cemmap
centre for microdata methods and practice

Exogeneity in semiparametric moment condition models

Paulo M.D.C. Parente
Richard J. Smith

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP30/12

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL
An ESRC Research Centre

Exogeneity in Semiparametric Moment Condition Models*

Paulo M.D.C. Parente
Department of Economics
University of Exeter
and
CReMic
University of Cambridge

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

This Revision: September 2012

Abstract

The primary concern of this article is the provision of definitions and tests for exogeneity appropriate for models defined through sets of conditional moment restrictions. These forms of exogeneity are expressed as additional conditional moment constraints and may be equivalently formulated as a countably infinite number of unconditional restrictions. Consequently, tests of exogeneity may be seen as tests for an additional set of infinite moment conditions. A number of test statistics are suggested based on GMM and generalized empirical likelihood. 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

*The authors are particularly grateful to M. Carrasco, X. Chen, V. Chernozhukov, V. Corradi, R. Crump, J.E.H. Davidson, J.-M. Dufour, S. Hemvanich, M. Jansson, J.L. Powell, J.M.C. Santos Silva and M.B. Stewart for useful discussions and comments. Earlier versions were presented at the Universities of Amsterdam, Cambridge, Exeter and Warwick, U.C. Berkeley, Tilburg and Yale Universities, Université de Montréal, the 2006 ESRC Econometric Study Group Conference, University of Bristol, the 2006 European Meetings of the Econometric Society, Vienna, the 2006 European Winter Meetings of the Econometric Society, Turin, the First Meeting of the *Portuguese Economic Journal*, Ponta Delgada, 2007, and the 2012 Banco de Portugal-CEFAGE Conference on "Robust Econometric Methods for Modelling Economic and Financial Variables". The Fundação para a Ciência e a Tecnologia and a 2002 Leverhulme Major Research Fellowship provided financial support for Parente and Smith respectively. Smith is also grateful for the hospitality of the Department of Economics, University of California at Berkeley, where earlier versions of the paper were written.

Key-words: Additional Conditional Moment Restrictions; Generalized Empirical Likelihood; GMM; Series Approximations; Specification Tests.

1 Introduction

The primary focus of this article is issues of exogeneity appropriate for models defined by a set of semiparametric conditional moment restrictions in the cross-sectional data context. More specifically its particular concerns are to provide definitions of and to propose tests for exogeneity in this setting. A second contribution of the paper is to detail practically efficacious GMM and generalised empirical likelihood (GEL) test statistics for additional conditional moment restrictions which include the exogeneity hypotheses considered here as special cases.

Numerous definitions of exogeneity have been provided in the literature; see, e.g., the discussion in section 2 of Deaton (2010). Engle, Hendry and Richard (1983), henceforth EHR, consider classical parametric maximum likelihood estimation. Consider the random vectors y and x and suppose that y is the target variate of interest. The random vector x is said to be (weakly) exogenous for the parameters characterising the conditional distribution of y given x if no loss of information results by disregarding the marginal distribution of x, i.e., conditional maximum likelihood is asymptotically efficient.[1] From a policy perspective, the exogeneity of x assumes a central importance in this context. If in addition the conditional distribution of y given x describes the behavioural relationship for y in terms of x and if x is also a vector of control variables for the policy maker then knowledge of the conditional distribution of y given x enables the policy maker to predict accurately the effect of a change in policy effected through x without knowledge of the joint distribution of y and x, or, more precisely, the marginal distribution of x.[2]

Other definitions of exogeneity have been formulated that are primarily concerned with the consistency of a particular parameter estimator. Hausman (1978, Section 2, pp.1252-1261) discusses exogeneity for a linear regression model in terms of a two equation

[1]See EHR, Definition 2.5, p.282. Technically this definition of exogeneity is that of statistical partial or S-ancillarity; see Basu (1977, Definition 8, p.357, and Case V, p.358).

[2]For the dynamic linear simultaneous equations model with normally distributed errors, EHR, Theorem 4.3 (a) and (b), p.298, provides sufficient conditions for exogeneity expressed in terms of the uncorrelatedness of particular structural error terms. More specifically, for classical normal linear regression, the exogeneity assumption is equivalent to the uncorrelatedness of regression error and covariates, i.e., the assumption commonly made when the objective is to estimate the best linear predictor.

triangular system. The definition of exogeneity provided there, however, is in fact more widely applicable for models defined via unconditional moment restrictions. Consider a linear regression model with scalar dependent variable y, covariate vector x and error term u uncorrelated with instrument vector w. Then x is exogenous if it is also uncorrelated with u. This particular definition of exogeneity is useful if interest centres on consistent estimation of the best linear predictor of y in terms of x. Moreover, it implies that the instrument vector w plays no role in the best linear predictor of y expressed in terms of x *and* w. However, for the policy maker, the best linear predictor may be difficult to interpret and thus its practical relevance difficult to justify. From this perspective the conditional mean of y rather than the best linear predictor may be of more importance and interest. Consequently the central concern of this paper when considering notions of exogeneity is with particular conditional expectations in the semiparametric moment condition setting.

More recently, Blundell and Horowitz (2007), hereafter BH, discuss exogeneity when the nonparametric estimation of a structural function $g(x)$ of the dependent variable y defined in terms of the covariate vector x is of interest. The conditional expectation of the structural error term $y - g(x)$ given a set of identifying instruments w is maintained to be zero. In this setting x is exogenous if the conditional expectation of y given x almost surely coincides with $g(x)$, i.e., the conditional expectation of the structural error term given x is zero. This definition has the advantage that standard nonparametric regression of y on x is then appropriate for consistent estimation of $g(x)$ and thus may be regarded as being a natural counterpart of the exogeneity definition concerned with estimator consistency in the parametric framework considered in, e.g., Hausman (1978). Because the maintained instruments w are now ignored this definition may be characterised as a partial form of exogeneity which we term as *marginal* exogeneity in mean below. Importantly, however, this definition of exogeneity may be inadequate in particular circumstances since, for the policy maker, if the instruments w are control variables, the effect of changes in the instruments w given x on y will in general be unknown without further knowledge of the conditional distribution of w given x. Of course, if the covariate vector x is itself under the control of the policy maker, then the effect of changes in x on y would be perfectly predictable.[3,4]

Therefore, the first concern of the paper is to clarify and provide an alternative definition of exogeneity to that of BH for general nonlinear models specified by conditional moment restrictions. A covariate vector x is said to be *conditionally* exogenous in mean if the expectation of the conditional moment indicator vector given both x *and* maintained instruments w is zero. In particular, if regression covariates x are conditionally exogenous then the instruments w are necessarily redundant as additional explanators. From the viewpoint of the policy maker such information is useful since to effect a change in the conditional mean of y given x *and* w, only x need now be varied. The constraints imposed by this definition of exogeneity are of course stricter than those arising through that of BH. Consequently, estimators which efficiently incorporate this information should dominate those which only make use of the marginal exogeneity restriction.

The paper also provides tests for additional moment restrictions in the conditional moment framework. These tests are then adapted for both marginal and conditional exogeneity in mean hypotheses. Most tests for exogeneity proposed in the literature focus on the best linear predictor since their primary concern is with linear regression settings. The most popular of these tests is probably the Durbin-Wu-Hausman test [Durbin (1954),

[3]Similar concerns and considerations to these apply more generally in separable nonlinear (latent variable) models such as the instrumental variable quantile regression model defined by $\mathcal{P}\{y \leq g(x)|w\} = \theta$, where w is a vector of instruments, i.e., the conditional θ-quantile $Q_\theta(y|w)$ of y given w is $g(x)$; see Chernozhukov and Hong (2003) and Honore and Hu (2003). Let $I(\cdot)$ be the indicator function. Hence the θ-quantile constraint $Q_\theta(y|w) = g(x)$ is equivalent to the moment condition $E[\theta - I(y \leq g(x))|w] = 0$. Marginal exogeneity in mean corresponds to $E[\theta - I(y \leq g(x))|x] = 0$, i.e., the conditional θ-quantile of y $Q_\theta(y|x)$ given x is also $g(x)$. If w is a vector of controls, policy interest would concern the conditional θ-quantile $Q_\theta(y|w, x)$ of y given w *and* x, i.e., whether given x changes in w affect $Q_\theta(y|w, x)$. In particular, if the conditional exogeneity in mean hypothesis $Q_\theta(y|w, x) = g(x)$ holds, then, given x, instruments w play no role in determining $Q_\theta(y|w, x)$.

[4]White and Chalak (2010), see also Chalak and White (2011), provides another definition of conditional exogeneity for the nonseparable model framework $y = r(x, u)$, where $r(\cdot)$ is an unknown structural function and y, x and excluded w are observable but u is not. Let $x = (x_a', x_b')'$. Here interest centres on the effect of x_a on y. White and Chalak (2010) defines x_a to be conditionally exogenous if x_a and u are conditionally independent given w and x_b. In separable models $y = g(x) + u$, the less restrictive conditional exogeneity or conditional mean independence constraint $E[y - g(x)|x, w] = E[y - g(x)|x_b, w] = 0$ discussed here allows identification of the effect of x_a on y given w and x_b; see also Chalak and White (2006).

Wu (1973), Hausman (1978)] which contrasts instrumental variable estimators obtained assuming orthogonality conditions between errors and instruments and errors and covariates (and instruments) respectively.[5] These types of tests, however, are inappropriate for models defined by conditional moment constraints. As noted by Bierens (1990), such orthogonality tests are generally inconsistent against some alternatives implied by conditional moment conditions as only a finite number of unconditional restrictions are used to formulate these tests.

The particular approach for test formulation taken here is based on an infinite number of unconditional moment restrictions that are designed to overcome the aforementioned test inconsistency difficulty. While tests of exogeneity have received relatively little attention in models defined by conditional moment restrictions, there is a vast related literature on tests of goodness of fit in regression models. See, for example, Eubank and Spiegelman (1990) in the nonlinear regression context. Other tests have also been proposed for this set-up by *inter alia* De Jong and Bierens (1994), Hong and White (1995) and Jayasuriya (1996). Donald, Imbens and Newey (2003), henceforth DIN, extends these ideas to the conditional moment restriction setting for GMM [Hansen (1982)] and GEL [Newey and Smith (2004), Smith (1997, 2011)]. This paper adapts these methods to formulate tests for additional moment restrictions in the conditional moment model framework and then specialises them for conditional and marginal exogeneity in mean hypotheses.[6] The basic underlying idea is to approximate conditional moment restrictions by a finite set of unconditional moment restrictions, the number of which is then allowed to grow with but at a slower rate than sample size. Both marginal and conditional exogeneity in mean hypotheses involve two sets of conditional moment restrictions with the second set implying the first. These sets of conditional moment conditions are replaced by corresponding sets of unconditional moment restrictions with the first set a subset of the second, cf.

[5]Lagrange multiplier or score tests are suggested in Engle (1982). Smith (1994) proposes efficient limited information classical test statistics for the sufficient exogeneity conditions in the dynamic simultaneous equations model discussed in EHR, Section 4, pp. 294-300.

[6]Alternative tests for exogeneity could also be based *inter alia* on the approaches of Bierens (1982, 1990), Wooldridge (1992), Yatchew (1992), Härdle and Mammen (1993), Fan and Li (1996), Zheng (1996,1998), Lavergne and Vuong (2000), Ellison and Ellison (2000) and Domínguez and Lobato (2004).

DIN. As a consequence, these exogeneity tests may be interpreted as tests for additional moment restrictions similar to those proposed by Newey (1985b) and Tauchen (1985) in the classical parametric setting, by Newey (1985a), Eichenbaum, Hansen and Singleton (1988) and Ruud (2000) for GMM and by Smith (1997, 2011) for GEL. After appropriate standardization, the test statistics converge in distribution to a standard normal variate rather than the usual chi-square distributed variate, intuitively, since, from an asymptotic standpoint, the statistics are based on an infinite number of unconditional moments. Furthermore, unlike orthogonality test statistics, efficient parameter estimators are not required for the formulation of these tests.[7]

The paper is organized as follows. Section 2 provides a detailed discussion of exogeneity appropriate for models defined by conditional moment restrictions. The test problem is then specified in section 3 together with some additional notation and requisite assumptions; GMM and GEL test statistics for marginal and conditional exogeneity in mean are also detailed there. Section 4 details the limiting distribution of these statistics under the null hypothesis of exogeneity whereas section 5 considers their asymptotic distribution under a suitable sequence of local alternatives. Section 6 discusses some issues concerning the computation of the test statistics. Section 7 presents a set of simulation results on the size and power of the test statistics. Section 8 concludes. Proofs of the results in the text and certain subsidiary lemmata are given in Appendix A. Tables associated with the simulation experiments are collected in Appendices B and C.

2 Exogeneity

2.1 Some Preliminaries

The standard definition of exogeneity in the classical linear regression setting is that of an absence of correlation between a covariate and the model error term. This definition,

[7] The succeeding theoretical analysis may in principle be straightforwardly adapted and extended for models defined by nonsmooth moment conditions 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).

however, may be rather restrictive if purposes other than estimation of the best linear predictor are of primary interest.

More generally, the conditional mean of the dependent variable is likely to be of more relevance for exogeneity considerations. Recently, BH proposes a definition for exogeneity in nonparametric regression. BH consider the model

$$y = g(x) + u, \tag{2.1}$$

where $g(\cdot)$ is an unknown structural function of inferential interest and x is a vector of covariates. BH maintain the identifying conditional moment restriction $E[u|w] = 0$ where w is a vector of instruments and define the covariate vector x to be exogenous if the conditional moment restriction $E[u|x] = 0$ holds. Therefore, $E[y|x] = g(x)$.

Examples 2.1 and 2.2 below demonstrate the well-known result that even though structural errors may be uncorrelated with instruments and covariates covariates are not necessarily exogenous in the BH sense.

Example 2.1: Consider the linear regression model

$$y = \beta_0 x + u, \tag{2.2}$$

where β_0 is an unknown parameter, the covariate $x = w + v_1$, $u = \sqrt{-2\ln(v_2)}\cos(2\pi v_1)$ and the instrument w and v_1, v_2 are independent uniform random variables on the unit interval $[0, 1]$; hence, u is standard normal being defined by the Box-Müller transformation. In this example, it is easily shown that instrument and covariate are uncorrelated with the structural error term u, i.e., $E[wu] = 0$, $E[xu] = 0$, although $E[x^2u] = (2\pi)^{-3/2}$. Therefore $E[u|x] \neq 0$.

Example 2.2: [Stock and Watson, 2007, Exercise 2.23, p.63.] Suppose now that $x = w + z$ and $u = z^2 - 1$ in (2.2) hold where w and z are independent standard normal random variates. In this case u is a centered chi-squared random variable with 1 degree of freedom. As in Example 2.1, $E[wu] = 0$ and $E[xu] = 0$, but $E[x^2u] = 2$ and thus

$E[u|x] \neq 0$.

The next example is constructed to illustrate a potential limitation of the BH definition of exogeneity and forms the basis of the simulation experiments considered in section 7.

Example 2.3: Consider the following revision to regression model (2.2)

$$y = \beta_0 x + f(w, x) + \upsilon, \tag{2.3}$$

where υ is standard normal and statistically independent of x and w. In this example the parameter β_0 may no longer be of sole interest but the form of $f(w, x)$ may be relevant too. Assume that w and x are jointly normally distributed each with mean zero, unit variance and correlation coefficient ρ where $\rho \in (-1, 1)$ and $\rho \neq 0$. Suppose

$$f(w, x) = x^2 + w^2 - (\frac{1 + \rho^2}{\rho})wx - (1 - \rho^2).$$

If $f(w, x)$ is erroneously omitted from (2.3) and the regression model (2.2) again estimated but with $u = f(w, x) + \upsilon$, β_0 may be consistently estimated by instrumental variables (IV) using w as instrument since $E[u|w] = 0$. Although the omitted variable $f(w, x)$ depends on x the covariate x is in fact exogenous in the BH sense as $E[u|x] = 0$. Consequently β_0 can also be consistently estimated using least squares (LS). In particular, under the BH exogeneity hypothesis $E[u|x] = 0$, the effect of changes of x on y are predictable since $E[y|x] = \beta_0 x$.

In general, however, regression model (2.3) is of relevance rather than (2.2); in particular, $f(w, x)$ is of importance if the instrument w is a control variable for the policy maker. The impact of altering w requires additional information concerning the conditional distribution of x given w, namely ρ here. To see this, from (2.3), $E[y|w] = \beta_0 E[x|w] + E[f(w, x)|w]$, i.e., $\beta_0 \rho w$ under the above assumptions. Moreover, if x is also a control variable for the policy maker, since $E[u|w, x] \neq 0$, the effect of changing w while keeping x unaltered requires examination of $E[y|w, x] = \beta_0 x + f(w, x)$.

Therefore, more generally, an appropriate but more restrictive definition of exogeneity than that of BH requires $E[u|w, x] = 0$ implying, for model (2.1), that $E[y|w, x] = g(x)$. Hence, when x is unaltered, changes of w have no effect on the conditional mean $E[y|w, x]$.

2.2 Definitions

We consider the more general conditional moment context with error vector defined by $u = u(z, \beta_0)$, where $u(z, \beta)$ is a known J-vector of functions of the random vector of observables z and the unknown p-vector of parameters β_0 which constitute the object of inferential interest.

Like BH we assume that there exists an observable vector of instruments w such that

$$E[u(z, \beta_0)|w] = 0. \tag{2.4}$$

Since the BH definition does not involve the maintained instrument vector w we view it as a partial or *marginal* form of exogeneity; *viz.*

Definition 2.1 *(Marginal Exogeneity in Mean.) The random vector x is marginally exogenous in mean (*MEM*) for β_0 if*

$$E[u(z, \beta_0)|x] = 0. \tag{2.5}$$

EXAMPLE 2.3 (cont.): Here $u(z, \beta_0) = y - \beta_0 x$ and x MEM for β_0 implies that β_0 may be consistently estimated by LS. LS is in general inefficient not only because it neglects the maintained constraint $E[u(z, \beta_0)|w] = 0$ (2.4) but also because the conditional nature of MEM (2.5) is ignored. An IV estimator for β_0 based on the joint conditional moment conditions (2.4) and (2.5) should be at least as efficient as LS or IV using only (2.4). Thus, although $E[y|x]$ is correctly specified as $\beta_0 x$, a more efficient estimator for β_0 than LS is possible.

If x MEM for β_0 and if x is a control variable, the average effect of changes in x by the policy maker on y is predictable. In contrast, if w is also a policy control variable, the likely impact on y occasioned by changes in w cannot be determined without further knowledge of the conditional distribution of x given w, namely ρ, i.e., x MEM for β_0 is uninformative. Moreover, x MEM for β_0 is unhelpful in determining the effect of changes in w on y while keeping x unaltered which requires knowledge of the conditional mean $E[y|w,x] = \beta_0 x + f(w,x)$.

In Example 2.1 $f(w,x) = 0$. However, because instrument w is excluded, x MEM for β_0 implies $E[y|w,x] = E[y|x] = \beta_0 x$, i.e., y is conditionally mean independent of w given x. Therefore, if x is kept unchanged, alterations in w have no effect on y, i.e., instruments w contribute no information in addition to that provided by x to the conditional expectation of y.

In general, therefore, MEM (2.5) may represent an incomplete definition of exogeneity from a practical perspective in the conditional moment context. To deal with this issue, the following definition of exogeneity revises that of BH incorporating the maintained instruments w and necessarily taking a *conditional* form.

Definition 2.2 *(Conditional Exogeneity in Mean.) The random vector x is conditionally exogenous in mean (*CEM*) for β_0 given w if*

$$E\left[u\left(z,\beta_0\right)|w,x\right] = 0. \tag{2.6}$$

CEM (2.6) not only implies the maintained conditional moment restriction (2.4) but also MEM (2.5). Thus, CEM is a more stringent requirement than MEM. Therefore, estimators using CEM are in general more efficient than those solely exploiting (2.4) and MEM. Note, however, that in Example 2.3 the marginal effect of w on y remains the same under both CEM and MEM, i.e., $E[y|w] = \beta_0\rho w$.

The next sections develop tests for both MEM and CEM and analyse their large sample properties.

3 GMM and GEL Test Statistics

3.1 Test Problem

The conditional moment constraints $E[u(z,\beta_0)|w] = 0$ (2.4) are maintained throughout. The null hypothesis is

$$H_0 : E[u(z,\beta_0)|s] = 0, E[u(z,\beta_0)|w] = 0 \tag{3.1}$$

with the alternative hypothesis

$$H_1 : E[u(z,\beta_0)|s] \neq 0, E[u(z,\beta_0)|w] = 0. \tag{3.2}$$

The use of the generic random vector s permits circumstances in which w may or may not be strictly included as a conditioning variate. Indeed the null hypothesis (3.1) allows both the definitions of exogeneity given in section 2.1 as special cases with $s = x$ and $s = (w', x')'$ as MEM (2.5) and CEM (2.6) respectively. The definition of s will be made explicit in each particular instance.

3.2 Approximating Conditional Moment Restrictions

Conditional moment conditions of the form given in (3.1) and (2.4) 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.

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

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

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

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

The next result, DIN Lemma 2.1, p.58, shows formally the equivalence between conditional moment restrictions 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 $E\left[u\left(z,\beta_0\right)|s\right]=0$, then $E[u(z,\beta_0)\otimes q^K(s)]=0$ for all K. Furthermore, if $E\left[u\left(z,\beta_0\right)|s\right]\neq 0$, 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 $g(z,\beta)=u(z,\beta)\otimes q^K(s)$. By considering the moment conditions $E[g(z,\beta_0)]=0$, 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 these estimators are consistent and achieve the semiparametric efficiency lower bound. To do so, however, requires the imposition of a normalization condition on the approximating functions, DIN Assumption 2, p.59, as now follows.

Let $\mathcal{S}$ denote the support of the random vector s.

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

Hence the null hypothesis (3.1) may be re-interpreted in terms of a sequence of additional unconditional moment restrictions. In particular, to test either MEM (2.5) and CEM (2.6) requires that their constituent conditional moment constraints and the maintained (2.4) are replaced by suitably defined unconditional moment restrictions based on Assumptions 3.1 and 3.2.

The maintained conditional moment restrictions (2.4) are consequently re-expressed as the sequence of unconditional moment restrictions

$$E[u(z,\beta_0)\otimes q^K_{\text{MA}}(w)]=0, K\rightarrow\infty, \tag{3.3}$$

for approximating functions $q^K_{\text{MA}}(\cdot)$ satisfying Assumptions 3.1 and 3.2 with $s=w$.

Let $f(K)$ be a function of K that yields positive integer numbers and satisfies $f(K) = O(K)$; for simplicity we set $f(K) = MK$, where M is a positive integer.[8] Also let $q_{\text{S}}^K(s)$ be a $f(K)$-vector of approximating functions that depends on s, with S=M or C and $s = x$ or $s = (w', x')'$ corresponding to MEM or CEM respectively. Additionally define $q^K(w,x) = (q_{\text{MA}}^K(w)', q_{\text{S}}^K(s)')'$. Therefore, if Assumptions 3.1 and 3.2 are satisfied for $q^K(w,x)$, the null hypothesis (3.1) is equivalent to the sequence of unconditional moments

$$E[u(z,\beta_0) \otimes q^K(w,x)] = 0, K \to \infty. \tag{3.4}$$

For MEM (2.5), $q_{\text{S}}^K(s)$ depends only on functions of x whereas for CEM (2.6) it involves additional functions of both w and x.

3.3 Basic Assumptions and Notation

We impose the following standard conditions to derive the asymptotic distributions of the test statistics discussed below.

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

Unlike DIN Assumption 6(b), it is unnecessary to impose $E\left[\left\|\sup_{\beta\in\mathcal{B}} \|u(z,\beta)\|^\gamma\right\|\right] < \infty$ for some $\gamma > 2$ for GEL. As noted in Guggenberger and Smith (2005), 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 A10 in DIN. Therefore, γ may be set as 2 in the succeeding Lemmata and Theorems in DIN concerned with GEL. Note that only root-n consistency rather than efficiency is required for the estimator $\hat{\beta}$. Moreover, since under the null hypothesis (3.1) $E[u(z,\beta_0)|s] = E[u(z,\beta_0)|w] = 0$, for $s = x$ or $s = (w', x')'$, only a single estimator is needed for β_0.

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

[8] The requirement that $f(K) = O(K)$ arises because of local power considerations; see section 5.

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

3.4 Test Statistics

Let $g_i(\beta)=u(z_i,\beta)\otimes q^K_{\text{MA}}(w_i)$ and $h_i(\beta)=u(z_i,\beta)\otimes q^K(w_i,x_i)$, $(i=1,...,n)$. Also let $\hat{g}(\beta)=\sum_{i=1}^n g_i(\beta)/n$ and $\hat{h}(\beta)=\sum_{i=1}^n h_i(\beta)/n$.

Conditional GMM statistics appropriate for tests of maintained and null hypotheses take the standard form

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

and

$$\mathcal{T}^h_{GMM}=n\hat{h}(\hat{\beta})'\hat{\Xi}^{-1}\hat{h}(\hat{\beta}) \tag{3.6}$$

where $\hat{\Omega}=\sum_{i=1}^n g_i(\hat{\beta})g_i(\hat{\beta})'/n$ and $\hat{\Xi}=\sum_{i=1}^n h_i(\hat{\beta})h_i(\hat{\beta})'/n$. See for example DIN, section 4, pp.63-64.

In the remainder of the paper tests that incorporate the information contained in the maintained hypothesis (3.2) are referred to as *restricted* tests whereas those that ignore it are *unrestricted* tests.

A restricted GMM statistic appropriate for testing the null hypothesis (3.1) comprising either MEM (2.5) or CEM (2.6) hypotheses against the maintained hypothesis (3.2) may be based on the difference of GMM criterion function statistics (3.6) and (3.5) for the revised hypotheses (3.4) and (3.3) respectively. For fixed and finite K, standard asymptotic theory for tests of the validity of additional moment restrictions [Newey (1985a)] yields test statistics that are chi-square distributed with JMK degrees of freedom. It is well known, however, that when the number of degrees of freedom is very large a chi-square random variable can be approximated, after standardization by subtraction of its mean and division by its standard deviation, by a standard normal random variable. The resultant GMM statistic is therefore defined as

$$\mathcal{J} = \frac{\mathcal{T}^h_{GMM} - \mathcal{T}^g_{GMM} - JMK}{\sqrt{2JMK}}. \tag{3.7}$$

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

$$\begin{aligned} \tilde{P}_n(\beta,\eta) &= \sum_{i=1}^n [\rho(\eta' h_i(\beta)) - \rho_0]/n, \\ \hat{P}_n(\beta,\lambda) &= \sum_{i=1}^n [\rho(\lambda' g_i(\beta)) - \rho_0]/n \end{aligned} \tag{3.8}$$

under null and alternative hypotheses where η and λ are the corresponding $J(M+1)K$- and JK-vectors of Lagrange multipliers associated with the unconditional moment constraints (3.4) and (3.3). Let $\rho_j(v) = \partial^j \rho(v)/\partial v^j$ and $\rho_j = \rho_j(0)$, $(j = 0, 1, 2, ...)$ where, without loss of generality, we impose the normalisation $\rho_1 = \rho_2 = -1$.

Let $\hat{\Lambda}_n(\beta) = \{\lambda : \lambda' g_i(\beta) \in \mathcal{V}, i = 1, ..., n\}$ and $\tilde{\Lambda}_n(\beta) = \{\eta : \eta' h_i(\beta) \in \mathcal{V}, i = 1, ..., n\}$. Given β, the respective Lagrange multiplier estimators for λ and η are given by

$$\hat{\lambda}(\beta) = \arg\max_{\lambda \in \hat{\Lambda}_n(\beta)} \hat{P}_n(\beta,\lambda), \tilde{\eta}(\beta) = \arg\max_{\eta \in \tilde{\Lambda}_n(\beta)} \tilde{P}_n(\beta,\eta).$$

Suppose that $\hat{\beta}$ is a root-n consistent estimator for β_0 under either null or alternative hypothesis. The corresponding respective Lagrange multiplier estimators for λ and η are then defined as $\hat{\lambda} = \hat{\lambda}(\hat{\beta})$ and $\tilde{\eta} = \tilde{\eta}(\hat{\beta})$.

Let $\hat{\eta} = S_{\text{MA}}\hat{\lambda}$ where $S_{\text{MA}} = I_J \otimes (I_K, 0_{MK})'$ is a $J(M+1)K \times JMK$ selection matrix. Additionally let $s(z,\beta) = S_{\text{S}}' h(z,\beta)$ where $S_{\text{S}} = I_J \otimes (0_K, I_{MK})'$ is a $J(M+1)K \times JMK$ selection matrix. Hence, $s(z,\beta) = u(z,\beta) \otimes q^K_{\text{S}}(w,x)$. Write $s_i(\beta) = s(z_i,\beta)$, $(i = 1, ..., n)$.

Similarly to the restricted GMM statistic $\mathcal{J}$, a restricted form of GEL likelihood ratio (LR) statistic for testing either MEM (2.5) or CEM (2.6) hypotheses against the maintained hypothesis (3.2) may be based on the difference of GEL criterion function (3.8) statistics; *viz.*

$$\mathcal{LR} = \frac{2n[\tilde{P}_n(\hat{\beta}, \tilde{\eta}) - \hat{P}_n(\hat{\beta}, \hat{\lambda})] - JMK}{\sqrt{2JMK}}. \tag{3.9}$$

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

$$\mathcal{LM} = \frac{n(\tilde{\eta} - \hat{\eta})'\hat{\Xi}(\tilde{\eta} - \hat{\eta}) - JMK}{\sqrt{2JMK}}, \tag{3.10}$$

$$\mathcal{S} = \frac{\sum_{i=1}^{n} \rho_1(\hat{\lambda}' g_i(\hat{\beta})) s_i(\hat{\beta})' S_{\rm s}' \hat{\Xi}^{-1} S_{\rm s} \sum_{i=1}^{n} \rho_1(\hat{\lambda}' g_i(\hat{\beta})) s_i(\hat{\beta})/n - JMK}{\sqrt{2JMK}} \tag{3.11}$$

and

$$\mathcal{W} = \frac{n\tilde{\eta}' S_{\rm s}(S_{\rm s}'\hat{\Xi}^{-1} S_{\rm s})^{-1} S_{\rm s}'\tilde{\eta} - JMK}{\sqrt{2JMK}}. \tag{3.12}$$

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

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

4 Asymptotic Null Distribution

The following theorem provides a statement of the limiting distribution of the restricted GMM statistic $\mathcal{J}$ (3.7) under the null hypothesis (3.1).

Theorem 4.1 *If Assumptions 3.1, 3.2, 3.3 and 3.4 hold for $s = w$ and $s = (w', x')'$ and if $K \to \infty$ and $\zeta(K)^2 K^2/n \to 0$, then $\mathcal{J} \xrightarrow{d} N(0, 1)$.*

Although this result is stated for a restricted GMM-based test of MEM or CEM it has a wider significance. It is also relevant and may be straightforwardly adapted with little alteration for constructing a test for the comparison of two sets of conditional moment restrictions where one set is nested within the other.

The next result details the limiting properties of the restricted GEL-based statistics for the exogeneity hypotheses (2.5) and (2.6) and their relationship to that of the GMM statistic $\mathcal{J}$ (3.7).

Theorem 4.2 *Let Assumptions 3.1, 3.2, 3.3, 3.4 and 3.5 hold for $s = w$ and $s = (w', x')'$ and in addition $K \to \infty$ and $\zeta(K)^2 K^3/n \to 0$. Then $\mathcal{LR}$, $\mathcal{LM}$, $\mathcal{S}$ and $\mathcal{W}$ converge in distribution to a standard normal random variate. Moreover all of these statistics are asymptotically equivalent to $\mathcal{J}$.*

Similarly to the GMM statistic $\mathcal{J}$ (3.7) the GEL statistics $\mathcal{LR}$, $\mathcal{LM}$, $\mathcal{S}$ and $\mathcal{W}$ may be applied with little alteration to the general problem of testing nested conditional moment restrictions.

Alternative unrestricted statistics for testing MEM (2.5) and CEM (2.6) hypotheses may be also defined which ignore the information contained in the maintained hypothesis (3.2); *viz.* the unrestricted GEL-based statistics

$$\mathcal{LR}^h = \frac{2n\tilde{P}_n(\hat{\beta}, \tilde{\eta}) - J(M+1)K}{\sqrt{2J(M+1)K}}, \mathcal{LM}^h = \frac{n\tilde{\eta}'\hat{\Xi}\tilde{\eta} - J(M+1)K}{\sqrt{2J(M+1)K}} \tag{4.1}$$

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

$$\mathcal{S}^h = \frac{n\hat{h}(\hat{\beta})'\hat{\Xi}^{-1}\hat{h}(\hat{\beta}) - J(M+1)K}{\sqrt{2J(M+1)K}}. \tag{4.2}$$

It is straightforward to show from the analysis used to establish Theorems 4.1 and 4.2 similarly to DIN that these unrestricted statistics also each converge in distribution to a standard normal random variate and are mutually asymptotically equivalent but not to the restricted $\mathcal{J}$, $\mathcal{LR}$, $\mathcal{LM}$, $\mathcal{S}$ and $\mathcal{W}$. The statistics $\mathcal{LR}^h$ and $\mathcal{S}^h$ are forms of GMM and GEL statistics suggested in DIN, section 6, pp.67-71, adapted for testing the null hypothesis (3.1).

This section concludes with an asymptotic independence result between the restricted GMM statistic $\mathcal{J}$ for testing (3.1) and the corresponding statistic $\mathcal{J}^g$ for testing the maintained hypothesis (2.4) given by

$$\mathcal{J}^g = \frac{\mathcal{T}^g_{GMM} - JK}{\sqrt{2JK}}; \tag{4.3}$$

viz.

Theorem 4.3 *If Assumptions 3.1, 3.2, 3.3 and 3.4 hold for* $s = w$ *and* $s = (w', x')'$ *and if* $K \to \infty$ *and* $\zeta(K)^2 K^2/n \to 0$*, then* $\mathcal{J}$ *is asymptotically independent of* $\mathcal{J}^g \xrightarrow{d} N(0,1)$.

A similar result holds for the associated restricted GEL statistics $\mathcal{LR}$, $\mathcal{LM}$, $\mathcal{S}$ and $\mathcal{W}$ and their counterparts for testing (2.4) with the additional constraint $\zeta(K)^2 K^3/n \to 0$.

The practical import of Theorem 4.3 is that the overall asymptotic size of the test sequence may be controlled, e.g., (a) test (2.4) using $\mathcal{J}^g$; (b) given (2.4), test $E[u(z,\beta_0)|s] = 0$ using $\mathcal{J}$, with overall size $1-(1-\alpha_a)(1-\alpha_b)$, where α_a and α_b are the asymptotic sizes of the individual tests in (a) and (b) respectively.

5 Asymptotic Local Power

This section considers the asymptotic distribution of the above statistics under a suitable sequence of local alternatives. Recall that the dimension $f(K)$ of approximating functions $q_{\text{s}}^K(s)$ satisfies $f(K) = O(K)$ which for simplicity is assumed to be linear in K, i.e., $f(K) = MK$; see below (3.3). Essentially, the import of this restriction is that it ensures a difference in local power between the restricted statistics of section 3.4 and the unrestricted statistics of section 4 that ignore the maintained conditional moment information (2.4).

We follow the set-up in Eubank and Spielgeman (1990) and Hong and White (1995), see also Tripathi and Kitamura (2003), which utilise local alternatives to the null hypothesis (3.1) of the form

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

where $\beta_{n,0} \in \mathcal{B}$ is a non-stochastic sequence such that $\beta_{n,0} \to \beta_0$. We also assume that $E[\xi(w,x)|w] = 0$ to ensure that the maintained hypothesis $E[u(z,\beta_0)|w] = 0$ in (3.2) is not violated.

This sequence of local alternatives (5.1) is particularly apposite for CEM. It is also appropriate as a description of local alternatives to the MEM hypothesis $E[u(z,\beta_0)|x] = 0$ in which case local alternatives may be described by taking expectation of (5.1) conditional on x, i.e.,

$$E[u(z,\beta_{n,0})|x] = \frac{\sqrt[4]{JMK}}{\sqrt{n}}E[\xi(w,x)|x].$$

To obtain the asymptotic distribution of the statistics proposed in section 3.4 under local alternatives (5.1) we invoke the following assumption.

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

The next result summarises the limiting distribution of the restricted statistics $\mathcal{J}$, $\mathcal{LR}$, $\mathcal{LM}$, $\mathcal{S}$ and $\mathcal{W}$ under the sequence of local alternatives (5.1). Let $\Sigma(w,x) = \Sigma(w,x;\beta_0)$.[9]

Theorem 5.1 *Let Assumptions 3.1, 3.2, 3.3, 3.4 and 5.1 hold for $s = w$ or $s = (w',x')'$, $K \to \infty$ and $\zeta(K)^2K^2/n \to 0$. Then $\mathcal{J}$ converges in distribution to a $N(\mu/\sqrt{2},1)$ random variate, where*

$$\mu = E[\xi(w,x)'\Sigma(w,x)^{-1}\xi(w,x)].$$

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

Theorem 5.1 reveals that tests based on these statistics should be one-sided. Although not discussed here, a similar analysis to that underpinning Lemma 6.5, p.71, in DIN demonstrates the consistency of tests based on the statistics $\mathcal{J}$, $\mathcal{LR}$, $\mathcal{LM}$, $\mathcal{S}$ and $\mathcal{W}$.

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

[9] The function $\xi(w,x)$ is required to satisfy Assumption 3.1. Tests for MEM use approximating functions that depend solely on x or on w but not both requiring some restrictions to be placed on the form of $\xi(w,x)$. Close inspection of the Proof of Theorem 5.1 reveals that $\xi(w,x) = \Sigma(x,w,\beta_0)(\psi(x)+\varphi(w))$ satisfies Assumption 3.1 for arbitrary functions $\psi(x)$ and $\varphi(w)$. In this case it can be shown that $\mu = E[\psi(x)'\Sigma(x,\beta_0)\psi(x)] - E[\varphi(w)'\Sigma(w,\beta_0)\varphi(w)]$, where $\Sigma(s,\beta_0) = E[u(s,\beta_0)u(s,\beta_0)'|s]$.

Corollary 5.1 *Let Assumptions 3.1, 3.2, 3.3, 3.4 and 5.1 hold for* $s = (w', x')'$ *and* $\zeta(K)^2 K^2/n \rightarrow 0$. *Then* $\mathcal{S}^h$ *converges in distribution to a* $N(\mu_h/\sqrt{2}, 1)$ *random variate, where*

$$\mu_h = \sqrt{\frac{M}{M+1}}\mu.$$

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

This corollary provides a justification for restricting the number of elements in $q_0^K(s)$ to depend linearly on K. If M was permitted to approach infinity with K, μ_h would then differ little from μ with the consequence that unrestricted tests would have a similar discriminatory power as that of restricted tests to detect local departures from the null hypothesis H_0. Indeed, Corollary 5.1 indicates that M should be chosen as small as possible.

6 Some Computational Issues

This section describes how the vectors of approximating functions $q_{\text{S}}^K(s)$ and $q_{\text{MA}}^K(w)$ may be constructed using Bernstein polynomials. For expositional simplicity suppose that there is a single instrument w and a single covariate x each with unit interval $[0, 1]$ support; the univariate approach described below may straightforwardly be adapted for the vector instrument and covariate case.[10]

For the instrument w Bernstein polynomials of degree p are defined by

$$\mathcal{B}_{i,p}(w) = \frac{p!}{i!(p-i)!} w^i (1-w)^{p-i}, i = 0, ..., p. \tag{6.1}$$

Bernstein polynomials have the following properties: (a) $\sum_{i=0}^{p} \mathcal{B}_{i,p}(w) = 1$; (b) $\mathcal{B}_{i,p-1}(w) = \frac{p-i}{p}\mathcal{B}_{i,p}(w) + \frac{i+1}{p}\mathcal{B}_{i+1,p}(w)$. See, e.g., section 2 of Qian, Riedel and Rosenberg (2011). These properties have important consequences for the construction of the vectors of approximating functions.

[10]Bernstein polynomials for a variate a with unbounded support on the real line and sample a_i, $(i = 1, ..., n)$, may be obtained by substitution of the transformed variates $\Phi((a_i - \bar{a})/s_a)$, $(i = 1, ..., n)$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function and $\bar{a}$ and s_a are the sample mean and sample standard deviation respectively.

For the maintained hypothesis (2.4) consider the vector

$$q_{\text{MA}}^{K}(w) = (\mathcal{B}_{0,K-1}(w), \mathcal{B}_{1,K-1}(w), ..., \mathcal{B}_{K-1,K-1}(w))',$$

where $K \geq 2$.

For tests of CEM recall that $q_{\text{C}}^{K}(w,x)$ is the vector of additional approximating functions. Let $K^{\text{C}} = [(\text{A}_{\text{C}}K)^{1/2}]$, where A_{C} is a positive real constant, $[\cdot]$ denotes the integer part of $\cdot$ and $K^{\text{C}} < K$; hence A_{C} approximates M defined in section 3.2. The elements of $q_{\text{C}}^{K}(w,x)$ are defined as the cross products of each the elements of the vector

$$q^{K^{\text{C}}}(w) = (\mathcal{B}_{0,K^{\text{c}}-1}(w), \mathcal{B}_{1,K^{\text{c}}-1}(w), \mathcal{B}_{2,K^{\text{c}}-1}(w), ..., \mathcal{B}_{K^{\text{c}}-1,K^{\text{c}}-1}(w))'$$

with each element of the vector

$$q^{K^{\text{C}}}(x) = (\mathcal{B}_{0,K^{\text{c}}-1}(x), \mathcal{B}_{1,K^{\text{c}}-1}(x), \mathcal{B}_{2,K^{\text{c}}-1}(x), ..., \mathcal{B}_{K^{\text{c}}-2,K^{\text{c}}-1}(x))', \qquad (6.2)$$

where the Bernstein polynomials in terms of the covariate x are defined as in (6.1) with w replaced by x. Thus $q^{K}(w,x) = (q_{\text{MA}}^{K}(w)', q_{\text{C}}^{K}(w,x)')'$. Note that $\mathcal{B}_{K^{\text{c}}-1,K^{\text{c}}-1}(x)$ is excluded from $q^{K^{\text{C}}}(x)$ to avoid perfect multicollinearity between the elements of $q_{\text{MA}}^{K}(w)$ and $q_{\text{C}}^{K}(w,x)$.[11]

To define a test of MEM (2.5) let $K^{\text{M}} = [\text{A}_{\text{M}}K]$, where A_{M} is a positive real number and $K^{\text{M}} < K$. Now $q^{K}(w,x) = (q_{\text{MA}}^{K}(w)', q_{\text{M}}^{K}(x)')'$, where $q_{\text{M}}^{K}(x) = q^{K^{\text{M}}}(x)$ and $q^{K^{\text{M}}}(x)$ is

[11] To see this

$$\sum\nolimits_{i=0}^{K^{\text{C}}-1} \mathcal{B}_{i,K^{\text{c}}-1}(x)\mathcal{B}_{j,K^{\text{c}}-1}(w) = \mathcal{B}_{j,K^{\text{c}}-1}(w), (j = 0, ..., K^{\text{C}} - 1),$$

by the Bernstein polynomial Property (a) above. Now by Property (b), as $K^{\text{C}} < K$, $\mathcal{B}_{j,K^{\text{c}}-1}(w)$ is a linear function of Bernstein polynomials of order $K-1$, i.e.,

$$\sum\nolimits_{i=0}^{K^{\text{C}}-1} \mathcal{B}_{i,K^{\text{c}}-1}(x)\mathcal{B}_{j,K^{\text{c}}-1}(w) = a_j' q_1^K(w)$$

for some vector a_j, $(j = 0, ..., K^{\text{C}} - 1)$. Dropping $\mathcal{B}_{K^{\text{c}}-1,K^{\text{c}}-1}(x)$ from $q^{K^{\text{C}}}(x)$ solves the perfect multicollinearity problem as

$$\begin{aligned} \sum\nolimits_{i=0}^{K^{\text{C}}-2} \mathcal{B}_{i,K^{\text{c}}-1}(x)\mathcal{B}_{j,K^{\text{c}}-1}(w) &= \sum\nolimits_{i=0}^{K^{\text{C}}-1} \mathcal{B}_{i,K^{\text{c}}-1}(x)\mathcal{B}_{j,K^{\text{c}}-1}(w) - \mathcal{B}_{K^{\text{c}}-1,K^{\text{c}}-1}(x)\mathcal{B}_{j,K^{\text{c}}-1}(w) \\ &= \mathcal{B}_{j,K^{\text{c}}-1}(w) - \mathcal{B}_{K^{\text{c}}-1,K^{\text{c}}-1}(x)\mathcal{B}_{j,K^{\text{c}}-1}(w) \\ &= a_j' q_K(w) - \mathcal{B}_{K^{\text{c}}-1,K^{\text{c}}-1}(x)\mathcal{B}_{j,K^{\text{c}}-1}(w). \end{aligned}$$

defined as in (6.2). The Bernstein polynomial $\mathcal{B}_{K^{\text{M}}-1,K^{\text{M}}-1}(x)$ is excluded from $q^{K^{\text{M}}}(x)$ to avoid perfect multicollinearity between the elements of $q^K(w,x)$. As for A_{C} above, A_{M} approximates M for tests of MEM.[12]

7 Simulation Evidence

This section reports abbreviated results from an extensive set of simulation experiments undertaken to evaluate the behaviour and performance of the restricted tests of MEM (2.5) $E[u(z,\beta_0)|x]=0$ and CEM (2.6) $E[u(z,\beta_0)|w,x]=0$ based on various GMM and GEL statistics given in section 3.4. Results are also presented for the unrestricted test statistics discussed in section 4 that ignore the maintained hypothesis (2.4).

7.1 Experimental Design

All experiments concern the regression model

$$y=\beta_0 x+u, \tag{7.1}$$

where x is a scalar covariate and u an error term. For simplicity, the value of the parameter β_0 is set as 0. Consideration is restricted to the single parameter β_0 to ease the computation burden associated with GEL estimation. Cf. Example 2.3.

The data generating processes for the covariate x, instrument w and error term u are as follows. Let z_x and z_w be jointly normally distributed with mean zero, unit variance and correlation coefficient ρ, where $\rho\in(-1,1)$ and $\rho\neq 0$; in all experiments we set $\rho=0.7$. The covariate x and instrument w are generated according to $x=\Phi(z_x)$ and $w=\Phi(z_w)$, where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function; hence, x and w are marginally distributed as uniform random variates on the unit interval

[12] Other classes of approximating functions such as B-Splines are also possible choices but severe if not perfect multicollinearity between the elements of the approximating functions used to approximate the null and maintained hypotheses is likely to occur. To ascertain the collinearity properties of B-Spline approximating functions in this setting is difficult because the positioning of the knots depends on K, K^{M} and K^{C}. The use of a generalised inverse in computations in the place of the inverse would avoid these collinearity difficulties.

$[0, 1]$. Let v be defined by

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

where υ is standard normally distributed and independent of z_x and z_w. The error term $u = v/\sqrt{var[v]}$ in (7.1) is by definition restricted to have unit variance again for reasons of simplicity and ease of comparison. Note that $var[v] = a^2(1+\rho^2)(\rho^{-1}-\rho)^2 + \tau^2(1-\rho^2) + 1$.

This model is characterised by the following properties: (a) the maintained hypothesis (2.4) is satisfied, i.e., $E[u|w] = 0$; (b) for the MEM (2.5) hypothesis

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

Thus $E[u|x] = 0$ if $\tau = 0$ and $E[u|x] \neq 0$ if $\tau \neq 0$; (c) for the CEM (2.6) hypothesis,

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

Hence $E[u|w,x] = 0$ if $a = \tau = 0$ and $E[u|w,x] \neq 0$ if $a \neq 0$ or $\tau \neq 0$.

Clearly, under the maintained hypothesis (2.4) $cov[u, w] = 0$ whereas under MEM (2.5) $cov[u, x] = 0$ if $\tau = 0$ for all values of a.

Sample sizes $n = 200$, 500, 1000 and 3000 are used in the experiments concerned with empirical size; nominal size is 0.05. Sample sizes of $n = 200$ and 500 only are considered for simplicity in the experiments examining empirical power. These experiments examine two designs, i.e., a varies and $\tau = 0$, i.e., MEM holds but CEM does not unless $a = 0$; $a = 0$ and τ varies, i.e., both MEM and CEM do not hold unless $\tau = 0$. Each experiment employs 5000 replications.

7.1.1 Estimators

Assumption 3.3(c) requires $\sqrt{n}$-consistent although not necessarily efficient estimation of β_0. We consider two stage least squares estimation (2SLQ) computed using the single instrument w, GMM (GMM), continuous updating (CUE), empirical likelihood (EL) and exponential tilting (ET) estimators computed under various hypotheses; GMM here refers to efficient two-step GMM where the weighting matrix is computed using 2SLQ. The subscripts MA, M and C indicate computation incorporating restrictions of the maintained, MEM and CEM hypotheses respectively.

GMM and GEL were computed using the simplex search algorithm of `MATLAB` to ensure a local optima is located. EL and ET require evaluation of $\hat{\lambda}(\beta)$ to construct the requisite profile GEL objective function; since the GEL objective function is twice continuously differentiable in λ the Newton method was used to locate $\hat{\lambda}(\beta)$ for given β. The computation of EL requires some care since the EL criterion involves the logarithm function and 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.[13]

7.1.2 Test Statistics

Tests for MEM (2.5) $E[u|x] = 0$ and CEM (2.6) $E[u|w,x] = 0$ based on unrestricted statistics that ignore the maintained hypothesis (2.4) are denoted respectively in the following by the superscripts DIN-M and DIN-C. Where relevant the test statistic subscripts CUE, EL and ET refer to the GEL criterion used to construct the test statistic with the argument of the statistic denoting the estimator at which it is evaluated whose subscript indicates the moment conditions MA (3.3), M or C (3.4) employed, *viz.* $\mathcal{I}^i(j)$ (3.7) where $i =$ DIN-M, DIN-C and $j =$ 2SLQ, GMM_k ($k =$ MA, M); $\mathcal{LR}^i_l(j)$ and $\mathcal{LM}^i_l(j)$ (4.1) where $i =$ DIN-M, DIN-C, $j =$ 2SLQ, EL_k, ET_k ($k =$ MA, M, C) and $l =$EL, ET; $\mathcal{S}^i(j)$ (4.2) where $i =$ DIN-M, DIN-C and $j =$2SLQ, CUE_k, EL_k, ET_k ($k =$MA, M, C).[14] Note that $\mathcal{J}^{\text{DIN-M}}(\text{2SLQ})$ and $\mathcal{S}^{\text{DIN-M}}_{\text{CUE}}(\text{2SLQ})$ are identical statistics as are $\mathcal{J}^{\text{DIN-C}}(\text{2SLQ})$ and $\mathcal{S}^{\text{DIN-C}}_{\text{CUE}}(\text{2SLQ})$.

Restricted tests for MEM (2.5) $E[u|x] = 0$ and CEM (2.6) $E[u|w,x] = 0$ that incorpo-

[13]See Owen (2001, (12.3), p.235). The code is available at `http://www-stat.stanford.edu/~owen/empirical/`

[14]A number of asymptotically equivalent test statistics for the maintained hypothesis were also investigated, *viz.* $\mathcal{J}^g(j)$ (4.3) where $j =$2SLQ, GMM_{MA}; $\mathcal{LR}^g_l(j)$ and $\mathcal{LM}^g_l(j)$ (4.1) where $j =$ 2SLQ, EL_{MA}, ET_{MA} and $l =$EL, ET; $\mathcal{S}^g_l(j)$ (4.2) where $j =$ 2SLQ, CUE_{MA}, EL_{MA}, ET_{MA} and $l =$CUE, EL, ET. The expressions for the latter statistics adapt (4.1) and (4.2) for the maintained hypothesis. The Hausman test based on an auxiliary regression as described in Davidson and Mackinnon (1993, section 7.9, p.237) or Wooldridge (2002, section 6.2.1, p.118) was also considered. Results are available on request from the authors.

rate the maintained hypothesis (2.4) $E[u|w]=0$ were also investigated. The following notation is adopted. Let

$$\mathcal{J}^i(j,k)=n\hat{h}(\hat{\beta}_j)'\hat{\Xi}^{-1}\hat{h}(\hat{\beta}_j)-n\hat{g}(\hat{\beta}_k)'\hat{\Omega}^{-1}\hat{g}(\hat{\beta}_k),$$

where $\hat{\Omega}=\sum_{i=1}^n g_i(\hat{\beta}_{2\text{SLQ}})g_i(\hat{\beta}_{2\text{SLQ}})'/n$ and $\hat{\Xi}=\sum_{i=1}^n h_i(\hat{\beta}_{2\text{SLQ}})h_i(\hat{\beta}_{2\text{SLQ}})'/n$; the indices i, j and k are defined below. Also let

$$\mathcal{LR}_l^i(j,k)=2n[\tilde{P}_n(\hat{\beta}_j,\tilde{\eta}_j)-\hat{P}_n(\hat{\beta}_k,\hat{\lambda}_k)]$$

where $\tilde{\eta}_j=\arg\max_{\eta\in\tilde{\Delta}_n(\hat{\beta}_j)}\tilde{P}_n(\hat{\beta}_j,\eta)$ and $\hat{\lambda}_k=\arg\max_{\lambda\in\hat{\Lambda}_n(\hat{\beta}_k)}\hat{P}_n(\hat{\beta}_k,\lambda)$. Corresponding Lagrange multiplier, score and Wald-type statistics are defined respectively as

$$\begin{aligned}\mathcal{LM}_l^i(j,k) &= n(\tilde{\eta}_j-\hat{\eta}_k)'\hat{\Xi}(\tilde{\eta}_j-\hat{\eta}_k),\\ \mathcal{S}_l^i(k) &= \textstyle\sum_{i=1}^n\rho_1(\hat{\lambda}_k'g_i(\hat{\beta}_k))s_i(\hat{\beta}_k)'S_{\text{S}}'\hat{\Xi}^{-1}S_{\text{S}}\displaystyle\sum_{i=1}^n\rho_1(\hat{\lambda}_k'g_i(\hat{\beta}_k))s_i(\hat{\beta}_k)/n,\\ \mathcal{W}_l^i(j) &= n\tilde{\eta}_j'S_{\text{S}}(S_{\text{S}}'\hat{\Xi}^{-1}S_{\text{S}})^{-1}S_{\text{S}}'\tilde{\eta}_j,\end{aligned}$$

where $\hat{\eta}_k=S_{\text{MA}}\hat{\lambda}_k$ and $S_{\text{MA}}=I_J\otimes(I_K,0_{MK})'$ and $S_{\text{S}}=I_J\otimes(0_K,I_{MK})'$ are $J(M+1)K\times JMK$ selection matrices; cf. section 3.4. Note that $\mathcal{J}^{\text{M}}$(2SLQ,2SLQ) and $\mathcal{LR}_{\text{CUE}}^{\text{M}}$(2SLQ,2SLQ) are identical statistics as are $\mathcal{J}^{\text{C}}$(2SLQ,2SLQ) and $\mathcal{LR}_{\text{CUE}}^{\text{C}}$(2SLQ,2SLQ).

The various indices are defined similarly to those of the unrestricted statistics, *viz.* the superscript i =M, C where M and C refer respectively to the MEM (2.5) $E[u|x]=0$ or CEM (2.6) $E[u|w,x]=0$ hypothesis under test, the arguments j,k =2SLQ, CUE$_m$, EL$_m$, ET$_m$ (m =MA, M, C) to the estimators employed and the subscript l =CUE,EL, ET to which member of the GEL class was used to construct the test.

The statistics $\mathcal{J}^i(j,k)$, $\mathcal{LR}_l^i(j,k)$, $\mathcal{LM}_l^i(j,k)$, $\mathcal{S}_l^i(j)$ and $\mathcal{W}_l^i(j)$ are calibrated against a chi-square distribution with JMK degrees of freedom and are referred to as *non-standardised* statistics. *Standardized* versions are defined as in (3.7), (3.9), (3.10), (3.11) and (3.12), e.g., $[\mathcal{J}^i(j,k)-JMK]/\sqrt{2JMK}$, and are calibrated against a standard normal distribution. Recall from section 5 that tests should be one sided.

We also consider the behaviour of restricted statistics that are robust to estimation effects suggested in Smith (1997, section II.2); see also Smith (2011, section 5). Let

$$\hat{\Psi} = \begin{pmatrix} 0 & \hat{H}' \\ \hat{H} & \hat{\Xi} \end{pmatrix},$$

where $\hat{H} = -\sum_{i=1}^{n} x_i \otimes q^K(w_i, x_i)/n$. Define the selection matrix S_ψ such that $S'_\psi(\beta', \eta')' = S'_{\rm s}\eta$. The corresponding non-standardised GEL score and Wald statistics are then defined by

$$\begin{aligned} \bar{\mathcal{S}}^i_l(k) &= \textstyle\sum_{i=1}^{n} \rho_1(\hat{\lambda}'_k g_i(\hat{\beta}_k)) s_i(\hat{\beta}_k)' S'_\psi \hat{\Psi}^{-1} S_\psi \sum_{i=1}^{n} \rho_1(\hat{\lambda}'_k g_i(\hat{\beta}_k)) s_i(\hat{\beta}_k)/n, \\ \bar{\mathcal{W}}^i_l(j) &= n\tilde{\eta}'_j S_{\rm s}(S'_\psi \hat{\Psi}^{-1} S_\psi)^{-1} S'_{\rm s}\tilde{\eta}_j, \end{aligned}$$

with their standardized counterparts given by $[\bar{\mathcal{S}}^i_l(k) - JMK]/\sqrt{2JMK}$ and $[\bar{\mathcal{W}}^i_l(k) - JMK]/\sqrt{2JMK}$.

7.2 Choice of the Number of Instruments

To implement the above tests requires a choice of the number of instruments to employ under the maintained hypothesis. The results in DIN, Table 1, p.71, apply for the choice of $O(K)$. Accordingly, K must satisfy $K^4/n \to 0$, e.g., $K = [Cn^{1/5}]$ for some $C > 0$ but then a choice of C is necessary. The use of the method described in Donald, Imbens and Newey (2009) for empirically determining K predominantly resulted in the choice $K = 2$. Consequently, to explore the robustness of the results to the choice of K, the additional alternatives $K = 3$ or 5 are also examined.[15]

To test the MEM (2.5) $E[u|x] = 0$ or CEM (2.6) $E[u|w, x] = 0$ hypotheses $K^{\rm M} = [{\rm A}_{\rm M} K]$ and $K^{\rm C} = [({\rm A}_{\rm C} K)^{1/2}]$ are required dependent on the *ad hoc* constants ${\rm A}_{\rm M}$ and ${\rm A}_{\rm C}$; see section 6. The choices ${\rm A}_{\rm M} = 1$ or 1.5 and ${\rm A}_{\rm C} = 2$ or 4.5 are considered. Recall from section 6 that ${\rm A}_{\rm M}$ and ${\rm A}_{\rm C}$ approximate M of section 3.2 and by Theorem 5.1 and Corollary 5.1 should be chosen as small as possible.

Tables 1 and 2 summarise the numbers of instruments used.

[15] Alternative possible criteria for the choice of K are information criteria such as AIC or BIC.

		$\mathrm{A}_{\mathrm{M}} = 1$	$\mathrm{A}_{\mathrm{M}} = 1.5$	
K	K^{M}	Total Number of Instruments	K^{M}	Total Number of Instruments
2	2	3	3	4
3	3	5	4	6
5	5	9	7	11

Table 1: MEM Intruments

		$\mathrm{A}_{\mathrm{C}} = 2$	$\mathrm{A}_{\mathrm{C}} = 4.5$	
K	K^{C}	Total Number of Instruments	K^{C}	Total Number of Instruments
2	2	4	3	8
3	2	5	3	9
5	3	11	4	17

Table 2: CEM Intruments

7.3 Empirical Size

The results on empirical size reported in Appendix B and discussed below are a substantially reduced subset of the simulation experiments undertaken. Overall these experiments revealed that nominal size is approximated relatively more closely by the empirical size of (a) the non-standardised tests; (b) tests based on efficient estimators, cf. Tripathi and Kitamura (2003); (c) the score-type statistic $\bar{\mathcal{S}}_l^i(k)$ robust to estimation effects. Consequently results for these forms of statistics only are presented. Both Wald versions $\mathcal{W}_l^i(j)$ (3.12) and $\bar{\mathcal{W}}_l^i(j)$ of test are also excluded as their empirical size properties are generally unsatisfactory. Results are presented for $K = 2$, 5; those for $K = 3$ closely resemble those for $K = 2$ and are therefore omitted.

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

7.3.1 MEM

Tables B.1-B.2 in Appendix B present the rejection frequencies for $K = 2$, 5 and $\mathrm{A}_{\mathrm{M}} = 1$, 1.5 for unrestricted DIN tests for MEM (2.5) $E[u|x] = 0$ that ignore the maintained hypothesis moment restrictions (2.4) $E[u|w] = 0$ whereas Tables B.3-B.4 report the corresponding results for restricted tests that incorporate these moment constraints.

In general, the empirical size of the non-standardised versions of the unrestricted $\mathcal{LR}^{\text{DIN-M}}_{\text{EL}}(\text{EL}_{\text{M}})$, $\mathcal{LR}^{\text{DIN-M}}_{\text{ET}}(\text{ET}_{\text{M}})$ and $\mathcal{LM}^{\text{DIN-M}}_{\text{EL}}(\text{EL}_{\text{M}})$, $\mathcal{LM}^{\text{DIN-M}}_{\text{ET}}(\text{ET}_{\text{M}})$ tests suffer from size distortions for moderate sample sizes $n = 200$ and 500. Of the remaining statistics, the non-standardised GMM statistic $\mathcal{J}^{\text{DIN-M}}(\text{GMM}_{\text{M}})$ (4.3) and GEL score statistics $\mathcal{S}^{\text{DIN-M}}_{\text{CUE}}(\text{CUE}_{\text{M}})$, $\mathcal{S}^{\text{DIN-M}}_{\text{EL}}(\text{EL}_{\text{M}})$, $\mathcal{S}^{\text{DIN-M}}_{\text{ET}}(\text{ET}_{\text{M}})$ (4.2) display satisfactory empirical size properties for most sample sizes. For a given sample size n there is a deterioration in performance to a lesser or greater degree for all statistics as K increases from 2 to 5 for fixed A_{M} and as A_{M} increases from 1 to 1.5 for fixed K, i.e., as the number of unconditional moments under test increases. In summary, the non-standardised GMM and GEL score forms of DIN-type test statistic for MEM appear to be the most reliable in terms of empirical size.

The overall conclusions for the restricted test statistics that incorporate the maintained moment restrictions (2.4) are similar. The performances of the non-standardised LR (3.9) and Lagrange multiplier (3.10) forms of test are more unsatisfactory at the smallest sample size $n = 200$; there is again a substantial deterioration for the larger $K = 5$ and to a lesser degree for $\text{A}_{\text{M}} = 1.5$. To summarise, in general the non-standardised GMM $\mathcal{J}^{\text{M}}(\text{GMM}_{\text{M}},\text{GMM}_{\text{MA}})$ (3.7) and CUE LR $\mathcal{LR}^{\text{M}}_{\text{CUE}}(\text{CUE}_{\text{M}},\text{CUE}_{\text{MA}})$ (3.9) forms of restricted test statistic for MEM appear to display the most satisfactory empirical size behaviour as do the robust score statistics $\bar{\mathcal{S}}^{\text{M}}_{\text{EL}}(\text{EL}_{\text{M}})$, $\bar{\mathcal{S}}^{\text{M}}_{\text{ET}}(\text{ET}_{\text{M}})$.

7.3.2 CEM

Tables B.5-B.6 in Appendix B present the rejection frequencies for $K = 2$, 5 and $\text{A}_{\text{C}} = 2$, 4.5 for unrestricted DIN tests of the CEM null hypothesis (2.6) $E[u|w, x] = 0$ whereas Tables B.7-B.8 report the corresponding results for the restricted tests.

The general conclusions are quite similar to those for the tests of MEM. Overall performance worsens for the larger $K = 5$ and $\text{A}_{\text{M}} = 4.5$ for all test versions. Of the DIN-type non-standardised tests the empirical size properties of the GMM statistic $\mathcal{J}^{\text{DIN-C}}(\text{GMM}_{\text{C}})$ (4.3) and GEL score statistics $\mathcal{S}^{\text{DIN-C}}_{\text{CUE}}(\text{CUE}_{\text{C}})$, $\mathcal{S}^{\text{DIN-C}}_{\text{EL}}(\text{EL}_{\text{C}})$, $\mathcal{S}^{\text{DIN-C}}_{\text{ET}}(\text{ET}_{\text{C}})$ (4.2) appear to be satisfactory for most sample sizes. For the non-standardised restricted tests, the GMM $\mathcal{J}^{\text{C}}(\text{GMM}_{\text{C}},\text{GMM}_{\text{MA}})$ (3.7) and CUE LR $\mathcal{LR}^{\text{C}}_{\text{CUE}}(\text{CUE}_{\text{C}},\text{CUE}_{\text{MA}})$ (3.9) forms and the robust versions of score statistic $\bar{\mathcal{S}}^{\mathrm{C}}_{\mathrm{EL}}(\mathrm{EL}_{\mathrm{MA}})$, $\bar{\mathcal{S}}^{\mathrm{C}}_{\mathrm{ET}}(\mathrm{ET}_{\mathrm{MA}})$ (3.11) display empirical size closest to the nominal 0.05.

7.4 Empirical Size-Adjusted Power

The reported results on empirical size-adjusted power in Appendix C like those for empirical size are a substantially reduced subset of the simulation experiments undertaken. Results are also only presented for $K = 2$ as *ceteris paribus* size-adjusted power tends to decline relatively and sometimes substantially for the larger $K = 5$. Typically power increases substantially as sample size n increases from 200 to 500. Full simulation results are available from the authors upon request.

7.4.1 $\tau = 0$

MEM

Recall that in this case the MEM hypothesis (2.5) $E[u|x] = 0$ holds but the CEM hypothesis (2.6) $E[u|w, x] = 0$ is violated unless $a = 0$. Unsurprisingly, results for tests of MEM, not reported here, indicate that size-adjusted power closely approximates nominal size.

CEM

Tables C.1-C.2 in Appendix C present the size-adjusted powers for $K = 2$ and different values of a and A_{C} for tests based on unrestricted DIN-type statistics of CEM (2.6) $E[u|w, x] = 0$ that ignore the maintained hypothesis moment restrictions (2.4) $E[u|w] = 0$ whereas Tables C.3-C.4 report the corresponding results for the restricted tests that incorporate these moment constraints.

At small values of a the unrestricted DIN-type GEL statistics $\mathcal{LM}^{\mathrm{DIN\text{-}C}}_{\mathrm{EL}}(\mathrm{EL}_{\mathrm{C}})$ and $\mathcal{LM}^{\mathrm{DIN\text{-}C}}_{\mathrm{ET}}(\mathrm{ET}_{\mathrm{C}})$ have maximum size-adjusted power with the DIN-type GEL statistics $\mathcal{LR}^{\mathrm{DIN\text{-}C}}_{\mathrm{EL}}(\mathrm{EL}_{\mathrm{C}})$ and $\mathcal{LR}^{\mathrm{DIN\text{-}C}}_{\mathrm{ET}}(\mathrm{ET}_{\mathrm{C}})$ slightly less powerful. These findings are lessened for larger a and for the larger sample size $n = 500$. However, these tests are precisely those that displayed the least satisfactory correspondence between empirical and nominal size.

The powers of the remaining statistics are broadly similar with the GEL score statistics $\mathcal{S}^{\text{DIN-C}}_{\text{EL}}(\text{EL}_{\text{C}})$ and $\mathcal{S}^{\text{DIN-C}}_{\text{ET}}(\text{ET}_{\text{C}})$ marginally superior. The effect of increasing A_{C} is to increase power contrary to the large sample theoretical prediction of section 5 for $K = 2$ with this finding reversed for the larger $K = 5$.

For the restricted tests, GEL Lagrange multiplier $\mathcal{LM}^{\text{C}}_{\text{EL}}(\text{EL}_{\text{C}},\text{EL}_{\text{MA}})$, $\mathcal{LM}^{\text{C}}_{\text{ET}}(\text{ET}_{\text{C}},\text{ET}_{\text{MA}})$ and LR $\mathcal{LR}^{\text{C}}_{\text{EL}}(\text{EL}_{\text{C}},\text{EL}_{\text{MA}})$, $\mathcal{LR}^{\text{C}}_{\text{ET}}(\text{ET}_{\text{C}},\text{ET}_{\text{MA}})$ tests dominate for small a but this result is ameliorated for larger values of a and sample size n. Again, as above, empirical and nominal size differences can be quite large for these statistics especially at the smaller sample size $n = 200$. Of the other tests, there is relatively little difference in power among the statistics but both GEL robustified score statistics $\bar{\mathcal{S}}^{\text{C}}_{\text{EL}}(\text{EL}_{\text{MA}})$, $\bar{\mathcal{S}}^{\text{C}}_{\text{ET}}(\text{ET}_{\text{MA}})$ appear marginally superior. Again the effect of increasing A_{C} on power when $K = 2$ runs counter to the theoretical prediction of section 5 but holds if $K = 5$.

Overall, incorporation of the maintained moment conditions leads to improvements in power as expected from the theoretical results of section 5 with the difference in power between DIN-type and restricted tests for the CEM hypothesis (2.6) $E[u|w, x] = 0$ larger when A_{C} is smaller.

7.4.2 $a = 0$

MEM

Tables C.5-C.6 in Appendix C present the size-adjusted powers for different values of τ and A_{M} for unrestricted DIN-type tests of MEM (2.5) $E[u|x] = 0$ whereas Tables C.7-C.8 report the corresponding results for the restricted tests.

In general, and in line with the theoretical prediction of section 5, power decreases with increased A_{M}.

For small τ and sample size $n = 200$ the differences in power between the various DIN-type tests are relatively small although the power associated with the GEL LM $\mathcal{LM}^{\text{DIN-M}}_{\text{EL}}(\text{EL}_{\text{M}})$, $\mathcal{LM}^{\text{DIN-M}}_{\text{ET}}(\text{ET}_{\text{M}})$ tests is somewhat less.

All restricted tests except those based on the GEL LM statistics $\mathcal{LM}^{\text{M}}_{\text{EL}}(\text{EL}_{\text{M}},\text{EL}_{\text{MA}})$, $\mathcal{LM}^{\text{M}}_{\text{ET}}(\text{ET}_{\text{M}},\text{ET}_{\text{MA}})$ appear to provide similar empirical power.

As discussed previously the power differences among tests are lessened for larger values of τ and sample size n. Likewise, restricted tests again appear more powerful than the unrestricted DIN-type tests that ignore the moment restrictions arising from the maintained hypothesis.

CEM

Tables C.9-C.10 in Appendix C present the size-adjusted powers for different values of τ and A_{C} for unrestricted DIN-type tests of CEM (2.6) $E[u|w,x]=0$ whereas Tables C.11-C.12 report the corresponding results for the restricted tests.

In general similarly to the tests of MEM power decreases with increases in A_{C} as expected from section 5.

Size-adjusted power is mostly similar except for the unrestricted GEL LM $\mathcal{LM}^{\text{DIN-C}}_{\text{EL}}(\mathrm{EL}_{\mathrm{C}})$, $\mathcal{LM}^{\text{DIN-C}}_{\text{ET}}(\mathrm{ET}_{\mathrm{C}})$ tests especially for smaller values of τ and the smaller sample size $n=200$.

Among the restricted tests the GMM statistics $\mathcal{J}^{\mathrm{C}}(\mathrm{GMM}_{\mathrm{C}},\mathrm{GMM}_{\mathrm{MA}})$ (3.7), the GEL statistic $\mathcal{LR}^{\mathrm{C}}_{\mathrm{CUE}}(\mathrm{CUE}_{\mathrm{C}},\mathrm{CUE}_{\mathrm{MA}})$ (3.9) and the robust GEL score statistics $\bar{\mathcal{S}}^{\mathrm{C}}_{\mathrm{EL}}(\mathrm{EL}_{\mathrm{MA}})$, $\bar{\mathcal{S}}^{\mathrm{C}}_{\mathrm{ET}}(\mathrm{ET}_{\mathrm{MA}})$ dominate in terms of size-adjusted power.

Again the restricted test statistics that incorporate the maintained alternative hypothesis display higher power than the unrestricted DIN-type tests. Interestingly both unrestricted and restricted tests for CEM appear more powerful than the corresponding tests for MEM when the MEM hypothesis is violated.

7.5 Summary

The empirical size of non-standardised tests more closely approximates nominal size than that of standardised tests. Restricted tests dominate those defined by unrestricted DIN-type statistics in terms of size-adjusted power. Power typically declines for increases in the constants A_{M} and A_{C} for tests of MEM and CEM respectively.

The MEM (2.5) $E[u|x]=0$ empirical sizes of tests based on the restricted GMM form $\mathcal{J}^{\mathrm{M}}(\mathrm{GMM}_{\mathrm{M}},\mathrm{GMM}_{\mathrm{MA}})$, the GEL statistic $\mathcal{LR}^{\mathrm{M}}_{\mathrm{CUE}}(\mathrm{CUE}_{\mathrm{M}},\mathrm{CUE}_{\mathrm{MA}})$ and the robust GEL score versions $\bar{\mathcal{S}}^{\mathrm{M}}_{\mathrm{EL}}(\mathrm{EL}_{\mathrm{M}})$ and $\bar{\mathcal{S}}^{\mathrm{M}}_{\mathrm{ET}}(\mathrm{ET}_{\mathrm{M}})$ most closely approximate nominal size. However, when testing against deviations from MEM restricted CEM tests dominate MEM tests in terms of size-adjusted power.

Of the restricted tests for the CEM null hypothesis (2.6) $E[u|w,x]=0$ those employing the GMM form $\mathcal{J}^{\mathrm{C}}(\mathrm{GMM_C},\mathrm{GMM_{MA}})$ (3.7), the GEL statistic $\mathcal{LR}^{\mathrm{C}}_{\mathrm{CUE}}(\mathrm{CUE_C},\mathrm{CUE_{MA}})$ (3.9), and the robust GEL score statistics $\bar{\mathcal{S}}^{\mathrm{C}}_{\mathrm{EL}}(\mathrm{EL_{MA}})$ and $\bar{\mathcal{S}}^{\mathrm{C}}_{\mathrm{ET}}(\mathrm{ET_{MA}})$ have empirical size closest to the nominal 0.05. The robust GEL score-based, $\bar{\mathcal{S}}^{\mathrm{C}}_{\mathrm{EL}}(\mathrm{EL_{MA}})$ and $\bar{\mathcal{S}}^{\mathrm{C}}_{\mathrm{ET}}(\mathrm{ET_{MA}})$, forms of test appear marginally superior to the GMM form $\mathcal{J}^{\mathrm{C}}(\mathrm{GMM_C},\mathrm{GMM_{MA}})$ (3.7) and the GEL statistic $\mathcal{LR}^{\mathrm{C}}_{\mathrm{CUE}}(\mathrm{CUE_C},\mathrm{CUE_{MA}})$ (3.9) in terms of size-adjusted power against deviations from the MEM or CEM hypotheses.

8 Conclusions

The primary focus of this article has concerned definitions of and tests for exogeneity appropriate for models defined by a set of semiparametric conditional moment restrictions where a finite dimensional parameter vector is the object of inferential interest. The paper argues that a definition of (marginal) exogeneity (in mean) (MEM) proposed in Blundell and Horowitz (2004) may not be adequate for particular circumstances. An alternative definition of (conditional) exogeneity (in mean) (CEM) is provided. The latter definition is quite closely related to that for classical parametric models.

A second contribution is to propose GMM- and GEL-based test statistics for additional conditional moment restrictions that include both MEM and CEM hypotheses as special cases. By reinterpreting the respective hypotheses as an infinite number of unconditional moment restrictions the corresponding tests may therefore be formulated as tests for additional sets of infinite numbers of unconditional moment restrictions. The limiting distributions of these test statistics are derived under the null hypotheses of marginal and conditional exogeneity and suitable sequences of local alternatives to these hypotheses. These results suggest that restricted tests that incorporate maintained moment constraints should dominate in terms of power unrestricted tests that ignore such information.

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

References

Basu, D. (1977): "On the Elimination of Nuisance Parameters", *Journal of the American Statistical Association*, 72, 355-366.

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

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

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

Chalak, K., and H. White (2006): "An Extended Class of Instrumental Variables for the Estimation of Causal Effects," working paper, U.C. San Diego.

Chalak, K., and H. White (2011): "Viewpoint: An Extended Class of Instrumental Variables for the Estimation of Causal Effects," *Canadian Journal of Economics*, 44, 1-51.

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

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

Chen, X., and D. Pouzo (2012): "Estimation of Nonparametric Conditional Moment

Models with Possibly Nonsmooth Generalized Residuals," *Econometrica*, 80, 277-321.

Chernozhukov, V., and H. Hong (2003): "An MCMC Approach to Classical Estimation," *Journal of Econometrics*, 115, 293-346.

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

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

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

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

Deaton, A. (2010): "Instruments, Randomization, and Learning about Development," *Journal of Economic Literature*, 48, 424-455.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Hayashi, F. (2000): *Econometrics.* Princeton: Princeton University Press.

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

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

Honore, B.E., and L. Hu (2004): "On the Performance of Some Robust Instrumental Variables Estimators," *Journal of Business and Economic Statistics*, 22, 30-39.

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

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

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

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

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

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

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

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

Qian, W., Riedel, M.D., and I. Rosenberg (2011): "Uniform Approximation and Bernstein Polynomials with Coefficients in the Unit Interval," *European Journal of Combinatorics,* 32, 448-463.

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

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

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

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

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

Smith, R.J. (2007a): "Efficient Information Theoretic Inference for Conditional Moment Restrictions,"*Journal of Econometrics*, 138, 430-460.

Smith, R.J. (2007b): "Local GEL Estimation with Conditional Moment Restrictions". Chapter 4 in G.D.A. Phillips and E. Tzavalis, eds., *The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis*, 100-122. Cambridge: Cambridge University Press.

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

Stock, J.H., and M.W. Watson (2007): *Introduction to Econometrics.* New York: Pearson Addison Wesley.

Tauchen, G. (1985): “Diagnostic Testing and Evaluation of Maximum Likelihood Models,” *Journal of Econometrics*, 30, 415-443.

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

White, H., and K. Chalak (2010): “Testing a Conditional Form of Exogeneity,” *Economics Letters*, 109, 88-90.

Wooldridge, J.M. (1990): “A Unified Approach to Robust, Regression-Based Specification Tests,” *Econometric Theory*, 6, 17-43.

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

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

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

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

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

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

Appendix A: Proofs

Throughout the Appendix, C will denote a generic positive constant that may be different in different uses, and CS, J, M, T and c_r Cauchy-Schwarz, Jensen, Markov, triangle and Loève c_r inequalities respectively.[16] Also we write w.p.a.1 for "with probability approaching 1".

A.1 Useful Lemmata

The following Lemma allows the relaxation of Assumption 6 in DIN for the GEL class of estimators.

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

Proof: Write $b_i = \sup_{\beta\in\mathcal{B}} \|u(z_i,\beta)\|^2$. By iterated expectations and 3.3(d), $E[b_i] = E[E[b_i|w]] < \infty$ for $1 \leq i \leq n$. Hence, it 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

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

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

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

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

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

[16] We use the general version of the Loève c_r inequality as stated in Davidson (1994, p.140).

Proof: Let $g_{z,n} = g(z, \beta_{0,n})$ and write $T = T_1 + T_2$ where

$$\begin{aligned} T_1 &= \sum_{i,j:i<j} \sqrt{\frac{2}{n^2 k}} g'_{z_i,n} C_n g_{z_j,n} \\ T_2 &= \frac{\sum_i g'_{z_i,n} C_n g_{z_i,n}/n - k}{\sqrt{2k}} \end{aligned}$$

Since $E[T_2] = 0$ and $var[T_2] \leq E[(g'_{i,n} C_n g_{i,n})^2]/2kn \to 0$, $T_2 \xrightarrow{p} 0$.

To prove the asymptotic normality of T_1 we verify the hypotheses of Hall (1984, Theorem 1, pp.3-4). Define

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

Then

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

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

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

On the other hand

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

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

$$\begin{aligned} \frac{1}{nk^2} E[(g'_{z_1,n} C_n g_{z_2,n})^4] &\leq \frac{1}{nk^2} E[(g'_{z_1,n} C_n g_{z_1,n})^2 (g'_{z_2,n} C_n g_{z_2,n})^2] \\ &= (\frac{1}{k\sqrt{n}} E[(g'_{z_1,n} C_n g_{z_1,n})^2])^2 \to 0. \end{aligned}$$

Since $E[G_n(z_1, z_2)^2]/E[H_n(z_1, z_2)^2]^2 = 1/k \to 0$, $T_1 \xrightarrow{d} N(0, 1)$ as required. ■

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

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

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

Proof: Let $g_{z_i} = g(z_i,\beta_0)$ and write $\mathcal{T} = \mathcal{T}_1 + \mathcal{T}_2$ where

$$\begin{aligned}
\mathcal{T}_1 &= \sum_{i,j:i<j} \sqrt{\frac{2}{n^2k}} g'_{z_i} Q g_{z_j} \\
\mathcal{T}_2 &= \frac{\sum_i g'_{z_i} Q g_{z_i}/n - ak}{\sqrt{2k}}
\end{aligned}$$

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

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

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

Then

$$\begin{aligned}
G_n(u,v) &\equiv E[H_n(z_1,u)H_n(z_1,v)] \\
&= \frac{2}{n^2k} E[g'_u Q g_{z_1} g'_{z_1} Q g_v] \\
&= \frac{2}{n^2k} g'_u Q\Omega Q g_v.
\end{aligned}$$

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

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

Also

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

Now, as Ω is positive definite, by CS

$$\begin{aligned} E[(g'_{z_1} Q g_{z_2})^4] &= E[(g'_{z_1} Q \Omega \Omega^{-1} g_{z_2})^4] \\ &\leq E[(g'_{z_1} Q \Omega Q g_{z_1,n})^2 (g'_{z_2} \Omega^{-1} g_{z_2})^2] \\ &= E\left[(g'_{z_1} Q \Omega Q g_{z_1})^2\right] E\left[(g'_{z_2} \Omega^{-1} g_{z_2})^2\right]. \end{aligned}$$

Hence, by (f),

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

Moreover, by (d),

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

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

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

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

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

Proof: The proof is similar to that of DIN Lemma A.3. Let $F_{i,n}$ be a symmetric square root of $U_{i,n}$, $P_{i,n} = F_{i,n} \otimes q'_i$, $P_n = (P'_{1,n}, ..., P'_{n,n})'$, $A_{i,n} = F_{i,n}^{-1} a_{i,n}$, $A_n =$

$(A_{1,n}', ..., A_{n,n}')'$, $\bar{A}_{i,n} = E[A_{i,n}|x_i] = F_{i,n}^{-1}\bar{a}_{i,n}$ and $\bar{A}_n = (\bar{A}_{1,n}', ..., \bar{A}_{n,n}')'$. Note that $P_n'P_n = \sum_i U_{i,n} \otimes q_i q_i'$ and

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

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

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

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

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

$$\begin{aligned} E[T_A] &= E[tr(Q_n E[(A_n - \bar{A}_n)\left(A_n - \bar{A}_n\right)'/n)|s]] \\ &\leq CE[tr(Q_n)]/n \leq CK/n \to 0 \end{aligned}$$

as $tr(\Omega_n) \leq CK$, using (b) and (f). Thus $T_A \xrightarrow{p} 0$ by M.

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

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

By c_r,

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

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

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

By T and CS

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

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

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

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

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

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

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

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

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

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

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

■

The following Lemma is similar to DIN Lemma A.4.

Lemma A.5 *If $q(\cdot)$ satisfies Assumption 3.1, (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) $E[\left\|U_{i,n}^{-1} - U_i^{-1}\right\|^2] \to 0$ and (e) $K \to \infty$ and $K^2/n \to 0$, then*

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

Proof: We prove the result by first showing that

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

and secondly that

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

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

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

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

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

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

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

since $E\left[B_nB_n'|x\right]$ 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). We use Serfling (2002, Corollary, p.32) to prove this result. We only need show that

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

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

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

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

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

from which (A.2) follows. ■

The following Lemmata are needed to prove the asymptotic normality of the test statistics under local alternatives.

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

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

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

Proof: The proof of these results are similar to that of Lemma A.6 of DIN. A major difference here is that some expectations are bounded for n large enough n rather than merely bounded as in DIN.

Using the same arguments as in DIN we have

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

where $M_{i,n} = \delta_i^2 + 2\delta_i\left\|u_{i,n}\right\|$ and $\delta_i = \delta(z_i)$. The final equality follows as $E[\left\|u(z, \beta_{0,n})\right\|^2]$ is bounded since $E[\delta(z)^2|x]$ is bounded and $E[\sup_{\beta\in\mathcal{B}}\left\|u(z,\beta)\right\|^2]$ is bounded by Assumption (3.3).

Now

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

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

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

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

For the third conclusion as in DIN

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

where the second inequality holds for n large enough.

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

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

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

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

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

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

Proof: The proof is as in that for DIN Lemma A.7. In fact the proof requires no stronger assumptions than those in DIN.

Let $u_{\beta i,n} = u_{\beta i}(\beta_{0,n})$, $\delta_i = \delta(z_i)$ and $\tilde{G}_n = \sum_i u_{\beta i,n} \otimes q_i/n$. Then by DIN Lemma A.2

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

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

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

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

The first conclusion follows by T.

In addition

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Now by independence

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

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

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

Moreover

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

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

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

Next, by CS and T,

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

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

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

It remains to prove that

$$\frac{n\hat{g}_n'(\hat{\Omega}^{-1}-\Omega_n^{-1})\hat{g}_n}{\sqrt{2JK}} = o_p\left(1\right).$$

From Lemma A.6,

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

■

A.2 Asymptotic Null Distribution

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

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

where $\Omega = E[g(z,\beta_0)g(z,\beta_0)']$ and $\Xi = E[h(z,\beta_0)h(z,\beta_0)']$. It also follows from Lemma A.7 of DIN that $\left\|\partial\hat{g}\left(\tilde{\beta}\right)/\partial\beta' - G\right\| \xrightarrow{p} 0$ and $\left\|\partial\hat{g}\left(\tilde{\beta}\right)/\partial\beta' - G\right\| \xrightarrow{p} 0$ for any $\tilde{\beta} = \beta_0 +$

$O_p(1/\sqrt{n})$. In addition, $G'\Omega^{-1}G$ and $H'\Xi^{-1}H$ are bounded, see the proof of Lemma A.8. Hence, the conditions of DIN Lemma 6.1 are met. Therefore,

$$\frac{n\hat{g}(\hat{\beta})'\hat{\Omega}^{-1}\hat{g}(\hat{\beta}) - n\hat{g}(\beta_0)'\Omega^{-1}\hat{g}(\beta_0)}{\sqrt{2JK}} \xrightarrow{p} 0.$$

and

$$\frac{n\hat{h}(\hat{\beta})'\hat{\Xi}^{-1}\hat{h}(\hat{\beta}) - n\hat{h}(\beta_0)'\Xi^{-1}\hat{h}(\beta_0)}{\sqrt{2J(M+1)K}} \xrightarrow{p} 0.$$

Now

$$\begin{aligned}\frac{n\hat{h}(\hat{\beta})'\hat{\Xi}^{-1}\hat{h}(\hat{\beta}) - n\hat{g}(\hat{\beta})'\hat{\Omega}^{-1}\hat{g}(\hat{\beta}) - JMK}{\sqrt{2JMK}} &= \frac{n\hat{h}(\hat{\beta})'\hat{\Xi}^{-1}\hat{h}(\hat{\beta}) - n\hat{h}(\beta_0)\Xi^{-1}\hat{h}(\beta_0)}{\sqrt{2JMK}} \\ &- \frac{n\hat{g}(\hat{\beta})'\hat{\Omega}^{-1}\hat{g}(\hat{\beta}) - n\hat{g}(\beta_0)'\Omega^{-1}\hat{g}(\beta_0)}{\sqrt{2JMK}} \\ &+ \frac{n\hat{h}(\beta_0)'\Xi^{-1}\hat{h}(\beta_0) - n\hat{g}(\beta_0)'\Omega^{-1}\hat{g}(\beta_0) - JMK}{\sqrt{2JMK}} \\ &= \frac{n\hat{h}(\beta_0)'\Xi^{-1}\hat{h}(\beta_0) - n\hat{g}(\beta_0)'\Omega^{-1}\hat{g}(\beta_0) - JMK}{\sqrt{2JMK}} + o_p(1).\end{aligned}$$

Define $S_{\text{MA}} = I_J \otimes (I_K, 0_{MK})'$ as a selection matrix such that $S'_{\text{MA}}\hat{h}(\beta_0) = \hat{g}(\beta_0)$. Therefore

$$\frac{n\hat{h}(\beta_0)'\Xi^{-1}\hat{h}(\beta_0) - n\hat{g}(\beta_0)'\Omega^{-1}\hat{g}(\beta_0) - JMK}{\sqrt{2JMK}} = \frac{n\hat{h}(\beta_0)'(\Xi^{-1} - S_{\text{MA}}\Omega^{-1}S'_{\text{MA}})\hat{h}(\beta_0) - JMK}{\sqrt{2JMK}}.$$

We use Lemma A.2 to obtain the conclusion of the theorem. First, $tr((\Xi^{-1} - S_{\text{MA}}\Omega^{-1}S'_{\text{MA}})\Xi) = tr(I_{J(M+1)K}) - tr(I_{JK}) = JMK$. Secondly, $(\Xi^{-1} - S_{\text{MA}}\Omega^{-1}S'_{\text{MA}})\Xi(\Xi^{-1} - S_{\text{MA}}\Omega^{-1}S'_{\text{MA}}) = \Xi^{-1} - S_{\text{MA}}\Omega^{-1}S'_{\text{MA}}$. Thirdly,

$$\begin{aligned}E[(h(z,\beta_0)'(\Xi^{-1} - S_{\text{MA}}\Omega^{-1}S'_{\text{MA}})h(z,\beta_0))^2] &\leq CE[\|h(z,\beta_0)\|^4] \\ &\leq CE[\|u(z,\beta_0)\|^4 \left\|q^K(w,x)\right\|^4] \\ &\leq CE[\left\|q^K(w,x)\right\|^4] \\ &\leq C\zeta(K)^2K.\end{aligned}$$

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

Proof of Theorem 4.2: First we focus on $\mathcal{LR}$ (3.9).

$$\begin{aligned}\mathcal{LR} &= \frac{2n[\tilde{P}_n(\hat{\beta},\tilde{\eta}) - \hat{P}_n(\hat{\beta},\hat{\lambda})] - JMK}{2\sqrt{JMK}} \\ &= \frac{\mathcal{T}^h_{GMM} - \mathcal{T}^g_{GMM} - JMK}{2\sqrt{JMK}} \\ &\quad + \frac{2n\tilde{P}_n(\hat{\beta},\tilde{\eta}) - \mathcal{T}^h_{GMM}}{\sqrt{2JMK}} \\ &\quad - \frac{2n\hat{P}_n(\hat{\beta},\hat{\lambda}) - \mathcal{T}^g_{GMM}}{\sqrt{2JMK}}.\end{aligned} \tag{A.3}$$

Write $\hat{g}_i = g_i(\hat{\beta})$, $(i = 1, ..., n)$, $\hat{g} = \hat{g}(\hat{\beta})$ and $\hat{g}_0 = \hat{g}(\beta_0)$. Using T and CS twice we have

$$\begin{aligned}\|\hat{g} - \hat{g}_0\| &\leq \sum\nolimits_{i=1}^n \left\|u(z_i,\hat{\beta}) - u(z_i,\beta_0)\right\| \|q(w_i)\|/n \\ &\leq (\sum\nolimits_{i=1}^n \delta(z_i)^2/n)^{1/2}(\sum\nolimits_{i=1}^n \|q(w_i)\|^2/n)^{1/2}\left\|\hat{\beta} - \beta_0\right\| = O_p(\sqrt{K/n})\end{aligned}$$

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

$$\frac{\partial \hat{P}_n(\hat{\beta},\hat{\lambda})}{\partial \lambda} = \sum\nolimits_{i=1}^n \rho_1(\hat{\lambda}'\hat{g}_i)\hat{g}_i/n = 0. \tag{A.4}$$

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

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

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

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

We deal with the third term in (A.3) first. Expanding $2n\hat{P}_n(\hat{\beta},\hat{\lambda})$ around $\lambda = 0$ and plugging in $\hat{\lambda}$ from (A.5),

$$2n\hat{P}_n(\hat{\beta},\hat{\lambda}) = 2n[-\hat{g}(\hat{\beta})'\hat{\lambda} - \hat{\lambda}'\ddot{\Omega}\;\hat{\lambda}/2] = n\hat{g}(\hat{\beta})'[2\dot{\Omega}^{-1} - \dot{\Omega}^{-1}\ddot{\Omega}\dot{\Omega}^{-1}]\hat{g}(\hat{\beta})$$

with $\ddot{\Omega} = -\sum_{i=1}^{n} \rho_2(\ddot{\lambda}'\hat{g}_i)\hat{g}_i\hat{g}_i'/n$ and $\ddot{\lambda}$ lies between $\hat{\lambda}$ and zero. Thus it remains to prove that

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

First notice that by DIN Lemma A.6 $\left\|\hat{\Omega} - \Omega\right\| = O_p(\zeta(K)\sqrt{K/n}) = o_p(1/\sqrt{K})$ and, thus, by Lemma A.1 we also have $\left\|\dot{\Omega} - \Omega\right\| = o_p(1/\sqrt{K})$ and $\left\|\ddot{\Omega} - \Omega\right\| = o_p(1/\sqrt{K})$. Hence $\left\|2\dot{\Omega} - \ddot{\Omega} - \Omega\right\| \xrightarrow{p} 0$. Consequently $\lambda_{\max}[(2\dot{\Omega} - \ddot{\Omega})^{-1}] \leq C$ w.p.a.1.. Thus, by T, as $\left(2\dot{\Omega}^{-1} - \dot{\Omega}^{-1}\ddot{\Omega}\dot{\Omega}^{-1}\right)^{-1} = \dot{\Omega}(2\dot{\Omega} - \ddot{\Omega})^{-1}\dot{\Omega}$,

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

On the other hand as $\lambda_{\max}(\Omega) \leq C$

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

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

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

By the same reasoning the second term in (A.3)

$$\frac{2n\tilde{P}_n(\hat{\beta},\tilde{\eta}) - \mathcal{T}^h_{GMM}}{\sqrt{2JMK}} \xrightarrow{p} 0.$$

Therefore, it follows from Theorem 3.6 that

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

We now turn to consider the Lagrange multiplier statistic

$$\mathcal{LM} = \frac{n(\tilde{\eta} - \hat{\eta})'\hat{\Xi}(\tilde{\eta} - \hat{\eta}) - JMK}{\sqrt{2JMK}}.$$

Write $\hat{h}_i = h_i(\hat{\beta})$, $(i = 1, ..., n)$, $\hat{h} = \hat{h}(\hat{\beta})$ and $\hat{h}_0 = \hat{h}(\beta_0)$. By a similar argument to that which established (A.5)

$$\tilde{\eta} = -\dot{\Xi}^{-1}\hat{h}(\hat{\beta})$$

where $\dot{\Xi} = -\sum_{i=1}^{n} \rho_1(\dot{\eta}'\hat{h}_i)\hat{h}_i\hat{h}_i'/n$ and $\dot{\eta}$ lies between $\tilde{\eta}$ and zero.

Let the $((M+1)JK)\times$ JK selection matrix $S_{\text{MA}} = I_J \otimes (I_K, 0_{MK})'$. Hence, $S_{\text{MA}}'\hat{h} = \hat{g}$ and $S_{\text{MA}}'\Xi S_{\text{MA}} = \Omega$. Also write $\hat{\eta} = S_{\text{MA}}\hat{\lambda}$. Thus, $\hat{\eta} = S_{\text{MA}}\hat{\lambda} = -S_{\text{MA}}\dot{\Omega}^{-1}\hat{g} = -S_{\text{MA}}\dot{\Omega}^{-1}S_{\text{MA}}'\hat{h}$.

Now

$$\begin{aligned} n(\tilde{\eta}-\hat{\eta})'\hat{\Xi}(\tilde{\eta}-\hat{\eta}) &= n\tilde{\eta}'\hat{\Xi}\tilde{\eta} - 2n\tilde{\eta}'\hat{\Xi}\hat{\eta} + n\hat{\eta}'\hat{\Xi}\hat{\eta} \\ &= n\hat{h}'\dot{\Xi}^{-1}\hat{\Xi}\dot{\Xi}^{-1}\hat{h} - 2n\hat{h}'\dot{\Xi}^{-1}\hat{\Xi}\ S_{\text{MA}}\dot{\Omega}^{-1}S_{\text{MA}}'\hat{h} + n\hat{h}'S_{\text{MA}}\dot{\Omega}^{-1}S_{\text{MA}}'\hat{\Xi}\ S_{\text{MA}}\dot{\Omega}^{-1}S_{\text{MA}}'\hat{h}. \end{aligned}$$

and therefore

$$\mathcal{LM} - \frac{\mathcal{T}_{GMM}^{h} - \mathcal{T}_{GMM}^{g} - JMK}{\sqrt{2JMK}} = \frac{n\hat{h}'(\dot{\Xi}^{-1}\hat{\Xi}\dot{\Xi}^{-1} - \hat{\Xi}^{-1})\hat{h}}{\sqrt{2JMK}}$$
$$+\frac{n\hat{h}'(S_{\text{MA}}\hat{\Omega}^{-1}S_{\text{MA}}' - 2\dot{\Xi}^{-1}\hat{\Xi}\ S_{\text{MA}}\dot{\Omega}^{-1}S_{\text{MA}}'\hat{h} + S_{\text{MA}}\dot{\Omega}^{-1}S_{\text{MA}}'\hat{\Xi}\ S_{\text{MA}}\dot{\Omega}^{-1}S_{\text{MA}}')\hat{h}}{\sqrt{2JMK}}.$$

We now demonstrate in turn that these terms are each $o_p(1)$.

By CS, the first term

$$\begin{aligned} n\hat{h}'(\dot{\Xi}^{-1}\hat{\Xi}\dot{\Xi}^{-1} - \hat{\Xi}^{-1})\hat{h}/\sqrt{K} &= n\hat{h}'\dot{\Xi}^{-1}(\hat{\Xi} - \dot{\Xi}\hat{\Xi}^{-1}\dot{\Xi})\dot{\Xi}^{-1}\hat{h}/\sqrt{K} \\ &\le n\left\|\tilde{\eta}\right\|^2 \left\|\hat{\Xi} - \dot{\Xi}\hat{\Xi}^{-1}\dot{\Xi}\right\|/\sqrt{K}. \end{aligned}$$

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

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

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

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

Therefore, by T $\left\|\hat{\Xi} - \dot{\Xi}\hat{\Xi}^{-1}\dot{\Xi}\right\| = o_p(1/\sqrt{K})$. As $\|\tilde{\eta}\| = O_p(\sqrt{K/n})$ by DIN Lemma A.11, $n\hat{h}'(\dot{\Xi}^{-1}\hat{\Xi}\dot{\Xi}^{-1} - \hat{\Xi}^{-1})\hat{h}/\sqrt{K} = nO_p(K/n)o_p(1/\sqrt{K})/\sqrt{K} = o_p(1)$.

For the second term, by CS

$$
\begin{aligned}
& n\hat{h}'(S_{\text{MA}}\hat{\Omega}^{-1}S'_{\text{MA}} - 2\dot{\Xi}^{-1}\hat{\Xi}\ S_{\text{MA}}\dot{\Omega}^{-1}S'_{\text{MA}} + S_{\text{MA}}\dot{\Omega}^{-1}S'_{\text{MA}}\hat{\Xi}\ S_{\text{MA}}\dot{\Omega}^{-1}S'_{\text{MA}})\hat{h}/\sqrt{K} \\
\leq \ & n\left\|\hat{h}\right\|^2 \left\|S_{\text{MA}}\hat{\Omega}^{-1}S'_{\text{MA}} - 2\dot{\Xi}^{-1}\hat{\Xi}S_{\text{MA}}\dot{\Omega}^{-1}S'_{\text{MA}} + S_{\text{MA}}\dot{\Omega}^{-1}S'_{\text{MA}}\hat{\Xi}\ S_{\text{MA}}\dot{\Omega}^{-1}S'_{\text{MA}}\right\|/\sqrt{K}.
\end{aligned}
$$

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

$$
\begin{aligned}
\left\|S_{\text{MA}}\hat{\Omega}^{-1}S'_{\text{MA}} - \dot{\Xi}^{-1}\hat{\Xi}S_{\text{MA}}\dot{\Omega}^{-1}S'_{\text{MA}}\right\| &\leq \left\|S_{\text{MA}}\dot{\Omega}^{-1}(\dot{\Omega}-\hat{\Omega})\hat{\Omega}^{-1}S'_{\text{MA}}\right\| + \left\|\dot{\Xi}^{-1}(\hat{\Xi}-\dot{\Xi})S_{\text{MA}}\dot{\Omega}^{-1}S'_{\text{MA}}\right\| \\
&= o_p(1/\sqrt{K}).
\end{aligned}
$$

Next by a similar argument

$$
\begin{aligned}
\left\|\dot{\Xi}^{-1}\hat{\Xi}\ S_{\text{MA}}\dot{\Omega}^{-1}S'_{\text{MA}} - S_{\text{MA}}\dot{\Omega}^{-1}S'_{\text{MA}}\hat{\Xi}\ S_{\text{MA}}\dot{\Omega}^{-1}S'_{\text{MA}}\right\| &\leq \left\|S_{\text{MA}}\dot{\Omega}^{-1}(\dot{\Omega}-\hat{\Omega})\dot{\Omega}^{-1}S'_{\text{MA}}\right\| \\
&\quad + \left\|\dot{\Xi}^{-1}(\hat{\Xi}-\dot{\Xi})S_{\text{MA}}\dot{\Omega}^{-1}S'_{\text{MA}}\right\| \\
&= o_p(1/\sqrt{K}).
\end{aligned}
$$

Therefore since $\left\|\hat{h}\right\| = O_p(\sqrt{K/n})$ by DIN Lemma A.14 of DIN $n\hat{h}'(S_{\text{MA}}\hat{\Omega}^{-1}S'_{\text{MA}} - 2\dot{\Xi}^{-1}\hat{\Xi}$ $S_{\text{MA}}\dot{\Omega}^{-1}S'_{\text{MA}} + S_{\text{MA}}\dot{\Omega}^{-1}S'_{\text{MA}}\hat{\Xi}\ S_{\text{MA}}\dot{\Omega}^{-1}S'_{\text{MA}})\hat{h}/\sqrt{K} = nO_p(K/n)o_p(1/\sqrt{K})/\sqrt{K} = o_p(1)$.

The score test statistic

$$
\mathcal{S} = \frac{\sum_{i=1}^{n}\rho_1(\hat{\lambda}'\hat{g}_i)\hat{s}'_i S'_{\text{s}}\hat{\Xi}^{-1}S_{\text{s}}\sum_{i=1}^{n}\rho_1(\hat{\lambda}'\hat{g}_i)\hat{s}_i/n - JMK}{\sqrt{2JMK}},
$$

where $\hat{s}_i = s_i(\hat{\beta})$, $(i = 1, ..., n)$ and $S_{\text{s}} = I_J \otimes (0_K, I_{MK})'$. Expanding the first order conditions $\sum_{i=1}^{n}\rho_1(\hat{h}'_i\tilde{\eta})\hat{h}_i/n = 0$ of (3.8) around $\hat{\eta}$ gives

$$
\sum_{i=1}^{n}\rho_1(\hat{h}'_i\hat{\eta})\hat{h}_i/n - \dot{\Xi}(\tilde{\eta}-\hat{\eta}) = 0
$$

w.p.a.1 where $\dot{\Xi} = -\sum_{i=1}^{n}\rho_2(\hat{h}'_i\dot{\eta})\hat{h}_i\hat{h}'_i/n$ and $\dot{\eta}$ lies between $\tilde{\eta}$ and $\hat{\eta}$. Since $\sum_{i=1}^{n}\rho_1(\hat{h}'_i\hat{\eta})\hat{h}_i/n = S_{\text{s}}\sum_{i=1}^{n}\rho_1(\hat{g}'_i\hat{\lambda})\hat{s}_i$,

$$
\sum_{i=1}^{n}\rho_1(\hat{\lambda}'\hat{g}_i)\hat{s}'_i S'_{\text{s}}\hat{\Xi}^{-1}S_{\text{s}}\sum_{i=1}^{n}\rho_1(\hat{\lambda}'\hat{g}_i)\hat{s}_i/n = n\left(\tilde{\eta}-\hat{\eta}\right)'\dot{\Xi}\hat{\Xi}^{-1}\dot{\Xi}\left(\tilde{\eta}-\hat{\eta}\right).
$$

Thus by CS and T

$$
\begin{aligned}
|\mathcal{S} - \mathcal{LM}| &= n\left|(\tilde{\eta}-\hat{\eta})'\left(\dot{\Xi}\hat{\Xi}^{-1}\dot{\Xi} - \hat{\Xi}\right)(\tilde{\eta}-\hat{\eta})\right|/\sqrt{2JMK} \\
&\leq n\left\|\dot{\Xi}\hat{\Xi}^{-1}\dot{\Xi} - \hat{\Xi}\right\|\left(\|\tilde{\eta}\| + \|\hat{\eta}\|\right)^2/\sqrt{2JMK} = o_p(1)
\end{aligned}
$$

as $\dot{\Xi}\hat{\Xi}^{-1}\dot{\Xi} - \hat{\Xi} = o_p(1/\sqrt{K})$ and $\|\tilde{\eta}\|$, $\|\hat{\eta}\|$ are both $O_p(\sqrt{K/n})$ by DIN Lemma A.11.

Finally we consider the Wald test statistic. From above, w.p.a.1

$$\tilde{\eta} - \hat{\eta} = \dot{\Xi}^{-1} S_{\mathrm{s}} \sum_{i=1}^{n} \rho_1(\hat{g}_i' \hat{\lambda}) \hat{s}_i / n$$

and thus

$$S_{\mathrm{s}}' \tilde{\eta} = S_{\mathrm{s}}' \dot{\Xi}^{-1} S_{\mathrm{s}} \sum_{i=1}^{n} \rho_1(\hat{g}_i' \hat{\lambda}) \hat{s}_i / n.$$

Therefore, w.p.a.1

$$\begin{aligned} |\mathcal{S} - \mathcal{W}| &= n \left| \tilde{\eta}' S_{\mathrm{s}} ((S_{\mathrm{s}}' \dot{\Xi}^{-1} S_{\mathrm{s}})^{-1} S_{\mathrm{s}}' \hat{\Xi}^{-1} S_{\mathrm{s}} (S_{\mathrm{s}}' \dot{\Xi}^{-1} S_{\mathrm{s}})^{-1} - S_{\mathrm{s}}' \hat{\Xi}^{-1} S_{\mathrm{s}}) S_{\mathrm{s}}' \tilde{\eta} \right| / \sqrt{2JMK} \\ &\le n \left\| S_{\mathrm{s}}' \tilde{\eta} \right\|^2 \left\| (S_{\mathrm{s}}' \dot{\Xi}^{-1} S_{\mathrm{s}})^{-1} S_{\mathrm{s}}' \hat{\Xi}^{-1} S_{\mathrm{s}} (S_{\mathrm{s}}' \dot{\Xi}^{-1} S_{\mathrm{s}})^{-1} - S_{\mathrm{s}}' \hat{\Xi}^{-1} S_{\mathrm{s}} \right\| / \sqrt{2JMK}. \end{aligned}$$

Since $\|S_{\mathrm{s}}' \tilde{\eta}\| = O_p(\sqrt{K/n})$ by DIN Lemma A.11 and by a similar argument to that which showed $\dot{\Xi}\hat{\Xi}^{-1}\dot{\Xi} - \hat{\Xi} = o_p(1/\sqrt{K})$, $(S_{\mathrm{s}}' \dot{\Xi}^{-1} S_{\mathrm{s}})^{-1} S_{\mathrm{s}}' \hat{\Xi}^{-1} S_{\mathrm{s}} (S_{\mathrm{s}}' \dot{\Xi}^{-1} S_{\mathrm{s}})^{-1} - S_{\mathrm{s}}' \hat{\Xi}^{-1} S_{\mathrm{s}} = o_p(1/\sqrt{K})$. Therefore, $|\mathcal{S} - \mathcal{W}| = o_p(1)$. ∎

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

$$\mathcal{J}^* = \gamma \mathcal{J} + \delta \mathcal{J}^g.$$

where the (arbitrary) scalars γ and δ are such that $\gamma^2 + \delta^2 > 0$. The desired result is proven if $\mathcal{J}^* \xrightarrow{d} N(0, \gamma^2 + \delta^2)$.

First, by Lemma 6.1 of DIN,

$$\mathcal{J} - \frac{n\hat{h}(\beta_0)' \Xi^{-1} \hat{h}(\beta_0) - n\hat{g}(\beta_0)' \Omega^{-1} \hat{g}(\beta_0) - JMK}{\sqrt{2JMK}} \xrightarrow{p} 0$$

and likewise

$$\mathcal{J}^g - \frac{n\hat{g}(\beta_0)' \Omega^{-1} \hat{g}(\beta_0) - JK}{\sqrt{2JK}} \xrightarrow{p} 0.$$

Therefore,

$$\mathcal{J}^* - \frac{1}{\sqrt{M}} \frac{n\hat{h}(\beta_0)' Q \hat{h}(\beta_0) - (\gamma M + \delta\sqrt{M}) JK}{\sqrt{2JK}} \xrightarrow{p} 0,$$

where $Q = \gamma \Xi^{-1} - (\gamma - \delta\sqrt{M}) S_{\mathrm{MA}} \Omega^{-1} S_{\mathrm{MA}}'$.

To prove $\sqrt{M}\mathcal{J}^* \xrightarrow{d} N(0, v)$, where $v = (\gamma^2 + \delta^2)M$, we verify conditions (a)-(f) of Lemma A.3.

Condition (a) is immediate.

For (b),

$$
\begin{aligned}
tr(Q\Xi) &= \gamma tr(I_{J(M+1)K}) - (\gamma - \delta\sqrt{M})tr(I_{JK}) \\
&= \gamma J(M+1)K - (\gamma - \delta\sqrt{M})JK \\
&= (\gamma M + \delta\sqrt{M})JK = aJK.
\end{aligned}
$$

To consider condition (c), note that

$$
\begin{aligned}
(Q\Xi)^2 &= (\gamma I_{J(M+1)K} - (\gamma M + \delta\sqrt{M})S_{\text{MA}}\Omega^{-1}S'_{\text{MA}}\Xi)^2 \\
&= \gamma^2 I_{J(M+1)K} - (\gamma^2 - M\delta^2)S_{\text{MA}}\Omega^{-1}S'_{\text{MA}}\Xi.
\end{aligned}
$$

Hence

$$
\begin{aligned}
tr[(Q\Xi)^2] &= \gamma^2 J(M+1)K - (\gamma^2 - M\delta^2)JK \\
&= (\gamma^2 + \delta^2)JMK = vJK.
\end{aligned}
$$

For (d),

$$
\begin{aligned}
(Q\Xi)^4 &= (\gamma^2 I_{J(M+1)K} - (\gamma^2 - M\delta^2)S_{\text{MA}}\Omega^{-1}S'_{\text{MA}}\Xi)^2 \\
&= \gamma^4 I_{J(M+1)K} - (\gamma^4 - M^2\delta^4)S_{\text{MA}}\Omega^{-1}S'_{\text{MA}}\Xi.
\end{aligned}
$$

Thus

$$
\begin{aligned}
tr[(Q\Xi)^4] &= \gamma^4 J(M+1)K - (\gamma^4 - M^2\delta^4)JK \\
&= (\gamma^4 + M\delta^4)JMK.
\end{aligned}
$$

From Lemma A.6 of DIN, $1/C \leq \lambda_{\min}(\Xi) \leq \lambda_{\max}(\Xi) \leq C$ and $1/C \leq \lambda_{\min}(\Omega) \leq \lambda_{\max}(\Omega) \leq C$. Therefore, condition (e) is satisfied using Assumption 3.2 as

$$
E[(h(z,\beta_0)'(\gamma\Xi^{-1} - (\gamma - \delta\sqrt{M})S_{\text{MA}}\Omega^{-1}S'_{\text{MA}})h(z,\beta_0))^2] \leq C\zeta(K)^2K.
$$

By a similar reasoning as for (e)

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

Also

$$\begin{aligned} Q\Xi Q &= (\gamma\Xi^{-1} - (\gamma - \delta\sqrt{M})S_{\text{MA}}\Omega^{-1}S'_{\text{MA}})\Xi(\gamma\Xi^{-1} - (\gamma - \delta\sqrt{M})S_{\text{MA}}\Omega^{-1}S'_{\text{MA}}) \\ &= \gamma^2(\Xi^{-1} - S_{\text{MA}}\Omega^{-1}S'_{\text{MA}}) + \delta^2 S_{\text{MA}}\Omega^{-1}S'_{\text{MA}}. \end{aligned}$$

Thus, condition (f) holds, as in (e),

$$E[(h(z,\beta_0)'Q\Xi Q h(z,\beta_0))^2] \le C\zeta(K)^2 K.$$

Consequently, by Lemma A.3, $\sqrt{M}\mathcal{J}^* \xrightarrow{d} N(0,v)$ which proves the desired result. ■

A.3 Asymptotic Local Alternative Distribution

Proof of Theorem 5.1: We prove the result for the GMM statistic $\mathcal{J}$. Proofs for GEL statistics $\mathcal{LR}$, $\mathcal{LM}$, $\mathcal{S}$ and $\mathcal{W}$ are omitted for brevity but essentially follow the same steps as in the proof of Theorem 4.2 above that demonstrates their asymptotic equivalence to the GMM statistic.

Let $\hat{g}_n = \hat{g}(\beta_{n,0})$ and $\hat{h}_n = \hat{h}(\beta_{n,0})$. Then, by Lemma A.8,

$$\frac{n\hat{h}(\hat{\beta})'\hat{\Xi}^{-1}\hat{h}(\hat{\beta}) - n\hat{h}'_n\Xi_n^{-1}\hat{h}_n}{\sqrt{2JMK}} \xrightarrow{p} 0, \quad \frac{n\hat{g}(\hat{\beta})'\hat{\Omega}^{-1}\hat{g}(\hat{\beta}) - n\hat{g}'_n\Omega_n^{-1}\hat{g}_n}{\sqrt{2JMK}} \xrightarrow{p} 0.$$

It then follows that $\mathcal{J} - (n\hat{h}'_n(\Xi_n^{-1} - S_{\text{MA}}\Omega_n^{-1}S'_{\text{MA}})\hat{h}_n - JMK)/\sqrt{2JMK} \xrightarrow{p} 0$.

Therefore it remains to prove that

$$\frac{n\hat{h}'_n(\Xi_n^{-1} - S_{\text{MA}}\Omega_n^{-1}S'_{\text{MA}})\hat{h}_n - JMK}{\sqrt{2JMK}} \xrightarrow{d} N(\mu/\sqrt{2}, 1).$$

We first consider the local alternative distribution under the CEM hypothesis when $s_i = (w'_i, x'_i)'$, $(i = 1, ..., n)$.

Let $h_{i,n} = h_i(\beta_{n,0})$, $\bar{h}_{i,n} = E[h_{i,n}|s_i]$ and $\tilde{h}_{i,n} = h_{i,n} - \bar{h}_{i,n}$, $(i = 1, ..., n)$. Also let $\bar{h}_n = \sum_{i=1}^n \bar{h}_{i,n}/n$ and $\tilde{h}_n = \sum_{i=1}^n \tilde{h}_{i,n}/n$. Write $P_n = \Xi_n^{-1} - S_{\text{MA}}\Omega_n^{-1}S'_{\text{MA}}$. Then,

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

First, we show that

$$\bar{h}'_n P_n \bar{h}_n = \frac{\sqrt{JMK}}{n}(\mu + o_p(1)).$$

Let $\xi_i = \xi(s_i)$ and $q_i = q^K(s_i)$, $(i = 1, ..., n)$. It follows by Lemma A.4 that

$$\begin{aligned} \bar{h}_n' \bar{\Xi}_n^{-1} \bar{h}_n &= \frac{\sqrt{JMK}}{n} \sum_{i,j=1}^{n} (\xi_i \otimes q_i)' \bar{\Xi}_n^{-1} (\xi_j \otimes q_j)/n^2 \\ &= \frac{\sqrt{JMK}}{n} (\mu + o_p(1)). \end{aligned}$$

Next, letting $q_{1i} = q_1^K(w_i)$, $(i = 1, ..., n)$, and, again using Lemma A.4,

$$\begin{aligned} \bar{h}_n' S_{\text{MA}} \bar{\Omega}_n^{-1} S_{\text{MA}}' \bar{h}_n &= \frac{\sqrt{JMK}}{n} \sum_{i,j=1}^{n} (\xi_i \otimes q_{1i})' \bar{\Omega}_n^{-1} (\xi_j \otimes q_{1j})/n^2 \\ &= \frac{\sqrt{JMK}}{n} o_p(1) \end{aligned}$$

as $E[\xi_i | w_i] = 0$ by hypothesis.

It therefore remains to show that

$$\frac{n}{\sqrt{2JMK}} \bar{h}_n' (\Xi_n^{-1} - \bar{\Xi}_n^{-1}) \bar{h}_n \xrightarrow{p} 0, \; \frac{n}{\sqrt{2JMK}} \bar{h}_n' S_{\text{MA}} (\Omega_n^{-1} - \bar{\Omega}_n^{-1}) S_{\text{MA}}' \bar{h}_n \xrightarrow{p} 0.$$

Similarly to the proof of Lemma 6.1 in DIN, from Lemma A.6,

$$\begin{aligned} \left| n \bar{h}_n' (\Xi_n^{-1} - \bar{\Xi}_n^{-1}) \bar{h}_n \right| / \sqrt{2JMK} &\leq n \left\| \Xi_n^{-1} \bar{h}_n \right\|^2 (\left\| \Xi_n - \bar{\Xi}_n \right\| + C \left\| \Xi_n - \bar{\Xi}_n \right\|^2) / \sqrt{2JMK} \\ &= n \left\| \Xi_n^{-1} \bar{h}_n \right\|^2 O_p(\zeta(K) \sqrt{K/n}) / \sqrt{2JMK} = o_p(1) \end{aligned}$$

since $\left\| \Xi_n^{-1} \bar{h}_n \right\|^2 = \bar{h}_n' \Xi_n^{-2} \bar{h}_n \leq C \bar{h}_n' \Xi_n^{-1} \bar{h}_n = O_p(\sqrt{K}/n)$. Likewise $\left| n \bar{h}_n' S_{\text{MA}} (\Xi_n^{-1} - \bar{\Xi}_n^{-1}) S_{\text{MA}}' \bar{h}_n \right| / \sqrt{2JMK} = o_p(1)$. Therefore,

$$\bar{h}_n' P_n \bar{h}_n = \frac{\sqrt{JMK}}{n} (\mu + o_p(1)).$$

Secondly, we demonstrate that

$$n \bar{h}_n' P_n \tilde{h}_n / \sqrt{2JMK} = o_p(1).$$

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

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

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

$$\begin{aligned} \bar{h}_n' \bar{\Xi}_n^{-1} \tilde{h}_n &= \frac{\sqrt[4]{JMK}}{n} \sum_{i,j=1}^{n} (\xi_i \otimes q_i)' \bar{\Xi}_n^{-1} \tilde{h}_{j,n} / n\sqrt{n} \\ &= O_p(\sqrt[4]{JMK}/n). \end{aligned}$$

Next, by hypothesis,

$$\begin{aligned} \left| n\bar{h}_n'(\Xi_n^{-1} - \bar{\Xi}_n^{-1})\tilde{h}_n \right| / \sqrt{2JMK} &\leq n \left\| \Xi_n^{-1} \bar{h}_n \right\| \left\| \Xi_n^{-1} \tilde{h}_n \right\| (\left\| \Xi_n - \bar{\Xi}_n \right\| + C \left\| \Xi_n - \bar{\Xi}_n \right\|^2) / \sqrt{2JMK} \\ &= n \left\| \Xi_n^{-1} \bar{h}_n \right\| \left\| \Xi_n^{-1} \tilde{h}_n \right\| O_p(\zeta(K)\sqrt{K/n}) / \sqrt{2JMK} = o_p(1) \end{aligned}$$

since $\left\| \Xi_n^{-1} \bar{h}_n \right\|^2 = O_p(\sqrt{K}/n)$ from above and $\left\| \Xi_n^{-1} \tilde{h}_n \right\| \leq \left\| \Xi_n^{-1} \hat{h}_n \right\| + \left\| \Xi_n^{-1} \bar{h}_n \right\| = O_p(\sqrt{K/n}) + O_p(\sqrt[4]{K/n^2})$. A similar analysis yields $n\bar{h}_n' S_{\text{MA}} \Omega_n^{-1} S_{\text{MA}}' \tilde{h}_n / \sqrt{2JMK} = o_p(1)$.

Finally, we require

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

To prove this, we invoke Lemma A.2. First, $tr(\Xi_n P_n) = JMK$. Secondly, we need to establish

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

By c_r

$$E[(\tilde{h}_{i,n}' P_n \tilde{h}_{i,n})^2] \leq 2E[(\tilde{h}_{i,n}' \Xi_n^{-1} \tilde{h}_{i,n})^2] + 2E[(\tilde{h}_{i,n}' S_{\text{MA}} \Omega_n^{-1} S_{\text{MA}}' \tilde{h}_{i,n})^2]$$

Again using c_r

$$E[(\tilde{h}_{i,n}' \Xi_n^{-1} \tilde{h}_{i,n})^2] \leq 3E[(h_{i,n}' \Xi_n^{-1} h_{i,n})^2] + 12E[(h_{i,n}' \Xi_n^{-1} \bar{h}_{i,n})^2] + 3E[(\bar{h}_{i,n}' \Xi_n^{-1} \bar{h}_{i,n})^2].$$

Now, for n large enough, $E[(h_{i,n}' \Xi_n^{-1} h_{i,n})^2] \leq CE[\|h_{i,n}\|^4]$. Since $\beta_{n,0} \in \mathcal{N}$ for n large enough, by Assumption 3.4 (c), similarly to the proof of Theorem 6.3 in DIN,

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

Next,

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

Lastly,

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

Hence, $E[(\tilde{h}'_{i,n}\Xi_n^{-1}\tilde{h}_{i,n})^2] = o_p(K\sqrt{n})$ as required. Likewise, $E[(\tilde{h}'_{i,n}S_{\text{MA}}\Omega_n^{-1}S'_{\text{MA}}\tilde{h}_{i,n})^2] = o_p(K\sqrt{n})$. Thirdly, $P_n\Xi_nP_n = P_n$. Therefore,

$$\frac{n\tilde{h}'_nP_n\tilde{h}_n - JMK}{\sqrt{2JMK}} \xrightarrow{d} N(0,1).$$

The conclusion of the theorem then follows. ■

Appendix B: Empirical Size

B.1 MEM

B.1.1 Unrestricted Tests

Table B.1 MEM Hypothesis DIN Test Rejection Frequencies: $\mathrm{A_M} = 1$

	Non-Standardized							
	$K=2$				$K=5$			
n	200	500	1000	3000	200	500	1000	3000
$\mathcal{J}^{\text{DIN-M}}(\text{GMM}_\text{M})$	4.84	4.60	5.06	4.98	4.12	4.14	4.82	4.82
$\mathcal{S}_{\text{CUE}}^{\text{DIN-M}}(\text{CUE}_\text{M})$	4.86	4.62	5.10	4.98	4.00	4.06	4.86	4.82
$\mathcal{LR}_{\text{EL}}^{\text{DIN-M}}(\text{EL}_\text{M})$	5.72	4.86	5.20	4.94	11.80	6.50	6.02	5.26
$\mathcal{LM}_{\text{EL}}^{\text{DIN-M}}(\text{EL}_\text{M})$	6.26	4.84	5.16	4.98	23.92	8.74	6.20	5.24
$\mathcal{S}_{\text{EL}}^{\text{DIN-M}}(\text{EL}_\text{M})$	4.84	4.60	5.06	5.00	4.38	4.16	4.86	4.80
$\mathcal{LR}_{\text{ET}}^{\text{DIN-M}}(\text{ET}_\text{M})$	5.72	4.98	5.28	5.00	10.28	6.44	6.14	5.40
$\mathcal{LM}_{\text{ET}}^{\text{DIN-M}}(\text{ET}_\text{M})$	7.62	5.50	5.58	5.14	25.56	10.72	7.92	6.18
$\mathcal{S}_{\text{ET}}^{\text{DIN-M}}(\text{ET}_\text{M})$	4.84	4.60	5.06	5.00	4.18	4.16	4.86	4.80

Table B.2 MEM Hypothesis DIN Test Rejection Frequencies: $\mathrm{A_M} = 1.5$

	Non-Standardised							
	$K=2$				$K=5$			
n	200	500	1000	3000	200	500	1000	3000
$\mathcal{J}^{\text{DIN-M}}(\text{GMM}_\text{M})$	4.68	4.56	4.66	4.82	3.64	4.12	4.52	5.26
$\mathcal{S}_{\text{CUE}}^{\text{DIN-M}}(\text{CUE}_\text{M})$	4.70	4.58	4.66	4.82	3.54	4.10	4.52	5.26
$\mathcal{LR}_{\text{EL}}^{\text{DIN-M}}(\text{EL}_\text{M})$	6.30	5.14	4.88	4.84	15.18	7.22	5.98	5.42
$\mathcal{LM}_{\text{EL}}^{\text{DIN-M}}(\text{EL}_\text{M})$	7.06	5.24	4.78	4.78	34.38	11.38	7.02	5.32
$\mathcal{S}_{\text{EL}}^{\text{DIN-M}}(\text{EL}_\text{M})$	4.82	4.56	4.66	4.82	3.90	4.18	4.56	5.26
$\mathcal{LR}_{\text{ET}}^{\text{DIN-M}}(\text{ET}_\text{M})$	6.26	5.14	4.96	4.88	12.26	6.90	6.18	5.58
$\mathcal{LM}_{\text{ET}}^{\text{DIN-M}}(\text{ET}_\text{M})$	8.74	6.32	5.56	5.04	34.18	13.56	9.60	6.48
$\mathcal{S}_{\text{ET}}^{\text{DIN-M}}(\text{ET}_\text{M})$	4.76	4.56	4.66	4.82	3.70	4.14	4.54	5.26

B.1.2 Restricted Tests

Table B.3 MEM Hypothesis Test Rejection Frequencies: $\mathrm{A}_{\mathrm{M}} = 1$

	Non-Standardized							
	$K=2$				$K=5$			
n	200	500	1000	3000	200	500	1000	3000
$\mathcal{J}^{\mathrm{M}}(\mathrm{GMM_M},\mathrm{GMM_{MA}})$	4.88	4.56	5.52	5.04	4.70	4.46	5.44	4.98
$\mathcal{LR}^{\mathrm{M}}_{\mathrm{CUE}}(\mathrm{CUE_M},\mathrm{CUE_{MA}})$	4.90	4.62	5.48	5.04	4.68	4.44	5.44	4.98
$\mathcal{LR}^{\mathrm{M}}_{\mathrm{EL}}(\mathrm{EL_M},\mathrm{EL_{MA}})$	5.68	4.78	5.50	5.08	11.82	6.42	6.40	5.26
$\mathcal{LM}^{\mathrm{M}}_{\mathrm{EL}}(\mathrm{EL_M},\mathrm{EL_{MA}})$	6.12	4.70	5.48	5.02	18.52	7.48	6.40	5.18
$\bar{\mathcal{S}}^{\mathrm{M}}_{\mathrm{EL}}(\mathrm{EL_M})$	5.14	4.74	5.54	5.06	7.66	5.66	6.00	5.14
$\mathcal{LR}^{\mathrm{M}}_{\mathrm{ET}}(\mathrm{ET_M},\mathrm{ET_{MA}})$	5.70	4.82	5.60	5.10	10.22	6.26	6.52	5.26
$\mathcal{LM}^{\mathrm{M}}_{\mathrm{ET}}(\mathrm{ET_M},\mathrm{ET_{MA}})$	6.74	5.16	5.88	5.18	19.54	9.02	7.54	5.86
$\bar{\mathcal{S}}^{\mathrm{M}}_{\mathrm{ET}}(\mathrm{ET_M})$	5.02	4.64	5.52	5.04	5.80	4.88	5.60	5.00

Table B.4 MEM Hypothesis Test Rejection Frequencies: $\mathrm{A}_{\mathrm{M}} = 1.5$

	Non-Standardized							
	$K=2$				$K=5$			
n	200	500	1000	3000	200	500	1000	3000
$\mathcal{J}^{\mathrm{M}}(\mathrm{GMM_M},\mathrm{GMM_{MA}})$	5.16	4.92	4.64	4.86	4.42	4.68	5.18	5.08
$\mathcal{LR}^{\mathrm{M}}_{\mathrm{CUE}}(\mathrm{CUE_M},\mathrm{CUE_{MA}})$	5.20	4.98	4.64	4.86	4.42	4.70	5.14	5.08
$\mathcal{LR}^{\mathrm{M}}_{\mathrm{EL}}(\mathrm{EL_M},\mathrm{EL_{MA}})$	6.22	5.36	4.78	4.98	15.70	7.56	6.70	5.40
$\mathcal{LM}^{\mathrm{M}}_{\mathrm{EL}}(\mathrm{EL_M},\mathrm{EL_{MA}})$	6.70	5.16	4.56	4.96	29.56	10.22	7.58	5.20
$\bar{\mathcal{S}}^{\mathrm{M}}_{\mathrm{EL}}(\mathrm{EL_M})$	5.68	5.12	4.76	4.90	7.52	5.92	5.84	5.26
$\mathcal{LR}^{\mathrm{M}}_{\mathrm{ET}}(\mathrm{ET_M},\mathrm{ET_{MA}})$	6.28	5.38	4.84	4.98	12.26	7.22	6.76	5.44
$\mathcal{LM}^{\mathrm{M}}_{\mathrm{ET}}(\mathrm{ET_M},\mathrm{ET_{MA}})$	8.04	6.16	5.02	5.08	28.44	12.00	9.20	6.00
$\bar{\mathcal{S}}^{\mathrm{M}}_{\mathrm{ET}}(\mathrm{ET_M})$	5.42	5.00	4.66	4.90	5.38	5.26	5.44	5.14

B.2 CEM

B.2.1 Unrestricted Tests

Table B.5 CEM Hypothesis DIN Test Rejection Frequencies: $A_C = 2$

	Non-Standardized							
	$K=2$				$K=5$			
n	200	500	1000	3000	200	500	1000	3000
$\mathcal{J}^{\text{DIN-C}}(\text{GMM}_{\text{C}})$	4.88	4.68	4.80	4.62	3.48	4.14	4.36	4.72
$\mathcal{S}^{\text{DIN-C}}_{\text{CUE}}(\text{CUE}_{\text{C}})$	4.92	4.66	4.76	4.60	3.38	4.10	4.32	4.78
$\mathcal{LR}^{\text{DIN-C}}_{\text{EL}}(\text{EL}_{\text{C}})$	6.58	5.34	4.94	4.66	17.14	8.64	6.58	5.36
$\mathcal{LM}^{\text{DIN-C}}_{\text{EL}}(\text{EL}_{\text{C}})$	7.96	5.52	4.76	4.50	42.82	16.9	9.66	5.94
$\mathcal{S}^{\text{DIN-C}}_{\text{EL}}(\text{EL}_{\text{C}})$	4.90	4.66	4.80	4.62	3.82	4.20	4.38	4.72
$\mathcal{LR}^{\text{DIN-C}}_{\text{ET}}(\text{ET}_{\text{C}})$	6.58	5.50	5.10	4.66	12.7	8.02	6.64	5.48
$\mathcal{LM}^{\text{DIN-C}}_{\text{ET}}(\text{ET}_{\text{C}})$	9.70	6.50	5.84	4.80	39.8	18.64	11.40	7.18
$\mathcal{S}^{\text{DIN-C}}_{\text{ET}}(\text{ET}_{\text{C}})$	4.88	4.66	4.80	4.62	3.58	4.16	4.36	4.72

Table B.6 CEM Hypothesis DIN Test Rejection Frequencies: $A_C = 4.5$

	Non-Standardized							
	$K=2$				$K=5$			
n	200	500	1000	3000	200	500	1000	3000
$\mathcal{J}^{\text{DIN-C}}(\text{GMM}_{\text{C}})$	3.62	4.24	4.50	4.46	2.58	3.72	4.06	4.14
$\mathcal{S}^{\text{DIN-C}}_{\text{CUE}}(\text{CUE}_{\text{C}})$	3.62	4.28	4.52	4.48	2.48	3.64	4.06	4.12
$\mathcal{LR}^{\text{DIN-C}}_{\text{EL}}(\text{EL}_{\text{C}})$	10.92	6.72	5.58	4.72	37.14	16.44	10.24	6.34
$\mathcal{LM}^{\text{DIN-C}}_{\text{EL}}(\text{EL}_{\text{C}})$	23.42	10.50	6.90	4.82	81.36	45.26	25.04	10.16
$\mathcal{S}^{\text{DIN-C}}_{\text{EL}}(\text{EL}_{\text{C}})$	3.72	4.30	4.54	4.46	3.28	3.84	4.08	4.14
$\mathcal{LR}^{\text{DIN-C}}_{\text{ET}}(\text{ET}_{\text{C}})$	9.36	6.46	5.74	4.84	18.70	12.02	8.82	6.16
$\mathcal{LM}^{\text{DIN-C}}_{\text{ET}}(\text{ET}_{\text{C}})$	23.84	12.02	8.36	5.64	63.28	39.48	24.82	11.40
$\mathcal{S}^{\text{DIN-C}}_{\text{ET}}(\text{ET}_{\text{C}})$	3.58	4.24	4.52	4.46	2.70	3.76	4.08	4.14

B.2.2 Restricted Tests

Table B.7 CEM Hypothesis Test Rejection Frequencies: $\mathrm{A_C} = 2$

	Non-Standardized							
	$K = 2$				$K = 5$			
n	200	500	1000	3000	200	500	1000	3000
$\mathcal{J}^{\mathrm{C}}(\mathrm{GMM_C},\mathrm{GMM_{MA}})$	4.88	4.80	4.84	4.90	3.66	4.58	4.42	4.84
$\mathcal{LR}^{\mathrm{C}}_{\mathrm{CUE}}(\mathrm{CUE_C},\mathrm{CUE_{MA}})$	5.06	4.82	4.86	4.90	3.70	4.52	4.44	4.86
$\mathcal{LR}^{\mathrm{C}}_{\mathrm{EL}}(\mathrm{EL_C},\mathrm{EL_{MA}})$	6.38	5.40	5.18	4.98	17.98	9.02	7.10	5.78
$\mathcal{LM}^{\mathrm{C}}_{\mathrm{EL}}(\mathrm{EL_C},\mathrm{EL_{MA}})$	7.58	5.54	5.12	4.90	39.56	16.66	9.90	6.46
$\bar{\mathcal{S}}^{\mathrm{C}}_{\mathrm{EL}}(\mathrm{EL_{MA}})$	5.38	5.02	5.04	4.92	6.98	5.60	4.98	5.06
$\mathcal{LR}^{\mathrm{C}}_{\mathrm{ET}}(\mathrm{ET_C},\mathrm{ET_{MA}})$	6.16	5.50	5.20	4.96	13.48	8.16	6.80	5.70
$\mathcal{LM}^{\mathrm{C}}_{\mathrm{ET}}(\mathrm{ET_C},\mathrm{ET_{MA}})$	8.42	6.30	5.72	5.06	35.72	17.08	11.00	7.42
$\bar{\mathcal{S}}^{\mathrm{C}}_{\mathrm{ET}}(\mathrm{ET_{MA}})$	5.18	4.92	4.90	4.90	4.70	5.08	4.72	4.96

Table B.8 CEM Hypothesis Test Rejection Frequencies: $\mathrm{A_C} = 4.5$

	Non-Standardized							
	$K = 2$				$K = 5$			
n	200	500	1000	3000	200	500	1000	3000
$\mathcal{J}^{\mathrm{C}}(\mathrm{GMM_C},\mathrm{GMM_{MA}})$	3.76	4.52	4.32	4.70	2.52	3.74	4.68	4.12
$\mathcal{LR}^{\mathrm{C}}_{\mathrm{CUE}}(\mathrm{CUE_C},\mathrm{CUE_{MA}})$	3.82	4.54	4.36	4.76	2.50	3.76	4.66	4.16
$\mathcal{LR}^{\mathrm{C}}_{\mathrm{EL}}(\mathrm{EL_C},\mathrm{EL_{MA}})$	10.96	7.24	5.78	5.00	38.90	16.80	11.56	6.60
$\mathcal{LM}^{\mathrm{C}}_{\mathrm{EL}}(\mathrm{EL_C},\mathrm{EL_{MA}})$	22.96	10.86	6.96	4.98	80.36	45.90	25.90	10.74
$\bar{\mathcal{S}}^{\mathrm{C}}_{\mathrm{EL}}(\mathrm{EL_{MA}})$	4.50	4.74	4.52	4.82	6.54	5.26	5.72	4.56
$\mathcal{LR}^{\mathrm{C}}_{\mathrm{ET}}(\mathrm{ET_C},\mathrm{ET_{MA}})$	9.52	6.84	5.72	5.08	19.30	12.46	9.94	6.42
$\mathcal{LM}^{\mathrm{C}}_{\mathrm{ET}}(\mathrm{ET_C},\mathrm{ET_{MA}})$	23.16	12.34	8.20	5.88	59.02	38.04	24.24	11.56
$\bar{\mathcal{S}}^{\mathrm{C}}_{\mathrm{ET}}(\mathrm{ET_{MA}})$	4.12	4.62	4.40	4.80	3.78	4.34	5.12	4.36

Appendix C: Empirical Size-Adjusted Power

C.1 CEM: $\tau = 0$

C.1.1 Unrestricted Tests

Table C.1 CEM Hypothesis DIN Test Size-Corrected Power: $\tau = 0$, $\mathrm{A_C} = 2$, $K = 2$

n	200					500				
a	0.2	0.4	0.6	0.8	1	0.2	0.4	0.6	0.8	1
$\mathcal{J}^{\text{DIN-C}}(\text{GMM}_\text{C})$	11.04	33.40	59.30	79.34	89.36	26.58	78.84	97.64	99.82	100.00
$\mathcal{S}^{\text{DIN-C}}_{\text{CUE}}(\text{CUE}_\text{C})$	10.92	32.76	58.50	78.66	88.78	26.60	78.74	97.62	99.82	100.00
$\mathcal{LR}^{\text{DIN-C}}_{\text{EL}}(\text{EL}_\text{C})$	12.58	37.90	65.42	84.60	92.88	27.14	80.52	98.14	99.86	100.00
$\mathcal{LM}^{\text{DIN-C}}_{\text{EL}}(\text{EL}_\text{C})$	14.54	42.16	71.26	87.58	94.38	28.88	82.92	98.52	99.90	99.96
$\mathcal{S}^{\text{DIN-C}}_{\text{EL}}(\text{EL}_\text{C})$	11.04	33.44	59.62	80.02	90.02	26.58	78.90	97.68	99.82	100.00
$\mathcal{LR}^{\text{DIN-C}}_{\text{ET}}(\text{ET}_\text{C})$	11.90	35.92	62.74	82.82	92.08	26.88	79.70	98.00	99.84	100.00
$\mathcal{LM}^{\text{DIN-C}}_{\text{ET}}(\text{ET}_\text{C})$	12.88	39.78	68.00	86.42	93.88	27.58	81.22	98.28	99.94	100.00
$\mathcal{S}^{\text{DIN-C}}_{\text{ET}}(\text{ET}_\text{C})$	11.04	33.40	59.36	79.42	89.42	26.60	78.84	97.66	99.82	100.00

Table C.2 CEM Hypothesis DIN Test Size-Corrected Power: $\tau = 0$, $\mathrm{A_C} = 4.5$, $K = 2$

n	200					500				
a	0.2	0.4	0.6	0.8	1	0.2	0.4	0.6	0.8	1
$\mathcal{J}^{\text{DIN-C}}(\text{GMM}_\text{C})$	23.30	70.96	94.60	98.76	99.60	63.92	99.84	100.00	100.00	100.00
$\mathcal{S}^{\text{DIN-C}}_{\text{CUE}}(\text{CUE}_\text{C})$	22.38	68.66	93.24	98.42	99.30	63.36	99.84	100.00	100.00	100.00
$\mathcal{LR}^{\text{DIN-C}}_{\text{EL}}(\text{EL}_\text{C})$	30.82	84.48	99.20	100.00	100.00	70.82	99.94	100.00	100.00	100.00
$\mathcal{LM}^{\text{DIN-C}}_{\text{EL}}(\text{EL}_\text{C})$	34.40	86.82	99.46	100.00	100.00	73.16	99.92	100.00	100.00	100.00
$\mathcal{S}^{\text{DIN-C}}_{\text{EL}}(\text{EL}_\text{C})$	23.46	72.02	95.30	99.14	99.76	64.04	99.84	100.00	100.00	100.00
$\mathcal{LR}^{\text{DIN-C}}_{\text{ET}}(\text{ET}_\text{C})$	28.24	81.76	98.90	99.96	100.00	69.48	99.92	100.00	100.00	100.00
$\mathcal{LM}^{\text{DIN-C}}_{\text{ET}}(\text{ET}_\text{C})$	33.56	87.34	99.60	100.00	100.00	74.68	99.94	100.00	100.00	100.00
$\mathcal{S}^{\text{DIN-C}}_{\text{ET}}(\text{ET}_\text{C})$	23.58	71.32	94.88	98.82	99.68	63.98	99.84	100.00	100.00	100.00

C.1.2 Restricted Tests

Table C.3 CEM Hypothesis Test Size-Corrected Power: $\tau = 0$, $\mathrm{A_C} = 2$, $K = 2$

n	200					500				
a	0.2	0.4	0.6	0.8	1	0.2	0.4	0.6	0.8	1
$\mathcal{J}^{\mathrm{C}}(\mathrm{GMM_C},\mathrm{GMM_{MA}})$	13.28	38.44	66.84	83.70	92.30	30.64	83.66	98.50	99.90	100.00
$\mathcal{LR}^{\mathrm{C}}_{\mathrm{CUE}}(\mathrm{CUE_C},\mathrm{CUE_{MA}})$	13.04	38.20	66.22	83.00	91.90	30.66	83.62	98.48	99.90	100.00
$\mathcal{LR}^{\mathrm{C}}_{\mathrm{EL}}(\mathrm{EL_C},\mathrm{EL_{MA}})$	14.26	42.12	72.20	88.10	94.70	32.40	85.18	98.86	99.98	100.00
$\mathcal{LM}^{\mathrm{C}}_{\mathrm{EL}}(\mathrm{EL_C},\mathrm{EL_{MA}})$	15.46	46.58	75.98	90.60	95.86	33.70	86.74	99.14	100.00	100.00
$\bar{\mathcal{S}}^{\mathrm{C}}_{\mathrm{EL}}(\mathrm{EL_{MA}})$	13.44	39.66	68.08	85.06	93.28	30.96	84.16	98.58	99.92	100.00
$\mathcal{LR}^{\mathrm{C}}_{\mathrm{ET}}(\mathrm{ET_C},\mathrm{ET_{MA}})$	13.98	41.08	70.98	87.26	94.38	31.84	84.78	98.82	99.92	100.00
$\mathcal{LM}^{\mathrm{C}}_{\mathrm{ET}}(\mathrm{ET_C},\mathrm{ET_{MA}})$	14.80	43.84	73.60	89.28	95.52	32.98	85.84	98.94	99.96	100.00
$\bar{\mathcal{S}}^{\mathrm{C}}_{\mathrm{ET}}(\mathrm{ET_{MA}})$	13.24	39.04	67.58	84.50	92.92	30.86	83.96	98.58	99.92	100.00

Table C.4 CEM Hypothesis Test Size-Corrected Power: $\tau = 0$, $\mathrm{A_C} = 4.5$, $K = 2$

n	200					500				
a	0.2	0.4	0.6	0.8	1	0.2	0.4	0.6	0.8	1
$\mathcal{J}^{\mathrm{C}}(\mathrm{GMM_C},\mathrm{GMM_{MA}})$	24.22	72.78	95.18	98.76	99.46	67.62	99.90	100.00	100.00	100.00
$\mathcal{LR}^{\mathrm{C}}_{\mathrm{CUE}}(\mathrm{CUE_C},\mathrm{CUE_{MA}})$	23.70	71.40	94.36	98.34	99.30	67.22	99.88	100.00	100.00	100.00
$\mathcal{LR}^{\mathrm{C}}_{\mathrm{EL}}(\mathrm{EL_C},\mathrm{EL_{MA}})$	32.46	86.52	99.32	100.00	100.00	73.30	99.94	100.00	100.00	100.00
$\mathcal{LM}^{\mathrm{C}}_{\mathrm{EL}}(\mathrm{EL_C},\mathrm{EL_{MA}})$	35.40	87.78	99.48	100.00	100.00	74.34	99.94	100.00	100.00	100.00
$\bar{\mathcal{S}}^{\mathrm{C}}_{\mathrm{EL}}(\mathrm{EL_{MA}})$	24.94	74.22	96.06	99.18	99.74	67.50	99.90	100.00	100.00	100.00
$\mathcal{LR}^{\mathrm{C}}_{\mathrm{ET}}(\mathrm{ET_C},\mathrm{ET_{MA}})$	29.80	83.82	98.98	99.94	100.00	72.14	99.94	100.00	100.00	100.00
$\mathcal{LM}^{\mathrm{C}}_{\mathrm{ET}}(\mathrm{ET_C},\mathrm{ET_{MA}})$	34.76	88.20	99.64	100.00	100.00	76.12	99.96	100.00	100.00	100.00
$\bar{\mathcal{S}}^{\mathrm{C}}_{\mathrm{ET}}(\mathrm{ET_{MA}})$	24.56	73.60	95.68	99.06	99.66	67.78	99.90	100.00	100.00	100.00

C.2 MEM: $a = 0$

C.2.1 Unrestricted Tests

Table C.5 MEM Hypothesis DIN Test Size-Corrected Power: $a = 0$, $A_M = 1$, $K = 2$

n	200					500				
τ	0.2	0.4	0.6	0.8	1	0.2	0.4	0.6	0.8	1
$\mathcal{J}^{\text{DIN-M}}(\text{GMM}_\text{M})$	32.82	88.42	99.84	100.00	100.00	72.82	99.94	100.00	100.00	100.00
$\mathcal{S}_{\text{CUE}}^{\text{DIN-M}}(\text{CUE}_\text{M})$	32.78	88.26	99.84	100.00	100.00	72.74	99.94	100.00	100.00	100.00
$\mathcal{LR}_{\text{EL}}^{\text{DIN-M}}(\text{EL}_\text{M})$	33.54	88.66	99.86	100.00	100.00	72.92	99.94	100.00	100.00	100.00
$\mathcal{LM}_{\text{EL}}^{\text{DIN-M}}(\text{EL}_\text{M})$	32.22	86.44	99.34	99.98	100.00	72.46	99.88	100.00	100.00	100.00
$\mathcal{S}_{\text{EL}}^{\text{DIN-M}}(\text{EL}_\text{M})$	32.84	88.52	99.84	100.00	100.00	72.84	99.94	100.00	100.00	100.00
$\mathcal{LR}_{\text{ET}}^{\text{DIN-M}}(\text{ET}_\text{M})$	33.10	88.80	99.86	100.00	100.00	73.12	99.94	100.00	100.00	100.00
$\mathcal{LM}_{\text{ET}}^{\text{DIN-M}}(\text{ET}_\text{M})$	28.74	86.22	99.62	100.00	100.00	71.04	99.94	100.00	100.00	100.00
$\mathcal{S}_{\text{ET}}^{\text{DIN-M}}(\text{ET}_\text{M})$	32.80	88.42	99.84	100.00	100.00	72.84	99.94	100.00	100.00	100.00

Table C.6 MEM Hypotesis DIN Test Size-Corrected Power: $a = 0$, $A_M = 1.5$, $K = 2$

n	200					500				
τ	0.2	0.4	0.6	0.8	1	0.2	0.4	0.6	0.8	1
$\mathcal{J}^{\text{DIN-M}}(\text{GMM}_\text{M})$	26.80	83.68	99.44	100.00	100.00	66.18	99.94	100.00	100.00	100.00
$\mathcal{S}_{\text{CUE}}^{\text{DIN-M}}(\text{CUE}_\text{M})$	26.64	83.50	99.46	100.00	100.00	66.20	99.92	100.00	100.00	100.00
$\mathcal{LR}_{\text{EL}}^{\text{DIN-M}}(\text{EL}_\text{M})$	27.38	83.70	99.38	100.00	100.00	66.00	99.94	100.00	100.00	100.00
$\mathcal{LM}_{\text{EL}}^{\text{DIN-M}}(\text{EL}_\text{M})$	26.06	80.56	98.86	99.86	100.00	64.28	99.82	100.00	100.00	100.00
$\mathcal{S}_{\text{EL}}^{\text{DIN-M}}(\text{EL}_\text{M})$	26.66	83.44	99.46	100.00	100.00	66.12	99.92	100.00	100.00	100.00
$\mathcal{LR}_{\text{ET}}^{\text{DIN-M}}(\text{ET}_\text{M})$	27.38	83.86	99.44	100.00	100.00	66.16	99.92	100.00	100.00	100.00
$\mathcal{LM}_{\text{ET}}^{\text{DIN-M}}(\text{ET}_\text{M})$	25.40	82.42	99.32	100.00	100.00	64.88	99.92	100.00	100.00	100.00
$\mathcal{S}_{\text{ET}}^{\text{DIN-M}}(\text{ET}_\text{M})$	26.82	83.56	99.44	100.00	100.00	66.12	99.92	100.00	100.00	100.00

C.2.2 Restricted Tests

Table C.7 MEM Hypothesis Test Size-Corrected Power: $a = 0$, $\mathrm{A}_{\mathrm{M}} = 1$, $K = 2$

n	200					500				
τ	0.2	0.4	0.6	0.8	1	0.2	0.4	0.6	0.8	1
$\mathcal{J}^{\mathrm{M}}(\mathrm{GMM}_{\mathrm{M}},\mathrm{GMM}_{\mathrm{MA}})$	42.50	93.54	99.92	100.00	100.00	82.28	100.00	100.00	100.00	100.00
$\mathcal{LR}^{\mathrm{M}}_{\mathrm{CUE}}(\mathrm{CUE}_{\mathrm{M}},\mathrm{CUE}_{\mathrm{MA}})$	42.46	93.56	99.92	100.00	100.00	82.36	100.00	100.00	100.00	100.00
$\mathcal{LR}^{\mathrm{M}}_{\mathrm{EL}}(\mathrm{EL}_{\mathrm{M}},\mathrm{EL}_{\mathrm{MA}})$	41.60	93.32	99.94	100.00	100.00	82.30	100.00	100.00	100.00	100.00
$\mathcal{LM}^{\mathrm{M}}_{\mathrm{EL}}(\mathrm{EL}_{\mathrm{M}},\mathrm{EL}_{\mathrm{MA}})$	40.26	92.10	99.78	100.00	100.00	81.78	100.00	100.00	100.00	100.00
$\bar{\mathcal{S}}^{\mathrm{M}}_{\mathrm{EL}}(\mathrm{EL}_{\mathrm{MA}})$	42.34	93.40	99.92	100.00	100.00	82.40	100.00	100.00	100.00	100.00
$\mathcal{LR}^{\mathrm{M}}_{\mathrm{ET}}(\mathrm{ET}_{\mathrm{M}},\mathrm{ET}_{\mathrm{MA}})$	42.48	93.56	99.94	100.00	100.00	82.28	100.00	100.00	100.00	100.00
$\mathcal{LM}^{\mathrm{M}}_{\mathrm{ET}}(\mathrm{ET}_{\mathrm{M}},\mathrm{ET}_{\mathrm{MA}})$	41.40	93.28	99.96	100.00	100.00	82.30	100.00	100.00	100.00	100.00
$\bar{\mathcal{S}}^{\mathrm{M}}_{\mathrm{ET}}(\mathrm{ET}_{\mathrm{MA}})$	42.46	93.46	99.92	100.00	100.00	82.26	100.00	100.00	100.00	100.00

Table C.8 MEM Hypothesis Test Size-Corrected Power: $a = 0$, $\mathrm{A}_{\mathrm{M}} = 1.5$, $K = 2$

n	200					500				
τ	0.2	0.4	0.6	0.8	1	0.2	0.4	0.6	0.8	1
$\mathcal{J}^{\mathrm{M}}(\mathrm{GMM}_{\mathrm{M}},\mathrm{GMM}_{\mathrm{MA}})$	31.14	87.18	99.82	100.00	100.00	71.46	99.98	100.00	100.00	100.00
$\mathcal{LR}^{\mathrm{M}}_{\mathrm{CUE}}(\mathrm{CUE}_{\mathrm{M}},\mathrm{CUE}_{\mathrm{MA}})$	31.24	87.36	99.78	100.00	100.00	71.30	99.98	100.00	100.00	100.00
$\mathcal{LR}^{\mathrm{M}}_{\mathrm{EL}}(\mathrm{EL}_{\mathrm{M}},\mathrm{EL}_{\mathrm{MA}})$	31.96	87.74	99.78	100.00	100.00	71.70	99.98	100.00	100.00	100.00
$\mathcal{LM}^{\mathrm{M}}_{\mathrm{EL}}(\mathrm{EL}_{\mathrm{M}},\mathrm{EL}_{\mathrm{MA}})$	32.10	86.08	99.36	99.96	100.00	71.12	99.94	100.00	100.00	100.00
$\bar{\mathcal{S}}^{\mathrm{M}}_{\mathrm{EL}}(\mathrm{EL}_{\mathrm{MA}})$	30.88	87.12	99.78	100.00	100.00	71.24	99.98	100.00	100.00	100.00
$\mathcal{LR}^{\mathrm{M}}_{\mathrm{ET}}(\mathrm{ET}_{\mathrm{M}},\mathrm{ET}_{\mathrm{MA}})$	31.42	87.62	99.80	100.00	100.00	71.90	99.98	100.00	100.00	100.00
$\mathcal{LM}^{\mathrm{M}}_{\mathrm{ET}}(\mathrm{ET}_{\mathrm{M}},\mathrm{ET}_{\mathrm{MA}})$	32.30	88.04	99.78	100.00	100.00	71.76	99.98	100.00	100.00	100.00
$\bar{\mathcal{S}}^{\mathrm{M}}_{\mathrm{ET}}(\mathrm{ET}_{\mathrm{MA}})$	30.86	87.18	99.80	100.00	100.00	71.64	99.98	100.00	100.00	100.00

C.3 CEM: $a = 0$

C.3.1 Unrestricted Tests

Table C.9 CEM Hypothesis DIN Test Size-Corrected Power: $a = 0$, $\mathrm{A_C} = 2$, $K = 2$

n	200					500				
τ	0.2	0.4	0.6	0.8	1	0.2	0.4	0.6	0.8	1
$\mathcal{J}^{\text{DIN-C}}(\text{GMM}_\text{C})$	26.42	83.08	99.40	100.00	100.00	66.40	99.90	100.00	100.00	100.00
$\mathcal{S}^{\text{DIN-C}}_{\text{CUE}}(\text{CUE}_\text{C})$	26.12	82.80	99.38	100.00	100.00	66.52	99.92	100.00	100.00	100.00
$\mathcal{LR}^{\text{DIN-C}}_{\text{EL}}(\text{EL}_\text{C})$	26.78	82.94	99.42	100.00	100.00	65.72	99.90	100.00	100.00	100.00
$\mathcal{LM}^{\text{DIN-C}}_{\text{EL}}(\text{EL}_\text{C})$	25.88	79.82	98.70	99.98	100.00	64.02	99.84	100.00	100.00	100.00
$\mathcal{S}^{\text{DIN-C}}_{\text{EL}}(\text{EL}_\text{C})$	26.44	83.02	99.40	100.00	100.00	66.44	99.90	100.00	100.00	100.00
$\mathcal{LR}^{\text{DIN-C}}_{\text{ET}}(\text{ET}_\text{C})$	26.80	83.14	99.48	100.00	100.00	66.06	99.92	100.00	100.00	100.00
$\mathcal{LM}^{\text{DIN-C}}_{\text{ET}}(\text{ET}_\text{C})$	26.42	82.44	99.34	100.00	100.00	65.70	99.90	100.00	100.00	100.00
$\mathcal{S}^{\text{DIN-C}}_{\text{ET}}(\text{ET}_\text{C})$	26.50	83.08	99.40	100.00	100.00	66.42	99.90	100.00	100.00	100.00

Table C.10 CEM Hypothesis DIN Test Size-Corrected Power: $a = 0$, $\mathrm{A_C} = 4.5$, $K = 2$

n	200					500				
τ	0.2	0.4	0.6	0.8	1	0.2	0.4	0.6	0.8	1
$\mathcal{J}^{\text{DIN-C}}(\text{GMM}_\text{C})$	18.84	69.58	97.54	99.98	100.00	50.34	99.48	100.00	100.00	100.00
$\mathcal{S}^{\text{DIN-C}}_{\text{CUE}}(\text{CUE}_\text{C})$	18.98	69.52	97.52	99.98	100.00	50.20	99.44	100.00	100.00	100.00
$\mathcal{LR}^{\text{DIN-C}}_{\text{EL}}(\text{EL}_\text{C})$	18.78	68.14	97.34	99.98	100.00	49.26	99.46	100.00	100.00	100.00
$\mathcal{LM}^{\text{DIN-C}}_{\text{EL}}(\text{EL}_\text{C})$	16.78	57.24	92.84	99.62	100.00	44.34	98.92	100.00	100.00	100.00
$\mathcal{S}^{\text{DIN-C}}_{\text{EL}}(\text{EL}_\text{C})$	19.04	69.58	97.54	99.98	100.00	50.28	99.46	100.00	100.00	100.00
$\mathcal{LR}^{\text{DIN-C}}_{\text{ET}}(\text{ET}_\text{C})$	18.62	69.00	97.46	99.98	100.00	50.14	99.46	100.00	100.00	100.00
$\mathcal{LM}^{\text{DIN-C}}_{\text{ET}}(\text{ET}_\text{C})$	17.42	63.56	96.10	99.98	100.00	47.26	99.28	100.00	100.00	100.00
$\mathcal{S}^{\text{DIN-C}}_{\text{ET}}(\text{ET}_\text{C})$	18.96	69.84	97.56	99.98	100.00	50.32	99.48	100.00	100.00	100.00

C.3.2 Restricted Tests

Table C.11 CEM Hypothesis Test Size-Corrected Power: $a = 0$, $\mathrm{A_C} = 2$, $K = 2$

n	200					500				
τ	0.2	0.4	0.6	0.8	1	0.2	0.4	0.6	0.8	1
$\mathcal{J}^{\mathrm{C}}(\mathrm{GMM_C},\mathrm{GMM_{MA}})$	31.52	87.60	99.78	100.00	100.00	72.12	99.98	100.00	100.00	100.00
$\mathcal{LR}^{\mathrm{C}}_{\mathrm{CUE}}(\mathrm{CUE_C},\mathrm{CUE_{MA}})$	31.30	87.60	99.74	100.00	100.00	72.16	99.98	100.00	100.00	100.00
$\mathcal{LR}^{\mathrm{C}}_{\mathrm{EL}}(\mathrm{EL_C},\mathrm{EL_{MA}})$	31.56	87.38	99.78	100.00	100.00	72.42	99.98	100.00	100.00	100.00
$\mathcal{LM}^{\mathrm{C}}_{\mathrm{EL}}(\mathrm{EL_C},\mathrm{EL_{MA}})$	29.94	84.28	99.38	100.00	100.00	70.62	99.98	100.00	100.00	100.00
$\bar{\mathcal{S}}^{\mathrm{C}}_{\mathrm{EL}}(\mathrm{EL_{MA}})$	31.72	87.64	99.80	100.00	100.00	72.26	99.98	100.00	100.00	100.00
$\mathcal{LR}^{\mathrm{C}}_{\mathrm{ET}}(\mathrm{ET_C},\mathrm{ET_{MA}})$	31.92	87.64	99.84	100.00	100.00	72.50	99.98	100.00	100.00	100.00
$\mathcal{LM}^{\mathrm{C}}_{\mathrm{ET}}(\mathrm{ET_C},\mathrm{ET_{MA}})$	31.26	87.06	99.82	100.00	100.00	72.12	99.98	100.00	100.00	100.00
$\bar{\mathcal{S}}^{\mathrm{C}}_{\mathrm{ET}}(\mathrm{ET_{MA}})$	31.60	87.74	99.78	100.00	100.00	72.22	99.98	100.00	100.00	100.00

Table C.12 CEM Hypothesis Test Size-Corrected Power: $a = 0$, $\mathrm{A_C} = 4.5$, $K = 2$

n	200					500				
τ	0.2	0.4	0.6	0.8	1	0.2	0.4	0.6	0.8	1
$\mathcal{J}^{\mathrm{C}}(\mathrm{GMM_C},\mathrm{GMM_{MA}})$	19.48	70.78	97.84	99.98	100.00	53.28	99.64	100.00	100.00	100.00
$\mathcal{LR}^{\mathrm{C}}_{\mathrm{CUE}}(\mathrm{CUE_C},\mathrm{CUE_{MA}})$	20.06	71.44	97.90	99.98	100.00	53.10	99.64	100.00	100.00	100.00
$\mathcal{LR}^{\mathrm{C}}_{\mathrm{EL}}(\mathrm{EL_C},\mathrm{EL_{MA}})$	20.26	70.56	97.66	100.00	100.00	51.68	99.58	100.00	100.00	100.00
$\mathcal{LM}^{\mathrm{C}}_{\mathrm{EL}}(\mathrm{EL_C},\mathrm{EL_{MA}})$	18.14	59.72	93.82	99.72	100.00	45.04	99.16	100.00	100.00	100.00
$\bar{\mathcal{S}}^{\mathrm{C}}_{\mathrm{EL}}(\mathrm{EL_{MA}})$	20.46	71.56	97.86	99.98	100.00	52.94	99.64	100.00	100.00	100.00
$\mathcal{LR}^{\mathrm{C}}_{\mathrm{ET}}(\mathrm{ET_C},\mathrm{ET_{MA}})$	20.16	70.94	97.86	99.98	100.00	51.80	99.54	100.00	100.00	100.00
$\mathcal{LM}^{\mathrm{C}}_{\mathrm{ET}}(\mathrm{ET_C},\mathrm{ET_{MA}})$	18.34	65.62	96.64	99.98	100.00	48.38	99.36	100.00	100.00	100.00
$\bar{\mathcal{S}}^{\mathrm{C}}_{\mathrm{ET}}(\mathrm{ET_{MA}})$	19.90	71.00	97.86	99.98	100.00	53.20	99.68	100.00	100.00	100.00