Additional Moment Condition Tests

Paulo M.D.C. Parente  Richard J. Smith
Department of Economics  cemmap
University of Exeter  U.C.L and I.F.S.
and
Faculty of Economics
University of Cambridge

This draft: September 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.

*The authors are grateful for comments on versions of this paper by participants at the CEMFI-UC3M Econometrics Seminar and the Universities of Cambridge and Surrey.
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. 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 asymptoptic 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 of interest include moment conditional homoskedasticity and instrument validity with particular definitions of exogeneity as special cases of the latter.

The approach of the paper exploits an equivalence between conditional moment constraints and a countably infinite number of unconditional restrictions, see Chamberlain (1987), with test statistics 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 approximate conditional moments by an appropriate finite set of unconditional moment conditions. 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 generally 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 the number of unconditional moments to grow with sample size at an appropriate rate.1 Likewise here

1Consistent 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).
both maintained and null hypothesis conditional moment constraints are approximated by corresponding sets of unconditional moment restrictions with the first a subset of the second 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, a result that intuitively 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 diverge to infinity. A similar result obtains for unrestricted statistics although the limit standard normal variate differs. Interestingly, unlike finite dimensional test statistics, efficient parameter estimation is no longer required for test implementation.\(^2\) Under a suitable sequence of local alternatives, restricted and unrestricted test statistics are asymptotically non-central standard normally distributed. Importantly, the non-centrality parameter of the restricted statistics exceeds that of the unrestricted statistics demonstrating the deficiency of the latter tests mirroring the results in the classical paramatric likelihood setting. Moreover, since both non-centrality parameters are non-negative one-sided tests of the additional conditional moment restrictions are apposite. The article also contains a number of subsidiary theoretical contributions that generalise some DIN results to address the null and local alternative asymptotic distributions of the test statistics.

The paper is organized as follows. Section 2 provides some initial definitions, details the test problem and describes moment conditional homoskedasticity and instrument validity examples that are used throughout the paper. GMM and GEL restricted 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 deficiency of the latter tests. 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.

2 Some Preliminaries

2.1 Definitions

The maintained hypothesis is defined in terms of the moment indicator vector given by $u_m(z, \beta_m)$, where $u_m(z, \beta_m)$ is a $J_m$-vector of known functions of the random vector of data observables $z$ and the $p_m$-vector of parameters $\beta_m$. 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

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

(2.1)

for some unknown value $\beta_{m0} \in B_m$ of the parameter vector $\beta_m$ where $B_m$ denotes the corresponding parameter space.

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

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

(2.2)

for some $\beta_{a0} \in B_a$. 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.
\(a\) with \(B_a\) the corresponding parameter space and \(s_a\) an observable vector of instruments. Together the parameter vectors \(\beta_{m0}\) and \(\beta_{a0}\) constitute the objects of inferential interest. Note that \(\beta_a\) may or may not be coincident with the maintained hypothesis parameter vector \(\beta_m\). 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
\[
E[u_m(z, \beta_{m0})|s_m] = 0 \quad (2.1)
\]
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.,
\[
H_0 : E[u_a(z, \beta_{a0})|s_a] = 0, E[u_m(z, \beta_{m0})|s_m] = 0 \quad (2.3)
\]
with the corresponding alternative hypothesis \(H_1\) given by
\[
H_1 : E[u_a(z, \beta_{a0})|s_a] \neq 0, E[u_m(z, \beta_{m0})|s_m] = 0. \quad (2.4)
\]

### 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
\[
u_a(z, \beta_a) = v(u_m(z, \beta_m)u_m(z, \beta_m)' - \Sigma)
\]

\(^3\)The theoretical analysis may in principle be straightforwardly adapted and extended for 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).
where \( v(\cdot) \) denotes the vectorised upper triangle of \( \cdot \). Thus \( J_a = J_m(J_m + 1)/2 \) and \( \beta_a = (\beta'_m, v(\Sigma)')' \) includes the maintained parameter vector \( \beta_m \). 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 \( H_1 : \Sigma_0(s_m) \neq \Sigma_0 \text{ for all } s_m \in S_m \), where \( S_m \) has non-zero probability mass, and \( E[u_m(z, \beta_m0)|s_m] = 0 \).

**Remark 2.1:** In the standard linear regression model \( u_m(z, \beta_m) = y - \beta_m x \) with maintained unconditional moment indicator vector \( s_m u_m(z, \beta_m) = s_m(y - \beta_m x) \), setting \( J_m = 1 \), and thus \( J_a = 1 \), CUE estimation of \( \beta_m \) would employ the inverse of the sample moment \( \sum_{i=1}^{n} s_m s_m' (y_i - \beta_m x_i)^2 / n \) as metric whereas, under conditional homoskedasticity, the LIML metric, i.e., the inverse of \( \sigma^2_n(\beta_m) \sum_{i=1}^{n} s_m s_m' / n \), where \( \sigma^2_n(\beta_m) = \sum_{i=1}^{n} (y_i - \beta_m x_i)^2 / n \), is apposite.

**Example 2.2 (Instrument Validity)**

Here 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 in this example is the validity of the additional instrument vector \( s_a \). The null hypothesis is therefore defined by

\[
H_0 : E[u_m(z, \beta_m0)|s_a] = 0, E[u_m(z, \beta_m0)|s_m] = 0
\]

with alternative hypothesis \( H_1 : E[u_m(z, \beta_m)|s_a] \neq 0 \text{ for all } \beta_m \in B_m \text{ and } s_a \in S_a \), where \( S_a \) has non-zero probability mass, and \( E[u_m(z, \beta_0)|s_m] = 0 \).

**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 but does not include the maintained instrument vector \( s_m \). In linear regression, see Remark 2.1 above, if \( x \) but not \( s_m \) is included in \( s_a \), then this hypothesis may be regarded as a marginal form of exogeneity hypothesis (ME) given by \( E[y - \beta_m0 x|s_a] = 0 \), i.e.,
\[ E[y|s_a] = \beta_{m0}x. \] Hence, LS estimation of \( \beta_{m0} \) is consistent but inefficient in the presence of conditional heteroskedasticity, see Cragg (1983), since only the conditional moment 
\[ E[y - \beta_{m0}x|x] = 0 \]
is used with the additional conditional moments 
\[ E[y - \beta_{m0}x|s_a] = 0 \]
partially and the maintained conditional moments 
\[ E[y - \beta_{m0}x|s_m] = 0 \]
neglected. Another interesting case is when the additional instrument vector \( s_a \) includes \( s_m \). For linear regression, this \textit{conditional} exogeneity hypothesis (CE) is given by 
\[ E[y - \beta_{m0}x|s_a] = 0, \]
i.e., 
\[ E[y|s_a] = \beta_{m0}E[x|s_a]. \]
When \( x \) is a strict subvector of \( s_a \), \( E[y|s_a] = \beta_{m0}x \) but again although LS is consistent for \( \beta_{m0} \) IV estimation using the conditional moments 
\[ E[y - \beta_{m0}x|s_a] = 0 \]
dominates LS. Note that CE is more stringent than ME. If \( x \) is a control variable for a policy maker its average effect on \( y \) for both ME and CE is predictable since 
\[ E[y|x] = \beta_{m0}x \]
when \( x \) is included in \( s_a \). If \( s_m \) is a control its impact on \( y \) requires 
\[ E[x|s_m] \]
since for both ME and CE 
\[ E[y|s_m] = \beta_{m0}E[x|s_m], \]
i.e., the condition 
\[ E[y - \beta_{m0}x|s_a] = 0 \]
is uninformative although under CE the effect of \( s_m \) on \( y \) given \( x \) is nil since 
\[ E[y|s] = \beta_{m0}x. \]
The model design in section 6 provides an example.

3 GMM and GEL Test Statistics

3.1 Approximating Conditional Moment Restrictions

Conditional moment conditions 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 \( q^K(s) = (q_{1K}(s), ..., q_{KK}(s))^t \) denote a \( K \)-vector of approximating functions.

\textbf{Assumption 3.1} For all \( K \), \( E[q^K(s)q^K(s)^t] \) is finite and for any \( a(s) \) with \( E[a(s)^2] < \infty \) there are \( K \)-vectors \( \gamma_K \) such that as \( K \to \infty \),
\[ E[(a(s) - q^K(s)^t\gamma_K)^2] \to 0. \]
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 $S$ 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 $s \in S$, $\sup_{s \in S} \left\| \tilde{q}^K(s) \right\| \leq \zeta(K)$, $E[\tilde{q}^K(s)\tilde{q}^K(s)']$ 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

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

for approximating functions $q^K_m(s_m)$ satisfying Assumptions 3.1 and 3.2. Likewise let $q^{MK}_a(s_a)$ 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

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

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

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

**Remark 3.1:** Strictly speaking the analysis below requires the dimension of $q^{MK}_a(\cdot)$ to be $O(K)$, i.e., of the same order as that of $q^K_m(\cdot)$. The choice $MK$ is adopted for simplicity and for ease of implementation. The restricted test statistics for (2.3) defined below are expressed as (or are asymptotically equivalent to) the difference of component statistics with their large sample behaviour determined by the relative number of approximating functions used to express the null and maintained hypotheses in unconditional form. If the dimension of $q^{MK}_a(\cdot)$ diverges at a rate different from $O(K)$, the component statistic with the dominant number of approximating functions governs the asymptotic behaviour of the restricted test statistic. More precisely, the $O(K)$ choice for the dimension of $q^{MK}_a(\cdot)$ guarantees that both component statistics are of the same asymptotic order and thus that neither set of moment restrictions is negligible asymptotically. In particular, in an asymptotic sense, the restricted tests thereby fully incorporate the maintained
hypothesis. If, however, the dimension of $q^M_K(\cdot)$ diverges at a rate greater than $O(K)$, the component statistic arising from the maintained hypothesis vanishes asymptotically and the resultant restricted test statistics takes the *unrestricted* form asymptotically. This assumption also has an important implication for the local power properties of both forms of test statistics ensuring that the restricted tests dominate the unrestricted forms of test; see section 5.

**Remark 3.2:** Note that the dimension $K$ is unrelated to those of $s_m$ or $s_a$. Section 6 discusses how the dimension of $q^M_K(\cdot)$ may be restricted to $O(K)$ in practice. Note also that Assumption 3.1 is satisfied with $K$ replaced by $MK$ or $O(K)$ for any finite choice of $s$.

**Example 2.1 (Conditional Homoskedasticity Cont.)**

Recall that $u_a(z, \beta_a) = v(u_m(z, \beta_m)u_m(z, \beta_m)' - \Sigma)$ with $\beta_a = (\beta'_m, v(\Sigma)')'$. In this case $s_a = s_m$ and thus the additional approximating functions are defined as $q^M_K(s_a) = q^K_m(s_m)$. Therefore $M = 1$. Hence, the null hypothesis $H_0 : \Sigma_0(s_m) = \Sigma_0, E[u_m(z, \beta_{m_0})|s_m] = 0$ is re-expressed in unconditional form as

$$E[u_a(z, \beta_{a_0}) \otimes q^K_m(s_m)] = 0, E[u_m(z, \beta_{m_0}) \otimes q^K_m(s_m)] = 0, K \to \infty.$$ 

**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 $q^M_K(s_a)$ with dimension $MK$. Thus, the null hypothesis $H_0 : E[u_m(z, \beta_{m_0})|s_a] = 0, E[u_m(z, \beta_{m_0})|s_m] = 0$ is re-expressed in unconditional form as

$$E[u_m(z, \beta_{m_0}) \otimes q^M_K(s_a)] = 0, E[u_m(z, \beta_{m_0}) \otimes q^K_m(s_m)] = 0, K \to \infty.$$ 

**Remark 3.3:** For regression the special cases $ME s_a = x$ with $q^M_K(s_a)$ functions of $x$ only and $CE s_a = (s_m, x)$ with $q^M_K(s_a)$ additional functions of $s_m$ and $x$ are of particular interest.
3.2 Basic Assumptions and Notation

Let $\beta$ denote the distinct elements of $\beta_m$ and $\beta_a$ with $\beta_0$ and the composite parameter space $\mathcal{B}$ defined similarly; $p$ as the number of parameters comprising $\beta$ in addition to those in $\beta_m$. 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 $q^K_m(s_m)$ and $q^K_a(s_a)$ respectively. It will be helpful to define a number of f.r.r. selection matrices $S^a_m$, $S^a_a$, $S^u_m$ and $S^u_a$; viz. $S^a_m q^K(s) = q^K_m(s_m)$, $S^a_a q^K(s) = q^K_a(s_a)$, $S^u_m u(z, \beta) = u_m(z, \beta_m)$ and $S^u_a u(z, \beta) = u_a(z, \beta_a)$. Correspondingly $S_m = S^u_m \otimes S^a_m$ and $S_a = S^u_a \otimes S^a_a$ 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 $q^K(s) = q^K_m(s_m)$. Hence $S^a_m = S^a_a = I_K$ and $S^u_m = (I_{J_m}, 0_{(J_m \times J_a)})$, $S^u_a = (0_{(J_a \times J_m)}, I_{J_a})$. 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

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

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

**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 $S^u_m = S^u_a = I_{J_m}$ and $S^a_m = (I_K, 0_{(K \times MK)})$, $S^a_a = (0_{(MK \times K)}, I_{MK})$. The unconditional moment indicator vector $u_m(z, \beta_m) \otimes (q^K_{s_m})', q^K_{s_a}(s_a)'$ corresponding to the null hypothesis

[10]
$H_0 : E[u_m(z, \beta_m)|s_m] = 0, E[u_m(z, \beta_m)|s_m] = 0$ may equivalently be re-arranged as

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

with that for the maintained hypothesis given by $S_m(u(z, \beta) \otimes q^K(s)) = u_m(z, \beta_m) \otimes q^K_m(s_m), K \to \infty$, 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 $\beta_0 \in \text{int}(B)$ such that $E[u_m(z, \beta_m)|s_m] = 0$ and $E[u_a(z, \beta_0)|s_a] = 0$; (c) $\sqrt{n}(\hat{\beta} - \beta_0) = O_p(1)$; (d) $E[\sup_{\beta \in B} \|u(z, \beta)\|^2 |s]$ is bounded.

Unlike DIN Assumption 6(b), p.67, it is unnecessary to impose $E[\sup_{\beta \in B} \|u(z, \beta)\|^{\gamma}] < \infty$ for some $\gamma > 2$ for GEL.\(^4\) 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 $u_{\beta \beta_j}(z, \beta) = \partial^2 u_j(z, \beta)/\partial \beta \partial \beta'$, $j = 1, \ldots, J$. Also let $\mathcal{N}$ denote a neighbourhood of $\beta_0$.

**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 $E[\|u_{\beta \beta_j}(z, \beta_0)\|^2 |s], (j = 1, \ldots, J)$, 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 $\beta \in \mathcal{N}$, $\|u(z, \beta) - u(z, \beta_0)\| \leq \delta(\beta) \|\beta - \beta_0\| \text{ and } E[\delta(\beta)^2 |s]$ is bounded; (e) $E[D(s)^T D(s)]$ is nonsingular.

### 3.3 Test Statistics

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

---

\(^4\)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 Lemmata and Theorems concerned with GEL.
GMM statistics appropriate for tests of maintained and null hypotheses expressed unconditionally in (3.1) and (3.3) take the standard forms

\[ T_{GMM}^g = n g_m(\hat{\beta}_m)' \hat{\Omega}_m^{-1} \hat{g}_m(\hat{\beta}_m) \]  

(3.4)

and

\[ T_{GMM}^g = n \hat{g}(\hat{\beta})' \hat{\Omega}^{-1} \hat{g}(\hat{\beta}) \]  

(3.5)

where \( \hat{\Omega}_m = \sum_{i=1}^n g_{mi}(\hat{\beta}_m)g_{mi}(\hat{\beta}_m)'/n \) 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 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 respective revised unconditional moment hypotheses; viz.

\[ J' = T_{GMM}^g - T_{GMM}^g - (J_aMK - p) \]

(3.6)

\[ \sqrt{2(J_aMK - p)} \]

where \( p \) is the number of additional parameters in \( \beta_a \) defining the additional conditional moment conditions (2.2) as compared with the maintained hypothesis (2.1) parameters \( \beta_m \).

Remark 3.4: For fixed and finite \( K \), GMM and GEL test statistics [Newey (1985a), Smith (2011)] for the validity of additional moment restrictions, e.g., \( T_{GMM}^g - T_{GMM}^g \), are under suitable conditions asymptotically chi-square distributed with \( J_aMK - p \) degrees of freedom where \( p \) denotes the number of parameters in \( \beta \) in addition to those in \( \beta_m \). The mean location \( J_aMK - p \) and standard deviation scale \( \sqrt{2(J_aMK - p)} \) standardisations of \( T_{GMM}^g - T_{GMM}^g \) in \( J' \) (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

$$
\hat{P}_g^\lambda(\beta, \lambda) = \frac{1}{n} \sum_{i=1}^{n} [\rho(\lambda' g_i(\beta)) - \rho_0]/n,
$$

$$
\hat{P}_{g_m}^\lambda_m(\beta_m, \lambda_m) = \frac{1}{n} \sum_{i=1}^{n} [\rho(\lambda_m' g_{mi}(\beta_m)) - \rho_0]/n,
$$

where $\lambda$ 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 null and maintained moment constraints. Let $\rho_j(v) = \partial^j \rho(v)/\partial v^j$ and $\rho_j = \rho_j(0)$, ($j = 0, 1, 2, \ldots$) where, without loss of generality, the normalisation $\rho_1 = \rho_2 = -1$ is imposed.$^5$

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

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

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

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

$$
\mathcal{L}R^r = \frac{2n(\hat{P}_g^\lambda(\hat{\beta}; \hat{\lambda}) - \hat{P}_{g_m}^\lambda_m(\hat{\beta}_m, \hat{\lambda}_m)) - (J_a M K - p)}{\sqrt{2(J_a M K - p)}}.
$$

Restricted Lagrange multiplier, score and Wald-type statistics are defined respectively

$^5$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.
As a result of
\[ L_M^r = \frac{n(\hat{\lambda} - S_m'\hat{\lambda}_m)\hat{\Sigma}(\hat{\lambda} - S_m'\hat{\lambda}_m) - (J_{aMK} - p)}{\sqrt{2(J_{aMK} - p)}}, \quad (3.8) \]

\[ S^r = \sum_{i=1}^{n} \rho_1(\hat{\lambda}'_m g_m(\hat{\beta}_m)) g_a(\hat{\beta}_a)' S_a \hat{\Sigma}^{-1} S_a' \sum_{i=1}^{n} \rho_1(\hat{\lambda}'_m g_m(\hat{\beta}_m)) g_a(\hat{\beta}_a)/n - (J_{aMK} - p) \]

\[ \sqrt{2(J_{aMK} - p)} \quad (3.9) \]

and

\[ W^r = \frac{n\hat{\lambda}' S_a'(S_a \hat{\Sigma}^{-1} S_a')^{-1} S_a \hat{\lambda} - (J_{aMK} - p)}{\sqrt{2(J_{aMK} - p)}}, \quad (3.10) \]

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 \( J^r \) (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 \( s_a \).

**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 
\[ J^r \overset{d}{\rightarrow} N(0,1). \]

6 Alternative restricted score and Wald statistics robust to estimation effects may be defined; viz.

\[ S^r = \sum_{i=1}^{n} \rho_1(\hat{\lambda}'_m g_m(\hat{\beta}_m)) g_a(\hat{\beta}_a)' (\hat{\Sigma}^{-1} - \hat{\Sigma}^{-1} \hat{G}' \hat{\Sigma}^{-1} \hat{G} \hat{\Sigma}^{-1} \hat{G}' \hat{\Sigma}^{-1} \hat{G} \hat{\Sigma}^{-1} \hat{G}') \sum_{i=1}^{n} \rho_1(\hat{\lambda}'_m g_m(\hat{\beta}_m)) g_a(\hat{\beta}_a)/n - (J_{aMK} - p) \]

\[ \sqrt{2(J_{aMK} - p)} \]

\[ W^r = \frac{n\hat{\lambda}'_a (S_a(\hat{\Sigma}^{-1} - \hat{\Sigma}^{-1} \hat{G}' \hat{\Sigma}^{-1} \hat{G} \hat{\Sigma}^{-1} \hat{G} \hat{\Sigma}^{-1} \hat{G}') S_a')^{-1} \hat{\lambda}_a - (J_{aMK} - p)}{\sqrt{2(J_{aMK} - p)}}. \]

See Smith (1997, section II.2, pp.511-514) and Smith (2011, section 5, pp.1209-1213).
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 \( J^r \) (3.6).

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

**Remark 4.1**: The large sample analysis for the behaviour of both restricted statistics and unrestricted discussed below under suitable sequences of local alternatives indicates that one-sided tests of the null hypothesis \( H_0 \) (2.3) are appropriate. E.g., the critical region \( \{ J^r \geq z_\alpha \} \) for the *standardised* GMM statistic \( J^r \) (3.6) has asymptotic size \( \alpha \) where \( \mathcal{P}\{N(0,1) \geq z_\alpha \} = \alpha \). Alternatively, critical regions based on *non-standardised* statistics may also be defined based on fixed and finite \( K \) asymptotics. E.g., for \( T^g_{GMM} - T^g_{GMM} \), the critical region \( \{ T^g_{GMM} - T^g_{GMM} \geq \chi^2_{JaMK-p,\alpha} \} \) is opposite where \( \{ \chi^2_{JaMK-p} \geq \chi^2_{JaMK-p,\alpha} \} = \alpha \) and \( \chi^2_{JaMK-p} \) denotes a chi-square distributed random variate with \( JaMK-p \) degrees of freedom. Note that \( p \) is negligible in the large \( K \), large \( n \) asymptotic analysis for 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

\[
LR^u = \frac{2n\hat{\beta} - ((J_a M + J_m)K - p - p_m)}{\sqrt{2((J_a M + J_m)K - p - p_m)}},
\]

\[
LM^u = \frac{n\hat{\lambda} - ((J_a M + J_m)K - p - p_m)}{\sqrt{2((J_a M + J_m)K - p - p_m)}},
\]

and the unrestricted GMM statistic based on \( T^g_{GMM} \) which takes the score form

\[
S^u = \frac{n\hat{\beta}'\hat{\Omega}^{-1}\hat{\beta} - ((J_a M + J_m)K - p - p_m)}{\sqrt{2((J_a M + J_m)K - p - p_m)}}.
\]
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 $J^r$, $LR^r$, $LM^r$, $S^r$ and $W^r$. The statistics $LR^u$ and $S^u$ are forms of GMM and GEL statistics suggested in DIN section 6, pp.67-71, adapted for testing the null hypothesis (2.3).

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

$$J^m = \frac{T_{GMM}^{g_m} - (J_m K - p_m)}{\sqrt{2(J_m K - p_m)}}; \quad (4.3)$$

viz.

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

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

**Remark 4.2:** Theorem 4.3 has the practical import 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 $J^m$; (b) given (2.1), test (2.2) using $J^r$, with overall asymptotic test size $1 - (1 - \alpha_m)(1 - \alpha_a)$, where $\alpha_m$ and $\alpha_a$ are the respective asymptotic sizes of the individual tests in (a) and (b).

## 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 emphasises the requirement that the dimension of the additional conditional moments approximating functions $q^M_K(s_a)$ should be of the same order as that of the maintained approximating functions
$q_m^K(s_m)$, i.e., $O(K)$, which without loss of generality is assumed to be linear in $K$, i.e., $MK$. As noted in Remark 3.1, the import of this restriction is that it ensures a contribution to local power for the restricted statistics from the maintained conditional moment information (2.1) and thus that restricted tests are more powerful than tests based on the unrestricted 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

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

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 $\xi_m(s) = S_m^m \xi(s)$, 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 $\beta = \beta_m = \beta_a$. For the CE case, i.e., $s = s_a$ and thus the maintained instruments $s_m$ are a subset of $s_a$, $E[\xi(s)|s_m] = 0$. Similarly for ME, 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.,

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

The following assumption is required to describe the asymptotic distribution under local alternatives (5.1) of the statistics in section 3.3.

**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 $\beta \in \mathcal{N}$, $\Sigma(s; \beta) = E[u(z, \beta)u(z, \beta)^T|s]$ and $\Sigma_m(s_m; \beta_m) = E[u_m(z, \beta_m)u_m(z, \beta_m)^T|s_m]$ 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 $J^r$, [17]
Let $\mathcal{L}\mathcal{R}^r, \mathcal{L}\mathcal{M}^r, S^r$ and $W^r$ 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 $J^r$ converges in distribution to a $N(\mu^r/\sqrt{2}, 1)$ random variate, where

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

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

**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 $J^r$, $\mathcal{L}\mathcal{R}^r$, $\mathcal{L}\mathcal{M}^r$, $S^r$ and $W^r$.

The following corollary to Theorem 5.1 details the limiting distribution of the unrestricted statistics $\mathcal{L}\mathcal{R}^u$, $\mathcal{L}\mathcal{M}^u$ and $S^u$ 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 $S^u$ 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 $\mathcal{L}\mathcal{R}^u$, $\mathcal{L}\mathcal{M}^u$ are asymptotically equivalent to $S^u$.

**Remark 5.3:** Corollary 5.1 further justifies restricting the dimension of $q_a^{MK}(s_a)$ to be $O(K)$ or linear in $K$. If $M$ grows with $K$, then $\mu^u$ should differ little from $\mu^r$ 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.
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 $W_r$ (3.10) and $W_r$, see fn.6, 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 in section 5. The full set of simulation results are available from the authors upon request.

All experiments concern the linear regression model

$$y = \beta_0 x + u,$$

(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 $\beta_0$ is set as 0. Consideration is restricted to the single parameter $\beta_0$ 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 $x$.

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) $E[u|s_m] = 0$, i.e., $J_m = 1$. 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., $s_a = (s_m, x)$. [19]
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 $\rho$, $\rho \in (-1, 1)$ and $\rho \neq 0$. Hence, $x$ and $s_m$ are marginally distributed as uniform random variates on the unit interval $[0, 1]$. The values $\rho = 0.3$ and 0.7 are examined although since the results are qualitatively similar only the former are reported here.

Let $v$ be defined by

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

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

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 $s_a = x$

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

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

$$E[u|s_a] = (a[\Phi^{-1}(x)^2 + \Phi^{-1}(s_m)^2 - \left(\frac{1 + \rho^2}{\rho}\right)\Phi^{-1}(x)\Phi^{-1}(s_m) - (1 - \rho^2)]$$

$$+ \tau[\Phi^{-1}(x) - \rho \Phi^{-1}(s_m)])/\text{var}[v],$$

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

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

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. The maintained conditional moment \( E[u(z, \beta)|s_m] \), cf. (2.1), is approximated using the vector of functions \( q^K_m(s_m) \) with elements \( s^k_m \), \( (k = 0, ..., K - 1) \). For ME \( E[u(z, \beta)|x] \) is approximated by a polynomial of order \( MK \) in \( x \), i.e., \( q^{MK}_a(s_a) \) 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 \( q^{MK}_a(s_a) \) with elements \( s^k_m x^l \), \( (k = 0, ..., (MK)^{1/2}, l = 1, ..., (MK)^{1/2}) \), where \( M \) and \( K \) are chosen to ensure that \( (MK)^{1/2} \) is an integer.

6.1.2 Estimators

Efficient estimation methods 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.

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.,

\[ \text{Theorem 8, p.90, in Lorenz (1986) establishes the requisite uniform converge for polynomial approximating functions; cf. Assumption 3.1.} \]

\[ \text{The simplex search algorithm of MATLAB is used to compute GMM and GEL hence ensuring a local optimum. The Newton method is used to locate } \hat{\lambda}(\beta) \text{ for given } \beta \text{ which is required for the profile GEL objective function, e.g., EL and ET. EL computation requires some care since the EL criterion involves the logarithm function which is undefined for negative arguments; this difficulty is avoided by employing the MATLAB code due to Owen in which logarithms are replaced by a function that is logarithmic for arguments larger than a small positive constant and quadratic below that threshold. See Owen (2001, (12.3), p.235); the code is available at http://www-stat.stanford.edu/~owen/empirical/} \]
the non-standardised restricted GEL LR statistic for CE based on EL criteria and estimators is denoted as \( LR_{el}^c = 2n(\hat{P}_{el}(\hat{\beta}_{elc}, \hat{\lambda}_{elc}) - \hat{P}_{el}^{\text{mm}}(\hat{\beta}_{elma}, \hat{\lambda}_{elma})) \), cf. (3.7). The non-standardised robustified score statistic \( \bar{S} \), see fn.6, is also examined. Restricted non-standardised test statistics are calibrated against chi-square distributions with \( MK \) (ME) and \( (MK)^{1/2}(MK)^{1/2} + 1 \) (CE) degrees of freedom respectively.\(^{10}\)

### 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 2 and 8 (CE) since \( (MK)^{1/2} \) is required to be integer.

To be completed

### 6.3 Summary

To be completed

### 7 Conclusions

To be completed

### References


\(^{10}\)A number of asymptotically equivalent test statistics for the maintained hypothesis 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.


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 \( \text{w.p.a.1} \) 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, \( \max_{\beta \in \mathcal{B}, \lambda \in \Lambda_n, 1 \leq i \leq n} |\lambda' g_i (\beta)| \to 0 \) and \( \text{w.p.a.1} \ \Lambda_n \subset \hat{\Lambda}_n (\beta) \) 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

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

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

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 \( k = \text{tr} (\Omega_n C_n) \) 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

\[
T = \frac{n \hat{g}(\beta_{0,n})'C_n \hat{g}(\beta_{0,n}) - k}{\sqrt{2k}} \overset{d}{\to} N(0,1).
\]

**Proof:** Let \( g_{i,n} = g(z_i, \beta_{0,n}), (i = 1, ..., n) \), and write \( T = T_1 + T_2 \) where

\[
T_1 = \sum_{i,j;i<j} \sqrt{\frac{2}{n^2k}} g_{i,n}' C_n g_{j,n},
\]

\[
T_2 = \frac{\sum_i g_{i,n}' C_n g_{i,n} / n - k}{\sqrt{2k}}.
\]

[28]
Since $E[\mathcal{T}_2] = 0$ and $\text{var}[\mathcal{T}_2] \leq E[(g_{1,n} C_n g_{i,n})^2]/2kn \to 0$, $\mathcal{T}_2 \xrightarrow{p} 0$.

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

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

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

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

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

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

On the other hand

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

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

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

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

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

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

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

[29]
Proof: Let \( g_i = g(z_i, \beta_0), \) \((i = 1, \ldots, n)\), and write \( T = T_1 + T_2 \) where

\[
T_1 = \sum_{i, j:i < j} \sqrt{\frac{2}{n^2k}} g'_i Q g_j \\
T_2 = \frac{\sum_i g'_i Q g_i/n - ak}{\sqrt{2k}}
\]

Since \( E[T_2] = 0 \) and \( \text{var}[T_2] \leq E[(g'_i Q g_i)^2]/2kn \to 0 \) by (e), \( T_2 \xrightarrow{P} 0. \)

To prove the asymptotic normality of \( T_1 \), as in the proof of Lemma A.2, we verify the hypotheses of Hall (1984, Theorem 1, pp.3-4). Define

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

Then, if \( u, v \neq 1 \),

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

Now \( E[H_n(z_1, z_2)|z_1] = \sqrt{\frac{2}{n^2k}} g'_1 Q E[g_2] = 0 \) and, by (c),

\[
E[H_n(z_1, z_2)^2] = \frac{2}{n^2k} E[(g'_1 Q g_2)^2] \\
= \frac{2}{n^2k} E[g'_1 Q Q g_1] = \frac{2}{n^2k} \text{tr}([Q Q]) = \frac{2v}{n^2}.
\]

Also

\[
\frac{E[H_n(z_1, z_2)^4]}{nE[H_n(z_1, z_2)^2]^2} = \frac{1}{nv^2k^2} E[(g'_1 Q g_2)^4].
\]

Now, as \( \Omega \) is positive definite, by CS

\[
E[(g'_1 Q g_2)^4] = E[(g'_1 Q \Omega^{-1} g_2)^4] \\
\leq E[(g'_1 Q Q g_1)^2(g'_2 \Omega^{-1} g_2)^2] \\
= E \left( (g'_1 Q Q g_1)^2 \right) E \left( (g'_2 \Omega^{-1} g_2)^2 \right).
\]

Hence, by (f),

\[
\frac{E[H_n(z_1, z_2)^4]}{nE[H_n(z_1, z_2)^2]^2} \leq \frac{1}{nv^2k^2} E \left( (g'_1 Q Q g_1)^2 \right) E \left( (g'_2 \Omega^{-1} g_2)^2 \right) = o(1).
\]
Moreover, by (d),

\[
E[G_n(z_1, z_2)^2] = \frac{4}{n^4 k^2} E[(g'_1 Q\Omega \Omega g'_2)^2] = \frac{4}{n^4 k^2} E[g'_1 \Omega Q \Omega \Omega g'_1] = \frac{4}{n^4 k^2} tr([\Omega]^4) = o(n^{-4}).
\]

Since \(E[G_n(z_1, z_2)^2]/E[H_n(z_1, z_2)^2] = o(1)\), \(T_1 \overset{d}{\rightarrow} N(0, v)\) as required. □

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

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

\[
(\sum_i a'_{i, n} \otimes q'_i) S' \left( S(\sum_i U_{i, n} \otimes q_i q'_i) S' \right)^{-1} S(\sum_i a_{i, n} \otimes q_i) / n - E[a'_i U_i^{-1} a_i] \overset{P}{\rightarrow} 0.
\]

**Proof:** The proof is similar to that of DIN Lemma A.3. Let \(F_i, n\) be a symmetric square root of \(U_i, n\), \(P_i, n = (F_i, n \otimes q'_i)^S\), \(P_n = (P'_{1, n}, \ldots, P'_{n, n})\)' , \(A_i, n = F_i^{-1} a_i, n\), \(A_n = (A'_{1, n}, \ldots, A'_{n, n})\)' , \(\bar{A}_i, n = E[A_i, n | s_i] = F_i^{-1} \bar{a}_i, n\) and \(\bar{A}_n = (\bar{A}'_{1, n}, \ldots, \bar{A}'_{n, n})\)' . Note that \(P_n P_n = S(\sum_i U_{i, n} \otimes q_i q'_i) S'\),

\[
(\sum_i a'_{i, n} \otimes q'_i) S' \left( S(\sum_i U_{i, n} \otimes q_i q'_i) S' \right)^{-1} S(\sum_i a_{i, n} \otimes q_i) = A'_n Q_n A_n
\]

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

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

\[
E[(A_n - \bar{A}_n)' (A_n - \bar{A}_n) | s] = \text{diag}(F_{1, n}^{-1} \text{var}[a_{1, n} | s_1] F_{1, n}^{-1}, \ldots, F_{n, n}^{-1} \text{var}[a_{n, n} | s_n] F_{n, n}^{-1})
\]

\[\leq CI\]

[31]
for $n$ large enough. Let $T_A = (A_n - \bar{A}_n)'Q_n(A_n - \bar{A}_n)/n$. Then,

$$E[T_A] = E[tr(Q_nE[(A_n - \bar{A}_n)(A_n - \bar{A}_n)'/n]s)]$$

$$\leq CE[tr(Q_n)]/n \leq CK/n \to 0$$

as $tr(Q_n) \leq CK$, using (b) and (f). Thus $T_A \to 0$ by M.

From Assumption 3.1, there exists a $K$ such that $E[U_1i_i a_i] \to 0$. Let $S'\gamma_K = vec(\Gamma_K'K)$. Now

$$\left\|\bar{A}_n - P_n\gamma_K\right\|^2/n = \sum_i \left\|F_i\bar{a}_i - (F_i \otimes q_i')S'\gamma_K\right\|^2/n$$

$$= \sum_i \|F_i\|^2 \left\|U_i^{-1}\bar{a}_i - (I_r \otimes q_i')S'\gamma_K\right\|^2/n$$

$$\leq C \sum_i \left\|U_i^{-1}\bar{a}_i - \Gamma_k q_i\right\|^2/n.$$ 

By $c_r$,

$$E[\left\|U_i^{-1}\bar{a}_i - \Gamma_K q_i\right\|^2] = E[\left\|(U_i^{-1} - U_i^{-1})\bar{a}_i + U_i^{-1}(\bar{a}_i - \bar{a}_i)\right\|^2]$$

$$\leq 3E[\left\|(U_i^{-1} - U_i^{-1})\bar{a}_i\right\|^2] + E[\left\|U_i^{-1}(\bar{a}_i - \bar{a}_i)\right\|^2]$$

$$+ E[\left\|U_i^{-1}\bar{a}_i - \Gamma_K q_i\right\|^2].$$

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

$$\left\|\bar{A}_n - P_n\gamma_K\right\|^2/n \to 0.$$

By T and CS

$$\left|A_n'Q_n(A_n - \bar{A}_n)/n\right| = \left|(A_n - \bar{A}_n)'Q_n(A_n - \bar{A}_n)/n\right|$$

$$+ 2\bar{A}_n'Q_n(A_n - \bar{A}_n)/n - \bar{A}_n'(I - Q_n)\bar{A}_n/n$$

$$\leq T_A + 2\sqrt{T_A}\sqrt{\bar{A}_n'\bar{A}_n}/n + T_A,$$
where \( T_A = A_n'(I - Q_n)A_n/n \). Now
\[
T_A = \left( A_n - P_n\tilde{g}_K \right)'(I - Q_n) \left( A_n - P_n\tilde{g}_K \right)/n \\
\leq \| A_n - P_n\tilde{g}_K \|^2/n \overset{p}{\to} 0.
\]

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

\[
\left| A_n'Q_nA_n/n - \tilde{A}_n'A_n/n \right| \overset{p}{\to} 0.
\]

To examine the large sample behaviour of \( \tilde{A}_n'\tilde{A}_n/n = \sum_i a_{i,n}U_i^{-1}a_{i,n}/n \), in particular, to show that \( \tilde{A}_n'\tilde{A}_n/n \overset{p}{\to} E[\tilde{a}_i'U_i^{-1}\tilde{a}_i] \), since \( a_{i,n} \) 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 \( \eta > 0 \),

\[
nP\{\frac{\tilde{a}_i'U_i^{-1}\tilde{a}_i}{n} > \eta \} \to 0.
\]

By M

\[
nP\{\frac{\tilde{a}_i'U_i^{-1}\tilde{a}_i}{n} > \eta \} \leq E\left[ \frac{\tilde{a}_i'U_i^{-1}\tilde{a}_i}{n} \right]^2 / (n\eta^2).
\]

For large enough \( n \), by (a) and (b), \( E[\tilde{a}_i'U_i^{-1}\tilde{a}_i]^2 \) is bounded. Therefore \( nP\{\frac{\tilde{a}_i'U_i^{-1}\tilde{a}_i}{n} > \eta \} \to 0 \). Secondly, for arbitrary \( \varepsilon > 0 \),

\[
n\text{var}\left[ \frac{\tilde{a}_i'U_i^{-1}\tilde{a}_i}{n} \right] 1\left( \frac{\tilde{a}_i'U_i^{-1}\tilde{a}_i}{n} < \varepsilon \right) \leq nE\left[ \frac{\tilde{a}_i'U_i^{-1}\tilde{a}_i}{n} \right]^2 1\left( \frac{\tilde{a}_i'U_i^{-1}\tilde{a}_i}{n} < \varepsilon \right) \\
\leq E\left[ \frac{\tilde{a}_i'U_i^{-1}\tilde{a}_i}{n} \right]^2/n \to 0.
\]

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

\[
E[\tilde{a}_i'U_i^{-1}\tilde{a}_i - \tilde{a}_iU_i^{-1}\tilde{a}_i] \leq E\left[ ||U_i^{-1}|| \tilde{a}_i - \tilde{a}_i \right]^2 \\
+ 2E\left[ ||U_i^{-1}|| \tilde{a}_i - \tilde{a}_i \right] + E\left[ ||U_i^{-1} - U_i^{-1}|| \tilde{a}_i \right]^2 \\
\leq C(E[\|\tilde{a}_i - \tilde{a}_i\|^2] + 2E[\|\tilde{a}_i - \tilde{a}_i\|] + E[\|U_i^{-1} - U_i^{-1}\|]) \\
\to 0.
\]

\[\boxed{\text{The following lemma is similar to DIN Lemma A.4, p.75.}}\]
Lemma A.5 If $q(.)$ 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) $E[\|U_{i,n}^{-1} - U_i^{-1}\|^2] \to 0$ and (e) $K \to \infty$ and $K^2/n \to 0$, then

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

**Proof:** The result is proved by first showing that

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

and secondly that

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

(A.1)

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

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

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

$$
\left|(A_n - \tilde{A}_n)'Q_nB_n/\sqrt{n}\right| \leq \sqrt{(A_n - \tilde{A}_n)'Q_n(A_n - \tilde{A}_n)\sqrt{B_n'Q_nB_n/\sqrt{n}}} = O_p(K/\sqrt{n}) \overset{p}{\to} 0.
$$

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

$$
E[\tilde{A}_n'(I - Q_n)B_n/\sqrt{n}] = E[\tilde{A}_n'(I - Q_n)E[B_nB_n'|s](I - Q_n)\tilde{A}_n]/n \leq CE[\tilde{A}_n'(I - Q_n)\tilde{A}_n]/n \to 0
$$

[34]
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

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

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

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

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

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

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

$$
\hat{\Omega} = \sum_i \hat{g}_i \hat{g}_i' / n, \tilde{\Omega}_n = \sum_i g_{i,n} g_{i,n}' / n,
\Omega_n = S(\sum_i \Sigma_{i,n} \otimes q_i q_i' )S' / n, \Omega_{0,n} = E[ g_{i,n} g_{i,n}' ].
$$

**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 $\tau_n \to 0$, $\|\hat{\Omega} - \Omega_{0,n}\| = O_p(\tau_n K)$, $\|\hat{\Omega}_n - \tilde{\Omega}_n\| = O_p(\zeta(K) \sqrt{K/n})$ and $\|\tilde{\Omega}_n - \Omega_{0,n}\| = O_p(\zeta(K) \sqrt{K/n})$. If Assumption 5.1(c) is satisfied then $1/C \leq \lambda_{\min}(\Omega_{0,n}) \leq \lambda_{\max}(\Omega_{0,n}) \leq C$ and, if $\tau_n K + \zeta(K) \sqrt{K/n} \to 0$, w.p.a.1 $1/C \leq \lambda_{\min}(\hat{\Omega}) \leq \lambda_{\max}(\hat{\Omega}) \leq C$, $1/C \leq \lambda_{\min}(\tilde{\Omega}_n) \leq \lambda_{\max}(\tilde{\Omega}_n) \leq C$. 

[35]
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

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

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 \( E[\sup_{\beta \in B} \|u(z, \beta)\|^2 \|s\|] \) is bounded by Assumption 3.3(d).

Now

\[
E[\|\hat{\Omega}_n - \hat{\Omega}_n\|^2] = E[\|S(\sum_i (u_{i,n}u_{i,n}' - \Sigma_{i,n}(s_i)) \otimes q_iq_i')S' / n\|^2].
\]

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

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

\[
E[\|\hat{\Omega}_n - \Omega_n\|^2] = E[\|S(\sum_i \Sigma_{i,n}(s_i) \otimes q_iq_i')S' / n - \Omega_n\|^2] \\
\leq Ctr(E[\Sigma_{i,n}(s_i)^2 \otimes (q_iq_i')^2]/n) \leq CE[\|q_i\|^4]/n \leq C\zeta(K)^2 K/n
\]

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 \( E[\sup_{\beta \in B} \|u(z, \beta)\|^2 |s] \) is bounded, it
follows that \( C^{-1}I_J \leq \Sigma_{i,n}(s_i) \leq CI_J \) and therefore

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

[36]
Hence $C^{-1} \leq \lambda_{\min}(\Omega_n) \leq \lambda_{\min}(\Omega) \leq C$. Note also that, if $\tau_n K + \zeta(K) \sqrt{K/n} \to 0$, we have $\|\hat{\Omega} - \hat{\Omega}_n\| = o_p(1)$ and $\|\hat{\Omega}_n - \Omega_n\| = o_p(1)$. Thus, by $T \|\hat{\Omega} - \Omega_n\| = o_p(1)$. Since $|\lambda(A) - \lambda(B)| \leq \|A - B\|$, where $\lambda(\cdot)$ denotes the minimum or maximum eigenvalue, $|\lambda_{\min}(\hat{\Omega}) - \lambda_{\min}(\Omega_n)| = o_p(1)$ and $|\lambda_{\max}(\hat{\Omega}) - \lambda_{\max}(\Omega_n)| = o_p(1)$. The final conclusion follows similarly.

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

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

**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 $\hat{G} - G_n = O_p(\tau_n \sqrt{K} + \sqrt{K/n})$ and $\|\hat{G}_n - G_n\| = O_p(\sqrt{K/n})$.

**Proof:** The proof is as in that for DIN Lemma A.7, p.79, and requires no stronger assumptions.

Let $u_{\beta_i}(\beta) = u_{\beta_i}(\beta_{0,n})$, $\delta_i = \delta(z_i)$ and $\hat{G}_n = S'(\sum_i u_{\beta_i} \otimes q_i)/n$. Then by DIN Lemma A.2, p.73,

$$E[\|\hat{G}_n - \bar{G}_n\|^2] = E[\|S'(\sum_i (u_{\beta_i}(z_i, \beta_{0,n}) - D_{i,n}) \otimes q_i)/n\|^2]
\leq CE[E[\|u_{\beta_i} \|^2 |s_i]\|q_i\|^2]/n \leq CK/n,$$

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 $M \|\hat{G}_n - G_n\|^2 = O_p(\sqrt{K/n})$.

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

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

The first conclusion follows by $T$.

In addition

$$E[\|\hat{G}_n - G_n\|^2] = E[\|S(\sum_i D_{i,n} \otimes q_i)/n - G_n\|^2]
\leq CE\left[\|D_{i,n}\|^2 \|q_i\|^2\right]/n \leq CK/n,$$
where the first inequality follows from $D_{t,n}$ 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

$$n \hat{g}(\hat{\beta})' \hat{\Omega}^{-1} \hat{g}(\hat{\beta}) - n \hat{g}(\beta_{n,0})' \Omega_n^{-1} \hat{g}(\beta_{n,0}) \xrightarrow{p} 0.$$  

**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 $\beta_{0,n}$

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

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

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

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

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

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

$$\leq C tr((\hat{G}_n - G_n)'(\hat{G}_n - G_n))$$

$$= C \left\| \hat{G}_n - G_n \right\|^2 \xrightarrow{p} 0.$$  

Similarly, $\left\| \hat{\Omega}^{-1}(\hat{\Omega} - \Omega_n) \right\|^2 \xrightarrow{p} 0$.

Now $G_n'\Omega_n^{-1}G_n$ is bounded for large enough $n$ as $\hat{G}_n'\hat{\Omega}_n^{-1}\hat{G}_n \xrightarrow{p} V^{-1}$ by Lemma A.4 where $V = (E[D(x)'\Sigma(x)^{-1}D(x)])^{-1}$ which exists from Assumptions 3.4(d)(e) as [38]
\[ E[D(x)'\Sigma(x)^{-1}D(x)] \geq CE[D(x)'D(x)]. \] Thus, \( \|\Omega_n^{-1}G_n\| \) is also bounded. Therefore, to prove that \( \|\hat{\Omega}^{-1}G_n\| = O_p(1) \), by T

\[
\|\hat{\Omega}^{-1}G_n - \Omega_n^{-1}G_n\| \leq \|\hat{\Omega}^{-1}(\hat{G}_n - G_n)\| + \|\hat{\Omega}^{-1}(\hat{\Omega} - \Omega_n)\Omega_n^{-1}G_n\|.
\]

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

Now by independence

\[
E[\hat{g}_n'\Omega_n^{-1}\hat{g}_n] = E[\hat{g}_{i,n}'\Omega_n^{-1}\hat{g}_{i,n}]/n = E[tr(\Omega_n^{-1}\hat{g}_{i,n}\hat{g}_{i,n}')] = K/n.
\]

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

\[
\|\hat{G}_n'\hat{\Omega}^{-1}\hat{g}_n - G_n'\Omega_n^{-1}\hat{g}_n\| \leq \|\hat{G}_n'\hat{\Omega}^{-1}(\hat{\Omega} - \Omega_n)\Omega_n^{-1}\hat{g}_n\| + \|(G_n - G_n)'\Omega_n^{-1}\hat{g}_n\|
\]
\[
\leq (\|\hat{G}_n'\hat{\Omega}^{-1}\| \|\hat{\Omega} - \Omega_n\| + \|\hat{G}_n - G_n\|) \|\Omega_n^{-1}\hat{g}_n\|
\]
\[
\leq (O_p(1)o_p(1) + o_p(1))O_p(\sqrt{K/n}) = o_p(\sqrt{K/n}).
\]

Moreover

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

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 \( \|\hat{G}_n'\hat{\Omega}^{-1}\hat{g}_n\| = o_p(\sqrt{K/n}) \). Therefore, by Assumption 3.3(c),

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

Next, by CS and T,

\[
\|\hat{G}_n'\hat{\Omega}^{-1}\hat{G}_n - G_n'\Omega_n^{-1}G_n\| \leq (\|\hat{G}_n'\hat{\Omega}^{-1}\| + \|\Omega_n^{-1}G_n\|) \|\hat{G}_n - G_n\|
\]
\[
+ \|\hat{G}_n'\hat{\Omega}^{-1}\| \|\hat{\Omega} - \Omega_n\| \|\Omega_n^{-1}G_n\|.
\]

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

\[
\frac{n(\hat{\beta} - \beta_{n,0})'\hat{G}_n'\hat{\Omega}^{-1}\hat{G}_n(\hat{\beta} - \beta_{n,0})}{\sqrt{2(J_nM + J_n)K}} = O_p(1/\sqrt{2(J_nM + J_n)K}) = o_p(1).
\]

[39]
It remains to prove that
\[
\frac{n\hat{g}'_m(\hat{\Omega}^{-1} - \Omega^{-1})\hat{g}_m}{\sqrt{2(J_a M + J_m)K}} = o_p(1).
\]

From Lemma A.6,
\[
\left|n\hat{g}'_m(\hat{\Omega}^{-1} - \Omega^{-1})\hat{g}_m\right| / \sqrt{2(J_a M + J_m)K} \leq n \left\|\Omega^{-1}\hat{g}_m\right\|^2 (\left\|\hat{\Omega} - \Omega\right\| + C \left\|\hat{\Omega} - \Omega\right\|^2) / \sqrt{2(J_a M + J_m)K}
\]
\[
= n(O_p(K/n)(O_p(\sqrt{K/n}) + O_p(\zeta(K)\sqrt{K/n}))/\sqrt{2(J_a M + J_m)K}
\]
\[
= O_p(\zeta(K)K/\sqrt{n}) = o_p(1).
\]

\[\square\]

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 \(\Omega_m = E[g_m(z, \beta_{m0})g_m(z, \beta_{0})]\) and \(\Omega = E[g(z, \beta_0)g(z, \beta_0)]\). It also follows from DIN Lemma A.7, p.79, that \(\left\|\partial \hat{g}_m(\hat{\beta}_m) / \partial \beta_m' - G_m\right\| \overset{p}{\to} 0\) and \(\left\|\partial \hat{g}(\hat{\beta}) / \partial \beta' - G\right\| \overset{p}{\to} 0\) for any \(\hat{\beta} = \beta_0 + O_p(1/\sqrt{n})\). 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,
\[
\frac{n\hat{g}_m(\hat{\beta}_m)'\hat{\Omega}_m^{-1}\hat{g}_m(\hat{\beta}_m) - n\hat{g}_m(\beta_{m0})'\Omega^{-1}\hat{g}_m(\beta_{m0})}{\sqrt{2J_mK}} \overset{p}{\to} 0.
\]
and
\[
\frac{n\hat{g}(\hat{\beta})'\hat{\Omega}^{-1}\hat{g}(\hat{\beta}) - n\hat{g}(\beta_0)'\Omega^{-1}\hat{g}(\beta_0)}{\sqrt{2(J_a M + J_m)K}} \overset{p}{\to} 0.
\]
Now
\[
\frac{n\hat{g}(\hat{\beta})'\hat{\Omega}^{-1}\hat{g}(\hat{\beta}) - n\hat{g}_m(\hat{\beta}_m)'\hat{\Omega}_m^{-1}\hat{g}_m(\hat{\beta}_m) - J_a MK}{\sqrt{2J_a MK}} = \frac{n\hat{g}(\hat{\beta})'\hat{\Omega}^{-1}\hat{g}(\hat{\beta}) - n\hat{g}(\beta_0)'\Omega^{-1}\hat{g}(\beta_0)}{\sqrt{2J_a MK}}
\]
\[
- \frac{n\hat{g}_m(\hat{\beta}_m)'\hat{\Omega}_m^{-1}\hat{g}_m(\hat{\beta}_m) - n\hat{g}_m(\beta_{m0})'\Omega^{-1}\hat{g}_m(\beta_{m0})}{\sqrt{2J_a MK}}
\]
\[
+ \frac{n\hat{g}(\beta_0)'\Omega^{-1}\hat{g}(\beta_0) - n\hat{g}_m(\beta_{m0})'\Omega^{-1}\hat{g}_m(\beta_{m0}) - J_a MK}{\sqrt{2J_a MK}}
\]
\[
= \frac{n\hat{g}(\beta_0)'\Omega^{-1}\hat{g}(\beta_0) - n\hat{g}_m(\beta_{m0})'\Omega^{-1}\hat{g}_m(\beta_{m0}) - J_a MK}{\sqrt{2J_a MK}}.
\]
Recall $S_m = S_m^u \otimes S_m^q$ and, thus, $S_m \hat{g}(\beta_0) = \hat{g}_m(\beta_{m0})$. Therefore
\[
\frac{n \hat{g}(\beta_0)' \Omega^{-1} \hat{g}(\beta_0) - n \hat{g}_m(\beta_{m0})' \Omega^{-1} \hat{g}_m(\beta_{m0}) - J_a MK}{\sqrt{2J_a MK}} = \frac{n \hat{g}(\theta_0)' (\Omega^{-1} - S_m' \Omega^{-1} S_m) \hat{g}(\theta_0) - J_a MK}{\sqrt{2J_a MK}}.
\]

Lemma A.2 provides the conclusion of the theorem. First, $tr((\Omega^{-1} - S_m' \Omega^{-1} S_m)\Omega) = tr(I_{J_a M + J_m K}) - tr(I_{J_a M}) = J_a MK$. Secondly, $(\Omega^{-1} - S_m' \Omega^{-1} S_m)\Omega(\Omega^{-1} - S_m' \Omega^{-1} S_m) = \Omega^{-1} - S_m' \Omega^{-1} S_m$. Thirdly,
\[
E[(g(z, \beta_0)'(\Omega^{-1} - S_m' \Omega^{-1} S_m)g(z, \beta_0))^2] \leq CE[\|g(z, \beta_0)\|^4] \\
\leq CE[\|u(z, \beta_0)\|^4 \|q^K(s)\|^4] \\
\leq CE[\|q^K(s)\|^4] \\
\leq C\zeta(K)^2 K.
\]
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.,
\[
\mathcal{LR}^r = \frac{2n(\hat{P}_\theta \hat{g}(\hat{\beta}, \hat{\lambda}) - \hat{P}_{\theta m}^g(\hat{\beta}_m, \hat{\lambda}_m)) - J_a MK}{2\sqrt{J_a MK}}
\]
(A.3)
\[
= \frac{T_{\theta M}^g - T_{GMM}^g - J_a MK}{2\sqrt{J_a MK}} + \frac{2n\hat{P}_\theta \hat{g}(\hat{\beta}, \hat{\lambda}) - T_{\theta M}^g}{2\sqrt{J_a MK}} - \frac{2n\hat{P}_{\theta m}^g(\hat{\beta}_m, \hat{\lambda}_m) - T_{GMM}^g}{2\sqrt{J_a MK}}.
\]
Write $\hat{g}_mi = g_m(\hat{\beta}_m)$, $(i = 1, ..., n)$, $\hat{g}_m = \hat{g}_m(\hat{\beta}_m)$ and $\hat{g}_{m0} = \hat{g}_m(\beta_{m0})$. Using T and CS twice
\[
\|\hat{g}_m - \hat{g}_{m0}\| \leq \sum_{i=1}^n \left\|u_m(z_i, \hat{\beta}_m) - u_m(z_i, \hat{\beta}_{m0})\right\| \left\|q^K(s_{mi})\right\| /n \\
\leq (\sum_{i=1}^n \delta(z_i)^2/n)^{1/2} (\sum_{i=1}^n \left\|q^K(s_{mi})\right\|^2 /n)^{1/2} \left\|\hat{\beta}_m - \beta_{m0}\right\| = O_p(\sqrt{K/n})
\]
where the second inequality follows from Assumption 3.4(d). Thus, from T and DIN Lemma A.9, p.81, $\|\hat{g}_m\| = O_p(\sqrt{K/n})$ and, therefore, $\left\|\hat{\lambda}_m\right\| = O_p(\sqrt{K/n})$ by DIN Lemma [41]
A.11, p.82. Consequently \( \hat{\lambda}_m \in \hat{\Lambda}_n(\hat{\beta}_m) \) w.p.a.1 and the first order conditions for \( \lambda_m \) are satisfied w.p.a.1, i.e.,

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

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

\[-\hat{g}_m(\hat{\beta}_m) - \hat{\Omega}_m \hat{\lambda}_m = 0 \]

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

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

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

\[
2n\tilde{P}^g_{\rho}(\hat{\beta}_m, \hat{\lambda}_m) = 2n(-\hat{g}_m(\hat{\beta}_m)\hat{\lambda}_m - \hat{\lambda}_m' \hat{\Omega}_m \hat{\lambda}_m/2) = n\hat{g}_m(\hat{\beta}_m)(2\hat{\Omega}_m^{-1} - \hat{\Omega}_m^{-1} \hat{\Omega}_m \hat{\Omega}_m^{-1}] \hat{g}_m(\hat{\beta}_m)
\]

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

\[
\frac{2n\tilde{P}^g_{\rho}(\hat{\beta}_m, \hat{\lambda}_m) - T_{gMM}^g}{\sqrt{2J_oMK}} = n\hat{g}_m(\hat{\beta}_m)[2\hat{\Omega}_m^{-1} - \hat{\Omega}_m^{-1} \hat{\Omega}_m \hat{\Omega}_m^{-1}] \hat{g}_m(\hat{\beta}_m)/\sqrt{2J_oMK} \to 0.
\]

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

\[
\|\hat{\Omega}_m (2\hat{\Omega}_m - \hat{\Omega}_m)^{-1} \hat{\Omega}_m - \Omega_m (2\hat{\Omega}_m - \hat{\Omega}_m)^{-1} \Omega_m \| \leq \|\hat{\Omega}_m - \Omega_m\|(2\hat{\Omega}_m - \hat{\Omega}_m)^{-1} (\hat{\Omega}_m - \Omega_m)\| + 2\|\hat{\Omega}_m (2\hat{\Omega}_m - \hat{\Omega}_m)^{-1} (\hat{\Omega}_m - \Omega_m)\|
\]

\[
\leq C(\|\hat{\Omega}_m - \Omega_m\|^2 + \|\hat{\Omega}_m - \Omega_m\|) = o_p(1/\sqrt{K}).
\]

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

\[
\|\Omega_m (2\hat{\Omega}_m - \hat{\Omega}_m)^{-1} \Omega_m - \Omega_m\| = \|\Omega_m (2\hat{\Omega}_m - \hat{\Omega}_m)^{-1} (\Omega_m - (2\hat{\Omega}_m - \hat{\Omega}_m))\|
\]

\[
\leq C\|\Omega_m - (2\hat{\Omega}_m - \hat{\Omega}_m)\| = o_p(1/\sqrt{K})
\]

[42]
yielding \( \| \hat{\Omega}^{-1}_m (2 \hat{\Omega}_m - \hat{\Omega}_m) \hat{\Omega}^{-1}_m - \Omega^{-1}_m \| = o_p(1/\sqrt{K}) \). Therefore, as \( \| \hat{\Omega}^{-1}_m - \Omega^{-1}_m \| = o_p(1/\sqrt{K}) \),

\[
\frac{2n \hat{P}_n \hat{g}_m (\hat{\beta}_m, \hat{\lambda}_m) - T_{GMM}^g}{\sqrt{2J_aMK}} = nO_p(K/n)O_p(1/\sqrt{K})/\sqrt{2JMK} = o_p(1).
\]

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

\[
\frac{2n \hat{P}_n (\hat{\beta}, \hat{\lambda}) - T_{GMM}^g}{\sqrt{2J_aMK}} \overset{p}{\to} 0.
\]

Therefore, from Theorem 4.1 it follows that

\[ \mathcal{L} \mathcal{R}^r \overset{d}{\to} N(0, 1). \]

Now consider the Lagrange multiplier statistic (3.8)

\[ \mathcal{L} \mathcal{M}^r = \frac{n(\hat{\lambda} - S_m' \hat{\lambda}_m)' \hat{\Omega}(\hat{\lambda} - S_m' \hat{\lambda}_m) - J_aMK}{\sqrt{2J_aMK}}. \]

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

\[ \hat{\lambda} = -\hat{\Omega}^{-1} \hat{g} \]

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

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

\[
n(\hat{\lambda} - S_m' \hat{\lambda}_m)' \hat{\Omega}(\hat{\lambda} - S_m' \hat{\lambda}_m) = n\hat{\lambda}' \hat{\Omega} \hat{\lambda} - 2n\hat{\lambda}' \hat{\Omega} S_m' \hat{\lambda}_m + n\hat{\lambda}' S_m \hat{\Omega} S_m' \hat{\lambda}_m
\]

\[
= n\hat{g}' \hat{\Omega}^{-1} \hat{\Omega} \hat{\Omega}^{-1} \hat{g} - 2n\hat{g}' \hat{\Omega}^{-1} \hat{\Omega} S_m' \hat{\Omega}^{-1} S_m \hat{g} + n\hat{g}' S_m' \hat{\Omega}^{-1} S_m \hat{\Omega} S_m' \hat{\Omega}^{-1} S_m \hat{g}.
\]

Therefore

\[
\mathcal{L} \mathcal{M}^r - \mathcal{T}_{GMM}^g - \mathcal{T}_{GMM}^g - J_aMK = \frac{n\hat{g}' (\hat{\Omega}^{-1} \hat{\Omega} \hat{\Omega}^{-1} - \hat{\Omega}^{-1}) \hat{g}}{\sqrt{2J_aMK}} + \frac{n\hat{g}' (S_m' \hat{\Omega}^{-1} S_m - 2\hat{\Omega} \hat{\Omega}^{-1} \hat{\Omega} S_m' \hat{\Omega}^{-1} S_m + S_m' \hat{\Omega}^{-1} S_m \hat{\Omega} S_m' \hat{\Omega}^{-1} S_m) \hat{g}}{\sqrt{2J_aMK}}.
\]

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

\[
n\hat{g}' (\hat{\Omega}^{-1} \hat{\Omega} \hat{\Omega}^{-1} - \hat{\Omega}^{-1}) \hat{g}/\sqrt{K} = n\hat{g}' \hat{\Omega}^{-1} (\hat{\Omega} \hat{\Omega}^{-1} \hat{\Omega}) \hat{\Omega}^{-1} \hat{g}/\sqrt{K}
\]

\[
\leq n \frac{\| \hat{\lambda} \|^2 \| \hat{\Omega} - \hat{\Omega} \hat{\Omega}^{-1} \hat{\Omega} \|}{\sqrt{K}}.
\]

[43]
By DIN Lemma A.6, p.78, \( \| \hat{\Omega} - \Omega \| = O_p(\zeta(K)\sqrt{K/n}) = o_p(1/\sqrt{K}) \). Thus \( \lambda_{\text{max}}(\hat{\Omega}^{-1}) \leq C \). Moreover

\[
\| \hat{\Omega} \hat{\Omega}^{-1} - \Omega \hat{\Omega}^{-1} \Omega \| \leq \| (\hat{\Omega} - \Omega)\hat{\Omega}^{-1}(\hat{\Omega} - \Omega) \| + \| 2\Omega \hat{\Omega}^{-1}(\hat{\Omega} - \Omega) \|
\]

\[
\leq C(\| \hat{\Omega} - \Omega \|^2 + \| \hat{\Omega} - \Omega \|)
\]

\[
= O_p(\zeta(K)\sqrt{K/n}) = o_p(1/\sqrt{K}).
\]

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

\[
\| \Omega \hat{\Omega}^{-1} - \Omega \| = \| \Omega \hat{\Omega}^{-1}(\hat{\Omega} - \Omega) \|
\]

\[
\leq \| \Omega \hat{\Omega}^{-1} \| \| \hat{\Omega} - \Omega \| = o_p(1/\sqrt{K}).
\]

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

For the second term, by CS

\[
n\hat{g}'(S_m'\hat{\Omega}^{-1}S_m - 2\hat{\Omega}^{-1}\hat{\Omega}S_m'\hat{\Omega}^{-1}S_m + S_m'\hat{\Omega}^{-1}S_m\hat{\Omega} S_m'\hat{\Omega}^{-1}S_m)\hat{g}/\sqrt{K}
\]

\[
\leq n\| \hat{g} \|^2 \| S_m'\hat{\Omega}^{-1}S_m - 2\hat{\Omega}^{-1}\hat{\Omega}S_m'\hat{\Omega}^{-1}S_m + S_m'\hat{\Omega}^{-1}S_m\hat{\Omega} S_m'\hat{\Omega}^{-1}S_m \| /\sqrt{K}.
\]

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

\[
\| S_m'\hat{\Omega}^{-1}S_m - \hat{\Omega}^{-1}\hat{\Omega}S_m'\hat{\Omega}^{-1}S_m \| \leq \| S_m'\hat{\Omega}^{-1}(\hat{\Omega}_m - \hat{\Omega}_m)\hat{\Omega}_m^{-1}S_m \| + \| \hat{\Omega}^{-1}(\hat{\Omega} - \hat{\Omega})S_m'\hat{\Omega}^{-1}S_m \|
\]

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

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

\[
\| \hat{\Omega}^{-1}\hat{\Omega} S_m'\hat{\Omega}^{-1}S_m - S_m'\hat{\Omega}^{-1}S_m\hat{\Omega} S_m'\hat{\Omega}^{-1}S_m \| \leq \| S_m'\hat{\Omega}^{-1}(\hat{\Omega}_m - \hat{\Omega}_m)\hat{\Omega}_m^{-1}S_m \|
\]

\[
+ \| \hat{\Omega}^{-1}(\hat{\Omega} - \hat{\Omega})S_m'\hat{\Omega}^{-1}S_m \|
\]

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

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

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

\[
= o_p(1).
\]
Let \( \hat{g}_{ai} = g_{ai}(\hat{\beta}_a) \), \((i = 1,\ldots, n)\). The score test statistic
\[
S^* = \frac{\sum_{i=1}^{n} \rho_1(\hat{\lambda}'m \hat{g}_{mi})g_{ai}'S_a\hat{\Omega}^{-1}S'_a \sum_{i=1}^{n} \rho_1(\hat{\lambda}'m \hat{g}_{mi})\hat{g}_{ai}}{\sqrt{2J_aMK}}.
\]

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
\[
\sum_{i=1}^{n} \rho_1(\hat{\lambda}'m \hat{g}_{mi})\hat{g}_i/n - \Omega(\hat{\lambda} - S'_m \hat{\lambda}_m) = 0 \quad (A.7)
\]
w.p.a.1 where \(\hat{\Omega} = -\sum_{i=1}^{n} \rho_2(\hat{\lambda}'\hat{g}_i)\hat{g}_i\hat{g}_i'/n\) and \(\hat{\lambda}\) lies between \(\hat{\lambda}\) and \(S'_m \hat{\lambda}_m\). Since \(\sum_{i=1}^{n} \rho_1(\hat{\lambda}'m \hat{g}_{mi})\hat{g}_i/n = S'_a \sum_{i=1}^{n} \rho_1(\hat{\lambda}'m \hat{g}_{mi})\hat{g}_{ai}/n\),
\[
\sum_{i=1}^{n} \rho_1(\hat{\lambda}'m \hat{g}_{mi})\hat{g}_ai'S_a\hat{\Omega}^{-1}S'_a \sum_{i=1}^{n} \rho_1(\hat{\lambda}'m \hat{g}_{mi})\hat{g}_{ai}/n = n(\hat{\lambda} - S'_m \hat{\lambda}_m)'\hat{\Omega}^{-1}\Omega(\hat{\lambda} - S'_m \hat{\lambda}_m).
\]
Thus by CS and T
\[
|S^* - \mathcal{L}\mathcal{M}'| = n \left| (\hat{\lambda} - S'_m \hat{\lambda}_m)'(\hat{\Omega}^{-1}\hat{\Omega} - \hat{\Omega})(\hat{\lambda} - S'_m \hat{\lambda}_m) \right| / \sqrt{2J_aMK} \\
\leq n \left\| \hat{\Omega}^{-1}\hat{\Omega} - \hat{\Omega} \right\| \left( \left\| \hat{\lambda} \right\| + \left\| \hat{\lambda}_m \right\| \right)^2 / \sqrt{2J_aMK} = o_p(1) \quad (1)
\]
as \(\hat{\Omega}^{-1}\hat{\Omega} - \hat{\Omega} = o_p(1/\sqrt{K})\) and \(\left\| \hat{\lambda} \right\|, \left\| \hat{\lambda}_m \right\|\) are both \(O_p(\sqrt{K/n})\) by DIN Lemma A.11, p.82.

For the Wald test statistic, from (A.7), w.p.a.1 \(\hat{\lambda} - S'_m \hat{\lambda}_m = \hat{\Omega}^{-1}S'_a \sum_{i=1}^{n} \rho_1(\hat{\lambda}'m \hat{g}_{mi})\hat{g}_i/n\).
Thus
\[
S_a\hat{\lambda} = S_a\hat{\Omega}^{-1}S'_a \sum_{i=1}^{n} \rho_1(\hat{\lambda}'m \hat{g}_{mi})\hat{g}_{ai}/n.
\]
Therefore, w.p.a.1
\[
|W^* - S^*| = n \left| \hat{\lambda}'S'_a((S_a\hat{\Omega}^{-1}S'_a)^{-1} - (S_a\hat{\Omega}^{-1}S'_a)^{-1}S_a\hat{\Omega}^{-1}S'_a(S_a\hat{\Omega}^{-1}S'_a)^{-1})S_a\hat{\lambda} \right| / \sqrt{2J_aMK} \\
\leq n \left\| S_a\hat{\lambda} \right\|^2 \left\| (S_a\hat{\Omega}^{-1}S'_a)^{-1} - (S_a\hat{\Omega}^{-1}S'_a)^{-1}S_a\hat{\Omega}^{-1}S'_a(S_a\hat{\Omega}^{-1}S'_a)^{-1} \right\| / \sqrt{2J_aMK}.
\]

Now \((S_a\hat{\Omega}^{-1}S'_a)^{-1} - (S_a\hat{\Omega}^{-1}S'_a)^{-1}S_a\hat{\Omega}^{-1}S'_a(S_a\hat{\Omega}^{-1}S'_a)^{-1} = o_p(1/\sqrt{K})\), cf. \(\left\| \hat{\Omega} - \hat{\Omega}\hat{\Omega}^{-1}\hat{\Omega} \right\| = o_p(1/\sqrt{K})\) above. Therefore, since \(\left\| S_a\hat{\lambda} \right\| = O_p(\sqrt{K/n})\) by DIN Lemma A.11, p.82, \(|W^* - S^*| = o_p(1)\).
Proof of Theorem 4.3: The proof uses the Cramér-Wold device. Consider the linear combination

\[ J^c = \alpha_r J^r + \alpha_m J^m. \]

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

First, by DIN Lemma 6.1, p.69,

\[ \frac{n\hat{g}(\beta_0)'\Omega^{-1}\hat{g}(\beta_0) - n\hat{g}_m(\beta_m)\Omega^{-1}_m\hat{g}_m(\beta_m) - J_aMK}{\sqrt{2J_aMK}} \xrightarrow{p} 0. \]

Likewise

\[ \frac{n\hat{g}_m(\beta_m)\Omega^{-1}_m\hat{g}_m(\beta_m) - J_mK}{\sqrt{2J_mK}} \xrightarrow{p} 0. \]

Therefore,

\[ \frac{n\hat{g}(\beta_0)'Q\hat{g}(\beta_0) - (\alpha_rJ_a + \alpha_mJ_m\sqrt{J_aM/J_m})K}{\sqrt{2JK}} \xrightarrow{p} 0, \]

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

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

Condition (a): immediate.

Condition (b):

\[ tr(Q\Omega) = \alpha_r tr(I_{(J_a+J_m)K}) - (\alpha_r - \alpha_m\sqrt{J_aM/J_m})tr(I_{J_mK}) \]

\[ = \alpha_r(J_a + J_m)K - (\alpha_r - \alpha_m\sqrt{J_aM/J_m})J_mK \]

\[ = \alpha_r(J_a + \alpha_mJ_m\sqrt{J_aM/J_m})K = vK. \]

Condition (c): note that

\[ (Q\Omega)^2 = (\alpha_rI_{(J_a+J_m)K}) - (\alpha_r - \alpha_m\sqrt{J_aM/J_m})S_m\Omega_m^{-1}S'_m\Omega^2 \]

\[ = \alpha_r^2I_{(J_a+J_m)K} - (\alpha_r^2 - \alpha_m^2(J_aM/J_m))S_m\Omega_m^{-1}S'_m\Omega. \]

Hence

\[ tr[(Q\Omega)^2] = (\alpha_r^2 + \alpha_m^2)J_aMK \]

\[ = vK. \]

[46]
Condition (d):

\[(Q\Omega)^4 = (\alpha_r^2 I_{(J_a M+J_m)K} - (\alpha_r^2 - \alpha_m^2(J_a M/J_m))S_m\Omega_m^{-1}S'_m\Omega)^2\]
\[= \alpha_r^4 I_{(J_a M+J_m)K} - (\alpha_r^4 - \alpha_m^4(J_a M/J_m)^2)S_m\Omega_m^{-1}S'_m\Omega.\]

Thus

\[tr[(Q\Omega)^4] = (\alpha^4 + \alpha_4^4 J_a M J_m)J_a MK\]
\[= o(K^2).\]

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

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

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

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

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

Also

\[Q\Omega Q = (\alpha_r \Omega^{-1} - (\alpha_r - \alpha_m \sqrt{J_a M/J_m})S_m\Omega_m^{-1}S')\Omega(\alpha_r \Omega^{-1} - (\alpha_r - \alpha_m \sqrt{J_a M/J_m})S_m\Omega_m^{-1}S')\]
\[= \alpha_r^2(\Omega^{-1} - S_m\Omega_m^{-1}S') + \alpha_m^2(J_a M/J_m)S_m\Omega_m^{-1}S'.\]

Thus, cf. Condition (e),

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

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 \(q_{mi} = q^K_m(s_{mi})\), \(\hat{g}_i = g_i(\hat{\beta}), \hat{g}_{mi} = g_{mi}(\hat{\beta}_m)\) and \(g_{i, n} = [47]\]
\( g_i(\beta_{0,n}), g_{mi,n} = g_m(\beta_{m0,n}) \), \((i = 1, \ldots, n)\). 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)u_{mi}(\beta')|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}], \((i = 1, \ldots, n)\), together with
\[
\hat{\Omega} = \sum_i \hat{g}_i \hat{g}_i'/n, \tilde{\Omega}_n = \hat{g}_{i,n} \hat{g}_i/n, \\
\hat{\Omega}_n = S(\sum_i \Sigma_{i,n} \otimes q_i')S'/n, \Omega_n = E[g_{i,n}g_i'],
\]

and
\[
\hat{\Omega}_m = \sum_i \hat{g}_m \hat{g}_m'/n, \tilde{\Omega}_{mn} = \sum_i g_{mi,n} g_{mi,n}'/n, \\
\hat{\Omega}_{mn} = \langle \sum_i \Sigma_{mi,n} \otimes q_{mi} q_{mi}' \rangle/n, \Omega_{mn} = E[g_{mi,n}g_{mi,n}'].
\]

**Proof of Theorem 5.1:** The result is established for the GMM statistic \( J^r. \) Proofs for the restricted GEL statistics \( LR^r, LM^r, S^r \) and \( W^r \) 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,
\[
\frac{n\hat{g}(\beta') \hat{\Omega}^{-1} \hat{g}(\beta) - n\hat{g}_n' \Omega_n^{-1} \hat{g}_n}{\sqrt{2J_aMK}} \xrightarrow{p} 0, \frac{n\hat{g}_m(\beta_m') \hat{\Omega}_m^{-1} \hat{g}_m(\beta_m) - n\hat{g}_{mn}' \Omega_{mn}^{-1} \hat{g}_{mn}}{\sqrt{2J_aMK}} \xrightarrow{p} 0.
\]

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

It remains to prove that
\[
\frac{n\hat{g}_n'(\Omega_n^{-1} - S_m' \Omega_{mn}^{-1} S_m) \hat{g}_n - J_a MK}{\sqrt{2J_aMK}} \xrightarrow{d} N(\mu^r/\sqrt{2}, 1).
\]

Let \( \bar{g}_{i,n} = E[g_{i,n}|s_i] \) and \( \bar{g}_{i,n} = g_{i,n} - \bar{g}_{i,n}, \((i = 1, \ldots, n)\). Also let \( \bar{g}_n = \sum_{i=1}^n \bar{g}_{i,n}/n \) and \( \tilde{g}_n = \sum_{i=1}^n \bar{g}_{i,n}/n. \) Write \( P_n = \Omega_n^{-1} - S_m' \Omega_{mn}^{-1} S_m. \) Then,
\[
\hat{g}_n P \hat{g}_n = \tilde{g}_n P \tilde{g}_n + 2\tilde{g}_n' P_n \tilde{g}_n + \tilde{g}_n' P_n \tilde{g}_n.
\]

The first step demonstrates
\[
\tilde{g}_n' P_n \tilde{g}_n = \frac{n \sqrt{J_aMK}}{n}(\mu^r + o_p(1)).
\]
Let $\xi_i = \xi(s_i)$ and $\xi_{mi} = \xi_m(s_i)$, $(i = 1, ..., n)$. It follows by Lemma A.4 that

$$g_n^0 \bar{\Omega}_n^{-1} \bar{g}_n = \frac{\sqrt{J_aMK}}{n} \sum_{i,j=1}^n (\xi_i \otimes q_i)^{S'} \bar{\Omega}_n^{-1} S(\xi_j \otimes q_j) / n^2$$

$$= \frac{\sqrt{J_aMK}}{n} (E[\xi(s)^{S'}(s)^{-1}\xi(s)] + o_p(1)).$$

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

$$g'_n S'_m \bar{\Omega}_{mn}^{-1} S_m \bar{g}_n = \frac{\sqrt{J_aMK}}{n} \sum_{i,j=1}^n (\xi_{mi} \otimes q_{mi})^{S'_m} \bar{\Omega}_{mn}^{-1} (\xi_{mj} \otimes q_{mj}) / n^2$$

$$= \frac{\sqrt{J_aMK}}{n} (E[E[\xi_m(s)|s_m]^{S'_m} \bar{\Sigma}_m(s_m)^{-1} E[\xi_m(s)|s_m] + o_p(1))

= \frac{\sqrt{J_aMK}}{n} o_p(1),$$

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

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

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

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

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

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

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

Secondly, it is shown that

$$n g'_n P \bar{g}_n / \sqrt{2J_aMK} = o_p(1).$$

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

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

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

$$\leq C$$

[49]
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,

\[
g'_n \hat{\Omega}_n^{-1} \hat{g}_n = \frac{\sqrt{J_a MK}}{n} \sum_{i,j=1}^n (\xi_i \otimes q_i) S' \hat{\Omega}_n^{-1} \hat{g}_{j,n}/n \sqrt{n} = O_p(\sqrt{J_a MK}/n).
\]

Next, by hypothesis,

\[
|n \hat{g}'_n (\Omega_n^{-1} - \hat{\Omega}_n^{-1}) \hat{g}_n| / \sqrt{2J_a MK} \leq n \| \Omega_n^{-1} \hat{g}_n \| \| \Omega_n^{-1} \hat{g}_n \| |(\| \Omega_n - \hat{\Omega}_n \| + C \| \Omega_n - \hat{\Omega}_n \|^2) / \sqrt{2J_a MK} = O_p(\sqrt{K/n})
\]

since \( \| \Omega_n^{-1} \hat{g}_n \|^2 = O_p(\sqrt{K/n}) \) from above and \( \| \Omega_n^{-1} \hat{g}_n \| \leq \| \Omega_n^{-1} \hat{g}_n \| + \| \Omega_n^{-1} \hat{g}_n \| = O_p(\sqrt{K/n}) + O_p(\sqrt{K/n^2}) \). A similar analysis yields \( n \hat{g}'_n S'_{m,n} \Omega^{-1}_{mn} S_m \hat{g}_n / \sqrt{2J_a MK} = o_p(1) \).

Finally, Lemma A.2 is invoked to prove

\[
\frac{n \hat{g}'_n P_n \hat{g}_n - J_a MK}{\sqrt{2J_a MK}} \overset{d}{\to} N(0, 1).
\]

First, \( tr(\Omega_n P_n) = J_a MK \). Secondly, to establish

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

by \( c_r \)

\[
E[(\hat{g}'_{i,n} P_n \hat{g}_{i,n})^2] \leq 2E[(\hat{g}'_{i,n} \Omega_n^{-1} \hat{g}_{i,n})^2] + 2E[(\hat{g}'_{i,n} S'_{m,n} \Omega^{-1}_{mn} S_m \hat{g}_{i,n})^2].
\]

Again using \( c_r \)

\[
E[(\hat{g}'_{i,n} \Omega_n^{-1} \hat{g}_{i,n})^2] \leq 3E[(g'_{i,n} \Omega_n^{-1} g_{i,n})^2] + 12E[(\hat{g}'_{i,n} \Omega_n^{-1} \hat{g}_{i,n})^2] + 3E[(\hat{g}'_{i,n} \Omega_n^{-1} \hat{g}_{i,n})^2].
\]

Now, for \( n \) large enough, \( E[(g'_{i,n} \Omega_n^{-1} g_{i,n})^2] \leq CE[\|g_{i,n}\|^4] \). 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,

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

Next,

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

[50]
Lastly,
\[ E[(\tilde{g}_{i,n}^{-1}\tilde{g}_{i,n})^2] \leq C(K/n^2)E[\|\xi_i\|^4\|q_i\|^4] \leq C\zeta(K)^2K^2/n^2. \]

Hence, \( E[(\tilde{g}_{i,n}^{-1}\tilde{g}_{i,n})^2] = o_p(K\sqrt{n}) \) as required. Likewise, \( E[(\tilde{g}_{i,n}^{-1}\tilde{g}_{i,n})^2] = o_p(K\sqrt{n}) \). Thirdly, \( P_n\Omega_nP_n = P_n \). Therefore,
\[
\frac{n\tilde{g}_n^T P_n \tilde{g}_n - J_n MK}{\sqrt{2J_n MK}} \overset{d}{\rightarrow} N(0, 1).
\]

The conclusion of the theorem then follows. ■