Additional Moment Condition Tests[*]

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

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

This draft: November 2014

Abstract

The primary focus of this article is 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 thus mirroring that of the classical parametric likelihood setting by defining *restricted* tests in contradistinction to standard *unrestricted* tests that ignore the maintained moment condition information. 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 relative practical efficacy of the various test statistics in terms of empirical size and size-adjusted power.

JEL Classification: C12, C14, C30
Key-words: GMM; Generalized Empirical Likelihood; Series Approximations; Restricted Tests; Unrestricted Tests; Local Power.

1   Introduction

The primary focus of this article is the provision of tests for additional conditional moment constraints in cross-section or short panel data contexts where a finite dimensional parameter vector is the object of inferential interest. The principal contribution is the explicit incorporation of conditional moment restrictions defining the maintained hypothesis in the formulation of the test statistics. Thus, our approach mirrors that of the classical parametric likelihood setting by defining *restricted* tests for the additional conditional moments in contradistinction to standard *unrestricted* tests that ignore the maintained moment condition information with a similar 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. Examples include moment conditional homoskedasticity and instrument validity with particular definitions of exogeneity as special cases of the latter. 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 of 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), henceforth DIN, 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 (1985a), Eichenbaum, Hansen and Singleton (1988) and Ruud (2000) for GMM [Hansen (1982)] and Smith (1997, 2011) for generalized empirical likelihood (GEL) [see also 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 DIN test statistics circumvent this difficulty by allowing

[1]

the number of unconditional moments to grow with sample size at an appropriate rate.[1] 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 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 although the limit standard normal variate differs.[2] 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 that of unrestricted statistics, a conclusion that extends to other statistics that partially or completely neglect the maintained conditional moment information, thereby demonstrating the deficiency of these latter tests and mirroring the results in the classical parametric likelihood setting. Importantly, these asymptotic local power results indicate that the rate of increase of the additional and maintained approximating unconditional moment information should be identical if full advantage is to be gained from the maintained hypothesis. Moreover, since non-centrality parameters are non-negative 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

---

[1]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).

[2]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).

validity examples that are used throughout the paper. GMM and GEL restricted and unrestricted 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 statistics under the null hypothesis of the additional conditional moment validity together with the large sample independence of the restricted test statistics and DIN GMM and GEL test statistics for the maintained hypothesis, thus permitting control of the overall test size if the maintained and then additional conditional moment restrictions are sequentially tested. 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 unrestricted and other tests that neglect the maintained conditional moment information. Section 6 presents a set of simulation results on the size and power of the test statistics. 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. The abbreviations f.r.r., n.s. and p.d. indicate "full row rank", "nonsingular" and "positive definite" respectively. Statistics are "asymptotically equivalent" if they differ by an $o_p(1)$ 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 random 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 vector of instruments s_m such that

$$E[u_m(z, \beta_{m0})|s_m] = 0 \qquad (2.1)$$

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

$$E[u_a(z, \beta_{a0})|s_a] = 0 \qquad (2.2)$$

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 the data observable vector z and the unknown p_a-vector of parameters β_a with $\mathcal{B}_a$ the corresponding parameter space and s_a an observable 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.[3]

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.,

$$H_0 : E[u_a(z, \beta_{a0})|s_a] = 0, E[u_m(z, \beta_{m0})|s_m] = 0 \qquad (2.3)$$

---

[3]In principle the theoretical analysis of the paper may be adapted and extended to deal with models defined by nonsmooth moment conditions and that include nonparametric components, e.g., semiparametric single index ordered choice models. See, e.g., Chen and Pouzo (2009, 2012) and Parente and Smith (2011).

with the corresponding alternative hypothesis H_1 given by

$$H_1 : E[u_a(z, \beta_{a0})|s_a] \neq 0, E[u_m(z, \beta_{m0})|s_m] = 0. \tag{2.4}$$

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

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

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

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

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

REMARK 2.2: Blundell and Horowitz (2007) defines a form of exogeneity hypothesis for non-parametric regression in which s_a coincides with the covariate vector x but does not include the maintained instrument vector s_m. The practical advantage of this exogeneity hypothesis is that the structural regression function is consistently estimated by nonparametric least squares (LS) thus avoiding the difficulties associated with non-parametric 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) with the inclusion of both s_m and x in s_a a *conditional* form (CE). Note that CE 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; see Cragg (1983). IV estimation fully incorporating both additional $E[y - \beta_{m0}x|s_a] = 0$ and maintained $E[y - \beta_{m0}x|s_m] = 0$ conditional moments dominates LS.

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, DIN Assumption 1, p.58, 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.

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 the maintained hypothesis (2.1) may be re-interpreted and re-expressed in terms of a sequence of unconditional moment restrictions as described above. Moreover, to formulate a test statistic appropriate for the null hypothesis (2.3) requires that its constituent conditional moment constraints, $E[u_a(z, \beta_{a0})|s_a] = 0$ (2.2) and $E[u_m(z, \beta_{m0})|s_m] = 0$ (2.1), are replaced by suitably defined 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 M and thus MK are positive integers. 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

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

REMARK 3.1: Strictly speaking, the dimension of $q_a^{MK}(\cdot)$ is required to be $O(K)$, i.e., the same order as that of $q_m^K(\cdot)$. The multiplicative choice MK is adopted for simplicity and for ease of implementation. *Restricted* test statistics for (2.3) defined in section 3.3 below are expressed as (or are asymptotically equivalent to) the difference of component statistics. Their large sample behaviour is 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 $O(K)$, the component statistic with the dominant number of approximating functions governs the asymptotic

[8]

behaviour of the restricted test statistic. More precisely, the $O(K)$ choice for the dimension of $q_a^{MK}(\cdot)$ guarantees that the component statistics are of the same asymptotic magnitude and thus that neither set of moment restrictions is negligible asymptotically. In particular, in an asymptotic sense, restricted tests thereby fully incorporate the maintained hypothesis (2.1). If, however, the dimension of $q_a^{MK}(\cdot)$ should diverge at a faster rate than $O(K)$, the component statistic arising from the maintained hypothesis is asymptotically negligible and therefore restricted test statistics take the *unrestricted* form asymptotically. This assumption has the important implication that in terms of asymptotic local power restricted tests dominate unrestricted and other forms of test; see section 5.

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 CE $s_a = (s_m, x)$ 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 and S_a^u; viz. $S_m^q q^K(s) = q_m^K(s_m)$, $S_a^q q^K(s) = q_a^{MK}(s_a)$, $S_m^u u(z, \beta) = u_m(z, \beta_m)$ and $S_a^u u(z, \beta) = u_a(z, \beta_a)$.[4] 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 $H_0 : \Sigma_0(s_m) = \Sigma_0$, $E[u(z, \beta_0)|s_m] = 0$ is then

$$S(u(z, \beta) \otimes q^K(s)) = \begin{pmatrix} u_m(z, \beta_m) \\ u_a(z, \beta_a) \end{pmatrix} \otimes q_m^K(s_m), K \to \infty,$$

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

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

---

[4]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.

may equivalently be re-arranged as

$$S(u(z,\beta) \otimes q^K(s)) = \begin{pmatrix} u_m(z,\beta_m) \otimes q_m^K(s_m) \\ u_m(z,\beta_m) \otimes q_a^{MK}(s_a) \end{pmatrix}, K \to \infty,$$

with that for the maintained hypothesis given by $S_m(u(z,\beta) \otimes q^K(s)) = u_m(z,\beta_m) \otimes q_m^K(s_m)$, $K \to \infty$, 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.[5] Note also that only root-n consistency rather than efficiency is required for the estimator $\hat{\beta}$.

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 $u(z,\beta)$. 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.*

---

[5] Guggenberger and Smith (2005) shows that if the sample data are `i.i.d.` only $\gamma = 2$ as in Assumption 3.3(d) is required; see Lemma 3 in Owen (1990). Indeed, 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.

3.3 Test Statistics

Let $g_{mi}(\beta_m) = S_m(u(z_i, \beta) \otimes q^K(s_i)) = u_m(z_i, \beta_m) \otimes q_m^K(s_{mi})$, $g_{ai}(\beta_a) = S_a(u(z_i, \beta) \otimes q^K(s_i)) = u_a(z_i, \beta) \otimes q_a^{MK}(s_{ai})$ 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

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

and

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

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), are referred to as *restricted* tests whereas those that ignore these constraints 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 revised hypotheses (3.3) and (3.1) respectively; *viz.*

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

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 3.3: For fixed and finite K, GMM and GEL test statistics [Newey (1985a), Smith (2011)] for the validity of additional moment restrictions, e.g., $\mathcal{T}_{GMM}^g - \mathcal{T}_{GMM}^{g_m}$, are under suitable conditions asymptotically chi-square distributed with $J_a MK - (p - p_m)$ degrees of freedom. The mean location $J_a MK - (p - p_m)$ and standard deviation scale

[12]

$\sqrt{2(J_a M K - (p - p_m))}$ standardisations of $\mathcal{T}^g_{GMM} - \mathcal{T}^{gm}_{GMM}$ in $\mathcal{J}^r$ (3.6) are introduced since similarly transformed chi-square random variates with large degrees of freedom are approximately standard normally distributed.

A number of alternative test statistics to GMM-based procedures for a finite number of additional moment restrictions using GEL [NS, Smith (1997, 2011)] may be adapted for the framework considered here. As in DIN and NS 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

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

where λ and $\lambda_m = S_m \lambda$ are the corresponding $(J_a M + J_m)K$- and $J_m K$-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.[6]

Let $\hat{\Lambda}^{gm}_n(\beta_m) = \{\lambda_m : \lambda'_m g_{mi}(\beta_m) \in \mathcal{V}, i = 1, ..., n\}$ 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

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

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

---

[6]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 et al. (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 NS. 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 NS, Section 2.1, pp.223-224.

(2.4) may be based on the difference of GEL criterion function (3.7) statistics; *viz.*

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

Restricted Lagrange multiplier, score and Wald-type statistics are defined respectively as[7]

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

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

and

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

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). Recall that the vector s collects the distinct elements of the maintained and additional instrument vectors s_m and s_a.

---

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

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

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

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

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

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 for the behaviour under suitable sequences of local alternatives of both 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 based on *non-standardised* statistics may also be defined based on fixed and finite K asymptotics. E.g., for $\mathcal{T}_{GMM}^g - \mathcal{T}_{GMM}^{g_m}$, the critical region $\{\mathcal{T}_{GMM}^g - \mathcal{T}_{GMM}^{g_m} \geq \chi^2_{J_aMK-(p-p_m),\alpha}\}$ is apposite where $\{\chi^2_{J_aMK-(p-p_m)} \geq \chi^2_{J_aMK-(p-p_m),\alpha}\} = \alpha$ and χ^2_k denotes a chi-square distributed random variable with k degrees of freedom. Note that $p - p_m$ is negligible in the large K, large n asymptotic analysis of Theorems 4.1 and 4.2.

Unrestricted statistics for testing the null hypothesis (2.3) may be also defined which ignore the information contained in the maintained hypothesis (2.1); *viz.* the unrestricted GEL-based statistics

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

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

[15]

and the unrestricted GMM statistic based on $\mathcal{T}_{GMM}^{g}$ which takes the score form

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

By a similar analysis to that used to establish Theorems 4.1 and 4.2 these unrestricted statistics also each converge in distribution to a standard normal random variate and are mutually asymptotically equivalent but not to the restricted $\mathcal{J}^{r}$, $\mathcal{LR}^{r}$, $\mathcal{LM}^{r}$, $\mathcal{S}^{r}$ and $\mathcal{W}^{r}$. 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).

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

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

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 or unrestricted statistics. It should be noted that statistics such as unrestricted statistics and these other forms of statistic are deficient in terms of asymptotic local power as compared to restricted tests; see section 5.

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.

[16]

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 $\zeta(K)^2 K^3/n \to 0$ is also made.

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 $\mathcal{J}^m$ is similar to that of classical and unconditional moment GMM and GEL tests for a sequence of parametric restrictions; cf. Newey (1985a) and Smith (2011). In particular, the unrestricted statistic $\mathcal{J}^u$ is the sum of (re-scaled) restricted $\mathcal{J}^r$ and maintained hypothesis $\mathcal{J}^m$ statistics mirroring the decomposition of unrestricted classical or GMM and GEL statistics for parametric restrictions.

5 Asymptotic Local Power

This section considers the asymptotic distribution of the statistics of section 3.3 under a suitable sequence of local alternatives. Critically, this discussion demonstrates the deficiency in terms of asymptotic local power of tests which fail to incorporate fully the maintained conditional information (2.1) and thereby the superiority of the restricted tests. Moreover, it also underlines the requirement that the dimension of the additional conditional moments approximating functions $q_a^{MK}(s_a)$ should be of the same order $O(K)$ as that of the maintained approximating functions $q_m^K(s_m)$. As noted in Remark 3.1, the import of this restriction is that the maintained conditional moment information (2.1) fully contributes to the asymptotic local power of the restricted statistics.

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 distributions of the statistics described in section 3.3 under local alternatives (5.1) are derived 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*

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

[18]

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 $\mathcal{W}^r$.

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

Corollary 5.1 *Let Assumptions 3.1-3.4 and 5.1 hold and $\zeta(K)^2 K^2/n \to 0$. 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 $\zeta(K)^2 K^3/n \to 0$, then $\mathcal{LR}^u$, $\mathcal{LM}^u$ are asymptotically equivalent to $\mathcal{S}^u$.

REMARK 5.3: Since $\mu^r > \mu^u$ restricted tests dominate unrestricted tests in terms of asymptotic local power. Other tests that partially or completely neglect the maintained conditional moment information (2.1) 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 may be shown to converge in distribution under the sequence of local alternatives (5.1) to a $N(\mu^a/\sqrt{2}, 1)$ random variable where $\mu^a = E[\xi_a(s)'\Sigma_a(s)^{-1}\xi_a(s)]$ and $\xi_a(s) = S_a^u \xi(s)$, $\Sigma_a(s; \beta_a) = E[u_a(z, \beta_a) u_a(z, \beta_a)'|s] = S_a^u \Sigma(s; \beta) S_a^{u\prime}$, $\Sigma_a(s) = \Sigma_a(s; \beta_{a0})$. Since $\mu^r - \mu^a = E[\xi(s)'(\Sigma(s)^{-1} - S_a^{u\prime}(S_a^u \Sigma(s) S_a^{u\prime})^{-1} S_a^u)\xi(s)] \geq 0$ tests based on these statistics are asymptotically less powerful relative to restricted tests.

REMARK 5.4: Corollary 5.1 further justifies restricting the dimension of $q_a^{MK}(s_a)$ to be $O(K)$ or linear in K. If M increases with K, then μ^u should differ little from μ^r with

consequential similar discriminatory power for both unrestricted and restricted tests for local departures from the null hypothesis H_0 (2.3). Moreover, Corollary 5.1 indicates that M should be chosen as small as possible in order to maximise the relative gain in asymptotic local power from the maintained conditional moment restrictions (2.1).

6   Simulation Evidence

This section reports the results from a subset of extensive simulation experiments undertaken to evaluate the behaviour and performance of restricted and unrestricted tests of ME and CE forms of instrument validity in the linear regression model based on the above GMM and GEL statistics; see Example 2.2. 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. Both forms of Wald test statistic $\mathcal{W}^r$ (3.10) and $\bar{\mathcal{W}}^r$, see fn.6, and the 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.[8]

All experiments concern the linear regression model

$$y = \beta_0 x + u, \tag{6.1}$$

where x is a scalar covariate and u a scalar error term. Thus $\beta = \beta_m = \beta_a$, $J_a = J_m = 1$ and $u(z,\beta) = u_m(z,\beta_m) = u_a(z,\beta_a)$ where $u(z,\beta) = y - \beta x$; see Example 2.2. For simplicity, the value of the parameter β_0 is set as 0. Consideration is restricted to the single parameter β_0 to ease the computational burden associated with GEL estimation. The maintained scalar instrument is again denoted as s_m with the additional scalar instrument the covariate x.

---

[8]The full set of simulation results are available from the authors upon request.

The regression design incorporates both ME and CE forms of additional conditional constraint restrictions (2.2); see Remark 2.2. Therefore the hypotheses of interest are as follows. First, the maintained 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 CE $E[u|s_m, x] = 0$, i.e., $s_a = (s_m, x)$.

6.1 Experimental Design

The covariate x and maintained instrument s_m are generated according to $x = \Phi(z_x)$ and $s_m = \Phi(z_m)$, where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function and z_x and z_m are jointly normally distributed with mean zero, unit variance and correlation coefficient ρ, $\rho \in (-1, 1)$ and $\rho \neq 0$. Hence, x and s_m are marginally distributed as uniform random variates on the unit interval $[0, 1]$. The value $\rho = 0.3$ is reported here; other experiments examine $\rho = 0.7$ with qualitatively similar results.

Let v be defined by

$$v = a[z_x^2 + z_m^2 - (\frac{1 + \rho^2}{\rho})z_x z_m - (1 - \rho^2)] + \tau(z_x - \rho z_m) + \upsilon,$$

where $\upsilon \sim N(0, 1)$ and is distributed independently of z_x and z_m. The error term $u = v/\sqrt{var[v]}$ in (6.1) is thus restricted to have unit variance for simplicity and ease of comparison. Note that $var[v] = a^2(1 + \rho^2)(\rho^{-1} - \rho)^2 + \tau^2(1 - \rho^2) + 1$.

Therefore the regression model (6.1) is characterised by the following properties. The maintained hypothesis (2.1) is always satisfied, i.e., $E[u|s_m] = 0$. For the additional instrumental validity restrictions (2.2): ME $s_a = x$

$$E[u|s_a] = \tau(1 - \rho^2)\Phi^{-1}(x)/var[v]$$

and, thus, $E[u|s_a] = 0$ if $\tau = 0$ and $E[u|s_a] \neq 0$ if $\tau \neq 0$; CE $s_a = (s_m, x)$

$$\begin{aligned} E[u|s_a] &= (a[\Phi^{-1}(x)^2 + \Phi^{-1}(s_m)^2 - (\frac{1 + \rho^2}{\rho})\Phi^{-1}(x)\Phi^{-1}(s_m) - (1 - \rho^2)] \\ &\quad + \tau[\Phi^{-1}(x) - \rho\Phi^{-1}(s_m)])/var[v], \end{aligned}$$

yielding $E[u|s_a] = 0$ if $a = \tau = 0$ and $E[u|s_a] \neq 0$ if $a \neq 0$ or $\tau \neq 0$.

[21]

The experiments discussed here consider two designs: ME $a = 0$, i.e., ME does not hold unless $\tau = 0$; CE $\tau = 0$, i.e., CE does not hold unless $a = 0$.[9]

Empirical test size is examined for sample sizes $n = 200$, 500, 1000 and 3000 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.

6.1.1 Approximating Functions

Polynomials are used to form the approximating functions in the simulations.[10] 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 s_m^k, $(k = 0, ..., K - 1)$. For ME $E[u(z, \beta)|x]$ is approximated by a polynomial of order MK in x, i.e., $q_a^{MK}(s_a)$ has elements x^k, $(k = 1, ..., MK)$. The CE case $E[u(z, \beta)|s_a]$, $s_a = (s_m, x)$, uses the vector of approximating functions $q_a^{MK}(s_a)$ with elements $s_m^k x^l$, $(k = 0, ..., (MK)^{1/2} - 1, l = 1, ..., (MK)^{1/2})$, where M and K are chosen to ensure that $(MK)^{1/2}$ is an integer.

6.1.2 Estimators

Efficient estimation methods examined include two stage least squares estimation (2SLQ) computed using the single instrument s_m, efficient two-step GMM (GMM) with the weighting matrix computed using 2SLQ, continuous updating (CUE), empirical likelihood (EL) and exponential tilting (ET). The estimator subscripts MA, ME and CE indicate computation incorporating maintained, ME and CE restrictions respectively.[11]

---

[9]The complete set of experiments examines (a) $\tau = 0$, i.e., ME holds but CE does not unless $a = 0$; (b) $a = 0$, i.e., neither ME nor CE hold unless $\tau = 0$.

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

[11]The simplex search algorithm of `MATLAB` is used to compute GMM and GEL hence ensuring a local optimum. The Newton method is used to locate $\hat{\lambda}(\beta)$ for given β which is required for the profile GEL objective function, e.g., EL and ET. EL computation requires some care since the EL criterion involves the logarithm function which is undefined for negative arguments; this difficulty is avoided by employing the `MATLAB` code due to Owen in which logarithms are replaced by a function that is logarithmic for arguments larger than a small positive constant and quadratic below that threshold. See Owen (2001, (12.3), p.235); the code is available at `http://www-stat.stanford.edu/~owen/empirical/`

6.1.3 Test Statistics

Restricted tests for ME $E[u|x] = 0$ and CE $E[u|s_m, x] = 0$ adopt the following notation. The superscripts M and C refer respectively to the ME or CE 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 statistic for CE based on EL criteria and estimators is denoted as $\mathcal{LR}_{\text{EL}}^{\text{C}} = 2n(\hat{P}_{\text{EL}}^{g}(\hat{\beta}_{\text{EL}_{\text{C}}}, \hat{\lambda}_{\text{EL}_{\text{C}}}) - \hat{P}_{\text{EL}}^{g_m}(\hat{\beta}_{\text{EL}_{\text{MA}}}, \hat{\lambda}_{\text{EL}_{\text{MA}}}))$, cf. (3.7). The non-standardised robustified score statistic $\bar{\mathcal{S}}$, see fn.6, evaluated at the corresponding efficient MA estimator is also examined. Restricted ME and CE non-standardised test statistics are calibrated against chi-square distributions with $MK - 1$ degrees of freedom.[12]

6.2 Choice of the Number of Instruments

Implementation of the above tests requires the choice of the number of instruments to employ under the maintained hypothesis. The method of Donald, Imbens and Newey (2009) predominantly indicated the choice $K = 2$ as did selection criteria such as SBC. The additional choice $K = 4$ was also examined to examine the effect of increasing K. To explore the influence of increasing M, cf. Corollary 5.1, M is chosen as 1 and 2 (ME) and, if $K = 2$, $M = 2$ and 8 and, if $K = 4$, $M = 1$ and 4 (CE) since $(MK)^{1/2}$ is required to be integer.

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.

---

[12] A number of asymptotically equivalent test statistics for the maintained hypothesis (2.1) were also investigated. The Hausman 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.

6.3.1 ME

Table B.1 in Appendix B presents the empirical rejection frequencies for $K = 2$ and 4 and $M = 1$ 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 $\mathcal{LR}_{\mathrm{EL}}^{\mathrm{M}}$, $\mathcal{LM}_{\mathrm{EL}}^{\mathrm{M}}$ and $\mathcal{LM}_{\mathrm{ET}}^{\mathrm{M}}$ and to a lesser extent $\mathcal{LR}_{\mathrm{ET}}^{\mathrm{M}}$ tests suffer from size distortions for moderate sample sizes $n = 200$ and 500 with a serious deterioration in performance for large M and K. Of the remaining statistics, the GMM $\mathcal{J}^{\mathrm{M}}$, the CUE LR $\mathcal{LR}_{\mathrm{CUE}}^{\mathrm{M}}$ and the GEL robust score $\bar{\mathcal{S}}_{\mathrm{EL}}^{\mathrm{M}}$, $\bar{\mathcal{S}}_{\mathrm{ET}}^{\mathrm{M}}$ statistics display satisfactory empirical size properties for most sample sizes. For a given sample size n there is a deterioration in performance to a lesser or greater degree for all statistics as M increases from 1 to 2 for fixed K and as K increases from 2 to 4 for fixed M, i.e., as the number of unconditional moments under test increases. In summary, tests for ME based on the statistics $\mathcal{J}^{\mathrm{M}}$, $\mathcal{LR}_{\mathrm{CUE}}^{\mathrm{M}}$ and $\bar{\mathcal{S}}_{\mathrm{EL}}^{\mathrm{M}}$, $\bar{\mathcal{S}}_{\mathrm{ET}}^{\mathrm{M}}$ appear to be the most reliable in terms of empirical size.

6.3.2 CE

Table B.2 in Appendix B presents the rejection frequencies for $K = 2$, $M = 2$ and 8 and $K = 4$, $M = 1$ and 4 for non-standardised restricted tests of the CE 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 $M = 8$ at $K = 2$ and for $M = 4$ at $K = 4$ for moderate sample sizes $n = 200$ and 500 for all test versions. The GMM $\mathcal{J}^{\mathrm{C}}$ (3.6) and CUE LR $\mathcal{LR}_{\mathrm{CUE}}^{\mathrm{C}}$ (3.7) forms and the robust score statistic versions $\bar{\mathcal{S}}_{\mathrm{EL}}^{\mathrm{C}}$, $\bar{\mathcal{S}}_{\mathrm{ET}}^{\mathrm{C}}$ display the most satisfactory empirical size at the nominal 0.05 level.

6.4 Size-Corrected Power

Typically size-corrected power increases substantially as sample size n increases from 200 to 500 whereas it declines with increased M and/or K.

6.4.1 ME

Table B.3 in Appendix B presents empirical rejection frequencies for non-standardised restricted ME tests for values $K = 2$ and 4 and $M = 1$ and 2 based on 0.05 level size-corrected critical values for deviations $\tau \neq 0$ from the ME hypothesis $E[u(z, \beta_0)|x] = 0$ with $a = 0$.

In general size-corrected empirical power differences between tests are less at larger τ and larger sample sizes n. Overall tests based on the ET statistics $\mathcal{LR}_{\text{ET}}^{\text{M}}$ and $\mathcal{LM}_{\text{ET}}^{\text{M}}$ are more powerful but these tests displayed an unsatisfactory correspondence between empirical and nominal size. For the smaller $M = 1$ and $K = 2$ indicated by Corollary 5.1 and selection criteria respectively the size-corrected empirical powers of all tests, not just those with reasonable empirical size characteristics, are very similar for both the smaller $n = 200$ and larger $n = 500$ sample sizes.

6.4.2 CE

Table B.4 in Appendix B presents empirical rejection frequencies for non-standardised restricted CE tests for values $K = 2$, $M = 2$ and 8 and $K = 4$, $M = 1$ and 4 based on 0.05 level size-corrected critical values for deviations $a \neq 0$ from the CE hypothesis $E[u(z, \beta_0)|s_m, x] = 0$ with $\tau = 0$.

Similar general conclusions to those for the ME tests above broadly follow. Interestingly, for given K empirical rejection frequencies decrease between $M = 2$ and 8 ($K = 2$) and $M = 1$ and 4 ($K = 4$) underlining the importance of a smaller choice of M; cf. Corollary 5.1. Given MK slight increases or decreases may occur for increases in K, e.g., between $M = 2$, $K = 2$ and $M = 1$, $K = 4$ or $M = 8$, $K = 2$ and $M = 4$, $K = 4$. At the smaller values $M = 2$ and $K = 2$ the ranking by empirical power for those tests with reasonable empirical size properties is, for $n = 200$, the group $\bar{\mathcal{S}}_{\text{EL}}^{\text{C}}$, $\bar{\mathcal{S}}_{\text{ET}}^{\text{C}}$ closely followed by $\mathcal{J}^{\text{C}}$, $\mathcal{LR}_{\text{CUE}}^{\text{C}}$ whereas, for $n = 500$, the group $\mathcal{LR}_{\text{EL}}^{\text{C}}$, $\mathcal{LR}_{\text{ET}}^{\text{C}}$ is followed by $\bar{\mathcal{S}}_{\text{EL}}^{\text{C}}$, $\bar{\mathcal{S}}_{\text{ET}}^{\text{C}}$ and then $\mathcal{J}^{\text{C}}$, $\mathcal{LR}_{\text{CUE}}^{\text{C}}$.

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 root n consistent estimators is recommended for test construction. Restricted dominate unrestricted tests in terms of empirical power. Empirical power typically declines for increases in the constant M for both ME and CE tests.

The ME $E[u(z, \beta_0)|x] = 0$ empirical sizes of tests based on the restricted GMM form $\mathcal{J}^{\mathrm{M}}$, the GEL statistic $\mathcal{LR}^{\mathrm{M}}_{\mathrm{CUE}}$ and the robust GEL score versions $\bar{\mathcal{S}}^{\mathrm{M}}_{\mathrm{EL}}$ and $\bar{\mathcal{S}}^{\mathrm{M}}_{\mathrm{ET}}$ most closely approximate nominal size. At small values of M and K for small sample size the size-corrected empirical power of the $\bar{\mathcal{S}}^{\mathrm{M}}_{\mathrm{EL}}$, $\bar{\mathcal{S}}^{\mathrm{M}}_{\mathrm{ET}}$ tests marginally dominates that of the $\mathcal{J}^{\mathrm{M}}$, $\mathcal{LR}^{\mathrm{M}}_{\mathrm{CUE}}$ tests whereas for moderate sample size the size-corrected empirical powers of all tests are broadly similar.

Of the restricted tests for the CE hypothesis $E[u(z, \beta_0)|s_m, x] = 0$ those employing the GMM form $\mathcal{J}^{\mathrm{C}}$, the GEL statistic $\mathcal{LR}^{\mathrm{C}}_{\mathrm{CUE}}$, and the robust GEL score statistics $\bar{\mathcal{S}}^{\mathrm{C}}_{\mathrm{EL}}$ and $\bar{\mathcal{S}}^{\mathrm{C}}_{\mathrm{ET}}$ have most satisfactory empirical size. Again for small M and K the robust GEL score-based, $\bar{\mathcal{S}}^{\mathrm{C}}_{\mathrm{EL}}$ and $\bar{\mathcal{S}}^{\mathrm{C}}_{\mathrm{ET}}$, forms of test appear marginally superior to the GMM form $\mathcal{J}^{\mathrm{C}}$ and the GEL statistic $\mathcal{LR}^{\mathrm{C}}_{\mathrm{CUE}}$ in terms of size-corrected empirical power against deviations from the CE hypothesis for small sample sizes whereas at moderate sample sizes the GEL LR statistics $\mathcal{LR}^{\mathrm{C}}_{\mathrm{EL}}$, $\mathcal{LR}^{\mathrm{C}}_{\mathrm{ET}}$ dominate $\bar{\mathcal{S}}^{\mathrm{C}}_{\mathrm{EL}}$, $\bar{\mathcal{S}}^{\mathrm{C}}_{\mathrm{ET}}$, $\mathcal{J}^{\mathrm{C}}$, $\mathcal{LR}^{\mathrm{C}}_{\mathrm{CUE}}$ tests.

7 Conclusions

The primary focus of this article has been 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

[26]

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 incorporate maintained moment constraints should dominate in terms of power unrestricted tests that ignore such information.

The simulation experiments undertaken to explore the efficacy of the various tests proposed in the paper indicate a number of 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,* $\max_{\beta \in \mathcal{B}, \lambda \in \Lambda_n, 1 \leq i \leq n} |\lambda' g_i(\beta)| \overset{p}{\to} 0$ *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 $1 \leq i \leq n$. It then follows from Owen (1990, Lemma 3, p.98) that $\max_{1 \leq i \leq n} b_i = o_p(n^{1/2})$. Therefore, by CS

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

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* $E[g(z,\beta_{0,n})] = 0$, $k \to \infty$, $E[(g(z,\beta_{0,n})'C_n g(z,\beta_{0,n}))^2]$ $\to 0$ *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

$$\mathcal{T}_1 = \sum_{i,j:i<j} \sqrt{\frac{2}{n^2 k}} g'_{i,n} C_n g_{j,n},$$

$$\mathcal{T}_2 = \frac{\sum_i g'_{i,n} C_n g_{i,n}/n - k}{\sqrt{2k}}.$$

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 Hall (1984, Theorem 1, pp.3-4). Define

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

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

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

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

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

On the other hand

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

[28]

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

$$\frac{1}{nk^2} E[(g'_{1,n} C_n g_{2,n})^4] \leq \frac{1}{nk^2} E[(g'_{1,n} C_n g_{1,n})^2 (g'_{2,n} C_n g_{2,n})^2]$$
$$= (\frac{1}{k\sqrt{n}} E[(g'_{1,n} C_n g_{1,n})^2])^2 \rightarrow 0.$$

Since $E[G_n(z_1, z_2)^2]/E[H_n(z_1, z_2)^2]^2 = 1/k \rightarrow 0$, $\mathcal{T}_1 \xrightarrow{d} N(0,1)$ 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) $E[(g(z, \beta_0)'Q\Omega Qg(z, \beta_0))^2] E[(g(z, \beta_0)'\Omega^{-1} g(z, \beta_0))^2] = o(nk^2)$ 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 $k \rightarrow \infty$ and $n \rightarrow \infty$.

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

$$\mathcal{T}_1 = \sum_{i,j:i<j} \sqrt{\frac{2}{n^2 k}} g'_i Q g_j$$
$$\mathcal{T}_2 = \frac{\sum_i g'_i Q g_i / n - ak}{\sqrt{2k}}$$

Since $E[\mathcal{T}_2] = 0$ and $var[\mathcal{T}_2] \leq E[(g'_i Q g_i)^2]/2kn \rightarrow 0$ 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 Hall (1984, Theorem 1, pp.3-4). Define

$$H_n(u, v) = \sqrt{\frac{2}{n^2 k}} g'_u Q g_v.$$

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

$$G_n(u, v) = E[H_n(z_1, u) H_n(z_1, v)|u, v]$$
$$= \frac{2}{n^2 k} E[g'_u Q g_1 g'_1 Q g_v|u, v]$$
$$= \frac{2}{n^2 k} g'_u Q\Omega Q g_v.$$

[29]

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

$$
\begin{aligned}
E[H_n(z_1, z_2)^2] &= \frac{2}{n^2 k} E[(g_1' Q g_2)^2] \\
&= \frac{2}{n^2 k} E[g_1' Q \Omega Q g_1] = \frac{2}{n^2 k} tr([Q\Omega]^2) = \frac{2v}{n^2}.
\end{aligned}
$$

Also

$$
\frac{E[H_n(z_1, z_2)^4]}{n E[H_n(z_1, z_2)^2]^2} = \frac{1}{n v^2 k^2} E[(g_1' Q g_2)^4].
$$

Now, as Ω is positive definite, by CS

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

Hence, by (f),

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

Moreover, by (d),

$$
\begin{aligned}
E[G_n(z_1, z_2)^2] &= \frac{4}{n^4 k^2} E[(g_1' Q \Omega Q g_2)^2] \\
&= \frac{4}{n^4 k^2} E[g_1' Q \Omega Q \Omega Q \Omega Q g_1] = \frac{4}{n^4 k^2} tr([Q\Omega]^4) = o(n^{-4}).
\end{aligned}
$$

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) $E[\left\|U_{i,n}^{-1} - U_i^{-1}\right\|^2] \to 0$, (e) $E[\|\bar{a}_{i,n} - \bar{a}_i\|^2] \to 0$, (f) $K \to \infty$ and $K/n \to 0$, then

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

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 $P'_n P_n = S(\sum_i U_{i,n} \otimes q_i q'_i)S'$ and

$$(\sum_i a'_{i,n} \otimes q'_i)S' \left(S(\sum_i U_{i,n} \otimes q_i q'_i)S' \right)^{-} S(\sum_i a_{i,n} \otimes q_i/n) = A'_n Q_n A_n$$

where $Q_n = P_n(P'_n P_n)^{-} P'_n$.

Let $s = (s_1, ..., s_n)$. As the data are `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,

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

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

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

[31]

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 Feller (1971, Theorem 1, p.316). 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

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

[32]

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

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

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

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

∎

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 $E[\|\varepsilon_{i,n}\|^4|s_i] \leq C$ for large enough n and $E[\|Y_i\|^2|s_i] \leq C$, (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) $E[\|U^{-1}_{i,n} - U^{-1}_i\|^2] \to 0$ and (e) $K \to \infty$ and $K^2/n \to 0$, then*

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

Proof: The result is proved by first showing that

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

and secondly that

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

[33]

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,

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

From the proof of Lemma A.4 $(A_n - \bar{A}_n)' Q_n (A_n - \bar{A}_n) = O_p(K)$ and $B'_n Q_n B_n = O_p(K)$, 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,

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

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,

$$E[\|\bar{A}'_n(I - Q_n)B_n/\sqrt{n}\|^2] = E[\bar{A}'_n(I - Q_n)E[B_n B'_n|s](I - Q_n)\bar{A}_n]/n$$
$$\le CE[\bar{A}'_n(I - Q_n)\bar{A}_n]/n \to 0$$

since $E[B_n B'_n|s]$ 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). Serfling (2002, Corollary, p.32) 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,

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

[34]

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

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

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

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

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 $1/C \leq \lambda_{\min}(\hat{\Omega}) \leq \lambda_{\max}(\hat{\Omega}) \leq C$, $1/C \leq \lambda_{\min}(\bar{\Omega}_n) \leq \lambda_{\max}(\bar{\Omega}_n) \leq C$.*

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

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

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

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

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

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

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

For the third conclusion as in DIN

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

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

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

Hence $C^{-1} \leq \lambda_{\min}(\Omega_n) \leq \lambda_{\min}(\Omega_n) \leq C$. Note also that, if $\tau_nK + \zeta(K)\sqrt{K/n} \to 0$, 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, $\left|\lambda_{\min}(\hat{\Omega}) - \lambda_{\min}(\Omega_n)\right| = o_p(1)$ and $\left|\lambda_{\max}(\hat{\Omega}) - \lambda_{\max}(\Omega_n)\right| = o_p(1)$. The final conclusion follows similarly. ∎

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

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

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,

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

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 $\|\tilde{G}_n - \bar{G}_n\|^2 = O_p(\sqrt{K/n})$.

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

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

The first conclusion follows by T.

In addition

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

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*

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

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

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

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

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

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

$$\left\|\hat{\Omega}^{-1}(\bar{G}_n - G_n)\right\|^2 = tr((\bar{G}_n - G_n)'\hat{\Omega}^{-2}(\bar{G}_n - G_n))$$

$$\leq Ctr((\bar{G}_n - G_n)'(\bar{G}_n - G_n))$$

$$= C\left\|\bar{G}_n - G_n\right\|^2 \overset{p}{\to} 0.$$

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

Now $G_n'\Omega_n^{-1}G_n$ is bounded for large enough n as $\bar{G}_n'\bar{\Omega}_n^{-1}\bar{G}_n \overset{p}{\to} V^{-1}$ 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, $\|\Omega_n^{-1}G_n\|$ 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 $\left\|\hat{\Omega}^{-1}(\bar{G}_n - G_n)\right\| \overset{p}{\to} 0$ by Lemma A.7. Secondly, $\left\|\hat{\Omega}^{-1}(\hat{\Omega} - \Omega_n)\Omega_n^{-1}G_n\right\| \leq \left\|\hat{\Omega}^{-1}(\hat{\Omega} - \Omega_n)\right\| \|\Omega_n^{-1}G_n\|$ by CS and Lemma A.6. Consequently, $\left\|\hat{\Omega}^{-1}\bar{G}_n\right\| = O_p(1)$.

Now by independence

$$E[\hat{g}_n'\Omega_n^{-1}\hat{g}_n] = E[g_{i,n}'\Omega_n^{-1}g_{i,n}]/n$$

$$= E[tr(\Omega_n^{-1}g_{i,n}g_{i,n}')/n] = K/n.$$

[38]

Hence, by M $\|\Omega_n^{-1}\hat{g}_n\| = O_p(\sqrt{K/n})$. 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, $\|G_n'\Omega_n^{-1}\hat{g}_n\| = O_p(1/\sqrt{n}) = o_p(\sqrt{K/n})$ 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,

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

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

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

It remains to prove that

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

From Lemma A.6,

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

∎

A.2 Asymptotic Null Distribution

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

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

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

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

Now

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

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, $(\Omega^{-1} - S_m'\Omega_m^{-1}S_m)\Omega(\Omega^{-1} - S_m'\Omega_m^{-1}S_m) =$

[40]

$\Omega^{-1} - S'_m \Omega_m^{-1} S_m$. Thirdly,

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

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

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

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

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

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

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, $\|\hat{\lambda}_m\| = O_p(\sqrt{K/n})$ 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.,

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

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

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

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 $\left\|\hat{\Omega}_m^{-1} - \Omega_m^{-1}\right\| = o_p(1/\sqrt{K})$,

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

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

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

[42]

Therefore, from Theorem 4.1 it follows that

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

Now consider the Lagrange multiplier statistic (3.8)

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

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

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

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

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

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 $\lambda_{\max}(\hat{\Omega}^{-1}) \leq C$. Moreover

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

[43]

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

$$
\left\|\Omega\hat{\Omega}^{-1}\Omega - \Omega\right\| = \left\|\Omega\hat{\Omega}^{-1}(\hat{\Omega} - \Omega)\right\|
$$

$$
\leq \left\|\Omega\hat{\Omega}^{-1}\right\|\left\|\hat{\Omega} - \Omega\right\| = o_p(1/\sqrt{K}).
$$

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

$$
n\hat{g}'(S_m'\hat{\Omega}_m^{-1}S_m - 2\dot{\Omega}^{-1}\hat{\Omega}S_m'\dot{\Omega}_m^{-1}S_m + S_m'\dot{\Omega}_m^{-1}S_m\hat{\Omega}\ S_m'\dot{\Omega}_m^{-1}S_m)\hat{g}/\sqrt{K}
$$

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

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

$$
\left\|S_m'\hat{\Omega}_m^{-1}S_m - \dot{\Omega}^{-1}\hat{\Omega}S_m'\dot{\Omega}_m^{-1}S_m\right\| \leq \left\|S_m'\dot{\Omega}_m^{-1}(\dot{\Omega}_m - \hat{\Omega}_m)\hat{\Omega}_m^{-1}S_m\right\| + \left\|\dot{\Omega}^{-1}(\hat{\Omega} - \dot{\Omega})S_m'\dot{\Omega}_m^{-1}S_m\right\|
$$

$$
= o_p(1/\sqrt{K}).
$$

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

$$
\left\|\dot{\Omega}^{-1}\hat{\Omega}\ S_m'\dot{\Omega}_m^{-1}S_m - S_m'\dot{\Omega}_m^{-1}S_m\hat{\Omega}S_m'\dot{\Omega}_m^{-1}S_m\right\| \leq \left\|S_m'\dot{\Omega}_m^{-1}(\dot{\Omega}_m - \hat{\Omega}_m)\dot{\Omega}_m^{-1}S_m\right\|
$$

$$
+ \left\|\dot{\Omega}^{-1}(\hat{\Omega} - \dot{\Omega})S_m'\dot{\Omega}_m^{-1}S_m\right\|
$$

$$
= o_p(1/\sqrt{K}).
$$

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

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

$$
= o_p(1).
$$

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

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

[44]

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 $\sum_{i=1}^n \rho_1(\hat{\lambda}'_m\hat{g}_{mi})\hat{g}_i/n = S'_a\sum_{i=1}^n \rho_1(\hat{\lambda}'_m\hat{g}_{mi})\hat{g}_{ai}/n$,

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

Thus by CS and T

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

as $\dot{\Omega}\hat{\Omega}^{-1}\dot{\Omega} - \hat{\Omega} = o_p(1/\sqrt{K})$ and $\|\hat{\lambda}\|$, $\|\hat{\lambda}_m\|$ 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,

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

[45]

Likewise

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

Therefore,

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

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

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

Condition (a): immediate.

Condition (b):

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

Condition (c): note that

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

Hence

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

Condition (d):

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

Thus

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

[46]

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),

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

∎

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

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

and

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

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 $\Omega_{mn} = S_m \Omega_n S'_m$. Then, by Lemma A.8,

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

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

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

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

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

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

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

[48]

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

$$\left| n\bar{g}_n'(\Omega_n^{-1} - \bar{\Omega}_n^{-1})\bar{g}_n \right| / \sqrt{2J_a M K} \leq n \left\| \Omega_n^{-1}\bar{g}_n \right\|^2 (\left\| \Omega_n - \bar{\Omega}_n \right\| + C \left\| \Omega_n - \bar{\Omega}_n \right\|^2) / \sqrt{2J_a M K}$$

$$= n \left\| \Omega_n^{-1}\bar{g}_n \right\|^2 O_p(\zeta(K)\sqrt{K/n}) / \sqrt{2J_a M K} = o_p(1)$$

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

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

Secondly, it is shown that

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

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

$$E[\|u_{i,n} - E[u_{i,n}|s_i]\|^4] \leq 8(E[\|u_{i,n}\|^4] + E[\|E[u_{i,n}|s_i]\|^4])$$

$$= 8(E[E[\|u_{i,n}\|^4 |s_i]) + E[\frac{J_a M K}{n^2}\|\xi_i\|^4]])$$

$$\leq C$$

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,

$$\bar{g}_n' \bar{\Omega}_n^{-1} \tilde{g}_n = \frac{\sqrt[4]{J_a M K}}{n} \sum_{i,j=1}^n (\xi_i \otimes q_i)' S' \bar{\Omega}_n^{-1} \tilde{g}_{j,n} / n\sqrt{n}$$

$$= O_p(\sqrt[4]{J_a M K}/n).$$

Next, by hypothesis,

$$\left| n\bar{g}_n'(\Omega_n^{-1} - \bar{\Omega}_n^{-1})\tilde{g}_n \right| / \sqrt{2J_a M K} \leq n \left\| \Omega_n^{-1}\bar{g}_n \right\| \left\| \Omega_n^{-1}\tilde{g}_n \right\| (\left\| \Omega_n - \bar{\Omega}_n \right\| + C \left\| \Omega_n - \bar{\Omega}_n \right\|^2) / \sqrt{2J_a M K}$$

$$= n \left\| \Omega_n^{-1}\bar{g}_n \right\| \left\| \Omega_n^{-1}\tilde{g}_n \right\| O_p(\zeta(K)\sqrt{K/n}) / \sqrt{2J_a M K} = o_p(1)$$

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 $n\bar{g}_n' S_m' \Omega_{mn}^{-1} S_m \tilde{g}_n / \sqrt{2J_a M K} = o_p(1)$.

Finally, Lemma A.2 is invoked to prove

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

[49]

First, $tr(\Omega_n P_n) = J_a M K$. 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,

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

Next,

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

Lastly,

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

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,

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

The conclusion of the theorem then follows. ∎

References

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

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.

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

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.

Delgado, M.A., Dominguez, M.A., and P. Lavergne (2006): "Consistent Tests of Conditional Moment Restrictions," *Annales d'Economie et de Statistique*, 81, 33-67.

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.

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.

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.

[52]

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.

Hirano, K., Imbens, G.W., and G. Ridder (2003): "Efficient Estimation of Average Treatment Effects using the Estimated Propensity Score," *Econometrica*, 71, 1161-1189.

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.

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

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.

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

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

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

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

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

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

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

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. (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.

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

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.

Table B.1. ME Tests Null Hypothesis Empirical Rejection Frequencies

K	M	n	$\mathcal{J}^{\mathrm{M}}$	$\mathcal{LR}^{\mathrm{M}}_{\mathrm{CUE}}$	$\mathcal{LR}^{\mathrm{M}}_{\mathrm{EL}}$	$\mathcal{LM}^{\mathrm{M}}_{\mathrm{EL}}$	$\mathcal{S}^{\mathrm{M}}_{\mathrm{EL}}$	$\mathcal{LR}^{\mathrm{M}}_{\mathrm{ET}}$	$\mathcal{LM}^{\mathrm{M}}_{\mathrm{ET}}$	$\mathcal{S}^{\mathrm{M}}_{\mathrm{ET}}$
		200	4.8	4.9	6.0	6.8	5.3	6.0	7.8	4.9
	1	500	4.7	4.8	5.4	5.5	4.9	5.5	6.6	4.8
		1000	4.2	4.2	4.4	4.4	4.4	4.5	4.9	4.3
2		3000	4.9	4.9	5.0	4.9	4.9	5.0	5.1	4.9
		200	4.4	4.5	8.2	12.5	5.1	4.5	6.9	4.7
	2	500	4.8	4.9	6.2	7.2	5.2	5.0	4.6	5.1
		1000	4.2	4.3	4.9	4.9	4.3	4.6	3.9	4.3
		3000	5.6	5.6	5.7	5.6	5.6	5.8	5.9	5.6
		200	4.6	4.6	9.8	15.0	6.2	5.1	8.0	5.5
	1	500	4.9	4.9	6.6	7.9	5.7	5.5	5.1	5.3
		1000	4.3	4.3	5.3	5.5	4.7	5.0	4.4	4.5
4		3000	5.5	5.5	5.8	5.6	5.8	5.9	6.2	5.6
		200	4.3	4.3	19.0	39.9	6.7	1.4	5.0	5.2
	2	500	4.3	4.2	9.0	15.1	5.2	0.8	1.7	4.7
		1000	4.3	4.4	6.2	7.8	5.0	0.6	0.9	4.6
		3000	5.7	5.7	6.1	5.9	5.8	1.2	1.2	5.7

Table B.2. CE Tests Null Hypothesis Empirical Rejection Frequencies

K	M	n	$\mathcal{J}^{\mathrm{C}}$	$\mathcal{LR}^{\mathrm{C}}_{\mathrm{CUE}}$	$\mathcal{LR}^{\mathrm{C}}_{\mathrm{EL}}$	$\mathcal{LM}^{\mathrm{C}}_{\mathrm{EL}}$	$\mathcal{S}^{\mathrm{C}}_{\mathrm{EL}}$	$\mathcal{LR}^{\mathrm{C}}_{\mathrm{ET}}$	$\mathcal{LM}^{\mathrm{C}}_{\mathrm{ET}}$	$\mathcal{S}^{\mathrm{C}}_{\mathrm{ET}}$
		200	4.6	4.7	8.7	13.9	5.5	8.2	14.9	5.1
	2	500	4.4	4.5	5.9	7.0	4.7	5.9	8.6	4.6
		1000	4.5	4.5	5.4	5.6	4.7	5.2	6.6	4.6
2		3000	5.0	5.0	5.2	5.2	5.1	5.3	5.7	5.0
		200	2.7	2.6	38.4	80.8	3.7	1.8	17.4	3.1
	8	500	3.7	3.6	15.8	40.0	4.1	1.6	5.7	3.8
		1000	4.1	4.0	8.9	18.6	4.4	1.9	3.5	4.2
		3000	5.3	5.3	6.8	8.1	5.4	1.9	2.3	5.3
		200	4.4	4.5	10.5	17.9	6.8	9.3	17.9	5.3
	1	500	4.6	4.7	6.7	8.3	5.5	6.6	9.8	5.2
		1000	4.5	4.5	5.6	6.2	5.1	5.7	7.2	4.8
4		3000	4.8	4.8	5.0	4.8	5.0	5.1	5.4	4.9
		200	2.6	2.3	51.2	91.0	5.9	1.7	18.2	3.7
	4	500	3.7	3.8	20.7	53.0	5.3	1.6	7.5	4.3
		1000	4.1	4.2	11.8	26.3	5.1	1.4	3.9	4.5
		3000	5.2	5.2	7.3	9.8	5.5	1.8	2.2	5.4

T.1

Table B.3. ME Tests Size-Corrected Power

n	K	M	τ	$\mathcal{J}^{\mathrm{M}}$	$\mathcal{LR}^{\mathrm{M}}_{\mathrm{CUE}}$	$\mathcal{LR}^{\mathrm{M}}_{\mathrm{EL}}$	$\mathcal{LM}^{\mathrm{M}}_{\mathrm{EL}}$	$\mathcal{S}^{\mathrm{M}}_{\mathrm{EL}}$	$\mathcal{LR}^{\mathrm{M}}_{\mathrm{ET}}$	$\mathcal{LM}^{\mathrm{M}}_{\mathrm{ET}}$	$\mathcal{S}^{\mathrm{M}}_{\mathrm{ET}}$
			0.05	7.6	7.5	7.6	7.3	7.5	7.8	7.7	7.7
		1	0.1	15.8	15.8	15.9	15.8	15.6	16.1	16.2	15.9
			0.15	30.4	30.3	30.2	29.6	29.7	30.6	30.3	30.3
	2		0.2	49.2	49.3	49.5	49.3	48.7	50.1	50.0	49.4
			0.05	6.3	6.4	6.5	6.5	6.4	6.9	6.7	6.3
		2	0.1	12.1	12.2	12.0	11.6	11.7	13.7	13.7	12.0
			0.15	22.2	23.0	22.9	21.3	21.8	26.4	27.0	22.3
200			0.2	38.8	39.6	39.8	36.1	38.1	44.4	45.1	38.8
			0.05	6.0	6.1	6.4	6.2	6.4	6.3	7.2	6.0
		1	0.1	11.6	11.8	12.2	11.2	11.6	12.9	14.1	11.4
			0.15	21.4	22.0	23.2	21.1	21.4	24.9	27.0	21.4
	4		0.2	37.1	37.4	39.5	35.7	37.4	42.1	44.5	37.1
			0.05	6.1	5.9	6.2	5.2	6.1	6.2	6.2	5.8
		2	0.1	8.0	8.3	9.2	6.8	8.3	11.8	11.2	8.1
			0.15	13.7	14.4	15.4	10.3	14.2	22.4	22.1	14.0
			0.2	23.9	25.1	26.3	15.2	23.4	38.4	37.8	23.9
			0.05	11.5	11.5	11.8	11.3	11.8	11.8	11.6	11.7
		1	0.1	34.5	34.3	34.3	33.3	34.6	34.4	34.1	34.4
			0.15	67.2	67.4	67.2	65.8	67.5	67.4	67.0	67.5
	2		0.2	90.8	90.8	90.8	90.3	90.9	90.9	90.7	90.8
			0.05	9.3	9.3	9.2	9.2	9.2	9.7	10.3	9.2
		2	0.1	26.4	26.5	26.6	25.7	26.4	28.4	30.6	26.3
			0.15	56.1	56.2	55.8	54.1	56.0	58.2	62.1	55.9
500			0.2	83.5	83.6	83.5	81.8	83.5	85.3	87.6	83.5
			0.05	9.3	9.5	9.2	8.5	8.8	9.6	10.3	9.1
		1	0.1	26.1	26.7	26.0	24.2	25.1	28.1	30.3	25.9
			0.15	55.5	56.2	55.0	52.8	55.0	57.8	61.5	55.7
	4		0.2	83.2	83.6	82.7	80.7	82.8	84.8	87.0	83.1
			0.05	7.8	7.6	7.9	7.3	7.8	8.7	9.3	7.9
		2	0.1	19.3	19.2	19.0	15.9	18.8	25.6	26.3	19.2
			0.15	43.5	43.8	42.7	34.9	43.1	54.8	55.3	43.5
			0.2	72.5	73.0	71.7	61.2	72.2	82.4	83.0	73.0

Table B.4. CE Tests Size-Corrected Power

n	K	M	a	$\mathcal{J}^{\mathrm{C}}$	$\mathcal{LR}^{\mathrm{C}}_{\mathrm{CUE}}$	$\mathcal{LR}^{\mathrm{C}}_{\mathrm{EL}}$	$\mathcal{LM}^{\mathrm{C}}_{\mathrm{EL}}$	$\mathcal{S}^{\mathrm{C}}_{\mathrm{EL}}$	$\mathcal{LR}^{\mathrm{C}}_{\mathrm{ET}}$	$\mathcal{LM}^{\mathrm{C}}_{\mathrm{ET}}$	$\mathcal{S}^{\mathrm{C}}_{\mathrm{ET}}$
			0.025	9.1	9.2	9.7	10.0	9.2	9.6	9.8	9.1
		2	0.05	24.7	24.5	27.3	28.8	25.0	26.3	28.8	24.9
			0.075	50.7	50.5	55.9	58.4	51.4	54.3	57.5	51.2
	2		0.1	78.4	78.4	82.9	83.7	79.3	81.8	83.8	79.2
			0.025	6.4	6.1	8.0	7.8	6.2	7.4	9.3	6.1
		8	0.05	12.7	12.2	17.5	17.3	12.5	20.3	23.2	12.1
			0.075	25.6	25.0	37.6	34.0	25.2	42.7	47.5	25.1
200			0.1	46.0	45.0	63.5	55.0	46.2	70.3	73.8	45.8
			0.025	9.8	9.9	10.6	10.5	9.8	10.3	10.8	10.1
		1	0.05	26.6	26.8	29.9	30.9	26.9	28.6	31.4	27.3
			0.075	53.9	53.4	60.7	62.0	55.1	58.7	62.4	55.2
	4		0.1	79.5	79.0	85.8	85.9	81.6	84.3	87.1	80.9
			0.025	5.9	5.7	7.4	7.9	6.3	7.7	9.6	6.1
		4	0.05	11.5	11.2	15.6	16.3	12.2	20.4	23.8	12.1
			0.075	23.8	23.3	34.1	31.4	24.2	43.6	49.4	24.2
			0.1	42.8	41.9	58.6	51.0	44.5	70.5	75.3	44.2
			0.025	18.9	18.7	19.5	19.6	19.4	19.2	19.7	19.1
		2	0.05	67.5	67.5	69.4	70.3	68.3	68.9	70.7	68.0
			0.075	95.8	95.8	96.3	96.9	95.9	96.0	96.7	95.9
	2		0.1	99.8	99.8	99.9	99.9	99.8	99.9	99.9	99.8
			0.025	11.1	11.1	13.1	12.0	11.4	15.0	17.0	11.3
		8	0.05	41.4	41.2	47.7	39.5	41.6	55.3	59.0	41.6
			0.075	81.5	81.4	88.3	79.2	81.9	92.0	93.7	81.7
500			0.1	98.1	98.0	99.1	96.6	98.2	99.8	99.8	98.2
			0.025	19.7	19.7	21.0	21.6	20.3	20.8	21.4	20.0
		1	0.05	69.9	69.9	73.5	73.2	71.4	72.7	74.2	70.8
			0.075	96.5	96.5	97.6	97.5	96.9	97.3	97.7	96.7
	4		0.1	99.9	99.9	100.0	99.9	100.0	100.0	100.0	99.9
			0.025	10.8	10.9	12.3	11.1	10.6	15.6	17.1	10.8
		4	0.05	38.7	38.5	45.8	37.0	38.7	56.0	58.7	39.2
			0.075	79.0	78.9	86.7	73.1	79.6	92.4	93.9	79.8
			0.1	97.4	97.3	98.9	94.4	97.6	99.6	99.7	97.7