Exogeneity in Semiparametric Models: Definitions and Tests*

Paulo M.D.C. Parente
Department of Economics
University of Warwick

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

First Draft: November 2005
This Revision: April 2006

Abstract

A major concern of this article is the provision of definitions for exogeneity appropriate for models defined through a set of conditional moment indicators. The paper also proposes tests for corresponding exogeneity hypotheses in the conditional moment restrictions framework. The basic idea underpinning test construction is the equivalence between a finite number of conditional moment restrictions and an infinite number of unconditional restrictions. Consequently, tests of exogeneity can be seen as tests for an additional set of infinite moment conditions. We suggest test statistics based on generalized method of moments and generalized empirical likelihood. The asymptotic distributions of our statistics are obtained under the null hypothesis and a suitable sequence of local alternatives.

JEL Classification: C12, C14, C30

Key-words: *Specification Tests, Series approximations, Conditional Moment Restrictions, GMM, GEL.*

*The authors are extremely grateful to R. Crump, S. Hemvanich, J.M.C. Santos Silva and M.B. Stewart for useful discussions. The Fundação para a Ciência e 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 his contribution to this revision was prepared. Any remaining errors are solely the responsibility of the authors.

1 Introduction

The primary concern of this paper is definitions and tests of (weak) exogeneity apposite for semiparametric moment condition models specified for cross-section data. Several definitions of exogeneity in different contexts have been given in the literature. For parametric models the version provided in Engle *et al.* (1983), henceforth referred to as EHR, would seem to be consensual; see Definition 2.5, p.282, in EHR. Essentially, their definition of exogeneity is that of statistical ancillarity. That is, a random vector x is exogenous for a vector of parameters if efficient inference for the parameter vector can be made based solely on the conditional distribution of observables conditional on the random vector x. In other words, no loss of information is incurred by disregarding the marginal distribution of an exogenous random vector. From a policy perspective, exogeneity also assumes a central importance. Let a given random vector x be exogenous for the parameter vector characterising the conditional distribution of a target vector y. Suppose also that the conditional distribution of y given x under exogeneity may be regarded as a behavioural relationship for the target vector y. If the exogenous random vector 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 accurately predict the effect of a change in policy expressed through changes in the exogenous random vector x without knowledge of the marginal distribution of x. For the dynamic linear simultaneous equations model with normally distributed errors, Theorem 4.3 (a) and (b), p.298, of EHR shows sufficient conditions for their exogeneity definition that involve the uncorrelatedness of particular error terms. More specifically, for classical linear regression this assumption is equivalent to that usually made when the objective is to estimate the best linear predictor, namely, that the regression error is uncorrelated with the vector of regressor variables.

More recently Blundell and Horowitz (2004) present a definition of exogeneity in a non-parametric regression model setting for a random variable y in terms of a vector of covariates x in which the expectation of the error term conditional on a set of identifying instruments w is maintained to be zero. Covariates x are said to be exogenous if the conditional expectation of the regression error term given x is also zero. Their definition has the advantage that standard nonparametric regression methods are then appropriate. Because the instruments are ignored in this definition it may be characterised as a partial form of exogeneity which we term below as *marginal* exogeneity. We demonstrate *via* an example that marginal exogeneity is not equivalent to that given by EHR for the parametric context. More importantly, however, if the instruments w are control variables for the policy maker, the effect of changes in w on the dependent variable y is unknown without further knowledge of the conditional distribution of the covariates x given w. Of course, if the covariates x themselves are under the control of the policy maker, then the effect of changes in x on the dependent variable y is perfectly predictable.

We therefore provide a different definition of exogeneity in the context of a general nonlinear conditional moment restrictions model. This definition more accords with that found in many econometrics textbooks; see, for example, Hayashi (2000) and Wooldridge (2002). A random vector is considered to be *conditionally* exogenous for the parameter vector if the expectation of the vector of moment indicators conditional on both the random vector x and maintained instruments w is zero. In particular, if covariates in a regression model are conditionally exogenous then as a consequence the instruments will be redundant as additional explanators, a property arising necessarily from ancillarity if covariates are exogenous. This definition is of course stricter than that of Blundell and Horowitz (2004). Consequently estimators which efficiently incorporate this assumption should dominate those which only make use of marginal exogeneity.

In addition, we also provide tests for both marginal and conditional exogeneity in the conditional moment framework. Most tests of exogeneity proposed in the literature focus on the best linear predictor setting. The most popular of these tests is probably the Hausman test [Hausman (1978)] which contrasts estimators obtained assuming orthogonality conditions between errors and instruments and errors and covariates (and instruments) respectively. Lagrange multiplier or score tests were suggested by Engle (1982). Smith (1994) gives limited information classical test statistics for exogeneity in the dynamic simultaneous equations model discussed in EHR, Section 4, pp. 294-300. However, these kinds of tests are inappropriate for models defined by conditional moment constraints. As noted out by Bierens (1990), orthogonality tests will generally be inconsistent against some alternatives implied by conditional moment conditions as only a finite number of unconditional restrictions are incorporated in such tests. Moreover, examples can be easily constructed in which although an orthogonality condition holds the conditional moment restrictions may not; see De Jong and Bierens (1994). While relatively little attention has been paid to tests of exogeneity in regression models defined by conditional moment restrictions, there is a vast literature on akin tests of goodness of fit in such models. Tests based on an infinite number of unconditional moment restrictions may be designed to overcome the aforementioned test inconsistency. See, for example, Eubank and Spiegelman (1990) in the non-linear regression context. Other tests based on this principle have been proposed, also for this set-up, by De Jong and Bierens (1994), Hong and White (1995) and Jayasuriya (1996). Donald, Imbens and Newey (2003), henceforth DIN, extend this idea to the conditional moment restriction setting for GMM [Hansen (1982)] and generalized empirical likelihood (GEL) [Newey and Smith (2004), Smith (1997, 2001)]. Our paper adapts these last methods developed in DIN to formulate tests for exogeneity in the conditional model framework.[1]

The approach in DIN approximates the conditional moment restrictions by a finite set of unconditional moment restrictions, the number of which is allowed to tend to infinity. Both marginal and conditional exogeneity hypotheses involve two sets of conditional moment restrictions where the second set implies the first. Likewise these sets of conditional moment conditions are replaced by corresponding sets of unconditional moment restrictions with the first set a subset of the second. Hence, our tests for exogeneity may be interpreted as tests for additional moment restrictions similar to those proposed by Newey (1985) and Tauchen (1985) in a fully parametric setting and by Smith (1997, 2001) for GEL. Since asymptotically we are dealing with an infinite number of moment

[1]Alternative tests for exogeneity could also be based 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) and Ellison and Ellison (2000) among others.

restrictions our test statistics do not converge to the usual chi-squared distribution but rather the standard normal distribution after appropriate standardization. Furthermore, unlike orthogonality test statistics, these tests do not require efficient parameter estimation under either conditional moment specification.

Also of particular interest and relevance are the recent papers by Tripathi and Kitamura (2003) and Delgado *et al.* (2005) which like DIN also propose tests for conditional moment restrictions. Tripathi and Kitamura (2003) propose a likelihood ratio (LR) test based on a smoothed conditional empirical likelihood (EL) objective function. Since the conventional EL ratio statistic does not allow one to test for conditional moment restrictions directly, Tripathi and Kitamura (2003) replace the empirical log-likelihood objective function by a kernel-based estimator for its conditional expected value. After a location-scale standardisation their empirical LR test statistic converges in distribution to a standard normal random variate under correctly specified conditional moment restrictions. Moreover, they prove their test is asymptotically optimal among a particular class of tests for over-identifying moment restrictions in the sense that it achieves maximum local average power in that class. More recently, Smith (2005a, 2005b) showed that this approach can be extended to obtain tests based on the Cressie-Read (1984) power divergence family of discrepancies and GEL. Delgado *et al.* (2005) propose two different methods. First, they replace the conditional moment restrictions by an equivalent finite number of non-arbitrary unconditional moment restrictions and test the validity of the latter. This approach is similar to those of Fan and Li (1996), Zheng (1996, 1998) and Lavergne and Vuong (2000) for regression models. Their test statistic also needs to resort to kernel methods as it is necessary to estimate certain conditional expectations. Like Tripathi and Kitamura's (2003) statistic their statistic also converges in distribution to a standard normal random variate under correct specification. Secondly, they show that their unconditional moment conditions are equivalent to some arbitrary set of orthogonality conditions between generalized residuals and a known function of the conditioning variables and some nuisance parameters. Since these moment conditions depend on nuisance parameters it is necessary to resort to a Kolmogorov-Smirnov or

Cramer-von-Mises form of test statistic. These test statistics are similar in spirit to those proposed by Bierens (1982, 1990) and have a non-standard limiting distribution.

The paper is organized as follows. Section 2 provides a detailed discussion of the exogeneity concept in models defined by conditional moment restrictions. The test problem is specified in section 3 together with some notation and assumptions. GMM and GEL test statistics for exogeneity are presented there also. Section 4 details the limiting distribution of these statistics under the exogeneity hypothesis whereas section 5 considers their asymptotic distribution under a suitable sequence of local alternatives. Section 5 concludes. Proofs of the results in the text and certain subsidiary lemmata are given in the Appendix.

2 Exogeneity

2.1 Some Preliminaries

We initially reconsider the landmark paper EHR which discusses exogeneity concepts for fully parametric models in the classical likelihood context and relates definitions of exogeneity to efficient parameter estimation.

Let the random vector z be partitioned as $z = (y', x')'$. It is assumed that the distribution of z is described by the density function $f(\cdot|\lambda)$ which is known up to the finite dimensional parameter vector $\lambda \in \Lambda$. Suppose that λ is partitioned as $\lambda = (\lambda_1', \lambda_2')'$, where $\lambda_1 \in \Lambda_1$ and $\lambda_2 \in \Lambda_2$. Then $[(y|x; \lambda_1), (x|\lambda_2)]$ operates a *sequential cut* on $f(z|\lambda)$ if and only if $f(z|\lambda) = f(y|x, \lambda_1) f(x|\lambda_2)$, where $f(\cdot|x, \lambda_1)$ and $f(\cdot|\lambda_2)$ denote the conditional and marginal densities of y given x and x respectively, and λ_1 and λ_2 are variation free, i.e. $(\lambda_1', \lambda_2')' \in \Lambda_1 \times \Lambda_2$; see EHR, Definition 2.4, p.282. The random vector x is (weakly) *exogenous* for the parameters of interest ψ if (i) ψ is a function of λ_1 and (ii) $[(y|x; \lambda_1), (x|\lambda_2)]$ operates a sequential cut on $f(z|\lambda)$; see EHR, Definition 2.5, p.282. Consequently, for efficient inference on ψ we may merely consider the conditional density $f(\cdot|x, \lambda_1)$ of y given x and ignore, without loss of information, the marginal density $f(\cdot|\lambda_2)$ of x.

To illustrate the consequences of this definition consider the following model

$$\begin{aligned} y &= x'\beta_0 + u, \\ x &= \Pi_x w + v, \end{aligned}$$

where y and x are scalar and vector endogenous variables. Assume that the joint distribution of (u, v) conditional on w is multivariate normal with mean vector zero and nonsingular conditionally heteroskedastic variance matrix

$$\Sigma(w) = \begin{pmatrix} \sigma_u^2 & \Sigma_{ux} \\ \Sigma_{ux} & \Sigma_{xx} \end{pmatrix} (w).$$

Thus, conditional on x and w, y is normally distributed with mean and variance given by

$$\begin{aligned} E[y|w,x] &= x'\beta_0 + \Sigma_{ux}(w)\Sigma_{xx}(w)^{-1}\left(x - \Pi_x w\right), \qquad (2.1) \\ var[y|w,x] &= \sigma_u^2(w) - \Sigma_{ux}(w)\Sigma_{xx}(w)^{-1}\Sigma_{ux}(w)'. \end{aligned}$$

The assumption $\Sigma_{ux}(w) = 0$ ensures the exogeneity of x for β_0 in the sense of EHR as then from (2.1) the density of y given x and w no longer depends on the parameter matrices Π_x which characterises the marginal distribution of x conditional on w. If u and v are independent of w this is precisely the condition in the best linear predictor framework when $E\left[y|w,x\right] = E\left[y|x\right] = x'\beta_0$.

More recently Blundell and Horowitz (2004) proposed a definition of exogeneity for a non-parametric regression set-up. They considered the model

$$y = g\left(x\right) + u,$$

where the function $g(\cdot)$ is unknown and thus needs to be nonparametrically estimated. The identifying conditional moment restriction $E[u|w] = 0$ is maintained where w is a vector of instruments. Blundell and Horowitz (2004) define the covariate vector x to be exogenous if $E[u|x] = 0$. This definition, however, is not equivalent to that given by EHR for the parametric setting. To see this reconsider the above example with $g\left(x\right) = x'\beta_0$ and rewrite (2.1) as

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

where $u = y - E[y|w,x] + \Sigma_{ux}(w)\Sigma_{xx}(w)^{-1}(x - \Pi_x w)$. Now

$$E[u|w,x] = \Sigma_{ux}(w)\Sigma_{xx}(w)^{-1}(x - \Pi_x w)$$

which if x and w are proper random vectors equals zero if and only if $\Sigma_{ux}(w) = 0$. However,

$$E[u|x] = E\left[\Sigma_{ux}(w)\Sigma_{xx}(w)^{-1}|x\right]x - E\left[\Sigma_{ux}(w)\Sigma_{xx}(w)^{-1}\Pi_x w|x\right].$$

Of course the condition $\Sigma_{ux}(w) = 0$ is now sufficient rather than necessary for $E[u|x] = 0$. More generally, the condition $E[u|x] = 0$ holds if and only if $E[\Sigma_{ux}(w)\Sigma_{xx}(w)^{-1}|x]x = E[\Sigma_{ux}(w)\Sigma_{xx}(w)^{-1}\Pi_x w|x]$ which requires additional constraints on the joint distribution of x and w.

Consequently, from a policy perspective, the particular definition of exogeneity given in Blundell and Horowitz (2004) may be inadequate. Suppose that the vector of instruments w is also a vector of control variables. Although under the exogeneity hypothesis $E[u|x] = 0$ the effect of changes of x on y are predictable as $E[y|x] = g(x)$, the impact of altering w is not. In particular, $E[y|w] = E[g(x)|w]$ which remains unchanged between maintained and exogeneity hypotheses. Knowledge of the conditional distribution of x given w is still required. A more apposite definition of exogeneity is given by $E[u|w,x] = 0$ in which case $E[y|w,x] = g(x)$. The effect on y of any change in w only occurs through any alteration consequently induced in x, the effect of which is predictable as above.[2]

2.2 Definitions

Rather than considering u to be a regression error as in the above example, we now deal with the more general form $u = u(z, \beta_0)$, 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 which constitute the object of inferential interest.

[2]Note that if $\Sigma(w) = \Sigma$, i.e. u and v are independent of w, then the conditions $E[u|w,x] = 0$, $E[u|x] = 0$ and $\Sigma_{ux} = 0$ are mutually equivalent. In this case, therefore, the distinction between EHR and Blundell and Horowitz (2004) does not apply.

Like Blundell and Horowitz (2004) we assume that there exists an observable vector of instruments w such that

$$E[u(z, \beta_0)|w] = 0. \tag{2.2}$$

The definition of exogeneity due to Blundell and Horowitz (2004) does not involve the maintained instrument vector w. Thus we consider it to be a partial or *marginal* form of exogeneity.

Definition 2.1 *(Marginal Exogeneity.) The random vector x is said to be marginally exogenous for β_0 if*

$$E[u(z, \beta_0)|x] = 0. \tag{2.3}$$

In the regression example of section 2.1 $u(z, \beta_0) = y - x'\beta_0$. Marginal exogeneity (2.3) of x for β_0 clearly implies that the moment condition $E[u\,(z, \beta_0)\,|x] = 0$ may be used to obtain a consistent estimator for β_0. However, such an estimator is unlikely to be efficient as it neglects the maintained information contained in the moment constraint (2.2). Furthermore, an estimator for β_0 based on the moment conditions (2.2) and (2.3) should be at least as efficient as an estimator which only used the conditions (2.2). Thus, if (2.3) holds, not only is $E[y|x]$ is correctly specified as $x'\beta_0$ but a more efficient estimator for β_0 is also possible.

In the regression context a consequence of marginal exogeneity (2.3) is that $E[y|x] = x'\beta_0$ and, thus, the regressor vector x may be used to control y. However, as noted above, if w is a policy control variable, the effect of changes in w on y is unknown as $E[y|w] = E[x|w]'\beta_0 = w'\Pi_x'\beta_0$. Without further knowledge of the conditional distribution of x given w, namely Π_x for this example, the impact of w on y is not predictable.

Consequently marginal exogeneity may represent an inadequate definition of exogeneity in some contexts. Revisiting the regression example the condition $E\,[u\,(z, \beta_0)\,|w, x] = 0$ implies $E[y|w, x] = E[y|x] = x'\beta_0$. Therefore, under this condition the instruments w contribute no additional information to the conditional expectation of y than that provided by x. This property of conditional mean independence of y and w given x might be considered inherently characteristic of an exogenous vector in the regression setting. Furthermore, as noted in section 2.1, it approximates more closely the definition of exogeneity given in EHR.

To deal with these issues, we revise the definition of exogeneity given in Blundell and Horowitz (2004) to incorporate the maintained instruments w which therefore takes a *conditional* form.

Definition 2.2 *(Conditional Exogeneity.) The random vector x is said to be conditionally exogenous for β_0 if*

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

Conditional exogeneity implies marginal exogeneity and is thus a more stringent requirement. Therefore, estimators which utilise conditional exogeneity (2.4) should be more efficient than those based on the marginal moment conditions (2.2) and (2.3).

The next sections develop and analyse the large sample properties of tests for both marginal and conditional types of exogeneity.

3 GMM and GEL Test Statistics

3.1 The Test Problem

Let s denote a generic random vector that will be made explicit in each particular instance. The null hypothesis is

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

with the alternative hypothesis as

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

where s is a random vector that may or may not include w. The null hypothesis (3.1) encompasses both the definitions of exogeneity given in section 2.2 with $s=x$ and $s=(w',x')'$ as marginal and conditional exogeneity respectively.

3.2 Approximating Conditional Moment Restrictions

It is well known [Chamberlain (1987)] that conditional moment conditions of the type given in both null and alternative hypotheses (3.1) and (3.2) are equivalent to a countable number of unconditional moment restrictions under certain regularity conditions. The following assumption, DIN Assumption 1, provides precise conditions.

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

Assumption 3.1 *For all K, $E[q^K(s)'q^K(s)]$ be 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, 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 $E[u(z, \beta_0)|s] = 0$, then $E[u(z, \beta_0) \otimes q^K(s)] = 0$ for all K. Furthermore, if $E[u(z, \beta_0)|s] \neq 0$, then $E[u(z, \beta_0) \otimes q^K(s)] \neq 0$ for K large enough.*

DIN define the unconditional moment indicator vector as $g(z, \beta) = u(z, \beta) \otimes q^K(s)$. By considering the moment conditions $E[g(z, \beta_0)] = 0$ and applying EL, IV, GMM or GEL, if K approaches infinity at an appropriate rate which will depend on n and the type of estimator considered, then DIN show that such estimators are consistent and achieve the semi-parametric efficiency bound. To do so, however, requires the imposition of a normalization condition on the approximating functions as in DIN Assumption 2 which follows.

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

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

Hence we may reinterpret the null hypothesis of marginal (2.3) or conditional (2.4) exogeneity as one consisting of additional moment restrictions. Therefore, we need to replace appropriately the conditional moment constraints (2.2) and either (2.3) or (2.4).

The maintained conditional moment restrictions (2.2) consequently become

$$E[u(z,\beta_0) \otimes q_1^K(w)] = 0 \tag{3.3}$$

for approximating functions $q_1^K(\cdot)$ that satisfy Assumptions 3.1 and 3.2 for $s = w$.

Let $q_0^K(w,x)$ be a MK-vector of approximating functions that depends only on the instruments w and on the random vector x, where M is some constant integer greater than or equal to 1. Additionally define $q^K(w,x) = (q_1^K(w)', q_0^K(w,x)')'$. Therefore, if Assumptions 3.1 and 3.2 are satisfied for $q^K(s)$ with $s = (w', x')'$, the moment conditions (2.2) together with either (2.3) or (2.4) are equivalent to

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

For the case of (2.3), $q_0^K(w,x)$ depends only on functions of x whereas for (2.4) $q_0^K(w,x)$ additionally depends on functions of both w and x. Therefore the null hypothesis H_0 (3.1) and the alternative hypothesis H_1 (3.2) may be replaced by the equivalent hypotheses (3.4) and (3.3) respectively, $K \to \infty$.

For technical reasons our analysis requires that the number of elements of $q_0^K(w,x)$ should depend linearly on K. Suppose a polynomial of order n_w in the elements of w is used to approximate $E[u(z,\beta_0)|w]$. If w is a r_w-vector, $q_1^K(w)$ will consist of

$$\sum_{i=0}^{n_w} \binom{i + r_w - 1}{i}$$

elements. Then, as in Hirano et al. (2003), we can set $K = (n_w + 1)^{r_w}$ which is an upper bound for this value.

To choose the MK additional functions corresponding to the conditional exogeneity case consider an approximating polynomial of order n_{wx} in the elements of x and w. Then, if x is a r_x-vector, removing all the terms that depend only on w, $q_0^K(w,x)$ consists of

$$\sum_{i=0}^{n_{wx}} \binom{i+r_w+r_x-1}{i} - \sum_{i=0}^{n_{wx}} \binom{i+r_w-1}{i}$$

elements.

Alternatively, consider a polynomial of order n_{wx} in the elements of x and w and remove all those terms that depend solely on w. Based on upper bounds for the elements of $q_1^K(w)$ and $q^K(w,x)$, $MK = (n_{wx}+1)^{r_w+r_x} - (n_{wx}+1)^{r_w}$. Hence, M is given by the next integer greater than

$$\frac{(n_{wx}+1)^{r_w+r_x} - (n_{wx}+1)^{r_w}}{(n_w+1)^{r_w}} \tag{3.5}$$

Let $n_{wx} = An_w^\gamma$, say, for some positive constants A and γ. Then, if $\gamma = r_w/(r_w+r_x)$, (3.5) converges to $A^{r_w+r_x}$ as $n_w \to \infty$. Therefore M is the next integer greater than $A^{r_w+r_x}$ given some A. If n_{wx} is chosen as the closest integer to $An_w^{r_w/(r_w+r_x)}$, an application of Lorenz (1986, Theorem 8, p.90) shows that Assumption (3.1) is satisfied for polynomials in x and w.

For marginal exogeneity, consideration of a polynomial of order n_x in x leads to $MK = (n_x+1)^{r_x}$ additional elements in $q^K(w,x)$. Thus, M is given by the next integer greater than

$$\frac{(n_x+1)^{r_x}}{(n_w+1)^{r_w}} \tag{3.6}$$

which is required to be bounded, e.g., if $n_x = An_w^\gamma$ similarly to above, $\gamma = r_w/r_x$.

3.3 Basic Assumptions and Notation

The following conditions are needed 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 not necessary to assume that $E\left[\left\|\sup_{\beta\in\mathcal{B}}\|u(z,\beta)\|^{\gamma}\right\|\right] < \infty$ for some $\gamma > 2$. As noted in Guggenberger and Smith (2005), if the sample data is i.i.d. one can set $\gamma = 2$ as in Assumption 3.3 by appeal to 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 to 2 in those succeeding Lemmata and Theorems in DIN which concern GEL. Notice that we only require root-n consistency for the estimator $\hat{\beta}$, not efficiency. Moreover, since under the null hypothesis $E[u(z,\beta_0)|s] = E[u(z,\beta_0)|w] = 0$, for $s = x$ or $s = (w', x')$, only a single estimator is needed for β_0.

Let $u_\beta(z,\beta) = \partial u(z,\beta)/\partial\beta'$, $D(s) = E[u_\beta(z,\beta)|s]$ 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.

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

3.4 Test Statistics

Let $g_i(\beta) = u(z_i,\beta)\otimes q_1^K(w_i)$, $h_i(\beta) = u(z_i,\beta)\otimes q^K(w_i,x_i)$, $(i = 1, ..., n)$. 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 both maintained and null hypotheses take the standard form

$$\mathcal{T}^g_{GMM} = n\hat{g}(\hat{\beta})'\hat{\Omega}^{-1}\hat{g}(\hat{\beta}) \tag{3.7}$$

and

$$\mathcal{T}^h_{GMM} = n\hat{h}(\hat{\beta})'\hat{\Xi}^{-1}\hat{h}(\hat{\beta}) \tag{3.8}$$

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.

A GMM statistic appropriate for testing the null hypothesis (3.1) comprising either the marginal (2.3) or conditional (2.4) exogeneity hypotheses against the maintained hypothesis (3.2) may be based on the difference of GMM criterion function statistics (3.8) and (3.7) 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 (1985)] 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 GMM statistic is therefore defined by

$$\mathcal{J} = \frac{\mathcal{T}^h_{GMM} - \mathcal{T}^g_{GMM} - JMK}{\sqrt{2JMK}}. \tag{3.9}$$

Smith (1997, 2001) proposed a number of alternative test statistics to GMM-based procedures based on GEL for a finite number of additional moment restrictions which may be adapted for the framework considered here.

As in DIN and Newey and Smith (2004) we define $\rho(v)$ to be a function of a scalar v that is concave on its domain, an open interval $\mathcal{V}$ containing zero. Let

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

define the respective GEL criteria 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 $\rho_j = \rho_j(0)$, $(j = 0, 1, 2, ...)$. We normalize $\rho_1 = \rho_2 = -1$ without loss of generality.

Given an estimator $\hat{\beta}$ for β_0 which is root-n consistent under either null or alternative hypotheses, Lagrange multiplier estimators may be computed for both η and λ. Let $\hat{\Lambda}_n(\beta) = \{\lambda : \lambda' g_i(\beta) \in \mathcal{V},\ i = 1, ..., n\}$ and $\tilde{\Delta}_n(\beta) = \{\eta : \eta' h_i(\beta) \in \mathcal{V},\ i = 1, ..., n\}$. The respective Lagrange multiplier estimators are then given by

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

Let $\hat{\eta} = S_g \hat{\lambda}$ where $S_g = I_J \otimes (I_K, 0_{MK})'$ is a $J(M+1)K \times JMK$ matrix. Additionally let $s(z, \beta) = S_0' h(z, \beta)$ where $S_0 = I_J \otimes (0_K, I_{MK})'$ is a $J(M+1)K \times JMK$ selection matrix. Hence, $s(z, \beta) = u(z, \beta) \otimes q_0^K(w, x)$. Define $s_i(\beta) = s(z_i, \beta)$, $(i = 1, ..., n)$.

Similar to the GMM statistic $\mathcal{J}$, a likelihood ratio (LR) form of GEL statistic for testing either the marginal (2.3) or conditional (2.4) exogeneity hypotheses against the maintained hypothesis (3.2) may be based on the difference of GEL criterion function statistics; *viz.*

$$\mathcal{LR} = \frac{2n[\tilde{P}_n(\hat{\beta}, \tilde{\eta}) - \hat{P}_n(\hat{\beta}, \hat{\lambda})] - JMK}{\sqrt{2JMK}}. \tag{3.10}$$

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

$$\begin{aligned} \mathcal{LM} &= \frac{n(\tilde{\eta} - \hat{\eta})'\hat{\Xi}(\tilde{\eta} - \hat{\eta}) - JMK}{\sqrt{2JMK}}, \\ \mathcal{S} &= \frac{\sum_{i=1}^n \rho_1(\hat{\lambda}' g_i(\hat{\beta})) s_i(\hat{\beta})' S_0' \hat{\Xi}^{-1} S_0 \sum_{i=1}^n \rho_1(\hat{\lambda}' g_i(\hat{\beta})) s_i(\hat{\beta})/n - JMK}{\sqrt{2JMK}}, \quad (3.11) \\ \mathcal{W} &= \frac{n\tilde{\eta}' S_0 (S_0' \hat{\Xi}^{-1} S_0)^{-1} S_0' \tilde{\eta} - JMK}{\sqrt{2JMK}}. \quad (3.12) \end{aligned}$$

We will require an additional assumption on $\rho(v)$ 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 rigorous statement of the limiting distribution of the GMM statistic $\mathcal{J}$ (3.9) 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 GMM-based test of the marginal or conditional exogeneity of x against the conditional moment restriction (2.2) given the instruments w of the maintained hypothesis (3.2) 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 GEL-based statistics for the exogeneity hypotheses (2.3) and (2.4) and their relationship to that of the GMM statistic $\mathcal{J}$ (3.9).

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 the GEL statistics* $\mathcal{LR}$*,* $\mathcal{LM}$*,* $\mathcal{S}$ *and* $\mathcal{W}$ *converge in distribution to a standard normal random variate. Moreover all of these statistics are asymptotically equivalent to the GMM statistic* $\mathcal{J}$*.*

Similar to the GMM statistic $\mathcal{J}$ (3.9) 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 the exogeneity hypotheses (2.3) and (2.4) may be also defined which ignore the information contained in the maintained hypothesis (3.2); *viz.* the GEL-based statistics

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

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

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

It is straightforward to show from the analysis used to establish Theorems 4.1 and 4.2 similarly to DIN that these statistics also each converge in distribution to a standard normal random variate and are mutually asymptotically equivalent but not to $\mathcal{J}$, $\mathcal{LR}$,

$\mathcal{LM}$, $\mathcal{S}$ and $\mathcal{W}$. The statistics $\mathcal{LR}^h$ and $\mathcal{S}^h$ are forms of those GMM and GEL statistics suggested in DIN, section 6, pp.67-71, appropriate for testing the unrestricted hypothesis $H_0 \cap H_1$.

5 Asymptotic Local Power

This section considers the asymptotic distribution of the above statistics under a suitable sequence of local alternatives.

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

$$H_{1n} : E[u(z, \beta_{n,0})|w, x] = \frac{\sqrt[4]{JMK}}{\sqrt{n}}\xi(w, x), \tag{5.13}$$

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

This sequence of local alternatives (5.13) is particularly apposite for conditional exogeneity. It is also appropriate as a description of local alternatives to the marginal exogeneity hypothesis (3.1) $E[u(z, \beta_0)|x] = 0$. In this case local alternatives may be described by taking the expectation of (5.13) conditional on x, that is,

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

In order to obtain the asymptotic distribution of the statistics proposed in section 3.4 under the local alternatives (5.13) we invoke the following assumption.

Assumption 5.1 *(a) $\beta_{n,0}$ is a non-stochastic sequence such that (5.13) 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 statistics $\mathcal{J}$, $\mathcal{LR}$, $\mathcal{LM}$, $\mathcal{S}$ and $\mathcal{W}$ under the sequence of local alternatives (5.13). Let $\Sigma(w,x) = \Sigma(w,x;\beta_0)$.

Theorem 5.1 *Let Assumptions 3.1, 3.2, 3.3, 3.4 and 5.1 hold for $s = w$ or $s = (w', x')'$ 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 $\mathcal{J}$.

Although not discussed here, one might use a similar analysis to that underpinning Lemma 6.5, p.71, in DIN to demonstrate the consistency of the statistics $\mathcal{J}$, $\mathcal{LR}$, $\mathcal{LM}$, $\mathcal{S}$ and $\mathcal{W}$.

The following corollary is a special case of the previous Theorem and presents the asymptotic distribution of $\mathcal{LR}^h$, $\mathcal{LM}^h$ and $\mathcal{S}^h$ under the same local alternative.

Corollary 5.1 *Let Assumptions 3.1, 3.2, 3.3, 3.4 and 5.1 hold for $s = (w', x')'$ and $\zeta(K)^2K^2/n \to 0$. 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 $\zeta(K)^2K^3/n \to 0$, then $\mathcal{LR}^h$, $\mathcal{LM}^h$ are asymptotically equivalent to $\mathcal{S}^h$.

Hence, for large M, μ_h will differ little from μ. Tests based on unrestricted statistics that ignore the maintained hypothesis H_1 will have a similar discriminatory power to detect the local departures from the null hypothesis H_0 as tests that incorporate H_1. Therefore, M should be chosen as small as possible.

6 Conclusions

This article discusses definitions of exogeneity for models given by a set of semiparametric conditional moment restrictions based on a vector of moment indicators. We assume that there exists an initial set of identifying instruments. A random vector is said to be marginally exogenous for the parameters specifying the conditional moment restrictions if the conditional expectation of the moment indicator vector given the random vector is zero. This definition accords with that proposed by Blundell and Horowitz (2004). We suggest that this definition may prove to be inadequate in particular circumstances. We therefore define an alternative exogeneity concept which we term conditional exogeneity. A random vector is said to be conditionally exogenous for the parameters specifying the conditional moment restrictions if the conditional expectation of the moment indicator vector given both instruments and the random vector is zero. This definition is more closely related to those for fully parametric models based on classical likelihood theory.

The paper also provides GMM- and GEL-based test statistics for both marginal and conditional exogeneity by reinterpreting the respective hypotheses as concerning an infinite number of unconditional moment restrictions. These tests may therefore be viewed as tests for additional sets of infinite moment restrictions. We derive the limiting distribution of these test statistics under the null hypotheses of marginal and conditional exogeneity and a suitable sequence of local alternatives.

Appendix: Proofs

Throughout the Appendix, C will denote a generic positive constant that may be different in different uses, and CS, M, T and c_r the Cauchy-Schwarz, Markov, triangle and Loève c_r inequalities respectively.[3] 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* $\beta \in \mathcal{B}$.

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

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 $1 \le i \le n$. Hence, it follows from Owen (1990, Lemma 3, p.98) that $\max_{1\le i\le n} b_i = o_p(n^{1/2})$. Therefore, by CS

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

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

The next Lemma is used in the proofs for asymptotic normality of the test statistics under both null and local alternatives.

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

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

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

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

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

To prove the asymptotic normality of T_1 we verify the hypotheses of Hall (1984, Theorem 1, pp.3-4). Let $g_{u,n} \equiv g(u,\beta_{0,n})$ and 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 $E[H_n(z_1,z_2)|z_1] = \sqrt{\frac{2}{n^2k}} g'_{z_i,n} C_n E[g_{z_2,n}] = 0$ and

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

On the other hand

$$\frac{E[H_n(z_1,z_2)^4]}{nE[H_n(z_1,z_2)^2]^2} = \frac{4}{n^5k^2} \frac{E[(g'_{z_1,n} C_n g_{z_2,n})^4]}{4/n^4}.$$

As $C_n = C_n \Omega_n C_n$ and Ω_n is positive definite, by CS

$$\frac{4}{n^5k^2} \frac{E[(g'_{z_1,n} C_n g_{z_2,n})^4]}{4/n^4} \leq \frac{1}{nk^2} E[(g'_{z_1,n} C_n g_{z_1,n})^2 (g'_{z_2,n} C_n g_{z_2,n})^2]$$
$$= (\frac{1}{\sqrt{n}k} E[(g'_{z_1,n} C_n g_{z_1,n})^2])^2 \to 0.$$

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. ■

The next Lemma mirrors DIN Lemma A3.

Lemma A.3 *Let $a_{i,n} = a_n(z_i)$, $\bar{a}_{i,n} = E[a_{i,n}|x_i]$, $a_i = a(z_i)$, $\bar{a}_i = E[a_i|x_i]$, $U_{i,n} = U_n(x_i)$ and $U_i = U(x_i)$. If Assumption (3.1) is satisfied, (i) $E[\|a_{i,n}\|^2 |x_i]$ is bounded for large enough n, (ii) $U_{i,n}$ is a $r \times r$ p.d. matrix that is bounded and has smallest eigenvalue bounded away from zero for large enough n, (iii) U_i is $r \times r$ p.d. matrix that is bounded and has smallest eigenvalue bounded away from zero, (iv) $E[\left\|U_{i,n}^{-1} - U_i^{-1}\right\|^2] \to 0$, (v) $E[\|\bar{a}_{i,n} - \bar{a}_i\|^2] \to 0$, (vi) $K \to \infty$ 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 A3. Let $F_{i,n} = U_{i,n}^{1/2}$, $P_{i,n} = F_{i,n} \otimes q_i$, $P_n = (P'_{1,n}, ..., P'_{n,n})'$, $A_{i,n} = F_{i,n}^{-1} a_{i,n}$, $A_n = (A'_{1,n}, ..., A'_{n,n})'$, $\bar{A}_{i,n} = E[A_{i,n}|x_i] = F_{i,n}^{-1} \bar{a}_{i,n}$ and $\bar{A}_n = E[A_n] = (\bar{A}'_{1,n}, ..., \bar{A}'_{n,n})'$. 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 $x = (x_1, ..., x_n)$. As the data are i.i.d. and by (ii)

$$\begin{aligned} E[\left(A_n - \bar{A}_n\right)\left(A_n - \bar{A}_n\right)' |x] &= diag(F_{1,n}^{-1} var[a_{1,n}|x_1] F_{1,n}^{-1}, ..., F_{n,n}^{-1} var[a_{n,n}|x_n] F_{n,n}^{-1}) \\ &\leq CI \end{aligned}$$

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

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

using (ii) and (vi). Thus $T_A \xrightarrow{p} 0$ by M.

Now by c_r

$$\begin{aligned} E[\left\|U_{i,n}^{-1}\bar{a}_{i,n} - \Gamma_K q_i\right\|^2] &= E[\left\|(U_{i,n}^{-1} - U_i^{-1})\bar{a}_{i,n} + U_i^{-1}(\bar{a}_{i,n} - \bar{a}_i) + U_i^{-1}\bar{a}_i - \Gamma_K q_i\right\|^2] \\ &\leq 3\left[E[\left\|(U_{i,n}^{-1} - U_i^{-1})\bar{a}_{i,n}\right\|^2] + E[\left\|U_i^{-1}(\bar{a}_{i,n} - \bar{a}_i)\right\|^2]\right. \\ &\quad \left. + E[\left\|U_i^{-1}\bar{a}_i - \Gamma_K q_i\right\|^2]\right] \to 0. \end{aligned}$$

This follows as from Assumption 1 $E[\left\|U_i^{-1}\bar{a}_i - \Gamma_K q_i\right\|^2] \to 0$. For the first term, since $\bar{a}_{i,n}$ is bounded for large enough n, by (iv) $E[\left\|(U_{i,n}^{-1} - U_i^{-1})\bar{a}_{i,n}\right\|^2] \to 0$ and for the second term $E[\left\|U_i^{-1}(\bar{a}_{i,n} - \bar{a}_i)\right\|^2] \to 0$ by (v) as U_i^{-1} is bounded from (iii). Then, for $\tilde{\gamma}_K = vec\,(\Gamma_K')$, by M

$$\begin{aligned} \left\|\bar{A}_n - P_n\tilde{\gamma}_K\right\|^2/n &= \sum_i \left\|F_{i,n}^{-1}\bar{a}_i - (F_{i,n} \otimes q_i')\tilde{\gamma}_K\right\|^2/n \\ &= \sum_i \|F_{i,n}\|^2 \left\|U_{i,n}^{-1}\bar{a}_i - (I \otimes q_i')\tilde{\gamma}_K\right\|^2/n \\ &= \sum_i \|F_{i,n}\|^2 \left\|U_{i,n}^{-1}\bar{a}_i - \Gamma_k q_i\right\|^2/n \\ &\leq C\sum_i \left\|U_{i,n}^{-1}\bar{a}_i - \Gamma_k q_i\right\|^2/n \xrightarrow{p} 0. \end{aligned}$$

Now

$$\begin{aligned} \left|A_n' Q_n A_n/n - \bar{A}_n'\bar{A}_n/n\right| &= \left|\left(A_n - \bar{A}_n\right)' Q_n \left(A_n - \bar{A}_n\right)/n\right. \\ &\quad \left. + 2\bar{A}_n' Q_n \left(A_n - \bar{A}_n\right)/n - \bar{A}_n'\left(I - Q_n\right)\bar{A}_n/n\right|. \end{aligned}$$

Notice that

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

Also by M $\bar{A}_n'\bar{A}_n/n = O_p(1)$. By T and CS

$$\left|A_n'Q_nA_n/n - \bar{A}_n'\bar{A}_n/n\right| \leq T_A + 2\sqrt{T_A}\sqrt{\bar{A}'\bar{A}/n} + \bar{T}_A \xrightarrow{p} 0.$$

To examine the large sample behaviour of $\bar{A}_n'\bar{A}_n/n = \sum_i \bar{a}_{i,n}U_{i,n}^{-1}\bar{a}_{i,n}/n$, 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, Chapter IX, section 9, 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 Chebyshev

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

For large enough n, as $U_{i,n}$ is p.d. and bounded from (ii) and as $\bar{a}_{i,n}$ is likewise from (i), $E[\left|\bar{a}_{i,n}'U_{i,n}^{-1}\bar{a}_{i,n}\right|^2]$ is also bounded. Therefore $n\mathcal{P}\{\left|\bar{a}_{i,n}'U_{i,n}^{-1}\bar{a}_{i,n}\right|/n > \eta\} \to 0$. Secondly, consider

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

However,

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

By CS and Jensen, since $\|\bar{a}_i\|$ and $\left\|U_{i,n}^{-1}\right\|$ are bounded from (i) and (ii), by (iv) and (v)

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

■

The following Lemma is similar to DIN Lemma A4.

Lemma A.4 *If Assumption 3.1 is satisfied, (i)* $\varepsilon_{i,n}$ *and* Y_i *are* $r \times 1$ *random vectors with* $E[\varepsilon_{i,n}|x_i] = 0$ *and* $E[\|\varepsilon_{i,n}\|^4 |x_i] \leq C$ *for large enough* n *and* $E[\|Y_i\|^2 |x_i] \leq C$*, (ii)* $U_{i,n} = U_n(x_i)$ *is* $r \times r$ *p.d. matrix that is bounded and has the smallest eigenvalue bounded away from zero for* n *large enough, (iii)* $U_i = U(x_i)$ *is* $r \times r$ *p.d. matrix that is bounded and has the smallest eigenvalue bounded away from zero, (iv)* $E[\left\|U_{i,n}^{-1} - U_i^{-1}\right\|^2] \to 0$ *and (v)* $K \to \infty$ *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

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

and secondly that

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

The proof structure of the first part is similar to that of DIN Lemma A4. Let $F_{i,n}$, P_n and thus Q_n be specified as in the proof of Lemma A.3, $A_{i,n} = F_{i,n}^{-1} Y_i$, $\bar{A}_{i,n} = E[A_{i,n}|x_i] = F_{i,n}^{-1} E[Y_i|x_i]$, $A_n = (A_{1,n}', ..., A_{n,n}')$, $\bar{A}_n = (\bar{A}_{1,n}', ..., \bar{A}_{n,n}')'$, $B_{i,n} = F_{i,n}^{-1} \varepsilon_{i,n}$ and $B_n = (B_{1,n}', ..., B_{n,n}')'$. By assumption $E[B_{i,n}|x_i] = 0$ and, consequently,

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

From the proof of Lemma A.3 $(A_n - \bar{A}_n)' Q_n (A_n - \bar{A}_n) = O_p(K)$ and $B_n' Q_n B_n = O_p(K)$, the latter holding by (i) as $E[\|\varepsilon_{i,n}\|^2 |x_i] \leq C$ for large enough n. Thus, for large enough n, by CS

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

Also, as in the proof of Lemma A.3, $E[\bar{A}_n'(I-Q_n)\bar{A}_n/n] \to 0$. Thus, by iterated expectations,

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

using $E\left[B_nB_n'|x\right]$ is bounded for large enough n by (i) and (ii). The conclusion follows by M and T.

It remains to prove $(A.1)$. We use the Corollary of section 1.9.3 of Serfling(2002, p.32) to prove this result.We only need to show that

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

where $b_n^2 = Var\left[E[Y_i|x_i]'U_{in}^{-1}\varepsilon_{i,n}\right]$. Now notice that by CS

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

as $E[\|\varepsilon_{i,n}\|^4|x_i]$, $E[Y_i|x_i]$ and $U_{i,n}^{-1}$ are bounded for n large enough. Also by Jensen inequality

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

Hence $(A.2)$ follows. ■

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

Let $u_i(\beta) = u(z_i,\beta)$, $q_i = q(s_i)$, $g_i(\beta) = u_i(\beta)\otimes q_i$, $\hat{g}_i = g_i(\hat{\beta})$ 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.5 *If Assumptions 3.2, 3.3 and 3.4 hold and* $\hat{\beta}-\beta_{0,n} = O(\tau_n)$ *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 A6 of DIN. However, rather than some expectations being bounded as in DIN, here they are bounded for n large enough.

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

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

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

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

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

For the third conclusion as in DIN

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

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.6 *If Assumptions 3.2 and 3.4 hold 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 A7. 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 A2

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

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 A7, 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.7 *Let Assumptions 3.1, 3.2, 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.5 and A.6

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

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.3 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.6. 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.5. 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),

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

Next, by CS and T,

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

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

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

It remains to prove that

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

From Lemma A.5,

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

■

A.2 Asymptotic Null Distribution

Proof of Theorem 4.1: By DIN Lemma A6 and $\zeta(K)^2 K^2/n \to 0$,

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

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 DIN Lemma A7 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.7. 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

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

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

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

$$\frac{n\hat{h}(\beta_0)'\Xi^{-1}\hat{h}(\beta_0) - n\hat{g}(\beta_0)'\Omega^{-1}\hat{g}(\beta_0) - JMK}{\sqrt{2JMK}} = \frac{n\hat{h}(\beta_0)'(\Xi^{-1} - S_g\Omega^{-1}S_g')\hat{h}(\beta_0) - JMK}{\sqrt{2JMK}}.$$

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

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

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 $\mathcal{LR}$ (3.10).

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

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

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

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.

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

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 $\lambda_{\max}[(2\dot{\Omega}-\ddot{\Omega})^{-1}] \le C$ w.p.a.1.. Thus, by T, as $\left(2\dot{\Omega}^{-1}-\dot{\Omega}^{-1}\ddot{\Omega}\dot{\Omega}^{-1}\right)^{-1} = \dot{\Omega}(2\dot{\Omega}-\ddot{\Omega})^{-1}\dot{\Omega}$,

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

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

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

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 3.8 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 $S_g = I_J\otimes(I_K, 0_{MK})'$. Hence, $S_g'\hat{h}=\hat{g}$ and $S_g'\Xi S_g=\Omega$. Also write $\hat{\eta}=S_g\hat{\lambda}$. Thus, $\hat{\eta}=S_g\hat{\lambda}=-S_g\dot{\Omega}^{-1}\hat{g}=-S_g\dot{\Omega}^{-1}S_g'\hat{h}$.

Now

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

and therefore

$$
\begin{aligned}
\mathcal{LM} - \frac{\mathcal{T}^h_{GMM} - \mathcal{T}^g_{GMM} - JMK}{\sqrt{2JMK}} &= \frac{n\hat{h}'(\dot{\Xi}^{-1}\hat{\Xi}\dot{\Xi}^{-1} - \hat{\Xi}^{-1})\hat{h}}{\sqrt{2JMK}} \\
&\quad + \frac{n\hat{h}'(S_g\hat{\Omega}^{-1}S_g' - 2\dot{\Xi}^{-1}\hat{\Xi}\ S_g\dot{\Omega}^{-1}S_g'\hat{h} + S_g\dot{\Omega}^{-1}S_g'\hat{\Xi}\ S_g\dot{\Omega}^{-1}S_g')\hat{h}}{\sqrt{2JMK}}.
\end{aligned}
$$

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

By CS, the first term

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

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

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

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

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

Therefore, by T $\left\|\hat{\Xi} - \dot{\Xi}\hat{\Xi}^{-1}\dot{\Xi}\right\| = o_p(1/\sqrt{K})$. As $\left\|\tilde{\eta}\right\| = O_p(\sqrt{K/n})$ 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

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

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

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

Next by a similar argument

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

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

The score test statistic

$$\mathcal{S} = \frac{\sum_{i=1}^n \rho_1(\hat{\lambda}' \hat{g}_i) \hat{s}_i' S_0' \hat{\Xi}^{-1} S_0 \sum_{i=1}^n \rho_1(\hat{\lambda}' \hat{g}_i) \hat{s}_i / n - JMK}{\sqrt{2JMK}},$$

where $\hat{s}_i = s_i(\hat{\beta})$, $(i = 1, ..., n)$ and $S_0 = I_J \otimes (0_K, I_{MK})'$. Expanding the first order conditions $\sum_{i=1}^n \rho_1(\hat{h}_i' \tilde{\eta}) \hat{h}_i / n = 0$ of (3.10) 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 $\sum_{i=1}^n \rho_1(\hat{h}_i' \hat{\eta}) \hat{h}_i / n = S_0 \sum_{i=1}^n \rho_1(\hat{g}_i' \hat{\lambda}) \hat{s}_i$,

$$\sum\nolimits_{i=1}^n \rho_1(\hat{\lambda}' \hat{g}_i) \hat{s}_i' S_0' \hat{\Xi}^{-1} S_0 \sum\nolimits_{i=1}^n \rho_1(\hat{\lambda}' \hat{g}_i) \hat{s}_i / n = n \left(\tilde{\eta} - \hat{\eta} \right)' \dot{\Xi} \hat{\Xi}^{-1} \dot{\Xi} \left(\tilde{\eta} - \hat{\eta} \right).$$

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

$$\tilde{\eta} - \hat{\eta} = \dot{\Xi}^{-1} S_0 \sum_{i=1}^n \rho_1(\hat{g}_i' \hat{\lambda}) \hat{s}_i / n$$

and thus

$$S_0'\tilde{\eta} = S_0'\dot{\Xi}^{-1}S_0 \sum_{i=1}^{n} \rho_1(\hat{g}_i'\hat{\lambda})\hat{s}_i/n.$$

Therefore, w.p.a.1

$$\begin{aligned} |\mathcal{S} - \mathcal{W}| &= n\left|\tilde{\eta}'S_0((S_0'\dot{\Xi}^{-1}S_0)^{-1}S_0'\hat{\Xi}^{-1}S_0(S_0'\dot{\Xi}^{-1}S_0)^{-1} - S_0'\hat{\Xi}^{-1}S_0)S_0'\tilde{\eta}\right|/\sqrt{2JMK} \\ &\leq n\left\|S_0'\tilde{\eta}\right\|^2 \left\|(S_0'\dot{\Xi}^{-1}S_0)^{-1}S_0'\hat{\Xi}^{-1}S_0(S_0'\dot{\Xi}^{-1}S_0)^{-1} - S_0'\hat{\Xi}^{-1}S_0\right\|/\sqrt{2JMK}. \end{aligned}$$

Since $\|S_0'\tilde{\eta}\| = O_p(\sqrt{K/n})$ by DIN Lemma A.11 and by a similar argument to that which showed $\dot{\Xi}\hat{\Xi}^{-1}\dot{\Xi} - \hat{\Xi} = o_p(1/\sqrt{K})$, $(S_0'\dot{\Xi}^{-1}S_0)^{-1}S_0'\hat{\Xi}^{-1}S_0(S_0'\dot{\Xi}^{-1}S_0)^{-1} - S_0'\hat{\Xi}^{-1}S_0 = o_p(1/\sqrt{K})$. Therefore, $|\mathcal{S} - \mathcal{W}| = o_p(1)$. ■

A.3 Asymptotic Local Alternative Distribution

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

Let $\hat{g}_n = \hat{g}(\beta_{n,0})$ and $\hat{h}_n = \hat{h}(\beta_{n,0})$. Then, by Lemma A.7,

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

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

Therefore it remains to prove that

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

We first consider the local alternative distribution under the conditional exogeneity 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 $P_n = \Xi_n^{-1} - S_g\Omega_n^{-1}S_g'$. 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.3 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.3,

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

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

It therefore remains to show that

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

Similarly to the proof of Lemma 6.1 in DIN, from Lemma A.5,

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

since $\left\|\bar{\Xi}_n^{-1}\bar{h}_n\right\|^2 = \bar{h}_n'\bar{\Xi}_n^{-2}\bar{h}_n \le C\bar{h}_n'\bar{\Xi}_n^{-1}\bar{h}_n = O_p(\sqrt{K}/n)$. Likewise $\left|n\bar{h}_n' S_g(\Xi_n^{-1} - \bar{\Xi}_n^{-1})S_g'\bar{h}_n\right|/\sqrt{2JMK} = o_p(1)$. 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, notice $\|\xi_i\|^2$ is bounded and that $\Sigma_{i,n}(s_i)^{-1}$ is bounded for n large enough. In addition by c_r

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

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

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

Next, by hypothesis,

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

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

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

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

Again using c_r

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

Now, for n large enough, $E[(h_{i,n}'\Xi_n^{-1}h_{i,n})^2] \le CE[\|h_{i,n}\|^4]$. 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,

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

Next,

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

Lastly,

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

Hence, $E[(\tilde{h}'_{i,n}\Xi_n^{-1}\tilde{h}_{i,n})^2] = o_p(K\sqrt{n})$ as required. Likewise, $E[(\tilde{h}'_{i,n}S_g\Omega_n^{-1}S'_g\tilde{h}_{i,n})^2] = o_p(K\sqrt{n})$. 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. ■

A References

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., and J.L. Horowitz (2004): "A Nonparametric Test of Exogeneity." CWP 15/04, cemmap, I.F.S. and U.C.L. http://cemmap.ifs.org.uk/wps/cwp1504.pdf

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

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.

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

Delgado, M.A., Dominguez, M.A., and P. Lavergne (2005): "Consistent Tests of Conditional Moment Restrictions." Forthcoming in *Annales d'Economie et de Statistique.*

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.

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. and Ridder, G. (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.

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.

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

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

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.

Powell, M.J.D. (1981): *Approximation Theory and Methods.* Cambridge: Cambridge 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 *Applications of Differential Geometry to Econometrics*, eds. M. Salmon and P. Marriott, 119-150. Cambridge: Cambridge University Press.

Smith, R. J. (2001): "GEL Methods for Moment Condition Models," working paper, University of Bristol. Revised version CWP 19/04, c*e*mmap, I.F.S. and U.C.L. http://cemmap.ifs.org.uk/wps/cwp0419.pdf

Smith, R.J. (2005a): "Efficient Information Theoretic Inference for Conditional Moment Restrictions." Forthcoming *Journal of Econometrics*. CWP 14/05, c*e*mmap, I.F.S. and U.C.L. http://cemmap.ifs.org.uk/wps/cwp1405.pdf

Smith, R.J. (2005b): "Local GEL Methods for Conditional Moment Restrictions." CWP 15/05, c*e*mmap, I.F.S. and U.C.L. http://cemmap.ifs.org.uk/wps/cwp1505.pdf

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

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. (2002): *Econometric Analysis of Cross Section and Panel Data*. Cambridge: The MIT Press.

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

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

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