more problematic because in some samples for particular BFGS EL estimates $\hat{\beta}_{\mathrm{EL}}$ the convex hull condition $\left\|\sum_{i=1}^n \hat{\pi}_i^{\mathrm{EL}} g(z_i, \hat{\beta}_{\mathrm{EL}})\right\| < 10^{-4}$ may not be satisfied where the EL implied probabilities $\hat{\pi}_i^{\mathrm{EL}} = 1/n(1 + \hat{\lambda}_{\mathrm{EL}}' g(z_i, \hat{\beta}_{\mathrm{EL}}))$, $(i = 1, ..., n)$, and the EL Lagrange multiplier $\hat{\lambda}_{\mathrm{EL}} = -\hat{\Omega}_\pi^{-1}\hat{g}(\hat{\beta}_{\mathrm{EL}})$ with $\hat{\Omega}_\pi = \sum_{i=1}^n \hat{\pi}_i^{\mathrm{EL}} g(z_i, \hat{\beta}_{\mathrm{EL}}) g(z_i, \hat{\beta}_{\mathrm{EL}})'$ and $\hat{g}(\beta) = \sum_{i=1}^n g(z_i, \beta)/n$; see Theorem 2.3, p.224, of Newey and Smith (2004). Hence EL is computed using the `matElike` `MATLAB` package with the optional Zip-solver package; see Zedlewski (2008).[18] In the case of non-convergence, EL is computed employing BFGS applied to the EL dual problem with the Lagrange multiplier obtained using `MATLAB` code based on eq. (12.3), p.235, of Owen (2001).[19] EL estimates obtained via this procedure are only considered to be valid

[16] Theorem 8, p.90, in Lorenz (1986) establishes the requisite uniform convergence for polynomial approximating functions; cf. Assumption 3.1.

[17] We are grateful to V. Chernozhukov for this suggestion.

[18] `matElike`, rather than solving the dual EL problem, solves the primal EL problem directly and is chosen as the default algorithm because it is faster on average than BFGS. Both BFGS and `matElike` solutions are identical if each converges to a solution in the convex hull.

[19] EL computation requires some care since the EL criterion involves the logarithm function which is undefined for negative arguments. This difficulty is avoided by replacing logarithms with a function that is logarithmic for

solutions if the convex hull condition is satisfied, otherwise no solution in the convex hull is reported. Note, however, that in the test size and power results reported in sections 6.3 and 6.4 the EL estimates satisfied the convex hull condition in all replications.[20]

6.1.3 Test Statistics

Restricted tests for ME $E[u|x]=0$ and JE $E[u|s_m,x]=0$ adopt the following notation. The superscripts M and J refer respectively to the ME or JE hypothesis under test with the subscripts CUE, EL, ET referring to which GEL criterion is used to construct the test and, as above, denoting the efficient estimator(s) employed. E.g., the non-standardised restricted GEL LR-type statistic for JE based on EL criteria and estimators is denoted as $\mathcal{LR}^{\text{J}}_{\text{EL}} = 2n(\hat{P}^{g}_{\text{EL}}(\hat{\beta}_{\text{EL}_{\text{J}}}, \hat{\lambda}_{\text{EL}_{\text{J}}}) - \hat{P}^{g_m}_{\text{EL}}(\hat{\beta}_{\text{EL}_{\text{MA}}}, \hat{\lambda}_{\text{EL}_{\text{MA}}}))$, cf. (3.7). LR-type CUE statistics evaluated at null and the maintained hypothesis EL and ET estimators are also computed using the subscript CUE(GEL) to denote the use of the CUE criterion and GEL estimators, e.g., for JE, $\mathcal{LR}^{\text{J}}_{\text{CUE(GEL)}} = 2n(\hat{P}^{g}_{\text{CUE}}(\hat{\beta}_{\text{GEL}_{\text{J}}}, \hat{\lambda}_{\text{GEL}_{\text{J}}}) - \hat{P}^{g_m}_{\text{CUE}}(\hat{\beta}_{\text{GEL}_{\text{MA}}}, \hat{\lambda}_{\text{GEL}_{\text{MA}}}))$. The non-standardised robustified score $\bar{\mathcal{S}}$ and Wald $\bar{\mathcal{W}}$ statistics, see fn. 11, evaluated at the corresponding efficient MA estimator are also examined. Restricted ME and JE non-standardised test statistics are calibrated against chi-square distributions with MK and $[(MK)^{1/2}]^2$ degrees of freedom respectively.[21]

GEL LM, score and Wald ME and JE test statistics require estimators of the variance matrix $\Omega = E[g(z,\beta_0 g(z,\beta_0)']$ and Jacobian $G = E[\partial g(z,\beta_0)/\partial\beta']$. The estimators considered for Ω and G are $\hat{\Omega} = n^{-1}\sum_{i=1}^{n} g(z_i,\hat{\beta}_{\text{GEL}})g(z_i,\hat{\beta}_{\text{GEL}})'$ and $\hat{G} = n^{-1}\sum_{i=1}^{n}\partial g(z_i,\hat{\beta}_{\text{GEL}})/\partial\beta'$ where $\hat{\beta}_{\text{GEL}}$ is the null hypothesis GEL estimator. Additional results are also presented for ME and JE LM tests based on the consistent estimator $\tilde{\Omega}_k = \hat{\Omega}_k\hat{\Omega}^{-1}\hat{\Omega}_k$ for Ω, see where [22]

$$\hat{\Omega}_k = \sum\nolimits_{i=1}^{n}\hat{k}_i g(z_i,\hat{\beta}_{\text{GEL}})g(z_i,\hat{\beta}_{\text{GEL}})', \hat{k}_i = -\frac{\rho_1(\hat{\lambda}'_{\text{GEL}}g(z_i,\hat{\beta}_{\text{GEL}}))+1}{n\hat{\lambda}'_{\text{GEL}}g(z_i,\hat{\beta}_{\text{GEL}})}, (i=1,...,n).$$

LM statistics based on $\hat{\Omega}_k$ are denoted $\widetilde{\mathcal{LM}}$.

6.2 Choice of the Number of Instruments

Implementation of the above tests requires a choice of K to employ under the maintained hypothesis. Because the Donald, Imbens and Newey (2009) method and selection criteria such as SBC predominantly indicated choices of K that varied relatively little with sample size, following Table 1, p. 71, of DIN, K

arguments larger than a small positive constant and quadratic below that threshold. The code is available at `http://www-stat.stanford.edu/~owen/empirical/`

[20]In a preliminary study the convex hull condition was found to be violated for values of K and M larger than those considered here. The adjusted EL estimator of Chen, Variyath Abraham (2008) offers an alternative to EL in such circumstances.

[21]A number of asymptotically equivalent test statistics for the maintained hypothesis (2.1) were also investigated. The Durbin (1954)-Wu (1973)-Hausman (1978) test based on an auxiliary regression as described in Davidson and Mackinnon (1993, section 7.9, p.237), see also Wooldridge (2002, section 6.2.1, p.118), was also considered. Results are available on request from the authors.

[22]Adapting Theorem 2.3, p.224, of Newey and Smith (2004), the LM statistic for overidentifying moment conditions based on $\hat{\Omega}_k$ is identical to the score statistic based on $\hat{\Omega}$, i.e., $n\hat{\lambda}'_{\text{GEL}}\hat{\Omega}_k\hat{\lambda}_{\text{GEL}} = n\hat{g}(\hat{\beta}_{\text{GEL}})'\hat{\Omega}^{-1}\hat{g}(\hat{\beta}_{\text{GEL}})$.

was chosen to satisfy $K^5/n \to 0$ according to the rule $K = [Cn^{0.19}]$ with $C = 2$ resulting in $K = 5$, 6, 7 and 8 corresponding to sample sizes $n = 200$, 500, 1000 and 1500 respectively. To explore the effect of increased M on test size and power the values $M = 1$ and 2 were examined.

6.3 Empirical Size

The results on empirical size reported here correspond to a nominal asymptotic level of 0.05; those results for nominal levels 0.01 and 0.10 are qualitatively similar and are therefore omitted.

6.3.1 ME

Table B.1 presents the empirical rejection frequencies for $M = 1$ and 2 for non-standardised restricted tests of the ME hypothesis $E[u(z, \beta_0)|x] = 0$ incorporating the maintained hypothesis moment restrictions $E[u(z, \beta_0)|s_m] = 0$.

In general, the empirical size of tests based on the Lagrange multiplier statistics $\mathcal{LM}^{\text{M}}_{\text{EL}}$, $\mathcal{LM}^{\text{M}}_{\text{ET}}$ and Wald statistics $\bar{\mathcal{W}}^{\text{M}}_{\text{EL}}$, $\bar{\mathcal{W}}^{\text{M}}_{\text{ET}}$ and to a lesser extent $\mathcal{LR}^{\text{M}}_{\text{EL}}$, $\mathcal{LR}^{\text{M}}_{\text{ET}}$ tests suffer from size distortions for moderate sample sizes $n = 200$ and 500 with a serious deterioration in performance as M increases from 1 to 2, i.e., as the number of unconditional moments under test increases. Of those remaining, the LR-type statistics $\mathcal{LR}^{\text{M}}_{\text{CUE}}$, $\mathcal{LR}^{\text{M}}_{\text{CUE(EL)}}$, $\mathcal{LR}^{\text{M}}_{\text{CUE(ET)}}$, the LM-type statistics $\widetilde{\mathcal{LM}}^{\text{M}}_{\text{EL}}$, $\widetilde{\mathcal{LM}}^{\text{M}}_{\text{ET}}$ and the ET robust score statistic $\bar{\mathcal{S}}^{\text{M}}_{\text{ET}}$ have good size properties. The 2SGMM criterion $\mathcal{J}^{\text{M}}$ statistic tends to be undersized and the EL robust score $\bar{\mathcal{S}}^{\text{M}}_{\text{EL}}$ statistic somewhat oversized except for the larger sample sizes.[23] Generally speaking, for a given sample size n and thus fixed K there is a deterioration in performance to a lesser or greater degree for larger M a finding also mirrored in other experiments by increasing K with fixed sample size.

In summary, tests for ME based on the statistics $\mathcal{LR}^{\text{M}}_{\text{CUE}}$ and $\mathcal{LR}^{\text{M}}_{\text{CUE(EL)}}$, $\mathcal{LR}^{\text{M}}_{\text{CUE(ET)}}$ and $\widetilde{\mathcal{LM}}^{\text{M}}_{\text{EL}}$, $\widetilde{\mathcal{LM}}^{\text{M}}_{\text{ET}}$ and $\bar{\mathcal{S}}^{\text{M}}_{\text{ET}}$ appear to be the most reliable in terms of empirical size.

6.3.2 JE

Table B.2 presents the rejection frequencies for $M = 1$ and 2 for non-standardised restricted tests of the JE null hypothesis $E[u(z, \beta_0)|s_m, x] = 0$.

The general conclusions are quite similar to those for the ME tests. Overall performance worsens substantially for the larger M for moderate sample sizes $n = 200$ and 500 for all test versions. The CUE LR-type forms $\mathcal{LR}^{\text{J}}_{\text{CUE}}$, $\mathcal{LR}^{\text{J}}_{\text{CUE(EL)}}$, $\mathcal{LR}^{\text{J}}_{\text{CUE(ET)}}$ evaluated at CUE, EL and ET estimators, the GEL LM-type statistics $\widetilde{\mathcal{LM}}^{\text{J}}_{\text{EL}}$, $\widetilde{\mathcal{LM}}^{\text{J}}_{\text{ET}}$ and the robust score statistic $\bar{\mathcal{S}}^{\text{J}}_{\text{ET}}$ display the most satisfactory empirical size at the nominal 0.05 level whereas as above the 2SGMM criterion $\mathcal{J}^{\text{J}}$ and the EL robust score $\bar{\mathcal{S}}^{\text{J}}_{\text{EL}}$ statistics are respectively undersized and oversized in the smaller sample sizes.

[23] Matsushita and Otsu (2013) obtained similar results for EL LR-type tests for overidentifying conditions to those reported here.

6.4 Empirical Power

Tables B.3 and B.4 present size-corrected (SC) and non size-corrected (NSC) empirical rejection frequencies at the 0.05 level of tests for the ME and JE hypotheses.[24] Given their poor size performance tests based on the Lagrange multiplier statistics $\mathcal{LM}_{\text{EL}}$, $\mathcal{LM}_{\text{ET}}$ and Wald statistics $\bar{\mathcal{W}}_{\text{EL}}$, $\bar{\mathcal{W}}_{\text{ET}}$ are not considered in this section.

Typically both rejection frequencies increase substantially as sample size n increases from 200 to 500 but decline with increased M although there are some exceptions for $n = 200$ and small ρ. In general the statistics that performed well in terms of empirical size yield similar rejection frequencies under the alternatives considered here.

6.4.1 ME

Table B.3 presents empirical rejection frequencies for non-standardised restricted ME tests for values $M = 1$ and 2 based on 0.05 level size-corrected and nominal non size-corrected critical values for deviations $\rho \neq 0$ from the ME hypothesis $E[u(z, \beta_0)|x] = 0$.

In general both rejection frequencies increase with deviation ρ and sample size n and decline with M with some exceptions at $\rho = 0.2$. Size-corrected empirical power differences between tests are less at higher values for the deviations ρ and sample sizes n. Overall tests based on the LR-type statistics $\mathcal{LR}^{\text{M}}_{\text{EL}}$ and $\mathcal{LR}^{\text{M}}_{\text{ET}}$ using the nominal 0.05 chi-square critical value are most powerful but it is precisely these tests that display an unsatisfactory correspondence between empirical and nominal size. Empirical power is relatively low at $\rho = 0.2$ for all tests employing size-corrected or non size-corrected critical values. Generally speaking empirical power for all tests employing size-corrected critical values, not just those with reasonable empirical size characteristics, is rather similar for both the smaller $n = 200$ and larger $n = 500$ sample sizes.

6.4.2 JE

Table B.4 presents empirical rejection frequencies for non-standardised restricted JE tests for values $M = 1$ and 2 based on 0.05 level size-corrected and nominal non size-corrected critical values for deviations $\rho \neq 0$ from the JE hypothesis $E[u(z, \beta_0)|s_m, x] = 0$.

Similar general conclusions to those for the ME tests above broadly follow. Interestingly, given M, sample size n and thus K, rejection frequencies are higher than those obtained for the ME hypothesis.

6.5 Summary

The empirical size of non-standardised tests more closely approximates nominal size than that of standardised tests. The use of efficient rather than root-n consistent estimators is recommended for test

[24] Horowitz and Savin (2000) argue that empirical rejection frequencies based on nominal critical values are the most relevant since size-correction is not realistically implementable in practice.

construction. Restricted dominate unrestricted tests in terms of empirical power. Empirical power typically declines for increases in M for both ME and JE tests.

For both the ME $E[u(z,\beta_0)|x] = 0$ and JE hypotheses $E[u(z,\beta_0)|s_m,x] = 0$ empirical sizes of restricted tests based on the restricted CUE LR-type statistics $\mathcal{LR}_{\text{CUE}}$, $\mathcal{LR}_{\text{CUE(EL)}}$, $\mathcal{LR}_{\text{CUE(ET)}}$, evaluated at CUE, EL and ET estimates, and the LM-type statistics $\widetilde{\mathcal{LM}}_{\text{EL}}$, $\widetilde{\mathcal{LM}}_{\text{ET}}$ and the robust ET score versions $\bar{\mathcal{S}}_{\text{ET}}$ most closely approximate nominal size. The differences in empirical power with size-corrected critical values between these tests are rather marginal.

7 Conclusions

The primary focus of this article has been concerned with the provision of tests for additional conditional moment constraints in cross-section or short panel data contexts. The principal contribution is the explicit incorporation of conditional moment restrictions defining the maintained hypothesis in the formulation of the test statistics mirroring test construction in the classical parametric likelihood setting. The approach reinterprets the respective conditional moment hypotheses as infinite numbers of unconditional moment restrictions with the corresponding tests formulated as tests for additional sets of infinite numbers of unconditional moment restrictions. The limiting distributions of these test statistics are derived under the null hypothesis and suitable sequences of local alternatives. These results suggest that restricted tests that fully incorporate maintained moment constraints in their construction should dominate in terms of power unrestricted tests that fail to do so.

The simulation experiments undertaken to explore the efficacy of the various tests proposed in the paper indicate a number of restricted tests possess both sufficiently satisfactory empirical size and power characteristics to allow their recommendation for econometric practice.

Appendix A: Proofs of Results

Throughout the Appendix, C will denote a generic positive constant that may be different in different uses with CS, J, M, T and c_r Cauchy-Schwarz, Jensen, Markov, triangle and Loève c_r, Davidson (1994, p.140), inequalities respectively. Also we write w.p.a.1 for "with probability approaching 1".

A.1 Useful Lemmata

The following lemma relaxes DIN Assumption 6, p.67, for the GEL class of estimators.

Lemma A.1. Let $\delta_n = o(n^{-1/2}\zeta(K)^{-1})$ and $\Lambda_n = \{\lambda : \|\lambda\| \leq \delta_n\}$. Then, if Assumption 3.3(d) is satisfied, $\max_{\beta\in\mathcal{B},\lambda\in\Lambda_n,1\leq i\leq n} |\lambda' g_i(\beta)| \xrightarrow{p} 0$ and w.p.a.1 $\Lambda_n \subset \hat{\Lambda}_n(\beta)$ for all $\beta \in \mathcal{B}$.

Proof. Write $b_i = \sup_{\beta\in\mathcal{B}} \|g(z_i,\beta)\|^2$. By iterated expectations and Assumption 3.3(d), $E[b_i] = E[E[b_i|w]] < \infty$ for $1 \le i \le n$. It then follows from Lemma 3, p.98, of Owen (1990) that $\max_{1\le i\le n} b_i = o_p(n^{1/2})$. Therefore, by CS

$$\max_{\beta\in\mathcal{B},\lambda\in\Lambda_n,1\le i\le n} |\lambda' g_i(\beta)| \le \delta_n \zeta(K) \max_{1\le i\le n} b_i \xrightarrow{p} 0.$$

Thus w.p.a.1 $\lambda' g_i(\beta) \in \mathcal{V}$ for all $\beta \in \mathcal{B}$ and $\lambda \in \Lambda_n$ giving the second conclusion. ■

The next two lemmata are used in the proofs of asymptotic normality for test statistics under both null and local alternative hypotheses and the asymptotic independence of test statistics under the null hypothesis.

Lemma A.2. Let $k = tr(\Omega_n C_n)$ where C_n and $\Omega_n = E[g(z,\beta_{0,n})g(z,\beta_{0,n})']$ are a symmetric and a positive definite matrix respectively. If $E[g(z,\beta_{0,n})] = 0$, $k \to \infty$, $E[(g(z,\beta_{0,n})'C_n g(z,\beta_{0,n}))^2]/k\sqrt{n} \to 0$ and $C_n\Omega_n C_n = C_n$, then

$$\mathcal{T} = \frac{n\hat{g}(\beta_{0,n})' C_n \hat{g}(\beta_{0,n}) - k}{\sqrt{2k}} \xrightarrow{d} N(0,1).$$

Proof. Let $g_{i,n} = g(z_i,\beta_{0,n})$, $(i = 1,...,n)$, and write $\mathcal{T} = \mathcal{T}_1 + \mathcal{T}_2$ where

$$\begin{aligned}\mathcal{T}_1 &= \sum\nolimits_{i,j:i<j} \sqrt{\frac{2}{n^2k}} g_{i,n}' C_n g_{j,n},\\ \mathcal{T}_2 &= \frac{\sum_i g_{i,n}' C_n g_{i,n}/n - k}{\sqrt{2k}}.\end{aligned}$$

Since $E[\mathcal{T}_2] = 0$ and $var[\mathcal{T}_2] \le E[(g_{i,n}' C_n g_{i,n})^2]/2kn \to 0$, $\mathcal{T}_2 \xrightarrow{p} 0$.

To establish the asymptotic normality of $\mathcal{T}_1$ we verify the hypotheses of Theorem 1, pp.3-4, of Hall (1984). Define

$$H_n(u,v) = \sqrt{\frac{2}{n^2k}} g_{u,n}' C_n g_{v,n}.$$

Then, if $u, v \ne 1$,

$$\begin{aligned}G_n(u,v) &= E[H_n(z_1,u)H_n(z_1,v)|u,v]\\ &= \frac{2}{n^2k} E[g_{u,n}' C_n g_{1,n} g_{1,n}' C_n g_{v,n}|u,v]\\ &= \frac{2}{n^2k} g_{u,n}' C_n \Omega_n C_n g_{v,n}\\ &= \sqrt{\frac{2}{n^2k}} H_n(u,v).\end{aligned}$$

Now $E[H_n(z_1,z_2)|z_1] = \sqrt{\frac{2}{n^2k}} g_{1,n}' C_n E[g_{2,n}] = 0$ and

$$\begin{aligned}E[H_n(z_1,z_2)^2] &= \frac{2}{n^2k} E[(g_{1,n}' C_n g_{2,n})^2]\\ &= \frac{2}{n^2k} E[g_{1,n}' C_n \Omega_n C_n g_{1,n}] = \frac{2}{n^2}.\end{aligned}$$

On the other hand

$$\frac{E[H_n(z_1,z_2)^4]}{nE[H_n(z_1,z_2)^2]^2} = \frac{1}{nk^2}E[(g_{1,n}'C_n g_{2,n})^4].$$

As $C_n = C_n\Omega_n C_n$, by CS

$$\begin{aligned}\frac{1}{nk^2}E[(g_{1,n}'C_n g_{2,n})^4] &\leq \frac{1}{nk^2}E[(g_{1,n}'C_n g_{1,n})^2(g_{2,n}'C_n g_{2,n})^2] \\ &= (\frac{1}{k\sqrt{n}}E[(g_{1,n}'C_n g_{1,n})^2])^2 \to 0.\end{aligned}$$

Since $E[G_n(z_1,z_2)^2]/E[H_n(z_1,z_2)^2]^2 = 1/k \to 0$, $\mathcal{T}_1 \xrightarrow{d} N(0,1)$ as required. ■

LEMMA A.3. If (a) $E[g(z,\beta_0)] = 0$, (b) $tr(Q\Omega) = ak$ for some finite $a \in \mathcal{R}\backslash\{0\}$, (c) $tr[(Q\Omega)^2] = vk$ for some finite $v > 0$, (d) $tr[(Q\Omega)^4] = o(k^2)$, (e) $E[(g(z,\beta_0)'Qg(z,\beta_0))^2] = o(nk)$ and (f) $E[(g(z,\beta_0)'Q\Omega Qg(z,\beta_0))^2]$ $\times E[(g(z,\beta_0)'\Omega^{-1}g(z,\beta_0))^2] = o(nk^2)$ are satisfied, then

$$\mathcal{T} = \frac{n\hat{g}(\beta_0)'Q\hat{g}(\beta_0) - ak}{\sqrt{2k}} \xrightarrow{d} N(0,v).$$

as $k \to \infty$ and $n \to \infty$.

PROOF. Let $g_i = g(z_i,\beta_0)$, $(i = 1,...,n)$, and write $\mathcal{T} = \mathcal{T}_1 + \mathcal{T}_2$ where

$$\begin{aligned}\mathcal{T}_1 &= \sum_{i,j:i<j}\sqrt{\frac{2}{n^2k}}g_i'Qg_j \\ \mathcal{T}_2 &= \frac{\sum_i g_i'Qg_i/n - ak}{\sqrt{2k}}\end{aligned}$$

Since $E[\mathcal{T}_2] = 0$ and $var[\mathcal{T}_2] \leq E[(g_i'Qg_i)^2]/2kn \to 0$ by (e), $\mathcal{T}_2 \xrightarrow{p} 0$.

To prove the asymptotic normality of $\mathcal{T}_1$, as in the proof of Lemma A.2, we verify the hypotheses of Theorem 1, pp.3-4, of Hall (1984). Define

$$H_n(u,v) = \sqrt{\frac{2}{n^2k}}g_u'Qg_v.$$

Then, if $u,v \neq 1$,

$$\begin{aligned}G_n(u,v) &= E[H_n(z_1,u)H_n(z_1,v)|u,v] \\ &= \frac{2}{n^2k}E[g_u'Qg_1g_1'Qg_v|u,v] \\ &= \frac{2}{n^2k}g_u'Q\Omega Qg_v.\end{aligned}$$

Now $E[H_n(z_1,z_2)|z_1] = \sqrt{\frac{2}{n^2k}}g_1'QE[g_2] = 0$ and, by (c),

$$\begin{aligned}E[H_n(z_1,z_2)^2] &= \frac{2}{n^2k}E[(g_1'Qg_2)^2] \\ &= \frac{2}{n^2k}E[g_1'Q\Omega Qg_1] = \frac{2}{n^2k}tr([Q\Omega]^2) = \frac{2v}{n^2}.\end{aligned}$$

Also

$$\frac{E[H_n(z_1,z_2)^4]}{nE[H_n(z_1,z_2)^2]^2} = \frac{1}{nv^2k^2}E[(g_1'Qg_2)^4].$$

Now, as Ω is positive definite, by CS

$$\begin{aligned}E[(g_1'Qg_2)^4] &= E[(g_1'Q\Omega\Omega^{-1}g_2)^4] \\ &\leq E[(g_1'Q\Omega Qg_1)^2(g_2'\Omega^{-1}g_2)^2] \\ &= E\left[(g_1'Q\Omega Qg_1)^2\right]E\left[(g_2'\Omega^{-1}g_2)^2\right].\end{aligned}$$

Hence, by (f),

$$\begin{aligned}\frac{E[H_n(z_1,z_2)^4]}{nE[H_n(z_1,z_2)^2]^2} &\leq \frac{1}{nv^2k^2}E\left[(g_1'Q\Omega Qg_1)^2\right]E\left[(g_2'\Omega^{-1}g_2)^2\right] \\ &= o(1).\end{aligned}$$

Moreover, by (d),

$$\begin{aligned}E[G_n(z_1,z_2)^2] &= \frac{4}{n^4k^2}E[(g_1'Q\Omega Qg_2)^2] \\ &= \frac{4}{n^4k^2}E[g_1'Q\Omega Q\Omega Q\Omega Qg_1] = \frac{4}{n^4k^2}tr([Q\Omega]^4) = o(n^{-4}).\end{aligned}$$

Since $E[G_n(z_1,z_2)^2]/E[H_n(z_1,z_2)^2]^2 = o(1)$, $\mathcal{T}_1 \xrightarrow{d} N(0,v)$ as required. ■

The next Lemma mirrors DIN Lemma A.3, p.73. Let $q_i = q(s_i)$, $(i = 1,...,n)$, where $q(\cdot)$ is a K-dimensional vector of functions of s.

Lemma A.4. Let $a_{i,n} = a_n(z_i)$, $\bar{a}_{i,n} = E[a_{i,n}|s_i]$, $a_i = a(z_i)$, $\bar{a}_i = E[a_i|s_i]$, $U_{i,n} = U_n(s_i)$ and $U_i = U(s_i)$. If $q(\cdot)$ satisfies Assumption 3.1, S is a finite n.s. matrix of column dimension rK, (a) $E[\|a_{i,n}\|^2|s_i]$ is bounded for large enough n, (b) $U_{i,n}$ is a $r\times r$ p.d. matrix that is bounded and has smallest eigenvalue bounded away from zero for large enough n, (c) U_i is $r\times r$ p.d. matrix that is bounded and has smallest eigenvalue bounded away from zero, (d) $E[\|U_{i,n}^{-1} - U_i^{-1}\|^2] \to 0$, (e) $E[\|\bar{a}_{i,n} - \bar{a}_i\|^2] \to 0$, (f) $K \to \infty$ and $K/n \to 0$, then

$$(\sum_i a_{i,n}'\otimes q_i')S'\left(S(\sum_i U_{i,n}\otimes q_iq_i')S'\right)^- S(\sum_i a_{i,n}\otimes q_i)/n - E[\bar{a}_i'U_i^{-1}\bar{a}_i] \xrightarrow{p} 0.$$

Proof. The proof is similar to that of DIN Lemma A.3. Let $F_{i,n}$ be a symmetric square root of $U_{i,n}$, $P_{i,n} = (F_{i,n}\otimes q_i')S'$, $P_n = (P_{1,n}',...,P_{n,n}')'$, $A_{i,n} = F_{i,n}^{-1}a_{i,n}$, $A_n = (A_{1,n}',...,A_{n,n}')'$, $\bar{A}_{i,n} = E[A_{i,n}|s_i] = F_{i,n}^{-1}\bar{a}_{i,n}$ and $\bar{A}_n = (\bar{A}_{1,n}',...,\bar{A}_{n,n}')'$. Note that $P_n'P_n = S(\sum_i U_{i,n}\otimes q_iq_i')S'$ and

$$(\sum_i a_{i,n}'\otimes q_i')S'\left(S(\sum_i U_{i,n}\otimes q_iq_i')S'\right)^- S(\sum_i a_{i,n}\otimes q_i/n) = A_n'Q_nA_n$$

where $Q_n = P_n(P_n'P_n)^-P_n'$.

Let $s = (s_1, ..., s_n)$. As the data are $\texttt{i.i.d.}$, by (a) and (b)

$$\begin{aligned} E[\left(A_n - \bar{A}_n\right)\left(A_n - \bar{A}_n\right)'|s] &= \texttt{diag}(F_{1,n}^{-1}var[a_{1,n}|s_1]F_{1,n}^{-1}, ..., F_{n,n}^{-1}var[a_{n,n}|s_n]F_{n,n}^{-1}) \\ &\leq CI \end{aligned}$$

for n large enough. Let $\mathcal{T}_A = (A_n - \bar{A}_n)'Q_n\left(A_n - \bar{A}_n\right)/n$. Then,

$$\begin{aligned} E[\mathcal{T}_A] &= E[tr(Q_nE[(A_n - \bar{A}_n)\left(A_n - \bar{A}_n\right)'/n)|s]] \\ &\leq CE[tr(Q_n)]/n \leq CK/n \to 0 \end{aligned}$$

as $tr(Q_n) \leq CK$, using (b) and (f). Thus $\mathcal{T}_A \xrightarrow{p} 0$ by M.

From Assumption 3.1, there exists a Γ_K such that $E[\left\|U_i^{-1}\bar{a}_i - \Gamma_K q_i\right\|^2] \to 0$. Let $S'\tilde{\gamma}_K = vec(\Gamma_K')$. Now

$$\begin{aligned} \left\|\bar{A}_n - P_n\tilde{\gamma}_K\right\|^2/n &= \sum\nolimits_i \left\|F_{i,n}^{-1}\bar{a}_i - (F_{i,n}\otimes q_i')S'\tilde{\gamma}_K\right\|^2/n \\ &= \sum\nolimits_i \left\|F_{i,n}\right\|^2\left\|U_{i,n}^{-1}\bar{a}_i - (I_r\otimes q_i')S'\tilde{\gamma}_K\right\|^2/n \\ &= \sum\nolimits_i \left\|F_{i,n}\right\|^2\left\|U_{i,n}^{-1}\bar{a}_i - \Gamma_k q_i\right\|^2/n \\ &\leq C\sum\nolimits_i \left\|U_{i,n}^{-1}\bar{a}_i - \Gamma_k q_i\right\|^2/n. \end{aligned}$$

By c_r,

$$\begin{aligned} E[\left\|U_{i,n}^{-1}\bar{a}_{i,n} - \Gamma_K q_i\right\|^2] &= E[\left\|(U_{i,n}^{-1} - U_i^{-1})\bar{a}_{i,n} + U_i^{-1}(\bar{a}_{i,n} - \bar{a}_i) + U_i^{-1}\bar{a}_i - \Gamma_K q_i\right\|^2] \\ &\leq 3(E[\left\|(U_{i,n}^{-1} - U_i^{-1})\bar{a}_{i,n}\right\|^2] + E[\left\|U_i^{-1}(\bar{a}_{i,n} - \bar{a}_i)\right\|^2] \\ &\quad + E[\left\|U_i^{-1}\bar{a}_i - \Gamma_K q_i\right\|^2). \end{aligned}$$

For the first term, by CS, $E[\left\|(U_{i,n}^{-1} - U_i^{-1})\bar{a}_{i,n}\right\|^2] \leq E[\left\|(U_{i,n}^{-1} - U_i^{-1})\right\|^2]E[\|\bar{a}_{i,n}\|^2] \to 0$ using (a) and (d). Secondly, $E[\left\|U_i^{-1}(\bar{a}_{i,n} - \bar{a}_i)\right\|^2] \leq CE[\|\bar{a}_{i,n} - \bar{a}_i\|^2] \to 0$ by (e) as U_i^{-1} is bounded by (c). Then, by M

$$\left\|\bar{A}_n - P_n\tilde{\gamma}_K\right\|^2/n \xrightarrow{p} 0.$$

By T and CS

$$\begin{aligned} \left|A_n'Q_nA_n/n - \bar{A}_n'\bar{A}_n/n\right| &= \left|(A_n - \bar{A}_n)'Q_n(A_n - \bar{A}_n)/n\right. \\ &\quad \left.+2\bar{A}_n'Q_n(A_n - \bar{A}_n)/n - \bar{A}_n'(I - Q_n)\bar{A}_n/n\right| \\ &\leq \mathcal{T}_A + 2\sqrt{\mathcal{T}_A}\sqrt{\bar{A}_n'\bar{A}_n/n} + \overline{\mathcal{T}}_A, \end{aligned}$$

where $\overline{\mathcal{T}}_A = \bar{A}_n'(I - Q_n)\bar{A}_n/n$. Now

$$\begin{aligned} \overline{\mathcal{T}}_A &= \left(\bar{A}_n - P_n\tilde{\gamma}_K\right)'(I - Q_n)\left(\bar{A}_n - P_n\tilde{\gamma}_K\right)/n \\ &\leq \left\|\bar{A}_n - P_n\tilde{\gamma}_K\right\|^2/n \xrightarrow{p} 0. \end{aligned}$$

Also, by M using (a) and (b), $\bar{A}_n'\bar{A}_n/n = O_p(1)$. Therefore,

$$\left|A_n'Q_nA_n/n - \bar{A}_n'\bar{A}_n/n\right| \xrightarrow{p} 0.$$

To examine the large sample behaviour of $\bar{A}_n'\bar{A}_n/n = \sum_i \bar{a}_{i,n}U_{i,n}^{-1}\bar{a}_{i,n}/n$, in particular, to show that $\bar{A}_n'\bar{A}_n/n \xrightarrow{p} E[\bar{a}_i'U_i^{-1}\bar{a}_i]$, since $\bar{a}_{i,n}$ and $U_{i,n}$ depend on n, we need to resort to a LLN for triangular arrays such as Theorem 1, p.316, of Feller (1971). Specifically, first we need to prove that, for each $\eta > 0$, $n\mathcal{P}\{\left|\bar{a}_{i,n}'U_{i,n}^{-1}\bar{a}_{i,n}\right|/n > \eta\} \to 0$. By M

$$n\mathcal{P}\{\left|\bar{a}_{i,n}'U_{i,n}^{-1}\bar{a}_{i,n}\right|/n > \eta\} \le E[\left|\bar{a}_{i,n}'U_{i,n}^{-1}\bar{a}_{i,n}\right|^2]/\left(n\eta^2\right).$$

For large enough n, by (a) and (b), $E[\left|\bar{a}_{i,n}'U_{i,n}^{-1}\bar{a}_{i,n}\right|^2]$ is bounded. Therefore $n\mathcal{P}\{\left|\bar{a}_{i,n}'U_{i,n}^{-1}\bar{a}_{i,n}\right|/n > \eta\} \to 0$. Secondly, for arbitrary $\varepsilon > 0$,

$$\begin{aligned} nvar[\frac{\left|\bar{a}_{i,n}'U_{i,n}^{-1}\bar{a}_{i,n}\right|}{n}1(\left|\bar{a}_{i,n}'U_{i,n}^{-1}\bar{a}_{i,n}\right| &< n\varepsilon)] \le nE[\frac{\left|\bar{a}_{i,n}'U_{i,n}^{-1}\bar{a}_{i,n}\right|^2}{n^2}1(\left|\bar{a}_{i,n}'U_{i,n}^{-1}\bar{a}_{i,n}\right| < n\varepsilon)] \\ &\le E[\left|\bar{a}_{i,n}'U_{i,n}^{-1}\bar{a}_{i,n}\right|^2]/n \to 0. \end{aligned}$$

Finally, $E[\bar{a}_{i,n}'U_{i,n}^{-1}\bar{a}_{i,n} - \bar{a}_iU_i^{-1}\bar{a}_i] = E[(\bar{a}_{i,n}-\bar{a}_i)'U_{i,n}^{-1}(\bar{a}_{i,n}-\bar{a}_i) + 2(\bar{a}_{i,n}-\bar{a}_i)'U_{i,n}^{-1}\bar{a}_i + \bar{a}_i'(U_{i,n}^{-1}-U_i^{-1})\bar{a}_i]$. Therefore, using T and CS, by (a), (b), (d) and (e),

$$\begin{aligned} E[\bar{a}_{i,n}'U_{i,n}^{-1}\bar{a}_{i,n} - \bar{a}_iU_i^{-1}\bar{a}_i] &\le E[\left\|U_{i,n}^{-1}\right\| \|\bar{a}_{i,n}-\bar{a}_i\|^2] \\ &\quad +2E[\left\|U_{i,n}^{-1}\right\| \|\bar{a}_{i,n}-\bar{a}_i\| \|\bar{a}_i\|] + E[\left\|U_{i,n}^{-1}-U_i^{-1}\right\| \|\bar{a}_i\|^2] \\ &\le C(E[\|(\bar{a}_{i,n}-\bar{a}_i)\|^2] + 2E[\|(\bar{a}_{i,n}-\bar{a}_i)\|] + E[\left\|U_{i,n}^{-1}-U_i^{-1}\right\|]) \\ &\to 0. \end{aligned}$$

■

The following lemma is similar to DIN Lemma A.4, p.75.

Lemma A.5. If $q(\cdot)$ satisfies Assumption 3.1, S is a finite n.s. matrix of column dimension rK, (a) $\varepsilon_{i,n}$ and Y_i are $r \times 1$ random vectors with $E[\varepsilon_{i,n}|s_i] = 0$ and $E[\|\varepsilon_{i,n}\|^4|s_i] \le C$ for large enough n and $E[\|Y_i\|^2|s_i] \le C$, (b) $U_{i,n} = U_n(s_i)$ is $r \times r$ p.d. matrix that is bounded and has the smallest eigenvalue bounded away from zero for n large enough, (c) $U_i = U(s_i)$ is $r \times r$ p.d. matrix that is bounded and has the smallest eigenvalue bounded away from zero, (d) $E[\left\|U_{i,n}^{-1}-U_i^{-1}\right\|^2] \to 0$ and (e) $K \to \infty$ and $K^2/n \to 0$, then

$$(\sum_i Y_i' \otimes q_i')S'\left(S(\sum_i U_{i,n} \otimes q_iq_i')S'\right)^{-1} S(\sum_i \varepsilon_{i,n} \otimes q_i)/\sqrt{n} = O_p(1).$$

Proof. The result is proved by first showing that

$$(\sum_i Y_i' \otimes q_i')S'\left(S(\sum_i U_{i,n} \otimes q_iq_i')S'\right)^{-1} S(\sum_i \varepsilon_{i,n} \otimes q_i)/\sqrt{n} - \sum_i E[Y_i|s_i]'U_{i,n}^{-1}\varepsilon_{i,n}/\sqrt{n} \xrightarrow{p} 0$$

and secondly that

$$\sum_i E[Y_i|s_i]'U_{i,n}^{-1}\varepsilon_{i,n}/\sqrt{n} = O_p(1). \tag{A.1}$$

The proof structure of the first part is similar to that of DIN Lemma A.4, p.75. Let $F_{i,n}$, P_n and thus Q_n be specified as in the proof of Lemma A.4, $A_{i,n} = F_{i,n}^{-1}Y_i$, $\bar{A}_{i,n} = E[A_{i,n}|s_i] = F_{i,n}^{-1}E[Y_i|s_i]$, $A_n = (A'_{1,n}, ..., A'_{n,n})'$, $\bar{A}_n = (\bar{A}'_{1,n}, ..., \bar{A}'_{n,n})'$, $B_{i,n} = F_{i,n}^{-1}\varepsilon_{i,n}$ and $B_n = (B'_{1,n}, ..., B'_{n,n})'$. By assumption $E[B_{i,n}|s_i] = 0$ and, consequently,

$$\begin{aligned} & (\sum_i Y_i' \otimes q_i')S'\left(S(\sum_i U_{i,n} \otimes q_i q_i')S'\right)^{-1} S(\sum_i \varepsilon_{i,n} \otimes q_i)/\sqrt{n} - E[Y_i|s_i]' U_{i,n}^{-1}\varepsilon_{i,n}/\sqrt{n} \\ = \ & A_n'Q_nB_n/\sqrt{n} - \bar{A}_n'B_n/\sqrt{n} = (A_n - \bar{A}_n)'Q_nB_n/\sqrt{n} - \bar{A}_n'(I - Q_n)B_n/\sqrt{n}. \end{aligned}$$

From the proof of Lemma A.4 $(A_n - \bar{A}_n)'Q_n(A_n - \bar{A}_n) = O_p(K)$ and $B_n'Q_nB_n = O_p(K)$, the latter holding by (a) as $E[\|\varepsilon_{i,n}\|^2|s_i] \le C$ for large enough n. Thus, for large enough n, by CS,

$$\left|(A_n - \bar{A}_n)'Q_nB_n/\sqrt{n}\right| \le \sqrt{(A_n - \bar{A}_n)'Q_n(A_n - \bar{A}_n)}\sqrt{B_n'Q_nB_n}/\sqrt{n} = O_p(K/\sqrt{n}) \xrightarrow{p} 0.$$

Also, as in the proof of Lemma A.4, $E[\bar{A}_n'(I - Q_n)\bar{A}_n/n] \to 0$. Thus, by iterated expectations,

$$\begin{aligned} E[\left\|\bar{A}_n'(I - Q_n)B_n/\sqrt{n}\right\|^2] &= E[\bar{A}_n'(I - Q_n)E[B_nB_n'|s](I - Q_n)\bar{A}_n]/n \\ &\le CE[\bar{A}_n'(I - Q_n)\bar{A}_n]/n \to 0 \end{aligned}$$

since $E[B_nB_n'|s]$ is bounded for large enough n by (a) and (b). The first part then follows by T and M.

It remains to prove the second part (A.1). Corollary, p.32, of Serfling (2002) is used to prove this result requiring only that

$$\lim_{n\to\infty} \frac{E[(E[Y_i|s_i]'U_{in}^{-1}\varepsilon_{i,n})^4]}{n^2 b_n^4} = 0, \tag{A.2}$$

where $b_n^2 = var[E[Y_i|s_i]'U_{in}^{-1}\varepsilon_{i,n}]$. Now, by CS, for large enough n,

$$\begin{aligned} E[(E[Y_i|s_i]'U_{in}^{-1}\varepsilon_{i,n})^4] &\le E[\|E[Y_i|s_i]\|^4 \left\|U_{in}^{-1}\right\|^4 \|\varepsilon_{i,n}\|^4] \\ &= E[\|E[Y_i|s_i]\|^4 \left\|U_{in}^{-1}\right\|^4 E[\|\varepsilon_{i,n}\|^4 |s_i]] \\ &\le C \end{aligned}$$

from (a) and (b). Also, by J,

$$\begin{aligned} b_n^2 &\le E[(E[Y_i|s_i]'U_{in}^{-1}\varepsilon_{i,n})^2] \\ &\le E[(E[Y_i|x_i]'U_{in}^{-1}\varepsilon_{i,n})^4]^{1/2} \le C \end{aligned}$$

from which (A.2) follows. ■

Let $u_i(\beta) = u(z_i, \beta)$, $g_i(\beta) = S(u_i(\beta) \otimes q_i)$, $\hat{g}_i = g_i(\hat{\beta})$ and $g_{i,n} = g_i(\beta_{0,n})$, $(i = 1, ..., n)$. Also let $u_{i,n} = u_i(\beta_{0,n})$, $\Sigma_i(\beta) = E[u_i(\beta)u_i(\beta)'|s_i]$, $\Sigma_{i,n} = \Sigma_i(\beta_{0,n}) = E[u_{i,n}u'_{i,n}|s_i]$, $(i = 1, ..., n)$, and

$$\begin{aligned}\hat{\Omega} &= \sum\nolimits_i \hat{g}_i\hat{g}_i'/n, \tilde{\Omega}_n = \sum\nolimits_i g_{i,n}g_{i,n}'/n, \\ \bar{\Omega}_n &= S(\sum\nolimits_i \Sigma_{i,n} \otimes q_iq_i')S'/n, \Omega_n = E[g_{i,n}g_{i,n}'].\end{aligned}$$

Lemma A.6. If $q(\cdot)$ satisfies Assumption 3.2, S a finite f.r.r. matrix and Assumptions 3.3 and 3.4 hold, then, if $\hat{\beta} - \beta_{0,n} = O_p(\tau_n)$ with $\tau_n \to 0$, $\left\|\hat{\Omega} - \tilde{\Omega}_n\right\| = O_p(\tau_n K)$, $\left\|\tilde{\Omega}_n - \bar{\Omega}_n\right\| = O_p(\zeta(K)\sqrt{K/n})$ and $\left\|\bar{\Omega}_n - \Omega_n\right\| = O_p(\zeta(K)\sqrt{K/n})$. If Assumption 5.1(c) is satisfied then $1/C \leq \lambda_{\min}(\Omega_n) \leq \lambda_{\max}(\Omega_n) \leq C$ and, if $\tau_n K + \zeta(K)\sqrt{K/n} \to 0$, w.p.a.1 $1/C \leq \lambda_{\min}\left(\hat{\Omega}\right) \leq \lambda_{\max}\left(\hat{\Omega}\right) \leq C$, $1/C \leq \lambda_{\min}\left(\bar{\Omega}_n\right) \leq \lambda_{\max}\left(\bar{\Omega}_n\right) \leq C$.

Proof. The proof of these results is similar to that of DIN Lemma A.6, p.78, with the major difference being that some expectations are assumed bounded for n large enough rather than being merely bounded.

Using the same arguments as in DIN

$$\begin{aligned}\left\|\hat{\Omega} - \tilde{\Omega}_n\right\| &\leq C\left\|\hat{\beta} - \beta_{0,n}\right\|\sum\nolimits_i M_{i,n}\left\|q_i\right\|^2/n \\ &= O_p(\tau_n E[M_{i,n}\left\|q_i\right\|^2]) \\ &= O_p(\tau_n K),\end{aligned}$$

where $M_{i,n} = \delta_i^2 + 2\delta_i\left\|u_{i,n}\right\|$ and $\delta_i = \delta(z_i)$. The final equality follows since $E[\delta(z)^2|s]$ is bounded and $E[\sup_{\beta\in\mathcal{B}}\left\|u(z,\beta)\right\|^2|s]$ is bounded by Assumption 3.3(d).

Now

$$E[\left\|\tilde{\Omega}_n - \bar{\Omega}_n\right\|^2] = E[\left\|S(\sum\nolimits_i(u_{i,n}u'_{i,n} - \Sigma_{i,n}(s_i)) \otimes q_iq_i')S'/n\right\|^2].$$

Since $\beta_{0,n} \to \beta_0$ and $\Sigma_i(\beta)$ is bounded for all $\beta \in \mathcal{N}$ it follows that for n large enough $\Sigma_{i,n}(s_i)$ is also bounded. Thus using similar arguments to those of DIN

$$E[\left\|\tilde{\Omega}_n - \bar{\Omega}_n\right\|^2] \leq CE[E[\left\|u_{i,n}\right\|^4|s_i]\left\|q_i\right\|^4]/n \leq C\zeta(K)^2K/n$$

as $E[\left\|u_{i,n}\right\|^4|s_i]$ is bounded for n large enough. Therefore the second conclusion follows by M.

For the third conclusion as in DIN

$$\begin{aligned}E[\left\|\bar{\Omega}_n - \Omega_n\right\|^2] &= E[\left\|S(\sum\nolimits_i \Sigma_{i,n}(s_i) \otimes q_iq_i')S'/n - \Omega_n\right\|^2] \\ &\leq Ctr(E[\Sigma_{i,n}(s_i)^2 \otimes (q_iq_i')^2]/n) \leq CE[\left\|q_i\right\|^4]/n \leq C\zeta(K)^2K/n\end{aligned}$$

where the second inequality holds for n large enough.

For the fourth conclusion, since, for all $\beta \in \mathcal{N}$, $\Sigma(s,\beta) = E[u(z,\beta)u(z,\beta)'|s]$ has smallest eigenvalue bounded away from zero and $E[\sup_{\beta\in\mathcal{B}} \|u(z,\beta)\|^2]$ is bounded, it follows that $C^{-1}I_J \leq \Sigma_{i,n}(s_i) \leq CI_J$ and therefore

$$C^{-1}I_{JK} = C^{-1}SE\left[I_J \otimes q_i q_i'\right]S' \leq \Omega_n \leq CSE\left[I_J \otimes q_i q_i'\right]S' = CI_{JK}.$$

Hence $C^{-1} \leq \lambda_{\min}(\Omega_n) \leq \lambda_{\min}(\Omega_n) \leq C$. Note also that, if $\tau_n K + \zeta(K)\sqrt{K/n} \to 0$, we have $\left\|\hat{\Omega} - \tilde{\Omega}_n\right\| = o_p(1)$ and $\left\|\tilde{\Omega}_n - \Omega_n\right\| = o_p(1)$. Thus, by T $\left\|\hat{\Omega} - \Omega_n\right\| = o_p(1)$. Since $|\lambda(A) - \lambda(B)| \leq \|A - B\|$, where $\lambda(\cdot)$ denotes the minimum or maximum eigenvalue, $\left|\lambda_{\min}\left(\hat{\Omega}\right) - \lambda_{\min}(\Omega_n)\right| = o_p(1)$ and $\left|\lambda_{\max}\left(\hat{\Omega}\right) - \lambda_{\max}(\Omega_n)\right| = o_p(1)$. The final conclusion follows similarly. ■

Let $u_{\beta i}(\beta) = \partial u(z_i,\beta)/\partial\beta'$, $D(s_i,\beta) = E[u_{\beta i}(\beta)|s_i]$, $D_{i,n} = D(s_i,\beta_{0,n})$,

$$\hat{G} = S'(\textstyle\sum_i u_{\beta i}(\beta) \otimes q_i)/n, \bar{G}_n = S'(\textstyle\sum_i D_{i,n} \otimes q_i)/n, G_n = S'E[D_{i,n} \otimes q_i].$$

Lemma A.7. If $q(\cdot)$ satisfies Assumption 3.2, S a finite n.s. matrix, Assumption 3.4 holds and $\hat{\beta} - \beta_{0,n} = O(\tau_n)$ with $\tau_n \to 0$, then $\left\|\hat{G} - \bar{G}_n\right\| = O_p(\tau_n\sqrt{K} + \sqrt{K/n})$ and $\left\|\bar{G}_n - G_n\right\| = O_p(\sqrt{K/n})$.

Proof. The proof is as in that for DIN Lemma A.7, p.79, and requires no stronger assumptions. Let $u_{\beta i,n} = u_{\beta i}(\beta_{0,n})$, $\delta_i = \delta(z_i)$ and $\tilde{G}_n = S'(\sum_i u_{\beta i,n} \otimes q_i)/n$. Then by DIN Lemma A.2, p.73,

$$\begin{aligned} E[\left\|\tilde{G}_n - \bar{G}_n\right\|^2] &= E[\left\|S(\textstyle\sum_i (u_\beta(z_i,\beta_{0,n}) - D_{i,n}) \otimes q_i)/n\right\|^2] \\ &\leq CE[E[\|u_{\beta i,n}\|^2 |s_i] \|q_i\|^2]/n \leq CK/n, \end{aligned}$$

where the last inequality follows for n large enough as $\beta_{0,n} \to \beta_0$ and $E[\sup_{\beta\in\mathcal{N}} \|u_\beta(z,\beta)\|^2 |s]$ is bounded. Hence, by M $\left\|\tilde{G}_n - \bar{G}_n\right\|^2 = O_p(\sqrt{K/n})$.

By the same arguments as in DIN Proof of Lemma A.7, pp.78-80, w.p.a.1

$$\begin{aligned} \left\|\hat{G} - \tilde{G}_n\right\| &\leq C\textstyle\sum_i \left\|u_{\beta i}(\hat{\beta}) - u_{\beta i,n}\right\| \|q_i\|/n \\ &\leq C\left\|\hat{\beta} - \beta_{0,n}\right\| \textstyle\sum_i \delta_i \|q_i\|/n = O_p(\tau_n\sqrt{K}). \end{aligned}$$

The first conclusion follows by T.

In addition

$$\begin{aligned} E[\left\|\bar{G}_n - G_n\right\|^2] &= E[\left\|S(\textstyle\sum_i D_{i,n} \otimes q_i)/n - G_n\right\|^2] \\ &\leq CE\left[\|D_{i,n}\|^2 \|q_i\|^2\right]/n \leq CK/n, \end{aligned}$$

where the first inequality follows from $D_{i,n}$ bounded for n large enough as $E[\sup_{\beta\in\mathcal{N}} \|u_\beta(z,\beta)\|^2 |s]$ is bounded from which the second conclusion follows. ■

The final lemma mirrors DIN Lemma 6.1, p.69.

Lemma A.8. Let $q(\cdot)$ satisfy Assumptions 3.1 and 3.2 and Assumptions 3.3, 3.4 and 5.1 hold. If $K \to \infty$ and $\zeta(K)^2 K^2/n \to 0$ then

$$\frac{n\hat{g}(\hat{\beta})'\hat{\Omega}^{-1}\hat{g}(\hat{\beta}) - n\hat{g}(\beta_{n,0})'\Omega_n^{-1}\hat{g}(\beta_{n,0})}{\sqrt{2(J_m + J_a M)K}} \xrightarrow{p} 0.$$

Proof. Let $g_{i,n} = g_i(\beta_{n,0})$, $(i = 1, ..., n)$, $\hat{g}_n = \hat{g}(\beta_{n,0})$ and $\hat{g} = \hat{g}(\hat{\beta})$. By an expansion of $\hat{g} = \hat{g}(\hat{\beta})$ around $\beta_{0,n}$

$$\hat{g} = \hat{g}_n + \bar{G}_n(\hat{\beta} - \beta_{n,0}),$$

where $\bar{G}_n = \partial\hat{g}(\bar{\beta}_n)/\partial\beta'$ and $\bar{\beta}_n$ is a mean value between $\hat{\beta}$ and $\beta_{n,0}$ which may differ from row to row. Thus

$$\begin{aligned}\frac{n\hat{g}(\hat{\beta})'\hat{\Omega}^{-1}\hat{g}(\hat{\beta}) - n\hat{g}_n'\Omega_n^{-1}\hat{g}_n}{\sqrt{2(J_m + J_a M)K}} &= \frac{n\hat{g}_n'(\hat{\Omega}^{-1} - \Omega_n^{-1})\hat{g}_n}{\sqrt{2(J_m + J_a M)K}} \\ &\quad + \frac{2n(\hat{\beta} - \beta_{n,0})'\bar{G}_n'\hat{\Omega}^{-1}\hat{g}_n}{\sqrt{2(J_m + J_a M)K}} \\ &\quad + \frac{n(\hat{\beta} - \beta_{n,0})'\bar{G}_n'\hat{\Omega}^{-1}\bar{G}_n(\hat{\beta} - \beta_{n,0})}{\sqrt{2(J_m + J_a M)K}}.\end{aligned}$$

To show that each term converges in probability to zero, we need to prove first some preliminary results.

Since $\lambda_{\min}\left(\hat{\Omega}\right) \geq C$ and $\lambda_{\min}\left(\bar{\Omega}_n\right) \geq C$ w.p.a.1, by Lemmata A.6 and A.7

$$\begin{aligned}\left\|\hat{\Omega}^{-1}(\bar{G}_n - G_n)\right\|^2 &= tr((\bar{G}_n - G_n)'\hat{\Omega}^{-2}(\bar{G}_n - G_n)) \\ &\leq Ctr((\bar{G}_n - G_n)'(\bar{G}_n - G_n)) \\ &= C\left\|\bar{G}_n - G_n\right\|^2 \xrightarrow{p} 0.\end{aligned}$$

Similarly, $\left\|\hat{\Omega}^{-1}(\hat{\Omega} - \Omega_n)\right\| \xrightarrow{p} 0$.

Now $G_n'\Omega_n^{-1}G_n$ is bounded for large enough n as $\bar{G}_n'\bar{\Omega}_n^{-1}\bar{G}_n \xrightarrow{p} V^{-1}$ by Lemma A.4 where $V = (E[D(x)'\Sigma(x)^{-1}D(x)])^{-1}$ which exists from Assumptions 3.4(d)(e) as $E[D(x)'\Sigma(x)^{-1}D(x)] \geq CE[D(x)'D(x)]$. Thus, $\left\|\Omega_n^{-1}G_n\right\|$ is also bounded. Therefore, to prove that $\left\|\hat{\Omega}^{-1}G_n\right\| = O_p(1)$, by T

$$\left\|\hat{\Omega}^{-1}\bar{G}_n - \Omega_n^{-1}G_n\right\| \leq \left\|\hat{\Omega}^{-1}(\bar{G}_n - G_n)\right\| + \left\|\hat{\Omega}^{-1}(\hat{\Omega} - \Omega_n)\Omega_n^{-1}G_n\right\|.$$

First, term $\left\|\hat{\Omega}^{-1}(\bar{G}_n - G_n)\right\| \xrightarrow{p} 0$ by Lemma A.7. Secondly, $\left\|\hat{\Omega}^{-1}(\hat{\Omega} - \Omega_n)\Omega_n^{-1}G_n\right\| \leq \left\|\hat{\Omega}^{-1}(\hat{\Omega} - \Omega_n)\right\| \left\|\Omega_n^{-1}G_n\right\|$by CS and Lemma A.6. Consequently, $\left\|\hat{\Omega}^{-1}\bar{G}_n\right\| = O_p(1)$.

Now by independence

$$\begin{aligned}E[\hat{g}_n'\Omega_n^{-1}\hat{g}_n] &= E[g_{i,n}'\Omega_n^{-1}g_{i,n}]/n \\ &= E[tr(\Omega_n^{-1}g_{i,n}g_{i,n}')/n] = K/n.\end{aligned}$$

Hence, by M $\left\|\Omega_n^{-1}\hat{g}_n\right\| = O_p(\sqrt{K/n})$. By T and CS

$$\begin{aligned}
\left\|\bar{G}_n'\hat{\Omega}^{-1}\hat{g}_n - G_n'\Omega_n^{-1}\hat{g}_n\right\| &\leq \left\|\bar{G}_n'\hat{\Omega}^{-1}(\hat{\Omega}-\Omega_n)\Omega_n^{-1}\hat{g}_n\right\| + \left\|(\bar{G}_n - G_n)'\Omega_n^{-1}\hat{g}_n\right\| \\
&\leq (\left\|\bar{G}_n'\hat{\Omega}^{-1}\right\|\left\|\hat{\Omega}-\Omega_n\right\| + \left\|\bar{G}_n - G_n\right\|)\left\|\Omega_n^{-1}\hat{g}_n\right\| \\
&\leq (O_p(1)o_p(1) + o_p(1))O_p(\sqrt{K/n}) = o_p(\sqrt{K/n}).
\end{aligned}$$

Moreover

$$E[\left\|G_n'\Omega_n^{-1}\hat{g}_n\right\|^2] = E[tr(\hat{g}_n'\Omega_n^{-1}G_nG_n'\Omega_n^{-1}\hat{g}_n)] = tr(G_n'\Omega_n^{-1}G_n)/n \leq C/n.$$

Thus, by M, $\left\|G_n'\Omega_n^{-1}\hat{g}_n\right\| = O_p(1/\sqrt{n}) = o_p(\sqrt{K/n})$ and, hence, by T $\left\|\bar{G}_n'\hat{\Omega}^{-1}\hat{g}_n\right\| = o_p(\sqrt{K/n})$. Therefore, by Assumption 3.3(c),

$$\frac{n(\hat{\beta}-\beta_{n,0})'\bar{G}_n'\hat{\Omega}^{-1}\hat{g}_n}{\sqrt{2JK}} = o_p(1).$$

Next, by CS and T,

$$\begin{aligned}
\left\|\bar{G}_n'\hat{\Omega}^{-1}\bar{G}_n - G_n'\Omega_n^{-1}G_n\right\| \leq\ & (\left\|\bar{G}_n'\hat{\Omega}^{-1}\right\| + \left\|\Omega_n^{-1}G_n\right\|)\left\|\bar{G}_n - G_n\right\| \\
&+ \left\|\bar{G}_n'\hat{\Omega}^{-1}\right\|\left\|\hat{\Omega}-\Omega_n\right\|\left\|\Omega_n^{-1}G_n\right\|.
\end{aligned}$$

Hence, $\bar{G}_n'\hat{\Omega}^{-1}\bar{G}_n = O_p(1)$ since $G_n'\Omega_n^{-1}G_n = O(1)$. Therefore

$$\frac{n(\hat{\beta}-\beta_{n,0})'\bar{G}_n'\hat{\Omega}^{-1}\bar{G}_n(\hat{\beta}-\beta_{n,0})}{\sqrt{2(J_m+J_aM)K}} = O_p(1/\sqrt{2(J_m+J_aM)K}) = o_p(1).$$

It remains to prove that

$$\frac{n\hat{g}_n'(\hat{\Omega}^{-1}-\Omega_n^{-1})\hat{g}_n}{\sqrt{2(J_m+J_aM)K}} = o_p(1).$$

From Lemma A.6,

$$\begin{aligned}
\left|n\hat{g}_n'\left(\hat{\Omega}^{-1}-\Omega_n^{-1}\right)\hat{g}_n\right|/\sqrt{2(J_m+J_aM)K} &\leq n\left\|\Omega_n^{-1}\hat{g}_n\right\|^2(\left\|\hat{\Omega}-\Omega_n\right\| + C\left\|\hat{\Omega}-\Omega_n\right\|^2)/\sqrt{2(J_m+J_aM)K} \\
&= n(O_p(K/n)(O_p(\sqrt{K/n}) + O_p(\zeta(K)\sqrt{K/n})))/\sqrt{2(J_m+J_aM)K} \\
&= O_p(\zeta(K)K/\sqrt{n}) = o_p(1).
\end{aligned}$$

■

A.2 Asymptotic Null Distribution

Proof of Theorem 4.1. By DIN Lemma A.6, p.78, and $\zeta(K)^2K^2/n \to 0$,

$$\left\|\hat{\Omega}_m - \Omega_m\right\|, \left\|\hat{\Omega}-\Omega\right\| = O_p((K^{3/2}/n^{1/2} + \zeta(K)K/n^{1/2})/\sqrt{K}) = o_p(1/\sqrt{K}),$$

where $\Omega_m = E[g_m(z,\beta_{m0})g_m(z,\beta_{m0})']$ and $\Omega = E[g(z,\beta_0)g(z,\beta_0)']$. It also follows from DIN Lemma A.7, p.79, that $\left\|\partial\hat{g}_m\left(\tilde{\beta}_m\right)/\partial\beta_m' - G_m\right\| \xrightarrow{p} 0$ and $\left\|\partial\hat{g}\left(\tilde{\beta}\right)/\partial\beta' - G\right\| \xrightarrow{p} 0$ for any $\tilde{\beta} = \beta_0 + O_p(1/\sqrt{n})$.

In addition, $G_m'\Omega_m^{-1}G_m$ and $G'\Omega^{-1}G$ are bounded; see Proof of Lemma A.8. Hence, the conditions of DIN Lemma 6.1, p.69, are met. Therefore,

$$\frac{n\hat{g}_m(\hat{\beta}_m)'\hat{\Omega}_m^{-1}\hat{g}_m(\hat{\beta}_m) - n\hat{g}_m(\beta_{m0})'\Omega_m^{-1}\hat{g}_m(\beta_{m0})}{\sqrt{2J_mK}} \xrightarrow{p} 0.$$

and

$$\frac{n\hat{g}(\hat{\beta})'\hat{\Omega}^{-1}\hat{g}(\hat{\beta}) - n\hat{g}(\beta_0)'\Omega^{-1}\hat{g}(\theta_0)}{\sqrt{2(J_m + J_aM)K)}} \xrightarrow{p} 0.$$

Now

$$\begin{aligned}\frac{n\hat{g}(\hat{\beta})'\hat{\Omega}^{-1}\hat{g}(\hat{\beta}) - n\hat{g}_m(\hat{\beta}_m)'\hat{\Omega}_m^{-1}\hat{g}_m(\hat{\beta}_m) - J_aMK}{\sqrt{2J_aMK}} &= \frac{n\hat{g}(\hat{\beta})'\hat{\Omega}^{-1}\hat{g}(\hat{\beta}) - n\hat{g}(\beta_0)'\Omega^{-1}\hat{g}(\beta_0)}{\sqrt{2J_aMK}} \\ &- \frac{n\hat{g}_m(\hat{\beta})'\hat{\Omega}_m^{-1}\hat{g}_m(\hat{\beta}_m) - n\hat{g}_m(\beta_{m0})'\Omega_m^{-1}\hat{g}_m(\beta_{m0})}{\sqrt{2J_aMK}} \\ &+ \frac{n\hat{g}(\beta_0)'\Omega^{-1}\hat{g}(\beta_0) - n\hat{g}_m(\beta_{m0})'\Omega_m^{-1}\hat{g}_m(\beta_{m0}) - J_aMK}{\sqrt{2J_aMK}} \\ &= \frac{n\hat{g}(\beta_0)'\Omega^{-1}\hat{g}(\beta_0) - n\hat{g}_m(\beta_{m0})'\Omega_m^{-1}\hat{g}_m(\beta_{m0}) - J_aMK}{\sqrt{2J_aMK}} \\ &\quad + o_p(1).\end{aligned}$$

Recall $S_m = S_m^u \otimes S_m^q$ and, thus, $S_m\hat{g}(\beta_0) = \hat{g}_m(\beta_{m0})$. Therefore

$$\frac{n\hat{g}(\beta_0)'\Omega^{-1}\hat{g}(\beta_0) - n\hat{g}_m(\beta_{m0})'\Omega_m^{-1}\hat{g}_m(\beta_{m0}) - J_aMK}{\sqrt{2J_aMK}} = \frac{n\hat{g}(\theta_0)'(\Omega^{-1} - S_m'\Omega_m^{-1}S_m)\hat{g}(\theta_0) - J_aMK}{\sqrt{2J_aMK}}.$$

Lemma A.2 provides the conclusion of the theorem. First, $tr((\Omega^{-1} - S_m'\Omega_m^{-1}S_m)\Omega) = tr(I_{(J_aM+J_m)K}) - tr(I_{J_mK}) = J_aMK$. Secondly, $(\Omega^{-1} - S_m'\Omega_m^{-1}S_m)\Omega(\Omega^{-1} - S_m'\Omega_m^{-1}S_m) = \Omega^{-1} - S_m'\Omega_m^{-1}S_m$. Thirdly,

$$\begin{aligned}E[(g(z,\beta_0)'(\Omega^{-1} - S_m'\Omega_m^{-1}S_m)g(z,\beta_0))^2] &\leq CE[\|g(z,\beta_0)\|^4] \\ &\leq CE[\|u(z,\beta_0)\|^4 \left\|q^K(s)\right\|^4] \\ &\leq CE[\left\|q^K(s)\right\|^4] \\ &\leq C\zeta(K)^2K.\end{aligned}$$

The result follows from Lemma A.2 as $\zeta(K)^2K/K\sqrt{n} = (\zeta(K)^2K^2/n)/\sqrt{K^4/n} \to 0$. ■

Proof of Theorem 4.2. Consider $\mathcal{LR}^r$ (3.7), i.e.,

$$\begin{aligned}\mathcal{LR}^r &= \frac{2n(\hat{P}_\rho^g(\hat{\beta},\hat{\lambda}) - \hat{P}_\rho^{g_m}(\hat{\beta}_m,\hat{\lambda}_m)) - J_aMK}{2\sqrt{J_aMK}} \\ &= \frac{\mathcal{T}_{GMM}^g - \mathcal{T}_{GMM}^{g_m} - J_aMK}{2\sqrt{J_aMK}} \\ &\quad + \frac{2n\hat{P}_\rho^g(\hat{\beta},\hat{\lambda}) - \mathcal{T}_{GMM}^g}{\sqrt{2J_aMK}} \\ &\quad - \frac{2n\hat{P}_\rho^{g_m}(\hat{\beta}_m,\hat{\lambda}_m) - \mathcal{T}_{GMM}^{g_m}}{\sqrt{2J_aMK}}.\end{aligned} \tag{A.3}$$

Write $\hat{g}_{mi} = g_{mi}(\hat{\beta}_m)$, $(i = 1, ..., n)$, $\hat{g}_m = \hat{g}_m(\hat{\beta}_m)$ and $\hat{g}_{m0} = \hat{g}_m(\beta_{m0})$. Using T and CS twice

$$\begin{aligned} \|\hat{g}_m - \hat{g}_{m0}\| &\leq \sum\nolimits_{i=1}^{n} \left\| u_m(z_i, \hat{\beta}_m) - u_m(z_i, \beta_{m0}) \right\| \left\| q_m^K(s_{mi}) \right\| / n \\ &\leq (\sum\nolimits_{i=1}^{n} \delta(z_i)^2/n)^{1/2} (\sum\nolimits_{i=1}^{n} \left\| q_m^K(s_{mi}) \right\|^2 /n)^{1/2} \left\| \hat{\beta}_m - \beta_{m0} \right\| = O_p(\sqrt{K/n}) \end{aligned}$$

where the second inequality follows from Assumption 3.4(d). Thus, from T and DIN Lemma A.9, p.81, $\|\hat{g}_m\| = O_p(\sqrt{K/n})$ and, therefore, $\left\|\hat{\lambda}_m\right\| = O_p(\sqrt{K/n})$ by DIN Lemma A.11, p.82. Consequently $\hat{\lambda}_m \in \hat{\Lambda}_n^m(\hat{\beta}_m)$ w.p.a.1 and the first order conditions for λ_m are satisfied w.p.a.1, i.e.,

$$\frac{\partial \hat{P}_{\rho}^{g_m}(\hat{\beta}_m, \hat{\lambda}_m)}{\partial \lambda_m} = \sum\nolimits_{i=1}^{n} \rho_1(\hat{\lambda}_m' \hat{g}_{mi}) \hat{g}_{mi}/n = 0. \tag{A.4}$$

Expanding (A.4) around $\lambda_m = 0$ gives

$$-\hat{g}_m(\hat{\beta}_m) - \dot{\Omega}_m \hat{\lambda}_m = 0$$

where $\dot{\Omega}_m = -\sum_{i=1}^{n} \rho_2(\dot{\lambda}_m' \hat{g}_{mi}) \hat{g}_{mi} \hat{g}_{mi}'/n$ and $\dot{\lambda}_m$ lies between $\hat{\lambda}_m$ and 0. Thus, w.p.a.1

$$\hat{\lambda}_m = -\dot{\Omega}_m^{-1} \hat{g}_m(\hat{\beta}_m). \tag{A.5}$$

For the third term in (A.3) expand $2n\hat{P}_{\rho}^{g_m}(\hat{\beta}_m, \hat{\lambda}_m)$ around $\lambda_m = 0$ and plug in $\hat{\lambda}_m$ from (A.5), i.e.,

$$2n\hat{P}_{\rho}^{g_m}(\hat{\beta}_m, \hat{\lambda}_m) = 2n(-\hat{g}_m(\hat{\beta}_m)'\hat{\lambda}_m - \hat{\lambda}_m' \ddot{\Omega}_m \hat{\lambda}_m/2] = n\hat{g}_m(\hat{\beta}_m)'(2\dot{\Omega}_m^{-1} - \dot{\Omega}_m^{-1} \ddot{\Omega}_m \dot{\Omega}_m^{-1}]\hat{g}_m(\hat{\beta}_m)$$

with $\ddot{\Omega}_m = -\sum_{i=1}^{n} \rho_2(\ddot{\lambda}_m' \hat{g}_{mi}) \hat{g}_{mi} \hat{g}_{ki}'/n$ and $\ddot{\lambda}_m$ lies between $\hat{\lambda}_m$ and 0. It remains to prove that

$$\frac{2n\hat{P}_{\rho}^{g_m}(\hat{\beta}_m, \hat{\lambda}_m) - \mathcal{T}_{GMM}^{g_m}}{\sqrt{2J_a MK}} = n\hat{g}_m(\hat{\beta}_m)'[2\dot{\Omega}_m^{-1} - \dot{\Omega}_m^{-1}\ddot{\Omega}_m\dot{\Omega}_m^{-1} - \hat{\Omega}_m^{-1}]\hat{g}_m(\hat{\beta}_m)/\sqrt{2J_a MK} \xrightarrow{p} 0.$$

By DIN Lemma A.6, p.78, $\left\|\hat{\Omega}_m - \Omega_m\right\| = O_p(\zeta(K)\sqrt{K/n}) = o_p(1/\sqrt{K})$. Similarly, by Lemma A.1, $\left\|\dot{\Omega}_m - \Omega_m\right\| = o_p(1/\sqrt{K})$ and $\left\|\ddot{\Omega}_m - \Omega_m\right\| = o_p(1/\sqrt{K})$. Hence $\left\|2\dot{\Omega}_m - \ddot{\Omega}_m - \Omega_m\right\| \xrightarrow{p} 0$. Consequently $\lambda_{\max}((2\dot{\Omega}_m - \ddot{\Omega}_m)^{-1}) \leq C$ w.p.a.1. Thus, by T, as $(2\dot{\Omega}_m^{-1} - \dot{\Omega}_m^{-1}\ddot{\Omega}_m\dot{\Omega}_m^{-1})^{-1} = \dot{\Omega}_m(2\dot{\Omega}_m - \ddot{\Omega}_m)^{-1}\dot{\Omega}_m$,

$$\begin{aligned} \left\|\dot{\Omega}_m(2\dot{\Omega}_m - \ddot{\Omega}_m)^{-1}\dot{\Omega}_m - \Omega_m(2\dot{\Omega}_m - \ddot{\Omega}_m)^{-1}\Omega\right\| &\leq \left\|(\dot{\Omega}_m - \Omega_m)(2\dot{\Omega}_m - \ddot{\Omega}_m)^{-1}(\dot{\Omega}_m - \Omega_m)\right\| \\ &\quad +2\left\|\Omega_m(2\dot{\Omega}_m - \ddot{\Omega}_m)^{-1}(\dot{\Omega}_m - \Omega_m)\right\| \\ &\leq C(\left\|\dot{\Omega}_m - \Omega_m\right\|^2 + \left\|\dot{\Omega}_m - \Omega_m\right\|) = o_p(1/\sqrt{K}). \end{aligned}$$

Also as $\lambda_{\max}(\Omega_m) \leq C$

$$\begin{aligned} \left\|\Omega_m(2\dot{\Omega}_m - \ddot{\Omega}_m)^{-1}\Omega_m - \Omega_m\right\| &= \left\|\Omega_m(2\dot{\Omega}_m - \ddot{\Omega}_m)^{-1}(\Omega_m - (2\dot{\Omega}_m - \ddot{\Omega}_m))\right\| \\ &\leq C\left\|\Omega_m - (2\dot{\Omega}_m - \ddot{\Omega}_m)\right\| = o_p(1/\sqrt{K}) \end{aligned}$$

yielding $\left\|\dot{\Omega}_m^{-1}(2\dot{\Omega}_m - \ddot{\Omega}_m)\dot{\Omega}_m^{-1} - \Omega_m^{-1}\right\| = o_p(1/\sqrt{K})$. Therefore, as $\left\|\hat{\Omega}_m^{-1} - \Omega_m^{-1}\right\| = o_p(1/\sqrt{K})$,

$$\frac{2n\hat{P}_{\rho}^{g_m}(\hat{\beta}_m, \hat{\lambda}_m) - \mathcal{T}_{GMM}^{g_m}}{\sqrt{2J_a MK}} = nO_p(K/n)o_p(1/\sqrt{K})/\sqrt{2JMK} = o_p(1).$$

The same reasoning for the second term in (A.3) yields

$$\frac{2n\hat{P}^{g}_{\rho}(\hat{\beta},\hat{\lambda})-\mathcal{T}^{g}_{GMM}}{\sqrt{2J_aMK}}\overset{p}{\to}0.$$

Therefore, from Theorem 4.1 it follows that

$$\mathcal{LR}^r\overset{d}{\to}N(0,1).$$

Now consider the Lagrange multiplier statistic (3.8)

$$\mathcal{LM}^r=\frac{n(\hat{\lambda}-S_m'\hat{\lambda}_m)'\hat{\Omega}(\hat{\lambda}-S_m'\hat{\lambda}_m)-J_aMK}{\sqrt{2J_aMK}}.$$

Write $\hat{g}_i=\hat{g}_i(\hat{\beta})$, $(i=1,...,n)$, $\hat{g}=\hat{g}(\hat{\beta})$ and $\hat{g}_0=\hat{g}(\beta_0)$. By a similar argument to that establishing (A.5), w.p.a.1

$$\hat{\lambda}=-\dot{\Omega}^{-1}\hat{g}$$

where $\dot{\Omega}=-\sum_{i=1}^{n}\rho_1(\dot{\lambda}'\hat{g}_i)\hat{g}_i\hat{g}_i'/n$ and $\dot{\lambda}$ lies between $\hat{\lambda}$ and 0.

Now $S_m\hat{g}=\hat{g}_m$ and $S_m\Omega S_m'=\Omega_m$. Thus, $S_m'\hat{\lambda}_m=-S_m'\dot{\Omega}_m^{-1}\hat{g}_m=-S_m'\dot{\Omega}_m^{-1}S_m\hat{g}$ and

$$\begin{aligned}n(\hat{\lambda}-S_m'\hat{\lambda}_m)'\hat{\Omega}(\hat{\lambda}-S_m'\hat{\lambda}_m) &= n\hat{\lambda}'\hat{\Omega}\hat{\lambda}-2n\hat{\lambda}'\hat{\Omega}S_m'\hat{\lambda}_m+n\hat{\lambda}_m'S_m\hat{\Omega}S_m'\hat{\lambda}_m\\ &= n\hat{g}'\dot{\Omega}^{-1}\hat{\Omega}\dot{\Omega}^{-1}\hat{g}-2n\hat{g}'\dot{\Omega}^{-1}\hat{\Omega}S_m'\dot{\Omega}_m^{-1}S_m\hat{g}+n\hat{g}'S_m'\dot{\Omega}_m^{-1}S_m\hat{\Omega}S_m'\dot{\Omega}_m^{-1}S_m\hat{g}.\end{aligned}$$

Therefore

$$\mathcal{LM}^r-\frac{\mathcal{T}^{g}_{GMM}-\mathcal{T}^{g_m}_{GMM}-J_aMK}{\sqrt{2J_aMK}}=\frac{n\hat{g}'(\dot{\Omega}^{-1}\hat{\Omega}\dot{\Omega}^{-1}-\hat{\Omega}^{-1})\hat{g}}{\sqrt{2J_aMK}} \tag{A.6}$$
$$+\frac{n\hat{g}'(S_m'\hat{\Omega}_m^{-1}S_m-2\dot{\Omega}^{-1}\hat{\Omega}S_m'\dot{\Omega}_m^{-1}S_m+S_m'\dot{\Omega}_m^{-1}S_m\hat{\Omega}\ S_m'\dot{\Omega}_m^{-1}S_m)\hat{g}}{\sqrt{2J_aMK}}.$$

Each term in (A.6) is $o_p(1)$. By CS, the first term

$$\begin{aligned}n\hat{g}'(\dot{\Omega}^{-1}\hat{\Omega}\dot{\Omega}^{-1}-\hat{\Omega}^{-1})\hat{g}/\sqrt{K} &= n\hat{g}'\dot{\Omega}^{-1}(\hat{\Omega}-\dot{\Omega}\hat{\Omega}^{-1}\dot{\Omega})\dot{\Omega}^{-1}\hat{g}/\sqrt{K}\\ &\le n\left\|\hat{\lambda}\right\|^2\left\|\hat{\Omega}-\dot{\Omega}\hat{\Omega}^{-1}\dot{\Omega}\right\|/\sqrt{K}.\end{aligned}$$

By DIN Lemma A.6, p.78, $\left\|\hat{\Omega}-\Omega\right\|=O_p(\zeta(K)\sqrt{K/n})=o_p(1/\sqrt{K})$. Thus $\lambda_{\max}(\hat{\Omega}^{-1})\le C$. Moreover

$$\begin{aligned}\left\|\dot{\Omega}\hat{\Omega}^{-1}\dot{\Omega}-\Omega\hat{\Omega}^{-1}\Omega\right\| &\le \left\|(\dot{\Omega}-\Omega)\hat{\Omega}^{-1}(\dot{\Omega}-\Omega)\right\|+\left\|2\Omega\hat{\Omega}^{-1}(\dot{\Omega}-\Omega)\right\|\\ &\le C(\left\|\dot{\Omega}-\Omega\right\|^2+\left\|\dot{\Omega}-\Omega\right\|)\\ &= O_p(\zeta(K)\sqrt{K/n})=o_p(1/\sqrt{K}).\end{aligned}$$

using DIN Lemma A.16, p.85. In addition, from CS and DIN Lemma A.6, p.78,

$$\begin{aligned}\left\|\Omega\hat{\Omega}^{-1}\Omega-\Omega\right\| &= \left\|\Omega\hat{\Omega}^{-1}(\hat{\Omega}-\Omega)\right\|\\ &\le \left\|\Omega\hat{\Omega}^{-1}\right\|\left\|\hat{\Omega}-\Omega\right\|=o_p(1/\sqrt{K}).\end{aligned}$$

Therefore, by T $\left\|\hat{\Omega}-\dot{\Omega}\hat{\Omega}^{-1}\dot{\Omega}\right\| = o_p(1/\sqrt{K})$. As $\left\|\hat{\lambda}\right\| = O_p(\sqrt{K/n})$ by DIN Lemma A.11, p.82, $n\hat{g}'(\dot{\Omega}^{-1}\hat{\Omega}\dot{\Omega}^{-1}-\hat{\Omega}^{-1})\hat{g}/\sqrt{K} = nO_p(K/n)o_p(1/\sqrt{K})/\sqrt{K} = o_p(1)$.

For the second term, by CS

$$\begin{aligned} & n\hat{g}'(S_m'\hat{\Omega}_m^{-1}S_m - 2\dot{\Omega}^{-1}\hat{\Omega}S_m'\dot{\Omega}_m^{-1}S_m + S_m'\dot{\Omega}_m^{-1}S_m\hat{\Omega}\ S_m'\dot{\Omega}_m^{-1}S_m)\hat{g}/\sqrt{K} \\ \leq\ & n\left\|\hat{g}\right\|^2\left\|S_m'\hat{\Omega}_m^{-1}S_m - 2\dot{\Omega}^{-1}\hat{\Omega}S_m'\dot{\Omega}_m^{-1}S_m + S_m'\dot{\Omega}_m^{-1}S_m\hat{\Omega}\ S_m'\dot{\Omega}_m^{-1}S_m\right\|/\sqrt{K}. \end{aligned}$$

Now by T and DIN Lemma A.6, p.78, since $\lambda_{\max}(\dot{\Omega}^{-1}) \leq C$ and $\lambda_{\max}(\hat{\Omega}^{-1}) \leq C$

$$\begin{aligned} \left\|S_m'\hat{\Omega}_m^{-1}S_m - \dot{\Omega}^{-1}\hat{\Omega}S_m'\dot{\Omega}_m^{-1}S_m\right\| &\leq \left\|S_m'\dot{\Omega}_m^{-1}(\dot{\Omega}_m-\hat{\Omega}_m)\hat{\Omega}_m^{-1}S_m\right\| + \left\|\dot{\Omega}^{-1}(\hat{\Omega}-\dot{\Omega})S_m'\dot{\Omega}_m^{-1}S_m\right\| \\ &= o_p(1/\sqrt{K}). \end{aligned}$$

Finally, since $\hat{\Omega}_m = S_m\hat{\Omega}S_m'$, by a similar argument

$$\begin{aligned} \left\|\dot{\Omega}^{-1}\hat{\Omega}\ S_m'\dot{\Omega}_m^{-1}S_m - S_m'\dot{\Omega}_m^{-1}S_m\hat{\Omega}S_m'\dot{\Omega}_m^{-1}S_m\right\| &\leq \left\|S_m'\dot{\Omega}_m^{-1}(\dot{\Omega}_m-\hat{\Omega}_m)\dot{\Omega}_m^{-1}S_m\right\| \\ &\quad + \left\|\dot{\Omega}^{-1}(\hat{\Omega}-\dot{\Omega})S_m'\dot{\Omega}_m^{-1}S_m\right\| \\ &= o_p(1/\sqrt{K}). \end{aligned}$$

Therefore, as $\|\hat{g}\| = O_p(\sqrt{K/n})$ by DIN Lemma A.14, p.84,

$$\begin{aligned} n\hat{g}'(S_m'\hat{\Omega}_m^{-1}S_m - 2\dot{\Omega}^{-1}\hat{\Omega}S_m'\dot{\Omega}_m^{-1}S_m + S_m'\dot{\Omega}_m^{-1}S_m\hat{\Omega}S_m'\dot{\Omega}^{-1}S_m)\hat{g}/\sqrt{K} &= nO_p(K/n)o_p(1/\sqrt{K})/\sqrt{K} \\ &= o_p(1). \end{aligned}$$

Let $\hat{g}_{ai} = g_{ai}(\hat{\beta}_a)$, $(i = 1, ..., n)$. The score test statistic

$$\mathcal{S}^r = \frac{\sum_{i=1}^n \rho_1(\hat{\lambda}_m'\hat{g}_{mi})\hat{g}_{ai}'S_a\hat{\Omega}^{-1}S_a'\sum_{i=1}^n \rho_1(\hat{\lambda}_m'\hat{g}_{mi})\hat{g}_{ai}/n - J_aMK}{\sqrt{2J_aMK}}.$$

An expansion of the first order conditions $\sum_{i=1}^n \rho_1(\hat{\lambda}'\hat{g}_i)\hat{g}_i/n = 0$ of (3.7) around $S_m'\hat{\lambda}_m$ gives

$$\sum\nolimits_{i=1}^n \rho_1(\hat{\lambda}_m'\hat{g}_{mi})\hat{g}_i/n - \dot{\Omega}(\hat{\lambda} - S_m'\hat{\lambda}_m) = 0 \tag{A.7}$$

w.p.a.1 where $\dot{\Omega} = -\sum_{i=1}^n \rho_2(\dot{\lambda}'\hat{g}_i)\hat{g}_i\hat{g}_i'/n$ and $\dot{\lambda}$ lies between $\hat{\lambda}$ and $S_m'\hat{\lambda}_m$. Since $\sum_{i=1}^n \rho_1(\hat{\lambda}_m'\hat{g}_{mi})\hat{g}_i/n = S_a'\sum_{i=1}^n \rho_1(\hat{\lambda}_m'\hat{g}_{mi})\hat{g}_{ai}/n$,

$$\sum\nolimits_{i=1}^n \rho_1(\hat{\lambda}_m'\hat{g}_{mi})\hat{g}_{ai}'S_a\hat{\Omega}^{-1}S_a'\sum\nolimits_{i=1}^n \rho_1(\hat{\lambda}_m'\hat{g}_{mi})\hat{g}_{ai}/n = n(\hat{\lambda} - S_m'\hat{\lambda}_m)'\dot{\Omega}\hat{\Omega}^{-1}\dot{\Omega}(\hat{\lambda} - S_m'\hat{\lambda}_m).$$

Thus by CS and T

$$\begin{aligned} |\mathcal{S}^r - \mathcal{LM}^r| &= n\left|(\hat{\lambda} - S_m'\hat{\lambda}_m)'(\dot{\Omega}\hat{\Omega}^{-1}\dot{\Omega} - \hat{\Omega})(\hat{\lambda} - S_m'\hat{\lambda}_m)\right|/\sqrt{2J_aMK} \\ &\leq n\left\|\dot{\Omega}\hat{\Omega}^{-1}\dot{\Omega} - \hat{\Omega}\right\|(\left\|\hat{\lambda}\right\| + \left\|\hat{\lambda}_m\right\|)^2/\sqrt{2J_aMK} = o_p(1) \end{aligned}$$

as $\dot{\Omega}\hat{\Omega}^{-1}\dot{\Omega} - \hat{\Omega} = o_p(1/\sqrt{K})$ and $\left\|\hat{\lambda}\right\|$, $\left\|\hat{\lambda}_m\right\|$ are both $O_p(\sqrt{K/n})$ by DIN Lemma A.11, p.82.

For the Wald test statistic, from (A.7), w.p.a.1 $\hat{\lambda} - S_m'\hat{\lambda}_m = \dot{\Omega}^{-1}S_a'\sum_{i=1}^n \rho_1(\hat{\lambda}_m'\hat{g}_{mi})\hat{g}_{ai}/n$. Thus

$$S_a\hat{\lambda} = S_a\dot{\Omega}^{-1}S_a'\sum_{i=1}^n \rho_1(\hat{\lambda}_m'\hat{g}_{mi})\hat{g}_{ai}/n.$$

Therefore, w.p.a.1

$$\begin{aligned} |\mathcal{W}^r - \mathcal{S}^r| &= n\left|\hat{\lambda}'S_a'((S_a\hat{\Omega}^{-1}S_a')^{-1} - (S_a\dot{\Omega}^{-1}S_a')^{-1}S_a\hat{\Omega}^{-1}S_a'(S_a\dot{\Omega}^{-1}S_a')^{-1})S_a\hat{\lambda}\right|/\sqrt{2J_aMK} \\ &\leq n\left\|S_a\hat{\lambda}\right\|^2\left\|(S_a\hat{\Omega}^{-1}S_a')^{-1} - (S_a\dot{\Omega}^{-1}S_a')^{-1}S_a\hat{\Omega}^{-1}S_a'(S_a\dot{\Omega}^{-1}S_a')^{-1}\right\|/\sqrt{2J_aMK}. \end{aligned}$$

Now $(S_a\hat{\Omega}^{-1}S_a')^{-1} - (S_a\dot{\Omega}^{-1}S_a')^{-1}S_a\hat{\Omega}^{-1}S_a'(S_a\dot{\Omega}^{-1}S_a')^{-1} = o_p(1/\sqrt{K})$, cf. $\left\|\hat{\Omega} - \dot{\Omega}\hat{\Omega}^{-1}\dot{\Omega}\right\| = o_p(1/\sqrt{K})$ above. Therefore, since $\left\|S_a\hat{\lambda}\right\| = O_p(\sqrt{K/n})$ by DIN Lemma A.11, p.82, $|\mathcal{W}^r - \mathcal{S}^r| = o_p(1)$. ■

Proof of Theorem 4.3. The proof uses the Cramér-Wold device. Consider the linear combination

$$\mathcal{J}^c = \alpha_r\mathcal{J}^r + \alpha_m\mathcal{J}^m.$$

where α_r and α_m are arbitrary finite scalars such that $\alpha_r^2 + \alpha_m^2 > 0$. The desired result obtains if $\mathcal{J}^c \xrightarrow{d} N(0, \alpha_r^2 + \alpha_m^2)$.

First, by DIN Lemma 6.1, p.69,

$$\mathcal{J}^r - \frac{n\hat{g}(\beta_0)'\Omega^{-1}\hat{g}(\beta_0) - n\hat{g}_m(\beta_{m0})'\Omega_m^{-1}\hat{g}_m(\beta_{m0}) - J_aMK}{\sqrt{2J_aMK}} \xrightarrow{p} 0.$$

Likewise

$$\mathcal{J}^m - \frac{n\hat{g}_m(\beta_{m0})'\Omega_m^{-1}\hat{g}_m(\beta_{m0}) - J_mK}{\sqrt{2J_mK}} \xrightarrow{p} 0.$$

Therefore,

$$\mathcal{J}^c - \frac{1}{\sqrt{J_aM}}\frac{n\hat{g}(\beta_0)'Q\hat{g}(\beta_0) - (\alpha_rJ_aM + \alpha_mJ_m\sqrt{J_aM/J_m})K}{\sqrt{2K}} \xrightarrow{p} 0,$$

where $Q = \alpha_r\Omega^{-1} - (\alpha_r - \alpha_m\sqrt{J_aM/J_m})S_m\Omega_m^{-1}S_m'$.

To prove $\sqrt{J_aM}\mathcal{J}^c \xrightarrow{d} N(0, v)$, where $v = (\alpha_r^2 + \alpha_m^2)J_aM$, the conditions Lemma A.3(a)-(f) are verified below.

Condition (a): immediate.

Condition (b):

$$\begin{aligned} tr(Q\Omega) &= \alpha_r tr(I_{(J_m+J_aM)K}) - (\alpha_r - \alpha_m\sqrt{J_aM/J_m})tr(I_{J_mK}) \\ &= \alpha_r(J_m + J_aM)K - (\alpha_r - \alpha_m\sqrt{J_aM/J_m})J_mK \\ &= \alpha_r(J_aM + \alpha_mJ_m\sqrt{J_aM/J_m})K = aK. \end{aligned}$$

Condition (c): note that

$$\begin{aligned} (Q\Omega)^2 &= (\alpha_rI_{(J_m+J_aM)K} - (\alpha_r - \alpha_m\sqrt{J_aM/J_m})S_m\Omega_m^{-1}S_m'\Omega)^2 \\ &= \alpha_r^2I_{(J_m+J_aM)K} - (\alpha_r^2 - \alpha_m^2(J_aM/J_m))S_m\Omega_m^{-1}S_m'\Omega. \end{aligned}$$

Hence

$$
\begin{aligned}
tr[(Q\Omega)^2] &= (\alpha_r^2+\alpha_m^2)J_aMK \\
&= vK.
\end{aligned}
$$

Condition (d):

$$
\begin{aligned}
(Q\Omega)^4 &= (\alpha_r^2 I_{(J_m+J_aM)K} - (\alpha_r^2-\alpha_m^2(J_aM/J_m))S_m\Omega_m^{-1}S_m'\Omega)^2 \\
&= \alpha_r^4 I_{(J_m+J_aM)K} - (\alpha_r^4-\alpha_m^4(J_aM/J_m)^2)S_m\Omega_m^{-1}S_m'\Omega.
\end{aligned}
$$

Thus

$$
\begin{aligned}
tr[(Q\Omega)^4] &= (\alpha_r^4+\alpha_m^4 J_aMJ_m)J_aMK \\
&= o(K^2).
\end{aligned}
$$

Condition (e): from DIN Lemma A.6, p.78, $1/C \leq \lambda_{\min}(\Xi) \leq \lambda_{\max}(\Xi) \leq C$ and $1/C \leq \lambda_{\min}(\Omega) \leq \lambda_{\max}(\Omega) \leq C$. Therefore, using Assumption 3.2

$$
E[(g(z,\beta_0)'(\alpha_r\Omega^{-1}-(\alpha_r-\alpha_m\sqrt{J_aM/J_m})S_m\Omega_m^{-1}S_m')g(z,\beta_0))^2] \leq C\zeta(K)^2K = o(nK)
$$

since $\zeta(K)^2K^2/n \to 0$.

Condition (f): by a similar reasoning to that for Condition (e)

$$
E[(g(z,\beta_0)'\Omega^{-1}g(z,\beta_0))^2] \leq C\zeta(K)^2K.
$$

Also

$$
\begin{aligned}
Q\Omega Q &= (\alpha_r\Omega^{-1}-(\alpha_r-\alpha_m\sqrt{J_aM/J_m})S_m\Omega_m^{-1}S_m')\Omega(\alpha_r\Omega^{-1}-(\alpha_r-\alpha_m\sqrt{J_aM/J_m})S_m\Omega_m^{-1}S_m') \\
&= \alpha_r^2(\Omega^{-1}-S_m\Omega_m^{-1}S_m') + \alpha_m^2(J_aM/J_m)S_m\Omega_m^{-1}S_m'.
\end{aligned}
$$

Thus, cf. Condition (e),

$$
E[(g(z,\beta_0)'Q\Omega Q g(z,\beta_0))^2] \leq C\zeta(K)^2K.
$$

■

A.3 Asymptotic Local Alternative Distribution

Let $u_i(\beta) = u(z_i,\beta)$, $u_{mi}(\beta_m) = S_m^u u_i(\beta) = u_m(z_i,\beta_m)$, $g_i(\beta) = S(u_i(\beta)\otimes q_i)$, $g_{mi}(\beta) = u_{mi}(\beta_m)\otimes q_{mi}$, where $q_i = q^K(s_i)$ and $q_{mi} = q_m^K(s_{mi})$, $\hat{g}_i = g_i(\hat{\beta})$, $\hat{g}_{mi} = g_{mi}(\hat{\beta}_m)$ and $g_{i,n} = g_i(\beta_{0,n})$, $g_{mi,n} = g_{mi}(\beta_{m0,n})$, $(i = 1,...,n)$. Also let $u_{i,n} = u_i(\beta_{0,n})$, $u_{mi,n} = u_{mi}(\beta_{m0,n})$, $\Sigma_i(\beta) = E[u_i(\beta)u_i(\beta)'|s_i]$,

$\Sigma_{mi}(\beta) = E[u_{mi}(\beta_m)u_{mi}(\beta_m)'|s_{mi}]$, $\Sigma_{i,n} = \Sigma_i(\beta_{0,n}) = E[u_{i,n}u_{i,n}'|s_i]$, $\Sigma_{mi,n} = \Sigma_{mi}(\beta_{m0,n}) = E[u_{mi,n}u_{mi,n}'|s_{mi}]$, $(i = 1, ..., n)$, together with

$$\begin{aligned}\hat{\Omega} &= \sum\nolimits_i \hat{g}_i\hat{g}_i'/n, \tilde{\Omega}_n = \sum\nolimits_i g_{i,n}g_{i,n}'/n,\\ \bar{\Omega}_n &= S(\sum\nolimits_i \Sigma_{i,n}\otimes q_iq_i')S'/n, \Omega_n = E[g_{i,n}g_{i,n}'].\end{aligned}$$

and

$$\begin{aligned}\hat{\Omega}_m &= \sum\nolimits_i \hat{g}_{mi}\hat{g}_{mi}'/n, \tilde{\Omega}_{mn} = \sum\nolimits_i g_{mi,n}g_{mi,n}'/n,\\ \bar{\Omega}_{mn} &= (\sum\nolimits_i \Sigma_{mi,n}\otimes q_{mi}q_{mi}')/n, \Omega_{mn} = E[g_{mi,n}g_{mi,n}'].\end{aligned}$$

Proof of Theorem 5.1. The result is established for the GMM statistic $\mathcal{J}^r$. Proofs for the restricted GEL statistics $\mathcal{LR}^r$, $\mathcal{LM}^r$, $\mathcal{S}^r$ and $\mathcal{W}^r$ essentially follow the same steps as in the Proof of Theorem 4.1 above but are omitted for brevity.

Let $\hat{g}_{mn} = \hat{g}_m(\beta_{mn,0})$ and $\hat{g}_n = \hat{g}(\beta_{n,0})$. Note $\Omega_{mn} = S_m\Omega_nS_m'$. Then, by Lemma A.8,

$$\frac{n\hat{g}(\hat{\beta})'\hat{\Omega}^{-1}\hat{g}(\hat{\beta}) - n\hat{g}_n'\Omega_n^{-1}\hat{g}_n}{\sqrt{2J_aMK}} \xrightarrow{p} 0, \frac{n\hat{g}_m(\hat{\beta}_m)'\hat{\Omega}_m^{-1}\hat{g}_m(\hat{\beta}_m) - n\hat{g}_{mn}'\Omega_{mn}^{-1}\hat{g}_{mn}}{\sqrt{2J_aMK}} \xrightarrow{p} 0.$$

Hence $\mathcal{J}^r - (n\hat{g}_n'(\Omega_n^{-1} - S_m'\Omega_{mn}^{-1}S_m)\hat{g}_n - J_aMK)/\sqrt{2J_aMK} \xrightarrow{p} 0$.

It remains to prove that

$$\frac{n\hat{g}_n'(\Omega_n^{-1} - S_m'\Omega_{mn}^{-1}S_m)\hat{g}_n - J_aMK}{\sqrt{2J_aMK}} \xrightarrow{d} N(\mu^r/\sqrt{2}, 1).$$

Let $\bar{g}_{i,n} = E[g_{i,n}|s_i]$ and $\tilde{g}_{i,n} = g_{i,n} - \bar{g}_{i,n}$, $(i = 1, ..., n)$. Also let $\bar{g}_n = \sum_{i=1}^n \bar{g}_{i,n}/n$ and $\tilde{g}_n = \sum_{i=1}^n \tilde{g}_{i,n}/n$. Write $P_n = \Omega_n^{-1} - S_m'\Omega_{mn}^{-1}S_m$. Then,

$$\hat{g}_n'P_n\hat{g}_n = \tilde{g}_n'P_n\tilde{g}_n + 2\bar{g}_n'P_n\tilde{g}_n + \bar{g}_n'P_n\bar{g}_n.$$

The first step demonstrates

$$\bar{g}_n'P_n\bar{g}_n = \frac{\sqrt{J_aMK}}{n}(\mu^r + o_p(1)).$$

Let $\xi_i = \xi(s_i)$ and $\xi_{mi} = \xi_m(s_i)$, $(i = 1, ..., n)$. It follows by Lemma A.4 that

$$\begin{aligned}\bar{g}_n'\bar{\Omega}_n^{-1}\bar{g}_n &= \frac{\sqrt{J_aMK}}{n}\sum\nolimits_{i,j=1}^n(\xi_i\otimes q_i)'S'\bar{\Omega}_n^{-1}S(\xi_j\otimes q_j)/n^2\\ &= \frac{\sqrt{J_aMK}}{n}(E[\xi(s)'\Sigma(s)^{-1}\xi(s)] + o_p(1)).\end{aligned}$$

Next, note $S_m(\xi_i\otimes q_i) = \xi_{mi}\otimes q_{mi}$, $(i = 1, ..., n)$. Thus, again using Lemma A.4,

$$\begin{aligned}\bar{g}_n'S_m'\bar{\Omega}_{mn}^{-1}S_m\bar{g}_n &= \frac{\sqrt{J_aMK}}{n}\sum\nolimits_{i,j=1}^n(\xi_{mi}\otimes q_{mi})'\bar{\Omega}_{mn}^{-1}(\xi_{mj}\otimes q_{mj})/n^2\\ &= \frac{\sqrt{J_aMK}}{n}(E[E[\xi_m(s)|s_m]'\Sigma_m(s_m)^{-1}E[\xi_m(s)|s_m]] + o_p(1))\\ &= \frac{\sqrt{J_aMK}}{n}o_p(1),\end{aligned}$$

since $E[\xi_m(s)|s_m] = 0$ by hypothesis. It remains to show that

$$\frac{n}{\sqrt{2J_a MK}}\bar{g}_n'(\Omega_n^{-1} - \bar{\Omega}_n^{-1})\bar{g}_n \xrightarrow{p} 0, \frac{n}{\sqrt{2J_a MK}}\bar{g}_n' S_m'(\Omega_{mn}^{-1} - \bar{\Omega}_{mn}^{-1})S_m\bar{g}_n \xrightarrow{p} 0.$$

Similarly to the Proof of DIN Lemma 6.1, pp.87-88, from Lemma A.6,

$$\begin{aligned}
\left|n\bar{g}_n'(\Omega_n^{-1} - \bar{\Omega}_n^{-1})\bar{g}_n\right|/\sqrt{2J_a MK} &\leq n\left\|\Omega_n^{-1}\bar{g}_n\right\|^2(\left\|\Omega_n - \bar{\Omega}_n\right\| + C\left\|\Omega_n - \bar{\Omega}_n\right\|^2)/\sqrt{2J_a MK} \\
&= n\left\|\Omega_n^{-1}\bar{g}_n\right\|^2 O_p(\zeta(K)\sqrt{K/n})/\sqrt{2J_a MK} = o_p(1)
\end{aligned}$$

since $\left\|\Omega_n^{-1}\bar{g}_n\right\|^2 = \bar{g}_n'\Omega_n^{-2}\bar{g}_n \leq C\bar{g}_n'\Omega_n^{-1}\bar{g}_n = O_p(\sqrt{K}/n)$. Likewise $\left|n\bar{g}_n' S_m'(\Omega_{mn}^{-1} - \bar{\Omega}_{mn}^{-1})S_m\bar{g}_n\right|/\sqrt{2J_a MK} = o_p(1)$. Therefore,

$$\bar{g}_n' P_n \bar{g}_n = \frac{\sqrt{J_a MK}}{n}(\mu^r + o_p(1)).$$

Secondly, it is shown that

$$n\bar{g}_n' P_n \tilde{g}_n/\sqrt{2J_a MK} = o_p(1).$$

Noting $\|\xi_i\|^2$ bounded and $\Sigma_{i,n}(s_i)^{-1}$ bounded for n large enough, by c_r

$$\begin{aligned}
E[\|u_{i,n} - E[u_{i,n}|s_i]\|^4] &\leq 8(E[\|u_{i,n}\|^4] + E[\|E[u_{i,n}|s_i]\|^4]) \\
&= 8(E[E[\|u_{i,n}\|^4|s_i]) + E[\frac{J_a MK}{n^2}\|\xi_i\|^4]] \\
&\leq C
\end{aligned}$$

for n large enough as $E[\|u_{i,n}\|^4|s_i] \leq C$ and $K/n^2 \to 0$. Hence, by Lemma A.5,

$$\begin{aligned}
\bar{g}_n'\bar{\Omega}_n^{-1}\tilde{g}_n &= \frac{\sqrt[4]{J_a MK}}{n}\sum\nolimits_{i,j=1}^{n}(\xi_i \otimes q_i)' S' \bar{\Omega}_n^{-1}\tilde{g}_{j,n}/n\sqrt{n} \\
&= O_p(\sqrt[4]{J_a MK}/n).
\end{aligned}$$

Next, by hypothesis,

$$\begin{aligned}
\left|n\bar{g}_n'(\Omega_n^{-1} - \bar{\Omega}_n^{-1})\tilde{g}_n\right|/\sqrt{2J_a MK} &\leq n\left\|\Omega_n^{-1}\bar{g}_n\right\|\left\|\Omega_n^{-1}\tilde{g}_n\right\|(\left\|\Omega_n - \bar{\Omega}_n\right\| + C\left\|\Omega_n - \bar{\Omega}_n\right\|^2)/\sqrt{2J_a MK} \\
&= n\left\|\Omega_n^{-1}\bar{g}_n\right\|\left\|\Omega_n^{-1}\tilde{g}_n\right\| O_p(\zeta(K)\sqrt{K/n})/\sqrt{2J_a MK} = o_p(1)
\end{aligned}$$

since $\left\|\Omega_n^{-1}\bar{g}_n\right\|^2 = O_p(\sqrt{K}/n)$ from above and $\left\|\Omega_n^{-1}\tilde{g}_n\right\| \leq \left\|\Omega_n^{-1}\hat{g}_n\right\| + \left\|\Omega_n^{-1}\bar{g}_n\right\| = O_p(\sqrt{K/n}) + O_p(\sqrt[4]{K/n^2})$. A similar analysis yields $n\bar{g}_n' S_m'\Omega_{mn}^{-1}S_m\tilde{g}_n/\sqrt{2J_a MK} = o_p(1)$.

Let $\bar{g}_{mi,n} = E[g_{mi,n}|s_i]$, $\tilde{g}_{mi,n} = g_{mi,n} - \bar{g}_{mi,n}$, $(i = 1, ..., n)$.

Finally to prove

$$\frac{n\tilde{g}_n' P_n \tilde{g}_n - J_a MK}{\sqrt{2J_a MK}} \xrightarrow{d} N(0, 1)$$

it is first established that

$$\frac{n\tilde{g}_n'(P_n - P_n^*)\tilde{g}_n}{\sqrt{2J_a MK}} = o_p(1)$$

where $P_n^* = \Omega_n^{*-1} - S_m'(\Omega_{mn}^*)^{-1}S_m$ with $\Omega_n^* = E[\tilde{g}_{i,n}\tilde{g}_{i,n}']$ and $\Omega_{mn}^* = E[\tilde{g}_{mi,n}\tilde{g}_{mi,n}']$. By T

$$|n\tilde{g}_n'(P_n - P_n^*)\tilde{g}_n| \le \left|n\tilde{g}_n'(\Omega_n^{-1} - (\Omega_n^*)^{-1})\tilde{g}_n'\right| + \left|n\tilde{g}_n'(S_m'\Omega_{mn}^{-1}S_m - S_m'(\Omega_{mn}^*)^{-1}S_m)\tilde{g}_n'\right|$$

The first term

$$\left|n\tilde{g}_n'(\Omega_n^{-1} - (\Omega_n^*)^{-1})\tilde{g}_n'\right| \le n\left\|\Omega_n^{-1}\tilde{g}_n\right\|^2(\|\Omega_n - \Omega_n^*\| + C\|\Omega_n - \Omega_n^*\|^2).$$

Therefore, noting $\Omega_n^* = \Omega_n - E[\bar{g}_{i,n}\bar{g}_{i,n}']$, from eq.(5.1)

$$\begin{aligned}\|\Omega_n - \Omega_n^*\| &= \frac{\sqrt[4]{J_aMK}}{\sqrt{n}}E[\|\xi_i\|^2\|q_i\|^2]^{1/2} \\ &= O_p(\frac{\sqrt[4]{K^3}}{\sqrt{n}}).\end{aligned}$$

Consequently, since $\left\|\Omega_n^{-1}\tilde{g}_n\right\| = O_p(\sqrt{K/n}) + O_p(\sqrt[4]{K/n^2})$,

$$\frac{\left|n\tilde{g}_n'(\Omega_n^{-1} - \Omega_n^{*-1})\tilde{g}_n'\right|}{\sqrt{2J_aMK}} \le \frac{O_p(K) + O_p(\sqrt{K})}{\sqrt{2J_aMK}}(O(\frac{\sqrt[4]{K^3}}{\sqrt{n}}) + \frac{\sqrt{K^3}}{n}) = o_p(1).$$

Similarly

$$\left|\frac{n\tilde{g}_n'S_m'(\Omega_{mn}^{-1} - (\Omega_{mn}^*)^{-1})S_m\tilde{g}_n'}{\sqrt{2J_aMK}}\right| = o_p(1).$$

Therefore

$$\frac{n\tilde{g}_n'(P_n - P_n^*)\tilde{g}_n}{\sqrt{2J_aMK}} = o_p(1)$$

Note that $1/C \le \lambda_{\min}(\Omega_n^*) \le \lambda_{\max}(\Omega_n^*) \le C$ because $|\lambda(A) - \lambda(B)| \le \|A - B\|$, $|\lambda_{\min}(\Omega_n^*) - \lambda_{\min}(\Omega_n)| = o(1)$ and $|\lambda_{\max}(\Omega_n^*) - \lambda_{\max}(\Omega_n)| = o(1)$. Similarly $1/C \le \lambda_{\min}(\Omega_{mn}^*) \le \lambda_{\max}(\Omega_{mn}^*) \le C$.

Lemma A.2 is now invoked to prove

$$\frac{n\tilde{g}_n'P_n^*\tilde{g}_n - J_aMK}{\sqrt{2J_aMK}} \xrightarrow{d} N(0,1).$$

First, $tr(\Omega_n^*P_n^*) = J_aMK$. Secondly, to establish

$$E[(\tilde{g}_{i,n}'P_n^*\tilde{g}_{i,n})^2] = o_p(K\sqrt{n}),$$

by c_r

$$E[(\tilde{g}_{i,n}'P_n^*\tilde{g}_{i,n})^2] \le 2E[(\tilde{g}_{i,n}'(\Omega_n^*)^{-1}\tilde{g}_{i,n})^2] + 2E[(\tilde{g}_{i,n}'S_m'(\Omega_{mn}^*)^{-1}S_m\tilde{g}_{i,n})^2].$$

Again using c_r

$$E[(\tilde{g}_{i,n}'(\Omega_n^*)^{-1}\tilde{g}_{i,n})^2] \le 3E[(g_{i,n}'(\Omega_n^*)^{-1}g_{i,n})^2] + 12E[(g_{i,n}'(\Omega_n^*)^{-1}\bar{g}_{i,n})^2] + 3E[(\bar{g}_{i,n}'(\Omega_n^*)^{-1}\bar{g}_{i,n})^2].$$

Now, for n large enough, $E[(g_{i,n}'(\Omega_n^*)^{-1}g_{i,n})^2] \le CE[\|g_{i,n}\|^4]$. Since $\beta_{n,0} \in \mathcal{N}$ for n large enough, by Assumption 3.4(c), similarly to the Proof of DIN Theorem 6.3, pp.89-90,

$$E[\|g_{i,n}\|^4] \le E[\|q_i\|^4 E[\|u_{i,n}\|^4 |s_i]] \le CE[\|q_i\|^4] \le C\zeta(K)^2K.$$

Next,

$$E[(g'_{i,n}(\Omega_n^*)^{-1}\bar{g}_{i,n})^2] \leq C(\sqrt{K}/n)E[\|\xi_i\|^2 \|q_i\|^2] \leq CK\sqrt{K}/n.$$

Lastly,

$$E[(\bar{g}'_{i,n}(\Omega_n^*)^{-1}\bar{g}_{i,n})^2] \leq C(K/n^2)E[\|\xi_i\|^4 \|q_i\|^4] \leq C\zeta(K)^2K^2/n^2.$$

Hence, $E[(\tilde{g}'_{i,n}(\Omega_n^*)^{-1}\tilde{g}_{i,n})^2] = o_p(K\sqrt{n})$ as required. Likewise, $E[(\tilde{g}'_{i,n}S'_m(\Omega_{mn}^*)^{-1}S_m\tilde{g}_{i,n})^2] = o_p(K\sqrt{n})$. Thirdly, $P_n^*\Omega_n^*P_n^* = P_n^*$. Therefore,

$$\frac{n\tilde{g}'_n P_n \tilde{g}_n - J_a MK}{\sqrt{2J_a MK}} \xrightarrow{d} N(0,1).$$

The conclusion of the theorem then follows. ■

References

Abramowitz, M., and I.A. Stegun, I. A. (1972): *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* Applied Mathematics Series 55, 10th. Edition. New York, NY: Dover Publications.

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

Banks, J., Blundell, R.W., and A. Lewbel (1997): "Quadratic Engel Curves, Indirect Tax Reform and Welfare Measurement," *Review of Economics and Statistics*, 79, 527–539.

Belloni, A., Chernozhukov, V., Chetverikov, D., and Kato, K. (2015): "Some New Asymptotic Theory for Least Squares Series: Pointwise and Uniform Results." Forthcoming in *Journal of Econometrics.*

Bierens, H.J. (1982): "Consistent Model Specification Tests," *Journal of Econometrics*, 20, 105-134.

Bierens, H.J. (1990): "A Consistent Conditional Moment Test of Functional Form," *Econometrica*, 58, 1443-1458.

Blundell, R.W., and J.L. Horowitz (2007): "A Non-Parametric Test of Exogeneity," *Review of Economic Studies*, 74, 1035-1058.

Carrasco, M., and J.-P. Florens (2000): "Generalization of GMM to a Continuum of Moment Conditions," *Econometric Theory*, 16, 797–834.

Chamberlain, G. (1987): "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," *Journal of Econometrics*, 34, 305-334.

Chen, J., Variyath, A.M., and B. Abraham (2008): "Adjusted Empirical Likelihood and Its Properties," *Journal of Computational and Graphical Statistics*, 17, 426-443.

Chen, X., and D. Pouzo (2009): "Efficient Estimation of Semiparametric Conditional Moment Models with Possibly Nonsmooth Residuals," *Journal of Econometrics*, 152, 46-60.

Chen, X., and D. Pouzo (2012): "Estimation of Nonparametric Conditional Moment Models with Possibly Nonsmooth Generalized Residuals," *Econometrica*, 80, 277-321.

Cragg, J.G. (1983): "More Efficient Estimation in the Presence of Heteroscedasticity of Unknown Form," *Econometrica*, 51, 751-763.

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

Davidson, J.E.H. (1994): *Stochastic Limit Theory.* Oxford: Oxford University Press.

Davidson, R., and J.G. MacKinnon (1993): *Estimation and Inference in Econometrics.* Oxford: Oxford University Press.

De Jong, R.M., and H.J. Bierens. (1994): "On the Limit Behavior of a Chi-Square Type Test if the Number of Conditional Moments Tested Approaches Infinity," *Econometric Theory*, 10, 70-90.

Domínguez, M.A., and I.N. Lobato (2004): "Consistent Estimation of Models Defined by Conditional Moment Restrictions," *Econometrica*, 72, 1601-1615.

Donald, S.G., Imbens, G.W., and W.K. Newey (2003): "Empirical Likelihood Estimation and Consistent Tests with Conditional Moment Restrictions," *Journal of Econometrics*, 117, 55-93.

Donald, S.G., Imbens, G.W., and W.K. Newey (2009): "Choosing Instrumental Variables in Conditional Moment Restriction Models," *Journal of Econometrics*, 152, 28-36.

Durbin, J. (1954): "Errors in variables," *Review of the International Statistical Institute*, 22, 23-32.

Eichenbaum, M.S., Hansen, L.P., and K.J. Singleton (1988): "A Time Series Analysis of Representative Agent Models of Consumption and Leisure Choice under Uncertainty," *Quarterly Journal of Economics*, 103, 51-78.

Ellison, G., and S.F. Ellison (2000): "A Simple Framework for Nonparametric Testing," *Journal of Econometrics*, 96, 1-23.

Engle, R.F. (1982): "A General Approach to Lagrange Multiplier Model Diagnostics," *Journal of Econometrics*, 20, 83-104.

Engle, R.F., Hendry, D.F., and J.-F. Richard (1983): "Exogeneity," *Econometrica*, 51, 277-304.

Eubank, R., and C. Spiegelman (1990): "Testing the Goodness of Fit of a Linear Model via Nonparametric Regression Techniques," *Journal of the American Statistical Association*, 85, 387-392.

Fan, Y., and Q. Li (1996): "Consistent Model Specification Tests: Omitted Variables and Semi-Parametric Functional Forms," *Econometrica*, 64, 865-890.

Feller, W. (1971): *An Introduction to Probability Theory and Its Applications*, Volume II, 2nd. Edition. New York: Wiley.

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

Hall, P. (1984): "Central Limit Theorem for Integrated Squared Error of Multivariate Nonparametric Density Estimators," *Journal of Multivariate Analysis*, 14, 1-16.

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

Hansen, L.P., J. Heaton and A. Yaron (1996): "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics*, 14, 262-280.

Härdle, W., and E. Mammen (1993): "Comparing Non-Parametric versus Parametric Regression Fits," *Annals of Statistics*, 21, 1926-1947.

Hausman, J.A. (1978): "Specification Tests in Econometrics," *Econometrica*, 46, 1251-1271.

Hong, Y., and H. White (1995): "Consistent Specification Testing via Nonparametric Series Regression," *Econometrica*, 63, 1133-1159.

Horowitz, J.L., and N.E. Savin (2000): "Empirically Relevant Critical Values for Hypothesis Tests: A Bootstrap Approach," *Journal of Econometrics*, 95, 375-389.

Hsu, S.-H., and C.-M. Kuan (2011): "Estimation of Conditional Moment Restrictions without Assuming Parameter Identifiability in the Implied Unconditional Moments," *Journal of Econometrics*, 165, 87–99.

Imbens, G.W. (1997): "One-Step Estimators for Over-Identified Generalized Method of Moments Models," *Review of Economic Studies*, 64, 359-383.

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

Jayasuriya, B.R. (1996): “Testing for Polynomial Regression Using Nonparametric Regression Techniques,” *Journal of the American Statistical Association*, 91, 1626-1631.

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

Kitamura, Y., Tripathi, G., and H. Ahn (2004): “Empirical Likelihood-Based Inference in Conditional Moment Restriction Models,” *Econometrica*, 72, 1667-1714.

Lavergne, P., and Q. Vuong (2000): “Nonparametric Significance Testing,” *Econometric Theory* 16, 576-601.

Lorentz, G.G. (1986): *Approximation of Functions.* New York: Chelsea Publishing Company.

Matsushita, Y., and T. Otsu (2013): “Second-Order Refinement Of Empirical Likelihood For Testing Overidentifying Restrictions,” *Econometric Theory*, 29, 324-353

Meullbauer, J. (1976): “Community Preferences and the Representative Consumer,” *Econometrica*, 44, 525–543.

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

Newey, W. K. (1994): “Series Estimation of Regression Functionals,” *Econometric Theory*, 10, 1-28.

Newey, W.K. (1997): “Convergence Rates and Asymptotic Normality for Series Estimators,” *Journal of Econometrics*, 79, 147-168.

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

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

Owen, A. (2001): *Empirical Likelihood.* New York: Chapman and Hall.

Parente, P.M.D.C., and R.J. Smith (2011): “GEL Methods for Nonsmooth Moment Indicators,” *Econometric Theory*, 27, 74-113.

Powell M. J. D. (1981): *Approximation Theory and Methods.* Cambridge University Press.

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

Ruud, P. (2000): *An Introduction to Classical Econometric Theory.* Oxford: Oxford University Press.

Serfling, R.J. (2002): *Approximation Theorems of Mathematical Statistics.* New York: Wiley.

Smith, R.J. (1994): "Asympotical Optimal Tests Using Limited Information and Testing for Exogeneity," *Econometric Theory*, 10, 53-69.

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

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

Smith, R. J. (2011): "GEL Methods for Moment Condition Models," *Econometric Theory*, 27, 1192-1235.

Tripathi, G., and Y. Kitamura (2003): "Testing Conditional Moment Restrictions," *Annals of Statistics*, 31, 2059-2095.

Wooldridge, J.M. (1992): "A Test of Functional Form against Nonparametric Alternatives," *Econometric Theory*, 8, 452-475.

Wooldridge, J.M. (2002): *Econometric Analysis of Cross Section and Panel Data.* Cambridge: MIT Press.

Wu, D.-M. (1973): "Alternative Tests of Independence between Stochastic Regressors and Disturbances," *Econometrica*, 41, 733-750.

Yatchew, A.J. (1992): "Nonparametric Regression Tests based on Least Squares," *Econometric Theory*, 8, 435-451.

Zedlewski, J. (2008): "Practical Empirical Likelihood Estimation using matElike." Manuscript, Harvard University.

Zheng, J.X. (1996): "A Consistent Test of Functional Form via Nonparametric Estimation Techniques," *Journal of Econometrics*, 75, 263-289.

Zheng, J.X. (1998): "A Consistent Nonparametric Test of Parametric Regression Model under Conditional Quantile Restrictions," *Econometric Theory*, 14, 123-138.