

Hahn, Jinyong; Newey, Whitney K.; Smith, Richard J.

**Working Paper**

## Tests for neglected heterogeneity in moment condition models

cemmap working paper, No. CWP26/11

**Provided in Cooperation with:**

The Institute for Fiscal Studies (IFS), London

*Suggested Citation:* Hahn, Jinyong; Newey, Whitney K.; Smith, Richard J. (2011) : Tests for neglected heterogeneity in moment condition models, cemmap working paper, No. CWP26/11, Centre for Microdata Methods and Practice (cemmap), London, <https://doi.org/10.1920/wp.cem.2011.2611>

This Version is available at:

<https://hdl.handle.net/10419/64703>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Tests for neglected heterogeneity in moment condition models

---

Jinyong Hahn  
Whitney K. Newey  
Richard J. Smith

The Institute for Fiscal Studies  
Department of Economics, UCL

**cemmap** working paper CWP26/11

# TESTS FOR NEGLECTED HETEROGENEITY IN MOMENT CONDITION MODELS\*

Jinyong Hahn  
Department of Economics  
U.C.L.A.

Whitney K. Newey  
Department of Economics  
M.I.T.

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

Preliminary  
This Draft: May 2011

## Abstract

The central concern of the paper is with the formulation of tests of neglected parameter heterogeneity appropriate for model environments specified by a number of unconditional or conditional moment conditions. We initially consider the unconditional moment restrictions framework. Optimal  $m$ -tests against moment condition parameter heterogeneity are derived with the relevant Jacobian matrix obtained as the second order derivative of the moment indicator in a leading case. GMM and GEL tests of specification based on generalized information matrix equalities appropriate for moment-based models are described and their relation to the optimal  $m$ -tests against moment condition parameter heterogeneity examined. A fundamental and important difference is noted between GMM and GEL constructions. The paper is concluded by a generalization of these tests to the conditional moment context.

---

\*Earlier versions of this paper were presented at the University of Nottingham and the Victoria University of Wellington and as a Keynote address by the third author at TEST 2010 International Symposium on Econometrics of Specification Tests in 30 Years held at WISE, Xiamen University. We are grateful to participants for their comments. Part of the research for the current version was undertaken while the third author was a Visiting Cambridge Fellow at the University of Canterbury and during research visits at the University of Otago and the Reserve Bank of New Zealand. He gratefully acknowledges the hospitality and support of these institutions.

**JEL Classification:** C13, C30

**Keywords:** GMM, GEL, Unconditional Moments, Conditional Moments, Score and LM Tests, Information Matrix Equality.

# 1 Introduction

For econometric estimation with cross-section and panel data the possibility of individual economic agent heterogeneity is a major concern. In particular, when parameters represent agent preferences investigators may wish to entertain the possibility that parameter values might vary across observational economic units. Although it may in practice be difficult to control for such parameter heterogeneity, the formulation and conduct of tests for parameter heterogeneity are often relatively straightforward. Indeed, in the classical parametric likelihood context, Chesher (1984) demonstrates that the well-known information matrix (IM) test due to White (1982) can be interpreted as a test against random parameter variation. In particular, the White (1980) test for heteroskedasticity in the classical linear regression model is a test for random variation in the regression coefficients. Such tests often provide useful ways of checking for unobserved individual heterogeneity.

The central concern of this paper is the development of optimal  $m$ -tests for parameter heterogeneity in models specified by moment conditions. We consider both unconditional and conditional model frameworks. Based on the results in Newey (1985a), to formulate an optimal  $m$ -test we find the linear combination of moment functions with maximal noncentrality parameter in the limiting noncentral chi-square distribution of a class of  $m$ -statistics under a local random parameter alternative. In a leading case, the optimal linear combination has a simple form, being expressed in terms of the second derivative of the moments with respect to those parameters that are considered possibly to be random, multiplied by the optimal weighting matrix. Thus, the moment conditions themselves provide all that is needed for the construction of test statistics for parameter heterogeneity.

We also consider generalized IM equalities associated with generalized method of moments (GMM) and generalized empirical likelihood (GEL) estimation. The GMM-based version of the generalized IM test statistic is identical to the optimal  $m$ -statistic employing the second derivative of the moments described above. The GEL form is

associated with a more general form of parameter heterogeneity test involving additional components that may be interpreted in terms of correlations between the sample Jacobian and the random variable driving potential parameter heterogeneity.

To provide a background for the subsequent discussion section 2 reconsiders the IM test of White (1982) and its interpretation as a test for parameter heterogeneity in Chesher (1984). We then consider the effect of parameter heterogeneity on the moment conditions in section 3 and derive the optimal linear combination to be used in constructing the tests in a leading case when the sample Jacobian is uncorrelated with the random heterogeneity variate. We give alternative Lagrange multiplier and score forms of the optimal  $m$ -statistic that, using the results of Newey (1985a), maximize local power. Section 4 of the paper provides moment specification tests obtained by consideration of generalized forms of the IM equality appropriate for GMM and GEL estimation. These statistics are then compared with those for moment condition parameter heterogeneity developed in section 3. The GMM form coincides with that of section 3 whereas the GEL statistic incorporates additional terms that implicitly allow for particular forms of correlation between the sample Jacobian and the random variate potentially driving parameter heterogeneity. These results are illustrated by consideration of empirical likelihood, a special case of GEL that allows a direct application of the classical likelihood-based approach to IM test construction discussed in section 2. The results of earlier sections are then extended in section 5 to deal with models specified in terms of conditional moment conditions. The Appendices contain relevant assumptions and proofs of results and assertions made in the main text.

Throughout the text  $(x_i, z_i)$ ,  $(i = 1, \dots, n)$ , will denote i.i.d. observations on the observable  $s$ -dimensional covariate or instrument vector  $x$  and the  $d$ -dimensional vector  $z$  that may include a sub-vector of  $x$ . The vector  $\beta$  denotes the parameters of interest with  $\mathcal{B}$  the relevant parameter space. Positive (semi-) definite is denoted as p.(s.)d. and f.c.r. is full column rank. Superscripted vectors denote the requisite element, e.g.,  $a^j$  is the  $j$ th element of vector  $a$ . UWL will denote a uniform weak law of large numbers such as Lemma 2.4 of Newey and McFadden (1994), and CLT will refer to the Lindeberg-Lévy

central limit theorem. “ $\xrightarrow{p}$ ” and “ $\xrightarrow{d}$ ” are respectively convergence in probability and distribution.

## 2 The Classical Information Matrix Test

We first consider the classical fully parametric likelihood context and briefly review the information matrix (IM) test initially proposed in the seminal paper White (1982). See, in particular, White (1982, section 4, pp. 9-12). The interpretation presented in Chesher (1984) of the IM test as a Lagrange multiplier (LM) or score test for neglected (parameter) heterogeneity is then discussed.

For the purposes of this section it is assumed that  $z$  has (conditional) distribution function  $F(\cdot, \beta)$  given covariates  $x$  known up to the  $p \times 1$  parameter vector  $\beta \in \mathcal{B}$ . We omit the covariates  $x$  from the exposition where there is no possibility of confusion. Suppose also that  $F(\cdot, \beta)$  possesses Radon-Nikodým conditional density  $f(z, \beta) = \partial F(z, \beta) / \partial v$  and that the density  $f(z, \beta)$  is twice continuously differentiable in  $\beta \in \mathcal{B}$ .

### 2.1 ML Estimation

The ML estimator  $\hat{\beta}_{ML}$  is defined by

$$\hat{\beta}_{ML} = \arg \max_{\beta \in \mathcal{B}} \frac{1}{n} \sum_{i=1}^n \log f(z_i, \beta).$$

Let  $\beta_0 \in \mathcal{B}$  denote the true value of  $\beta$  and  $E_0[\cdot]$  denote expectation taken with respect to  $f(z, \beta_0)$ . The IM  $\mathcal{I}(\beta_0)$  is then defined by  $\mathcal{I}(\beta_0) = -E_0[\partial^2 \log f(z, \beta_0) / \partial \beta \partial \beta']$ , its inverse defining the classical Cramér-Rao efficiency lower bound. Under standard regularity conditions, see, e.g., Newey and McFadden (1994),  $\hat{\beta}_{ML}$  is a root- $n$  consistent estimator of  $\beta_0$  with limiting representation

$$\sqrt{n}(\hat{\beta}_{ML} - \beta_0) = -\mathcal{I}(\beta_0)^{-1} n^{-1/2} \sum_{i=1}^n \partial \log f(z_i, \beta_0) / \partial \beta + O_p(n^{-1/2}). \quad (2.1)$$

Consequently the ML estimator  $\hat{\beta}_{ML}$  has an asymptotic normal distribution described by

$$\sqrt{n}(\hat{\beta}_{ML} - \beta_0) \xrightarrow{d} N(0, \mathcal{I}(\beta_0)^{-1}).$$

## 2.2 IM Equality and IM Specification Test

With  $E[\cdot]$  as expectation taken with respect to  $f(z, \beta)$ , twice differentiation of the identity  $E[1] = 1$  with respect to  $\beta$  demonstrates that the density function  $f(z, \cdot)$  obeys the familiar IM *equality*

$$\begin{aligned} E\left[\frac{1}{f(z, \beta)} \frac{\partial^2 f(z, \beta)}{\partial \beta \partial \beta'}\right] &= E\left[\frac{\partial^2 \log f(z, \beta)}{\partial \beta \partial \beta'}\right] + E\left[\frac{\partial \log f(z, \beta)}{\partial \beta} \frac{\partial \log f(z, \beta)}{\partial \beta'}\right] \\ &= 0. \end{aligned}$$

Therefore, under correct specification, i.e.,  $z$  distributed with density function  $f(z, \beta_0)$ , and given the consistency of  $\hat{\beta}_{ML}$  for  $\beta_0$ , by an i.i.d. UWL, the contrast with zero

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \frac{1}{f(z_i, \hat{\beta}_{ML})} \frac{\partial^2 f(z_i, \hat{\beta}_{ML})}{\partial \beta \partial \beta'} &= \frac{1}{n} \sum_{i=1}^n \left[ \frac{\partial^2 \log f(z_i, \hat{\beta}_{ML})}{\partial \beta \partial \beta'} \right. \\ &\quad \left. + \frac{\partial \log f(z_i, \hat{\beta}_{ML})}{\partial \beta} \frac{\partial \log f(z_i, \hat{\beta}_{ML})}{\partial \beta'} \right] \end{aligned}$$

consistently estimates a  $p \times p$  matrix of zeroes. The IM test of White (1982) is a (conditional) moment test [Newey (1985b)] for correct specification based on selected elements of the re-scaled moment vector<sup>1</sup>

$$n^{1/2} \sum_{i=1}^n \frac{1}{f(z_i, \hat{\beta}_{ML})} \text{vec}\left(\frac{\partial^2 f(z_i, \hat{\beta}_{ML})}{\partial \beta \partial \beta'}\right)/n. \quad (2.2)$$

## 2.3 Neglected Heterogeneity

The IM test may also be interpreted as a test for neglected heterogeneity; see Chesher (1984). To see this we now regard  $\beta$  as a random vector and the density  $f(z, \beta)$  as the *conditional* density of  $z$  given  $\beta$ . Absence of parameter heterogeneity corresponds to  $\beta = \beta_0$  almost surely.

Suppose that the marginal density of  $\beta$  is  $\eta^{-p/2} h((\beta - \beta_0)'(\beta - \beta_0)/\eta)$  where  $\eta \geq 0$  is a non-negative scalar, this density being a location-scale generalisation of the spherically symmetric class [Kelker (1970)]. Given the symmetry of  $h(\cdot)$  in  $\beta$ ,  $E[\beta] = \beta_0$ . Equivalently, writing  $\beta = \beta_0 + \eta^{1/2} w$ ,  $w$  has the symmetric continuous density  $h(w'w)$ . Thus,

<sup>1</sup>Apart from symmetry, in some cases there may be a linear dependence and, thus, a redundancy between the elements of  $\partial^2 f(z, \beta)/\partial \beta \partial \beta'$ , in particular, those associated with parametric models based on the normal distribution, e.g., linear regression, Probit and Tobit models.



likewise,  $E[w] = 0$ . The formulation of neglected heterogeneity *via* the scalar  $\eta = 0$ , rather than the matrix counterpart  $var[\eta^{1/2}w]$ , is adopted solely to simplify exposition. Absence of (parameter) heterogeneity corresponds to  $\eta = 0$  (rather than  $var[\eta^{1/2}w] = 0$ ) since then  $\beta = \beta_0$  almost surely.

The marginal density of the observation vector  $z$  is

$$\int f(z, \beta_0 + \eta^{1/2}w)h(w'w)dw$$

with consequent score associated with  $\eta$  given by

$$\frac{1}{2}\eta^{-1/2}\frac{1}{\int f(z, \beta_0 + \eta^{1/2}w)h(w'w)dw} \int w' \frac{\partial f(z, \beta_0 + \eta^{1/2}w)}{\partial \beta} h(w'w)dw.$$

Evaluation at  $\eta = 0$  yields the indeterminate ratio 0/0 suggesting the use of L'Hôpital's rule on the ratio<sup>2</sup>

$$\frac{1}{2}\eta^{1/2} \int w' \frac{\partial f(z, \beta_0 + \eta^{1/2}w)}{\partial \beta} h(w'w)dw/\eta.$$

Taking the limit  $\lim_{\eta \rightarrow 0+}$  gives the score for  $\eta$  as

$$\frac{1}{2}tr\left(\frac{1}{f(z, \beta_0)} \frac{\partial^2 f(z, \beta_0)}{\partial \beta \partial \beta'} var[w]\right). \quad (2.3)$$

Consequently, given the non-singularity of  $var[w]$ , cf. Chesher (1984, Assumption (ii), p.867), the expression (2.3) suggests a (conditional) moment or score test statistic [Newey (1985b)] for the absence of parameter heterogeneity based on the non-redundant elements of the moment indicator

$$\frac{1}{f(z, \beta_0)} \frac{\partial^2 f(z, \beta_0)}{\partial \beta \partial \beta'}; \quad (2.4)$$

cf. (2.2). See Chesher (1984, p.686).

---

<sup>2</sup>Alternatively specifying the marginal density of  $\beta$  as  $\kappa^{-p}h((\beta - \beta_0)/\kappa)$  with  $h(\cdot)$  symmetric and  $\kappa$  a non-negative scalar and writing  $\beta = \beta_0 + \kappa w$ , then  $w$  has continuous density  $h(w)$  with  $E[w] = 0$ . Thus the marginal density of  $z$  is  $\int f(z, \beta_0 + \kappa w)h(w)dw$  with score with respect to  $\kappa$

$$\frac{1}{\int f(z, \beta_0 + \kappa w)h(w)dw} \int w' \frac{\partial f(z, \beta_0 + \kappa w)}{\partial \beta} h(w)dw.$$

In this set-up the absence of (parameter) heterogeneity corresponds to  $\kappa = 0$  and evaluation of the score with respect to  $\kappa$  at  $\kappa = 0$  yields 0 since  $E[w] = 0$ , i.e., the score for  $\kappa$  is identically zero at  $\kappa = 0$ . This difficulty is resolved by the reparameterisation  $\eta = \kappa^2$ . Cf. Lee and Chesher (1986).

### 3 Moment Condition Models

In many applications, researchers find the requirement to provide a full specification for the (conditional) density  $f(z, \beta)$  of the observation vector  $z$  necessitated by ML to be unpalatable. The alternative environment we consider is one that is now standard, where the model is defined by a finite number of non-linear unconditional moment restrictions; cf. the seminal paper Hansen (1982).

Let  $g(z, \beta)$  denote an  $m \times 1$  vector of known functions of the data observation  $z$  and, as above,  $\beta$  a  $p \times 1$  parameter vector with  $m \geq p$ . In the absence of parameter heterogeneity, we assume there is a true parameter value  $\beta_0$  which uniquely satisfies the moment condition

$$E_z[g(z, \beta)] = 0, \tag{3.1}$$

where  $E_z[\cdot]$  denotes expectation taken with respect to the (unknown) distribution of  $z$ .

Given their first order asymptotic equivalence under correct specification, we adopt the generic notation  $\hat{\beta}$  for both GMM and GEL estimators for  $\beta_0$  obtained under the moment constraint (3.1) where there is no possibility of confusion; see sections 4.1 and 4.2 below where GMM and GEL are briefly described. For the convenience of the reader, we repeat the sufficient conditions for the consistency and asymptotic normality of GMM and GEL given in Newey and Smith (2004), henceforth NS, Assumptions 1 and 2, p.226, and Assumption 4, p.227, as Assumptions A.1-A.3 in Appendix A.

#### 3.1 Optimal $m$ -Tests

To describe the form of an optimal  $m$ -statistic relevant for testing moment condition neglected heterogeneity we initially consider a general hypothesis testing environment.

Write  $\theta = (\alpha', \beta')'$  where  $\alpha$  is an  $r$ -vector of additional parameters. Suppose that the maintained hypothesis is defined by a value  $\theta_0 = (\alpha'_0, \beta'_0)'$  satisfying the moment condition

$$E_z[g(z; \theta_0)] = 0.$$

Also suppose the null hypothesis under test is  $\alpha_0 = 0$ ; we then write the vector of moment functions under  $\alpha_0 = 0$ , cf. (3.1), as  $g(z, \beta) = g(z; 0, \beta)$ . Let  $g_i(\beta) = g(z_i, \beta)$ , ( $i = 1, \dots, n$ ), and  $\hat{g}(\beta) = \sum_{i=1}^n g_i(\beta)/n$ . Then, by a random sampling CLT, under the hypothesis  $\alpha_0 = 0$ ,  $\sqrt{n}\hat{g}(\beta_0) \xrightarrow{d} N(0, \Omega)$  where  $\Omega = E[g(z, \beta_0)g(z, \beta_0)']$  which is assumed to be non-singular.

In this general setting, tests for  $\alpha_0 = 0$  may be based on a linear combination  $L$  of the sample moments  $\hat{g}(\beta)$  evaluated at  $\hat{\beta}$ , i.e.,  $L'\hat{g}(\hat{\beta})$ ; see, e.g., Newey (1985a). Let  $G = (G_\alpha, G_\beta)$  be f.c.r.  $p + r$  where  $G_\alpha = E[\partial g(z; 0, \beta_0)/\partial \alpha']$  and  $G_\beta = E[\partial g(z, \beta_0)/\partial \beta']$ . The optimality concept employed here is defined in terms of asymptotic local power against local alternatives of the form  $\alpha_{0n} = \delta/\sqrt{n}$  where  $\delta \neq 0$ . Among the class of test statistics with a limiting chi-square null distribution with  $r$  degrees of freedom those statistics with largest non-centrality parameter are optimal. An optimal  $m$ -test for  $\alpha_0 = 0$  is then defined by setting  $L' = G'_\alpha \Omega^{-1}$ , see Proposition 3, p.241, of Newey (1985a). An asymptotically equivalent statistic to that given in Newey (1985a) is the Lagrange multiplier (LM) version of Newey and West (1987), i.e.,

$$n\hat{g}(\hat{\beta})'\hat{\Omega}^{-1}\hat{G}(\hat{G}'\hat{\Omega}^{-1}\hat{G})^{-1}\hat{G}'\hat{\Omega}^{-1}\hat{g}(\hat{\beta}),$$

where  $\hat{G}$  and  $\hat{\Omega}$  denote estimators for  $G$  and  $\Omega$  respectively consistent under the null hypothesis  $\alpha_0 = 0$ .

### 3.2 Neglected Heterogeneity

The approach adopted here is similar to that of Chesher (1984) in the likelihood context described above in section 2.3. As there, for ease of exposition, we centre  $\beta$  at  $\beta_0$  and write

$$\beta = \beta_0 + \kappa w,$$

in terms of the non-negative scalar parameter  $\kappa$ ,  $\kappa \geq 0$ , and the  $p$ -vector of random variables  $w$ .

**Assumption 3.1** (*Parameter Heterogeneity.*) *The parameter vector  $\beta$  is a random vector with (unconditional) mean  $\beta_0$ .*

Under Assumption 3.1,  $E_w[w] = 0$ , where  $E_w[\cdot]$  is expectation taken with respect to the marginal distribution of  $w$ . An absence of neglected heterogeneity corresponds to the hypothesis  $\kappa = 0$ ; cf. section 2.3 and Chesher (1984).

With parameter heterogeneity, since it often represents an economic-theoretic constraint, we re-interpret the moment condition (3.1) as being agent specific. Hence, we rewrite (3.1) in terms of expectation taken with respect to the distribution of  $z$  *conditional* on  $\beta$ , i.e.,  $w$ ,

$$E_z[g(z, \beta)|w] = 0, \tag{3.2}$$

where  $E_z[\cdot|w]$  is expectation conditional on  $w$ ; cf. section 2.3.

Let  $E_{z,w}[\cdot]$  be expectation with respect to the joint distribution of  $z$  and  $w$ . The Jacobian with respect to  $\kappa$  is then given by

$$\begin{aligned} G_\kappa(\beta_0, \kappa) &= E_{z,w}\left[\frac{\partial g(z, \beta)}{\partial \kappa}\right] \\ &= E_{z,w}\left[\frac{\partial g(z, \beta_0 + \kappa w)}{\partial \beta'} w\right] \\ &= E_w\left[E_z\left[\frac{\partial g(z, \beta_0 + \kappa w)}{\partial \beta'} |w\right] w\right]. \end{aligned}$$

Evaluation of the Jacobian  $G_\kappa(\beta_0, \kappa)$  at  $\kappa = 0$  results in

$$\begin{aligned} G_\kappa &= G_\kappa(\beta_0, 0) \\ &= E_{z,w}\left[\frac{\partial g(z, \beta_0)}{\partial \beta'} w\right] \\ &= E_w\left[E_z\left[\frac{\partial g(z, \beta_0)}{\partial \beta'} |w\right] w\right]. \end{aligned} \tag{3.3}$$

In general, of course, the difficulty that arises in the classical context described in section 2.3, is absent. That is, the null hypothesis Jacobian  $G_\kappa$  is not identically zero unless  $w$  and  $\partial g(z, \beta_0)/\partial \beta'$  are uncorrelated. However, the Jacobian expression (3.3) does not permit an optimal  $m$ -statistic to be constructed without further elaboration concerning the joint distribution of  $z$  and  $w$ .

The remainder of this section considers circumstances in which the null hypothesis Jacobian  $G_\kappa$  is identically zero, i.e., conditions under which  $w$  and  $\partial g(z, \beta_0)/\partial \beta'$  are

uncorrelated. We return to the general case in section 4 when we consider generalized IM statistics appropriate for the moment condition context; see, in particular, section 4.2.

First,  $G_\kappa$  is identically zero if the derivative matrix  $\partial g(z, \beta_0)/\partial \beta'$  is conditionally mean independent of  $w$  since from (3.3) then

$$G_\kappa = E_z\left[\frac{\partial g(z, \beta_0)}{\partial \beta'}\right]E_w[w] = 0$$

as  $E_w[w] = 0$  from Assumption 3.1. Such a situation would arise when random variation in the parameters is independent of the observed data. Indeed, this assumption may be reasonable for many applications, but is likely not to be satisfied in models with simultaneity, where the data are partly determined by the value of the parameters.

We now summarise the above discussion in the following results.

**Lemma 3.1** *Under Assumption 3.1, the Jacobian with respect to  $\kappa$  is identically zero in the absence of parameter heterogeneity, under  $\kappa = 0$ , i.e.,  $G_\kappa = 0$ , if  $w$  and  $\partial g(z, \beta_0)/\partial \beta'$  are uncorrelated.*

**Corollary 3.1** *If Assumption 3.1 is satisfied, the Jacobian with respect to  $\kappa$  is identically zero in the absence of parameter heterogeneity, under  $\kappa = 0$ , i.e.,  $G_\kappa = 0$ , if  $\partial g(z, \beta_0)/\partial \beta'$  is conditionally mean independent of  $w$ .*

To gain some further insight, consider a situation relevant in many applications in which the moment condition (3.1) arises from a set of moment restrictions conditional on a set of instruments or covariates  $x$ . Consequently, we re-interpret the moment condition under parameter heterogeneity (3.2) as being taken conditional on *both* instruments  $x$  and  $w$ , i.e.,

$$E_z[g(z, \beta)|w, x] = 0,$$

where  $E_z[\cdot|w, x]$  denotes expectation conditional on  $w$  and  $x$ . Assumption 3.1 is correspondingly revised as

**Assumption 3.2** (*Conditional Parameter Heterogeneity.*) *The parameter vector  $\beta$  is a random vector with conditional mean  $\beta_0$  given covariates  $x$ .*

Now  $E_w[w|x] = 0$  with  $E_w[\cdot|x]$  expectation taken with respect to  $w$  conditional on  $x$ . The conditional mean independence of  $w$  and  $x$  of Assumption 3.2 is rather innocuous as it may not be too unreasonable to hazard that the heterogeneity component  $w$  should not involve the instruments  $x$ . The Jacobian (3.3) with respect to  $\kappa$  is then

$$\begin{aligned} G_\kappa(\beta_0, \kappa) &= E_x[E_{z,w}[\frac{\partial g(z, \beta_0 + \kappa w)}{\partial \kappa} | x]] \\ &= E_x[E_{z,w}[\frac{\partial g(z, \beta_0 + \kappa w)}{\partial \beta'} w | x]]. \end{aligned} \tag{3.4}$$

Evaluation of the Jacobian  $G_\kappa(\beta_0, \kappa)$  at  $\kappa = 0$  results in

$$G_\kappa = E_x[E_{z,w}[\frac{\partial g(z, \beta_0)}{\partial \beta'} w | x]].$$

The next result is then immediate.

**Lemma 3.2** *Under Assumption 3.2, the Jacobian with respect to  $\kappa$  is identically zero in the absence of parameter heterogeneity, under  $\kappa = 0$ , i.e.,  $G_\kappa = 0$ , if  $w$  and  $\partial g(z, \beta_0)/\partial \beta'$  are conditionally uncorrelated given instruments  $x$ .*

The condition of Lemma 3.2 is satisfied in the following circumstances. Rewrite the Jacobian (3.4) using the law of iterated expectations as

$$\begin{aligned} G_\kappa &= E_x[E_w[E_z[\frac{\partial g(z, \beta_0)}{\partial \beta'} | w, x] w | x]] \\ &= E_x[E_z[\frac{\partial g(z, \beta_0)}{\partial \beta'} | x] E_w[w | x]], \end{aligned}$$

the second equality holding if the derivative matrix  $\partial g(z, \beta_0)'/\partial \beta$  is conditionally mean independent of  $w$  given  $x$ . We may therefore state

**Corollary 3.2** *Under Assumption 3.2, the Jacobian with respect to  $\kappa$  is identically zero in the absence of parameter heterogeneity, under  $\kappa = 0$ , i.e.,  $G_\kappa = 0$ , if  $\partial g(z, \beta_0)'/\partial \beta$  is conditionally mean independent of  $w$  given covariates  $x$ .*

Cf. section 2.3. Such a situation would arise if the derivative matrix  $\partial g(z, \beta_0)' / \partial \beta$  is solely a function of  $x$ . Examples include static (nonlinear) panel data models but the conditions of Lemma 3.2 and Corollary 3.2 would generally not be satisfied for dynamic panel data or simultaneous equation models.

To deal with the general case of identically zero Jacobian with respect to  $\kappa$  identified in Lemma 3.1, like Lee and Chesher (1986), as in other cases considered there, the simple reparametrisation  $\eta = \kappa^2$  suffices to fix the problem, i.e.,  $\beta = \beta_0 + \eta^{1/2}w$ ; see also Chesher (1984, pp.867-868) and section 2.3. The Jacobian with respect to  $\eta$  is

$$\begin{aligned} G_\eta(\beta_0, \eta) &= E\left[\frac{\partial g(z, \beta)}{\partial \eta}\right] \\ &= \frac{1}{2}\eta^{-1/2}E\left[\frac{\partial g(z, \beta_0 + \eta^{1/2}w)}{\partial \beta'}w\right], (j = 1, \dots, m). \end{aligned}$$

Evaluation at  $\eta = 0$  results in the indeterminate ratio 0/0. Define  $G_\eta^j(\beta, \eta) = \partial g^j(z, \beta) / \partial \beta$ , ( $j = 1, \dots, m$ ). Applying L'Hôpital's rule to the ratio

$$\frac{1}{2}\eta^{1/2}E\left[\frac{\partial g(z, \beta_0 + \eta^{1/2}w)}{\partial \beta'}w\right]/\eta, \quad (3.5)$$

and taking the limits  $\lim_{\eta \rightarrow 0^+}$  of numerator and denominator in (3.5), results in the following expression for the Jacobian with respect to  $\eta$  at  $\eta = 0$

$$\begin{aligned} G_\eta^j &= G_\eta^j(\beta_0) = \lim_{\eta \rightarrow 0^+} G_\eta^j(\beta_0, \eta) \\ &= \frac{1}{2}E_{z,w}[w' \lim_{\eta \rightarrow 0^+} \frac{\partial g^j(z, \beta_0 + \eta^{1/2}w)}{\partial \beta \partial \beta'} w] \\ &= \frac{1}{2}tr(E_{z,w}[\frac{\partial^2 g^j(z, \beta_0)}{\partial \beta \partial \beta'} ww']), (j = 1, \dots, m). \end{aligned}$$

See Appendix C.1.

If  $\partial^2 g^j(z, \beta_0) / \partial \beta \partial \beta'$ , ( $j = 1, \dots, m$ ), are conditionally mean independent of  $w$ , then, under Assumption 3.1,

$$\begin{aligned} G_\eta^j &= \frac{1}{2}tr(E_w[E_z[\frac{\partial^2 g^j(z, \beta_0)}{\partial \beta \partial \beta'} |w] ww']) \\ &= \frac{1}{2}tr(E_z[\frac{\partial^2 g^j(z, \beta_0)}{\partial \beta \partial \beta'}] var_w[w]), (j = 1, \dots, m). \end{aligned} \quad (3.6)$$

Cf. Corollary 3.1.

Alternatively, if  $\partial^2 g^j(z, \beta_0)/\partial\beta\partial\beta'$ , ( $j = 1, \dots, m$ ), are conditionally mean independent of  $w$  given instruments or covariates  $x$ , then

$$E_z[\partial^2 g^j(z, \beta_0)/\partial\beta\partial\beta'|w, x] = E_z[\partial^2 g^j(z, \beta_0)/\partial\beta\partial\beta'|x], (j = 1, \dots, m).$$

Hence, under Assumption 3.2, since  $E_w[w|x] = 0$ , using the law of iterated expectations,

$$\begin{aligned} G_\eta^j &= \frac{1}{2} \text{tr}(E_x[E_w[E_z[\frac{\partial^2 g^j(z, \beta_0)}{\partial\beta\partial\beta'}|w, x]ww'|x]]) \\ &= \frac{1}{2} \text{tr}(E_x[E_z[\frac{\partial^2 g^j(z, \beta_0)}{\partial\beta\partial\beta'}|x] \text{var}_w[w|x]]), (j = 1, \dots, m). \end{aligned}$$

Cf. Corollary 3.2. Moreover, if the random variation in  $\beta$ , i.e.,  $w$ , is also second moment independent of  $x$ ,  $\text{var}_w[w|x] = \text{var}_w[w]$ ,

$$G_\eta^j = \frac{1}{2} \text{tr}(E_z[\frac{\partial^2 g^j(z, \beta_0)}{\partial\beta\partial\beta'}] \text{var}_w[w]), (j = 1, \dots, m). \quad (3.7)$$

We summarise the above development in the following result.

**Theorem 3.1** *Either (a) under Assumption 3.1, if (a)  $\partial g(z, \beta_0)'/\partial\beta$  and  $\partial^2 g^j(z, \beta_0)/\partial\beta\partial\beta'$ , ( $j = 1, \dots, m$ ), are conditionally mean independent of  $w$ , or (b) under Assumption 3.2, if  $\partial g(z, \beta_0)/\partial\beta'$  and  $\partial^2 g^j(z, \beta_0)/\partial\beta\partial\beta'$ , ( $j = 1, \dots, m$ ), are conditionally mean independent of  $w$  given instruments or covariates  $x$  and  $w$  is second moment independent of  $x$ , the Jacobian of an optimal  $m$ -test against neglected parameter heterogeneity consists of the non-redundant elements of*

$$E_z[\frac{\partial^2 g(z, \beta_0)}{\partial\beta_k\partial\beta_l}], (k \leq l, l = 1, \dots, p).$$

### 3.3 Test Statistics

Let  $\hat{\Omega}(\beta) = \sum_{i=1}^n g_i(\beta)g_i(\beta)'/n$ . Define

$$G_{\beta i}(\beta) = \frac{\partial g_i(\beta)}{\partial\beta'}, [G_{\eta i}(\beta)]_{rs} = \frac{\partial^2 g_i(\beta)}{\partial\beta_k\partial\beta_l}, (k \leq l, l = 1, \dots, p).$$

We stack the vectors  $[G_{\eta i}(\beta)]_{kl}$ , ( $k \leq l, l = 1, \dots, p$ ), as columns of the  $m \times p(p+1)/2$  matrix  $G_{\eta i}(\beta)$ , ( $i = 1, \dots, n$ ).



Let  $G_\beta = E_z[\partial g(z, \beta)/\partial \beta']$  and  $[G_\eta]_{kl} = E_z[\partial^2 g(z, \beta)/\partial \beta_k \partial \beta_l]$ , ( $k \leq l, l = 1, \dots, p$ ), that are stacked similarly to  $[G_{\eta_i}(\beta)]_{kl}$ , ( $k \leq l, l = 1, \dots, p$ ), as the columns of the  $m \times p(p+1)/2$  matrix  $G_\eta$ . As in the classical case, there may be a linear dependence among the columns of the population matrix  $G_\eta$  taken together with  $G_\beta$ . Moreover, for economic theoretic reasons, parameter heterogeneity may only be suspected in a subset of the elements of  $\beta$ . Therefore, we adopt the notation  $G_\eta^c$  for those non-redundant  $r$  columns chosen from  $G_\eta$  with  $G_{\eta_i}^c(\beta)$ , ( $i = 1, \dots, n$ ), their sample counterparts.

To define the requisite GMM and GEL statistics, define the sample moment estimators  $\hat{G}_\beta(\beta) = \sum_{i=1}^n G_{\beta i}(\beta)/n$ ,  $\hat{G}_\eta^c(\beta) = \sum_{i=1}^n G_{\eta_i}^c(\beta)/n$  and write  $\hat{G}(\beta) = (\hat{G}_\beta(\beta), \hat{G}_\eta^c(\beta))$  with the consequent  $\hat{\Sigma}(\beta) = (\hat{G}(\beta)' \hat{\Omega}(\beta)^{-1} \hat{G}(\beta))^{-1}$ .

The optimal GMM or GEL LM-type statistic for neglected heterogeneity is

$$\mathcal{LM}_n = n \hat{g}(\hat{\beta})' \hat{\Omega}(\hat{\beta})^{-1} \hat{G}(\hat{\beta}) \hat{\Sigma}(\hat{\beta}) \hat{G}(\hat{\beta})' \hat{\Omega}(\hat{\beta})^{-1} \hat{g}(\hat{\beta}); \quad (3.8)$$

see Newey and West (1987) and Smith (2010). Given the optimal GMM or GEL estimator  $\hat{\beta}$ , define  $\hat{\lambda} = \arg \sup_{\lambda \in \hat{\Lambda}_n(\hat{\beta})} \hat{P}_n(\hat{\beta}, \lambda)$  where  $\hat{P}_n(\beta, \lambda)$  is the GEL criterion stated in (4.5) below and the set  $\hat{\Lambda}_n(\hat{\beta})$  given in section 4.2. Since  $n^{1/2} \hat{\lambda} = -\hat{\Omega}(\hat{\beta})^{-1} n^{1/2} \hat{g}(\hat{\beta}) + O_p(n^{-1/2})$  under local alternatives to (3.1), a score-type test asymptotically equivalent to (3.8) may also be defined

$$\mathcal{S}_n = n \hat{\lambda}' \hat{G}(\hat{\beta}) \hat{\Sigma}(\hat{\beta}) \hat{G}(\hat{\beta})' \hat{\lambda}. \quad (3.9)$$

Cf. the first order conditions defining the GEL estimator  $\hat{\beta}$ ; see section 4.2 below.

The limiting distributions of the statistics  $\mathcal{LM}_n$  (3.8) and  $\mathcal{S}_n$  (3.9) in the absence of parameter heterogeneity may then be described. Let  $\mathcal{N}$  denote a neighbourhood of  $\beta_0$ .

**Theorem 3.2** *If Assumptions A.1, A.2 and A.3 of Appendix A are satisfied together with  $E[\sup_{\beta \in \mathcal{N}} \|\partial^2 g(z, \beta)/\partial \beta_k \partial \beta_l\|] < \infty$ , ( $k \leq l, l = 1, \dots, p$ ),  $\text{rank}(E[(G_\beta, G_\eta^c)]) = p + r$  and  $p + r \leq m$ , then*

$$\mathcal{S}_n, \mathcal{LM}_n \xrightarrow{d} \chi_r^2.$$

See, e.g., Newey and West (1987) and Smith (2010).

Note that the Jacobian estimator  $\hat{G}(\beta)$  may equivalently be replaced by its GEL counterpart  $\tilde{G}(\beta) = (\tilde{G}_\beta(\beta)', \tilde{G}_\eta(\beta)) = \sum_{i=1}^n \hat{\pi}_i(\beta, \hat{\lambda}(\beta))(G_{\beta i}(\beta)', G_{\eta i}^c(\beta)')$ , where the implied probabilities  $\hat{\pi}_i(\beta, \hat{\lambda}(\beta))$ , ( $i = 1, \dots, n$ ), are defined in (4.7) below. Likewise the variance matrix estimator  $\hat{\Omega}(\beta)$  may be replaced by  $\tilde{\Omega}(\beta) = \sum_{i=1}^n \hat{\pi}_i(\beta, \hat{\lambda}(\beta))g_i(\beta)g_i(\beta)'$ .

The above development critically relies on an assumption of (unconditional or conditional) uncorrelatedness of the heterogeneity variate  $w$  and the sample Jacobian  $\partial g(z, \beta_0)/\partial \beta'$ , i.e.,

$$E_{z,w}\left[\frac{\partial g(z, \beta_0)}{\partial \beta'}w\right] = 0, \quad (3.10)$$

necessitating the use of L'Hôpital's rule to obtain the Jacobian with respect to  $\eta$  evaluated at  $\eta = 0$ . Cf. Lemmata 3.1 and 3.2 and Corollaries 3.1 and 3.2. The next section, in particular, section 4.2, develops an alternative approach to the construction of test statistics against moment condition parameter heterogeneity that potentially permits an implicit correlation between  $w$  and  $\partial g(z, \beta_0)/\partial \beta'$ .

## 4 Generalized Information Matrix Tests

Optimal GMM or GEL tests for neglected heterogeneity based on the moment indicator second derivative  $\partial^2 g(z, \beta_0)/\partial \beta_k \partial \beta_l$ , ( $k \leq l, l = 1, \dots, p$ ), described in section 3.3, may also be interpreted in terms of GMM and GEL versions of a generalized IM equality. As is well known, see, e.g., Tauchen (1985), the GMM objective function satisfies a generalized form of the IM equality described in (2.2). As described below a similar relation is revealed for GEL.

Let  $G_\beta = E[\partial g(z, \beta_0)/\partial \beta']$  and  $\Omega = E[g(z, \beta_0)g(z, \beta_0)']$ .

### 4.1 GMM

The standard estimator of  $\beta$  is the efficient two step (2S) GMM estimator due to Hansen (1982). Suppose  $\tilde{\beta}$  is a preliminary consistent estimator for  $\beta_0$ . The 2SGMM estimator

is defined as

$$\hat{\beta}_{2S} = \arg \min_{\beta \in \mathcal{B}} \hat{g}(\beta)' \hat{\Omega}(\tilde{\beta})^{-1} \hat{g}(\beta). \quad (4.1)$$

Under (3.1) and, in particular, Assumptions A.1-A.3 it is straightforward to show that  $\hat{\beta}_{2S} \xrightarrow{p} \beta_0$  and that  $\hat{\beta}_{2S}$  is asymptotically normally distributed, i.e.,  $\sqrt{n}(\hat{\beta}_{2S} - \beta_0) \xrightarrow{d} N(0, (G'_\beta \Omega^{-1} G_\beta)^{-1})$ . See, e.g., Newey and McFadden (1994). The matrix  $G'_\beta \Omega^{-1} G_\beta$  may be thought of as a generalized IM appropriate for the moment condition context. Cf. the classical information matrix  $\mathcal{I}(\beta_0)$  defined in section 2.1. Indeed, its inverse, i.e., the asymptotic variance of efficient 2SGMM estimator  $\hat{\beta}_{2S}$ , corresponds to the semiparametric efficiency lower bound, see Chamberlain (1987).

Although similar in structure to GMM, the continuous updating estimator (CUE) criterion of Hansen, Heaton, and Yaron (1996) differs by requiring that the 2SGMM criterion is also simultaneously minimized over  $\beta$  in  $\hat{\Omega}(\beta)$ , i.e., the CUE is given by

$$\hat{\beta}_{CUE} = \arg \min_{\beta \in \mathcal{B}} \hat{g}(\beta)' \hat{\Omega}(\beta)^{-1} \hat{g}(\beta), \quad (4.2)$$

where  $A^-$  now denotes any generalized inverse of a matrix  $A$ , satisfying  $AA^-A = A$ .

Now consider the rescaled GMM objective function

$$\hat{Q}_n(\beta) = \hat{g}(\beta)' \hat{\Omega}(\tilde{\beta})^{-1} \hat{g}(\beta) / 2. \quad (4.3)$$

To describe a generalized IM equality similar to (2.2) for the GMM criterion  $\hat{Q}_n(\beta)$ , first, under Assumptions A.1-A.3, by a UWL and a CLT, the limiting normal distribution associated with the score  $\partial \hat{Q}_n(\beta_0) / \partial \beta = \hat{G}_\beta(\beta_0)' \hat{\Omega}(\tilde{\beta})^{-1} \hat{g}(\beta_0)$  obtained from (4.3) may be stated as

$$\sqrt{n} \frac{\partial \hat{Q}_n(\beta_0)}{\partial \beta} \xrightarrow{d} N(0, G'_\beta \Omega^{-1} G_\beta).$$

Secondly, the asymptotic variance and generalized IM  $G'_\beta \Omega^{-1} G_\beta$  is equal to the asymptotic limit of the Hessian matrix  $\partial^2 \hat{Q}_n(\beta_0) / \partial \beta \partial \beta'$ , *viz.*

$$\begin{aligned} \frac{\partial^2 \hat{Q}_n(\beta_0)}{\partial \beta_k \partial \beta_l} &= [\hat{G}_\beta(\beta_0)' \hat{\Omega}(\tilde{\beta})^{-1} \hat{G}_\beta(\beta_0)]_{kl} + \frac{\partial^2 \hat{g}(\beta_0)'}{\partial \beta_k \partial \beta_l} \hat{\Omega}(\tilde{\beta})^{-1} \hat{g}(\beta_0) \\ &= [G'_\beta \Omega^{-1} G_\beta]_{kl} + O_p(n^{-1/2}), \quad (k \leq l, l = 1, \dots, p), \end{aligned}$$

with the first term by application of a UWL and the second term by a UWL and CLT.

Hence, analogously with the classical IM test statistic of White (1982), see section 2.2, a GMM-based IM test for the moment specification  $E_z[g(z, \beta_0)] = 0$  may be based on the contrast between an estimator of the asymptotic variance of  $\sqrt{n}\partial\hat{Q}(\theta_0)/\partial\theta$ , i.e., the generalized IM  $G'_\beta\Omega^{-1}G_\beta$ , with the Hessian evaluated at the 2SGMM estimator  $\hat{\beta}$ ,

$$\frac{\partial^2\hat{Q}(\hat{\beta})}{\partial\beta_k\partial\beta_l} = [\hat{G}_\beta(\hat{\beta})'\hat{\Omega}(\tilde{\beta})^{-1}\hat{G}_\beta(\hat{\beta})]_{kl} + \frac{\partial^2\hat{g}(\hat{\beta})'}{\partial\beta_k\partial\beta_l}\hat{\Omega}(\tilde{\beta})^{-1}\hat{g}(\hat{\beta}), (k \leq l, l = 1, \dots, p).$$

A standard estimator for the generalized IM  $G'_\beta\Omega^{-1}G_\beta$  is  $\hat{G}_\beta(\hat{\beta})'\hat{\Omega}(\tilde{\beta})^{-1}\hat{G}_\beta(\hat{\beta})$ . This estimator also has an interpretation as an outer product form of estimator based on the “scores”  $\hat{G}_\beta(\hat{\beta})'\hat{\Omega}(\tilde{\beta})^{-1}g_i(\tilde{\beta})$ , ( $i = 1, \dots, n$ ); cf. the score  $\partial\hat{Q}(\beta_0)/\partial\beta = \hat{G}_\beta(\beta_0)'\hat{\Omega}(\tilde{\beta})^{-1}\hat{g}(\beta_0)$ . The generalized GMM IM specification test statistic is therefore based on the non-redundant “scores” from

$$\frac{\partial^2\hat{g}(\hat{\beta})'}{\partial\beta_k\partial\beta_l}\hat{\Omega}(\tilde{\beta})^{-1}\hat{g}(\hat{\beta}), (k \leq l, l = 1, \dots, p);$$

cf. the optimal LM form of neglected heterogeneity test statistic  $\mathcal{LM}_n$  (3.8) above.

Recall though from section 3.2 that this formulation of  $\mathcal{LM}_n$  implicitly incorporates the (unconditional or conditional) uncorrelatedness of  $w$  and  $\partial g(z, \beta_0)'/\partial\beta$ ; cf. Lemmata 3.1 and 3.2 and Corollaries 3.1 and 3.2. The implicit Jacobian is therefore constructed from

$$E_z\left[\frac{\partial^2 g(z, \beta_0)}{\partial\beta_k\partial\beta_l}\right], (k \leq l, l = 1, \dots, p); \quad (4.4)$$

cf. Theorem 3.1.

## 4.2 GEL

An alternative class of criteria relevant for the estimation of models defined in terms of the moment condition (3.1) is the GEL class; see, e.g., NS and Smith (1997, 2010). Indeed CUE (4.2) is included as a special case of GEL; see fn.3 below.

GEL estimation is based on a scalar function  $\rho(v)$  of a scalar  $v$  that is concave on its domain, an open interval  $\mathcal{V}$  containing zero. Without loss of generality, it is convenient

to normalize  $\rho(\cdot)$  with  $\rho_1 = \rho_2 = -1$  where  $\rho_j(v) = \partial^j \rho(v) / \partial v^j$  and  $\rho_j = \rho_j(0)$ , ( $j = 0, 1, 2, \dots$ ). Let  $\hat{\Lambda}_n(\beta) = \{\lambda : \lambda' g_i(\beta) \in \mathcal{V}, i = 1, \dots, n\}$ . The GEL criterion is defined as

$$\hat{P}_n(\beta, \lambda) = \sum_{i=1}^n [\rho(\lambda' g_i(\beta)) - \rho(0)] / n \quad (4.5)$$

with the GEL estimator of  $\beta$  given as the solution to a saddle point problem; *viz.*<sup>3</sup>

$$\hat{\beta} = \arg \min_{\beta \in \mathcal{B}} \sup_{\lambda \in \hat{\Lambda}_n(\beta)} \hat{P}_n(\beta, \lambda). \quad (4.6)$$

Let  $\hat{\lambda}(\beta) = \arg \sup_{\lambda \in \hat{\Lambda}_n(\beta)} \hat{P}_n(\beta, \lambda)$  and  $\hat{\lambda} = \hat{\lambda}(\hat{\beta})$ . Under Assumption B.1 of the Appendix, by NS Theorem 3.1, p.226,  $\hat{\beta} \xrightarrow{p} \beta_0$  and  $\hat{\lambda} \xrightarrow{p} 0$  and, together with the additional Assumption B.2,  $\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, (G'_\beta \Omega^{-1} G_\beta)^{-1})$  and  $\sqrt{n}\hat{\lambda} \xrightarrow{d} N(0, \Omega^{-1} - \Omega^{-1} G_\beta \Sigma_{\beta\beta} G'_\beta \Omega^{-1})$ , see NS Theorem 3.2, p.226.

Similarly to Back and Brown (1993), empirical or implied GEL probabilities may be defined for a given GEL function  $\rho(\cdot)$  as

$$\hat{\pi}_i(\beta, \hat{\lambda}(\beta)) = \frac{\rho_1(\hat{\lambda}(\beta)' g_i(\beta))}{\sum_{j=1}^n \rho_1(\hat{\lambda}(\beta)' g_j(\beta))}, (i = 1, \dots, n); \quad (4.7)$$

cf. NS and Brown and Newey (1992, 2002).<sup>4</sup>

A similar analysis to that described above in section 4.1 for GMM may be based on the GEL criterion with its re-interpretation as a pseudo-likelihood function to obtain a generalized IM equality. To do so consider the profile GEL criterion obtained from (4.5) after substituting out  $\lambda$  with  $\hat{\lambda}(\beta)$ , i.e.,

$$\hat{P}_n(\beta) = \hat{P}_n(\beta, \hat{\lambda}(\beta)). \quad (4.8)$$

---

<sup>3</sup>Both EL and exponential tilting (ET) estimators are included in the GEL class with  $\rho(v) = \log(1-v)$  and  $\mathcal{V} = (-\infty, 1)$ , [Qin and Lawless (1994), Imbens (1997) and Smith (1997)] and  $\rho(v) = -\exp(v)$ , [Kitamura and Stutzer (1997), Imbens, Spady and Johnson (1998) and Smith (1997)], respectively, as is the CUE, as indicated above, if  $\rho(v)$  is quadratic [NS]. Minimum discrepancy estimators based on the Cressie and Read (1984) family  $h(\pi) = [\gamma(\gamma + 1)]^{-1} [(n\pi)^{\gamma+1} - 1] / n$  are also members of the GEL class [NS].

<sup>4</sup>The GEL empirical probabilities  $\hat{\pi}_i(\beta, \hat{\lambda}(\beta))$ , ( $i = 1, \dots, n$ ), sum to one by construction, satisfy the sample moment conditions  $\sum_{i=1}^n \hat{\pi}_i(\beta, \hat{\lambda}(\beta)) g_i(\beta) = 0$  that define the first order conditions for  $\hat{\lambda}(\beta)$ , and are positive when  $\hat{\lambda}(\beta)' \hat{g}_i(\beta)$  is small uniformly in  $i$ . As in Brown and Newey (1998),  $\sum_{i=1}^n \hat{\pi}_i(\hat{\beta}, \hat{\lambda}(\hat{\beta})) a(z_i, \hat{\beta})$  is a semiparametrically efficient estimator of  $E_z[a(z, \beta_0)]$ .

Hence, by the envelope theorem, the score with respect to  $\beta$  is

$$\frac{\partial \hat{P}_n(\beta)}{\partial \beta} = \sum_{i=1}^n \rho_1(\hat{\lambda}(\beta)' g_i(\beta)) G_{\beta i}(\beta) \hat{\lambda}(\beta) / n. \quad (4.9)$$

The corresponding Hessian with respect to  $\beta$  from the profile GEL criterion  $\hat{P}_n(\beta)$  (4.8) is

$$\begin{aligned} & \sum_{i=1}^n \rho_2(\hat{\lambda}(\beta)' g_i(\beta)) G_{\beta i}(\beta) \hat{\lambda}(\beta) [\hat{\lambda}(\beta)' G_{\beta i}(\beta) + g_i(\beta)' \frac{\partial \hat{\lambda}(\beta)}{\partial \beta'}] / n \\ & + \sum_{i=1}^n \rho_1(\hat{\lambda}(\beta)' g_i(\beta)) \left[ \sum_{j=1}^m \frac{\partial^2 g_i^j(\beta)}{\partial \beta \partial \beta'} \hat{\lambda}_j(\beta) + G_{\beta i}(\beta)' \frac{\partial \hat{\lambda}(\beta)}{\partial \beta'} \right] / n. \end{aligned}$$

The derivative matrix  $\partial \hat{\lambda}(\beta) / \partial \beta'$  is given by application of the implicit function theorem to the first order conditions defining  $\hat{\lambda}(\beta)$ ; see (C.1) in Appendix C.2.

Let  $\hat{\lambda}_0 = \hat{\lambda}(\beta_0)$ ,  $g_i = g_i(\beta_0)$ ,  $\rho_{1i} = \rho_1(\hat{\lambda}'_0 g_i)$ ,  $\rho_{2i} = \rho_2(\hat{\lambda}'_0 g_i)$ ,  $G_{\beta i} = G_{\beta i}(\beta_0)$  and  $G_{\beta_k i}(\beta) = \partial g_i(\beta) / \partial \beta_k$ , ( $k = 1, \dots, p$ ), ( $i = 1, \dots, n$ ).

Evaluating the Hessian (4.10) at  $\beta_0$ , Appendix C.2 demonstrates that the first term is  $O_p(n^{-1})$  whilst the second and third terms are both  $O_p(n^{-1/2})$ . The fourth term consists of the  $O_p(1)$  component

$$- \sum_{i=1}^n \rho_{1i} G'_{\beta i} / n \left[ \sum_{i=1}^n \rho_{2i} g_i g'_i / n \right]^{-1} \sum_{i=1}^n \rho_{1i} G_{\beta i} / n,$$

a consistent estimator for generalized IM  $G'_\beta \Omega^{-1} G_\beta$ , and the  $O_p(n^{-1/2})$  component

$$- \sum_{i=1}^n \rho_{1i} G'_{\beta i} / n \left[ \sum_{i=1}^n \rho_{2i} g_i g'_i / n \right]^{-1} \sum_{i=1}^n \rho_{2i} g_i \hat{\lambda}'_0 G_{\beta i} / n.$$

Let  $\hat{g}_i = g_i(\hat{\beta})$ ,  $\hat{\rho}_{1i} = \rho_1(\hat{\lambda}' \hat{g}_i)$ ,  $\hat{\rho}_{2i} = \rho_2(\hat{\lambda}' \hat{g}_i)$ ,  $\hat{G}_{\beta i} = G_{\beta i}(\hat{\beta})$  and  $\hat{G}_{\beta_k i} = G_{\beta_k i}(\hat{\beta})$ , ( $k = 1, \dots, p$ ), ( $i = 1, \dots, n$ ). Similarly to GMM, a GEL IM test for the moment specification  $E_z[g(z, \beta_0)] = 0$  is based on the contrast between an estimator of the generalized IM  $G'_\beta \Omega^{-1} G_\beta$  and the GEL Hessian (4.10) evaluated at  $\hat{\beta}$ . A GEL estimator for  $G'_\beta \Omega^{-1} G_\beta$  is

$$- \sum_{i=1}^n \hat{\rho}_{1i} \hat{G}'_{\beta i} \left[ \sum_{i=1}^n \hat{\rho}_{2i} \hat{g}_i \hat{g}'_i / n \right]^{-1} \sum_{i=1}^n \hat{\rho}_{1i} \hat{G}_{\beta i} / n;$$

this estimator has the approximate interpretation as an outer product form of estimator based on the “scores”  $-\left[ \sum_{i=1}^n \hat{\rho}_{1i} \hat{G}'_{\beta i} / n \right] \left[ \sum_{i=1}^n \hat{\rho}_{2i} \hat{g}_i \hat{g}'_i / n \right]^{-1} \sqrt{-\hat{\rho}_{2i}} \hat{g}_i$ , ( $i = 1, \dots, n$ ); cf. (4.9) and the asymptotic representation (C.2) in Appendix C.2 for  $\sqrt{n} \hat{\lambda}_0$ .

Therefore in the GEL context the generalized IM equality gives rise to the score

$$\begin{aligned} & \left[ \sum_{i=1}^n \hat{\rho}_{1i} \frac{\partial^2 \hat{g}'_i}{\partial \beta_k \partial \beta_l} / n \right. \\ & + \check{G}'_{\beta_k} \check{\Omega}^{-1} \sum_{i=1}^n \hat{\rho}_{2i} \hat{g}'_i \hat{G}'_{\beta_{ki}} / n \\ & \left. + \check{G}'_{\beta_l} \check{\Omega}^{-1} \sum_{i=1}^n \hat{\rho}_{2i} \hat{g}'_i \hat{G}'_{\beta_{ki}} / n \right] \hat{\lambda}, \quad (k \leq l, l = 1, \dots, p), \end{aligned}$$

where  $\check{G}'_{\beta_k} = \sum_{i=1}^n \hat{\rho}_{1i} \hat{G}_{\beta_{ki}} / n$ , ( $k = 1, \dots, p$ ), and  $\check{\Omega} = -\sum_{i=1}^n \hat{\rho}_{2i} \hat{g}'_i \hat{g}'_i / n$ . Asymptotically, therefore, the implicit Jacobian is

$$\begin{aligned} & E \left[ \frac{\partial^2 g(z, \beta_0)}{\partial \beta_k \partial \beta_l} \right] \tag{4.10} \\ & + G'_{\beta_k} \Omega^{-1} E [g(z, \beta_0) G_{\beta_l}(z, \beta_0)'] \\ & + G'_{\beta_l} \Omega^{-1} E [g(z, \beta_0) G_{\beta_k}(z, \beta_0)'], \quad (k \leq l, l = 1, \dots, p). \end{aligned}$$

The first term in (4.10) is identical to the GMM Jacobian (4.4) but, interestingly, the second and third terms are absent for GMM. This occurs because of the use of the preliminary consistent estimator  $\tilde{\beta}$  to estimate  $\Omega$  in 2SGMM whereas GEL implicitly also optimises a variance component over  $\beta$ , cf. CUE (4.2). This first term might be regarded as arising from that component of the heterogeneity random variate  $w$  that is (unconditionally or conditionally) uncorrelated with the sample Jacobian  $\partial g(z, \beta_0) / \partial \beta'$ . The additional terms in (4.10) involve the covariances between the moment indicator derivative matrix  $G_{\beta_k}(z, \beta_0)$  and the “score”  $G'_{\beta_l} \Omega^{-1} g(z, \beta_0)$  and likewise  $G_{\beta_l}(z, \beta_0)$  and the “score”  $G'_{\beta_k} \Omega^{-1} g(z, \beta_0)$ , ( $k \leq l, l = 1, \dots, p$ ). Cf.  $G_\kappa$  (3.3). These terms thereby implicitly allow for a correlation between the heterogeneity variate  $w$  and the sample Jacobian  $\partial g(z, \beta_0) / \partial \beta'$ . Note that these terms are absent when the derivative matrix  $\partial g(z, \beta_0) / \partial \beta$  is solely a function of  $x$  and the moment indicator obeys the conditional moment constraint  $E_z[g(z, \beta_0) | x] = 0$ , e.g., static (nonlinear) panel data models. However, as noted above, these terms are likely to be relevant for dynamic panel data or simultaneous equation models.

After substitution of  $\hat{\beta}$  for  $\beta_0$ , the above score is expressed in terms of the Lagrange multiplier-type estimator  $\hat{\lambda} = \hat{\lambda}(\hat{\beta})$ . Hence the resultant statistic will be of the

LM type  $\mathcal{LM}_n$ . An equivalent score-type test, cf.  $\mathcal{S}_n$ , is obtained by substitution of  $-\hat{\Omega}(\hat{\beta})^{-1}\sqrt{n}\hat{g}(\hat{\beta})$  for  $\sqrt{n}\hat{\lambda}(\hat{\beta})$  since  $\sqrt{n}\hat{\lambda}(\hat{\beta}) = -\Omega^{-1}\sqrt{n}\hat{g}(\hat{\beta}) + O_p(n^{-1/2})$  in the absence of parameter heterogeneity.

### 4.3 An Example: Empirical Likelihood<sup>5</sup>

To illustrate the development above we consider empirical likelihood (EL), a special case of GEL; see fn.3. As is well-known, see *inter alia* Owen (1988, 2001) and Kitamura (2007), EL may be interpreted as non-parametric ML. Indeed EL *is* ML when  $z$  has discrete support. Hence, EL is an example of GEL where the classical ML-based (conditional) IM test moment indicators (2.2) and (2.4) may be applied directly to derive a test against parameter heterogeneity. The resultant EL-based statistic may then be compared with the GMM and GEL Jacobians (4.4) and (4.10) obtained in sections 4.1 and 4.2 respectively.

The EL implied probabilities, cf. (4.7), are

$$\hat{\pi}_i(\beta, \lambda) = \frac{1}{n(1 + \lambda'g_i(\beta))}, (i = 1, \dots, n),$$

where  $\lambda$  is a vector of Lagrange multipliers corresponding to imposition of the moment restrictions  $\sum_{i=1}^n \hat{\pi}_i(\beta, \lambda)g(z_i, \beta) = 0$ . The EL criterion is then defined as

$$\begin{aligned} \mathcal{EL}_n(\beta, \lambda) &= \frac{1}{n} \sum_{i=1}^n \log \hat{\pi}_i(\beta, \lambda) \\ &= -\frac{1}{n} \sum_{i=1}^n \log n(1 + \lambda'g_i(\beta)). \end{aligned}$$

Given  $\beta$  the Lagrange multiplier vector  $\lambda$  may be concentrated or profiled out using the solution  $\hat{\lambda}(\beta)$  to the likelihood equations

$$\begin{aligned} 0 &= -\sum_{i=1}^n \hat{\pi}_i(\beta, \hat{\lambda}(\beta))g(z_i, \beta) \\ &= -\sum_{i=1}^n \frac{1}{n(1 + \hat{\lambda}(\beta)'g_i(\beta))}g_i(\beta), \end{aligned} \tag{4.11}$$

---

<sup>5</sup>We are grateful to Y. Kitamura for suggesting this example.



obtained from setting  $\partial \mathcal{E} \mathcal{L}_n(\beta, \hat{\lambda}(\beta)) / \partial \lambda = 0$ . The resultant profile EL criterion is

$$\begin{aligned} \mathcal{E} \mathcal{L}_n(\beta) &= \frac{1}{n} \sum_{i=1}^n \log \hat{\pi}_i(\beta, \hat{\lambda}(\beta)) \\ &= -\frac{1}{n} \sum_{i=1}^n \log n(1 + \hat{\lambda}(\beta)' g_i(\beta)) \end{aligned}$$

with likelihood equations

$$\begin{aligned} 0 &= -\sum_{i=1}^n \hat{\pi}_i(\beta, \hat{\lambda}(\beta)) G_{\beta i}(\beta)' \hat{\lambda}(\beta) \\ &= -\sum_{i=1}^n \frac{1}{n(1 + \hat{\lambda}(\beta)' g_i(\beta))} G_{\beta i}(\beta)' \hat{\lambda}(\beta). \end{aligned}$$

Therefore, a classical EL-based IM test or, equivalently, test for the absence of parameter heterogeneity uses the non-redundant elements of the moment indicators

$$\frac{1}{n} \sum_{i=1}^n \frac{1}{\hat{\pi}_i(\beta_0, \hat{\lambda}_0)} \frac{\partial^2 \hat{\pi}_i(\beta_0, \hat{\lambda}_0)}{\partial \beta_k \partial \beta_l}, \quad (k \leq l, l = 1, \dots, p),$$

evaluated at the EL estimator  $\hat{\beta}$ ; cf. sections 2.2 and 2.3. As detailed in Appendix C.3

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \frac{1}{\hat{\pi}_i(\beta_0, \hat{\lambda}_0)} \frac{\partial^2 \hat{\pi}_i(\beta_0, \hat{\lambda}_0)}{\partial \beta_k \partial \beta_l} &= \left[ \sum_{i=1}^n \frac{1}{n(1 + \hat{\lambda}'_0 g_i)} G'_{\beta k i} \right] \check{\Omega}(\beta_0)^{-1} \left[ \sum_{i=1}^n \frac{1}{n(1 + \hat{\lambda}'_0 g_i)^2} g_i G'_{\beta l i} \right] \hat{\lambda}_0 \\ &\quad + \left[ \sum_{i=1}^n \frac{1}{n(1 + \hat{\lambda}'_0 g_i)} G'_{\beta l i} \right] \check{\Omega}(\beta_0)^{-1} \left[ \sum_{i=1}^n \frac{1}{n(1 + \hat{\lambda}'_0 g_i)^2} g_i G'_{\beta k i} \right] \hat{\lambda}_0 \\ &\quad - \left[ \sum_{i=1}^n \frac{1}{n(1 + \hat{\lambda}'_0 g_i)} \frac{\partial^2 g_i(\beta_0)'}{\partial \beta_k \partial \beta_l} \right] \hat{\lambda}_0 + O_p(n^{-1}), \end{aligned}$$

( $k \leq l, l = 1, \dots, p$ ), where

$$\check{\Omega}(\beta) = \sum_{i=1}^n \frac{1}{n(1 + \hat{\lambda}(\beta)' g_i(\beta))^2} g_i(\beta) g_i(\beta)'.$$

These terms are exactly those given in section 4.2 above for the GEL IM statistic specialised for EL since, defining  $\rho(v) = \log(1 - v)$ , see fn.3,  $\rho_{1i} = -1/n(1 - \hat{\lambda}'_0 g_i)$  and  $\rho_{2i} = -1/n(1 - \hat{\lambda}'_0 g_i)^2$ , ( $i = 1, \dots, n$ ).

## 5 Many Instruments

The development of earlier sections has been primarily concerned with unconditional moment restrictions. In our discussion of moment condition neglected heterogeneity in

section 3, it was noted that many models expressed in terms of unconditional moment restrictions arise from consideration of conditional moment constraints. This section adapts the above analysis of moment condition neglected heterogeneity to the conditional moment context. Like Appendix A, for ease of reference, Appendix B collects together assumptions given in Donald, Imbens and Newey (2003), DIN henceforth, sufficient for the consistency and asymptotic normality of GMM and GEL.

To provide an analysis for this setting, let  $u(z, \beta)$  denote a  $s$ -vector of known functions of the data observation  $z$  and  $\beta$ . The model is completed by the conditional moment restriction

$$E_z[u(z, \beta)|x] = 0 \text{ w.p.1,} \tag{5.1}$$

satisfied uniquely at true parameter value  $\beta_0 \in \text{int}(\mathcal{B})$ . In many applications, the conditional moment function  $u(z, \beta)$  would be a vector of residuals.

It is well known [Chamberlain (1987)] that conditional moment conditions of the type (5.1) are equivalent to a countable number of unconditional moment restrictions under certain regularity conditions. Assumption 1, p.58, in DIN, repeated as Assumption B.1 of Appendix B, provides precise conditions. To summarise, for each positive integer  $K$ , if  $q^K(x) = (q_{1K}(x), \dots, q_{KK}(x))'$  denotes a  $K$ -vector of approximating functions, then we require  $q^K(x)$  such that for all functions  $a(x)$  with  $E[a(x)^2] < \infty$  there are  $K$ -vectors  $\gamma_K$  such that as  $K \rightarrow \infty$ ,  $E[(a(x) - q^K(x)' \gamma_K)^2] \rightarrow 0$ . Possible approximating functions are splines, power series and Fourier series. See *inter alia* DIN and Newey (1997) for further discussion. DIN Lemma 2.1, p.58, formally shows the equivalence between conditional moment restrictions and a sequence of unconditional moment restrictions of the type considered in this section.

Like DIN we define an unconditional moment indicator vector as

$$g(z, \beta) = u(z, \beta) \otimes q(x),$$

where  $q(x) = q^K(x)$  omitting the index  $K$  where there can be no possibility of confusion; thus, from earlier sections,  $m = s \times K$ . Assumption B.2, i.e., Assumption 2, p.59, of

DIN, imposes the normalisation requirement that, for each  $K$ , there exists a constant scalar  $\zeta(K)$  and matrix  $B_K$  such that  $\tilde{q}^K(x) = B_K q^K(x)$  for all  $x \in \mathcal{X}$ , where  $\mathcal{X}$  denotes the support of the random vector  $x$ , with  $\sup_{x \in \mathcal{X}} \|\tilde{q}^K(x)\| \leq \zeta(K)$  and  $\sqrt{K} \leq \zeta(K)$ .

GMM and GEL are applied based on the consequent unconditional moment condition  $E_z[g(z, \beta_0)] = 0$ ; cf. (3.1). Define the conditional Jacobian matrix  $D_\beta(x, \beta) = E_z[\partial u(z, \beta)/\partial \beta' | x]$  and conditional second moment matrix  $V(x, \beta) = E_z[u(z, \beta)u(z, \beta)' | x]$ . By stipulating that  $K$  approaches infinity at an appropriate rate, dependent on  $n$  and the type of estimator considered, then DIN, Theorems 5.4, p.66, and 5.6, p.67, respectively, shows that GMM and GEL are root- $n$  consistent and achieve the semi-parametric efficiency lower bound  $\mathcal{I}(\beta_0)^{-1}$  where  $\mathcal{I}(\beta) = E_x[D_\beta(x, \beta)'V(x, \beta)^{-1}D_\beta(x, \beta)]$ , i.e.,<sup>6</sup>

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, \mathcal{I}(\beta_0)^{-1}).$$

Let  $u_\beta(z, \beta) = \partial u(z, \beta)/\partial \beta'$  and  $u_i(\beta) = u(z_i, \beta)$ ,  $u_{\beta_i}(\beta) = u_\beta(z_i, \beta)$ , ( $i = 1, \dots, n$ ). Also let  $g_i(\beta) = u_i(\beta) \otimes q_i$ , where  $q_i = q(x_i)$ , ( $i = 1, \dots, n$ ). Write  $G_{\beta_i}(\beta) = u_{\beta_i}(\beta) \otimes q_i$ , ( $i = 1, \dots, n$ ),  $\hat{G}_\beta(\beta) = \sum_{i=1}^n G_{\beta_i}(\beta)/n$  and, likewise,  $\hat{\Omega}(\beta) = \sum_{i=1}^n g_i(\beta)g_i(\beta)'/n = \sum_{i=1}^n u_i(\beta)u_i(\beta)' \otimes q_i q_i'/n$ .

## 5.1 Neglected Heterogeneity

The relevant Jacobian terms follow directly from the analysis for the unconditional moment case. Thus, for GMM, define

$$\begin{aligned} [G_{\eta_i}(\beta)]_{kl} &= \frac{\partial^2 g_i(\beta)}{\partial \beta_k \partial \beta_l} \\ &= \frac{\partial^2 u_i(\beta)}{\partial \beta_k \partial \beta_l} \otimes q_i, \quad (k \leq l, l = 1, \dots, p). \end{aligned}$$

---

<sup>6</sup>GMM and GEL require Assumptions B.1-B.5 and Assumptions B.1-B.6 respectively of Appendix B. The respective rates for the scalar normalisation  $\zeta(K)$  for GMM and GEL are  $\zeta(K)^2 K/n \rightarrow 0$  and  $\zeta(K)^2 K^2/n \rightarrow 0$ . See DIN, Theorems 5.4, p.66, and 5.6, p.67, respectively.

or, for GEL,

$$\begin{aligned}
[G_{\eta_i}(\beta)]_{kl} &= \frac{\partial^2 g_i(\beta)}{\partial \beta_k \partial \beta_l} \\
&\quad + G_{\beta_k i}(\beta) g_i(\beta)' \Omega^{-1} G_{\beta_l} \\
&\quad + G_{\beta_l i}(\beta) g_i(\beta)' \Omega^{-1} G_{\beta_k} \\
&= \frac{\partial^2 u_i(\beta)}{\partial \beta_k \partial \beta_l} \otimes q_i \\
&\quad + \left[ \frac{\partial u_i(\beta)}{\partial \beta_k} u_i(\beta)' \otimes q_i q_i' \right] \Omega^{-1} G_{\beta_l} \\
&\quad + \left[ \frac{\partial u_i(\beta)}{\partial \beta_l} u_i(\beta)' \otimes q_i q_i' \right] \Omega^{-1} G_{\beta_k}, \quad (k \leq l, l = 1, \dots, p).
\end{aligned} \tag{5.2}$$

Cf. sections 4.1 and 4.2.

## 5.2 Test Statistics

As previously stack the vectors  $[G_{\eta_i}(\beta)]_{kl}$ , ( $k \leq l, l = 1, \dots, p$ ), defined in (5.2) or (5.2), as columns of the  $m \times p(p+1)/2$  matrix  $G_{\eta_i}(\beta)$ , ( $i = 1, \dots, n$ ). Since there may be linear dependencies among the population counterparts of these columns taken together with those of  $G_{\beta_i}(\beta)$ , ( $i = 1, \dots, n$ ), let  $G_{\eta_i}^c(\beta)$  denote those  $r$  non-redundant columns chosen from  $G_{\eta_i}(\beta)$ , ( $i = 1, \dots, n$ ). Appropriate estimators for  $G_{\beta_k}$ , ( $k = 1, \dots, p$ ), and  $\Omega$  are defined below.

To define the requisite GMM and GEL statistics, define the sample moment estimators  $\hat{G}_\beta(\beta) = \sum_{i=1}^n G_{\beta_i}(\beta)/n$ ,  $\hat{G}_\eta^c(\beta) = \sum_{i=1}^n G_{\eta_i}^c(\beta)/n$  and write  $\hat{G}(\beta) = (\hat{G}_\beta(\beta), \hat{G}_\eta^c(\beta))$  with the consequent  $\hat{\Sigma}(\beta) = (\hat{G}(\beta)' \hat{\Omega}(\beta)^{-1} \hat{G}(\beta))^{-1}$ .

The respective optimal GMM or GEL score and LM statistics for neglected heterogeneity are defined exactly as in the unconditional case above, i.e.,

$$\mathcal{S}_n = n \hat{\lambda}' \hat{G}(\hat{\beta}) \hat{\Sigma}(\hat{\beta}) \hat{G}(\hat{\beta})' \hat{\lambda},$$

and

$$\mathcal{LM}_n = n \hat{g}(\hat{\beta})' \hat{\Omega}(\hat{\beta})^{-1} \hat{G}(\hat{\beta}) \hat{\Sigma}(\hat{\beta}) \hat{G}(\hat{\beta})' \hat{\Omega}(\hat{\beta})^{-1} \hat{g}(\hat{\beta}).$$

We employ Lemmata A.3, p.73, and A.4, p.75, of DIN in our proofs for the limiting distributions of  $\mathcal{S}_n$  and  $\mathcal{LM}_n$  in the absence of parameter heterogeneity. This requires

the strengthening of the assumptions therein, in particular, Assumption B.5 of Appendix B.

Let  $u_{\beta\beta}^j(z, \beta) = \partial^2 u^j(z, \beta) / \partial \beta \partial \beta'$ , ( $j = 1, \dots, s$ ), and  $u_{\beta\beta\beta_k}^j(z, \beta) = \partial^3 u^j(z, \beta) / \partial \beta \partial \beta' \partial \beta_k$ , ( $k = 1, \dots, p$ ), ( $j = 1, \dots, s$ ). Also let  $\mathcal{N}$  denote a neighbourhood of  $\beta_0$  and  $D_\beta(x) = E_z[u_\beta(z, \beta_0)|x]$ . We write  $D_\eta^c(x, \beta)$  as the non-redundant components selected from either  $E_z[u_{\beta_k\beta_l}(z, \beta)|x]$  or  $E_z[u_{\beta_k\beta_l}(z, \beta)|x] + E_z[u_{\beta_k}(z, \beta)u(z, \beta)'|x]V(x, \beta)^{-1}D_{\beta_l}(x, \beta) + E_z[u_{\beta_l}(z, \beta)u(z, \beta)'|x]V(x, \beta)^{-1}D_{\beta_k}(x, \beta)$ , ( $k \leq l, l = 1, \dots, p$ ), with  $D(x, \beta) = (D_\beta(x, \beta), D_\eta^c(x, \beta))$  and  $D(x) = D(x, \beta_0)$ .

**Assumption 5.1** (a)  $u(z, \beta)$  is thrice differentiable in  $\mathcal{N}$ ; (b)  $E_z[\sup_{\beta \in \mathcal{N}} \|u_{\beta\beta}^j(z, \beta)\|^2 |x]$  and  $E_z[\sup_{\beta \in \mathcal{N}} \|u^j(z, \beta)u_\beta(z, \beta)\|^2 |x]$ , ( $j = 1, \dots, s$ ), are bounded; (c)  $E_z[\|u_{\beta\beta\beta_k}^j(z, \beta_0)\|^2 |x]$ , ( $k = 1, \dots, p$ ),  $E_z[\|u_\beta(z, \beta_0)\|^4 |x]$  and  $E_z[\|u^j(z, \beta_0)u_{\beta\beta}^k(z, \beta_0)\|^2 |x]$ , ( $j, k = 1, \dots, s$ ), are bounded; (d)  $E_x[D(x)'D(x)]$  is nonsingular.

Consequently, a similar result to that in the unconditional case may be stated for the LM statistic  $\mathcal{LM}_n$ .

**Theorem 5.1** Let Assumptions B.1-B.5,  $\tilde{\beta} = \beta_0 + O_p(1/\sqrt{n})$  and  $\zeta(K)^2 K/n \rightarrow 0$  be satisfied for GMM or, for GEL, let Assumptions B.1-B.6 and  $\zeta(K)^2 K^2/n \rightarrow 0$  hold, where the scalar  $\zeta(K)$  is defined in Assumption B.2 of Appendix B. Then, under Assumption 5.1, in the absence of parameter heterogeneity,

$$\mathcal{LM}_n \xrightarrow{d} \chi_r^2.$$

Likewise, there is a corresponding result for the score statistic  $\mathcal{S}_n$ ; viz.

**Theorem 5.2** Let Assumptions B.1-B.5,  $\tilde{\beta} = \beta_0 + O_p(1/\sqrt{n})$  and  $\zeta(K)^2 K/n \rightarrow 0$  be satisfied for GMM or, for GEL, let Assumptions B.1-B.6 and  $\zeta(K)^2 K^2/n \rightarrow 0$  hold, where the scalar  $\zeta(K)$  is defined in Assumption B.2 of Appendix B. Then, under Assumption 5.1, in the absence of parameter heterogeneity,

$$\mathcal{S}_n \xrightarrow{d} \chi_r^2.$$

Indeed, as the proofs of these theorems attest,  $\mathcal{LM}_n$  and  $\mathcal{S}_n$  are asymptotically equivalent in the absence of parameter heterogeneity, i.e.,  $\mathcal{LM}_n - \mathcal{S}_n \xrightarrow{p} 0$ .

### 5.3 Test Consistency

Similarly to Lemma 6.5, p.71, in DIN, we may obtain a test consistency result for the LM statistic  $\mathcal{LM}_n$ .

**Theorem 5.3** *Suppose  $\hat{\beta} \xrightarrow{p} \beta_*$  such that  $E_z[u(z, \beta_*)|x] \neq 0$ . Under Assumption 5.1 and Assumptions B.1-B.6 with  $\beta_*$  replacing  $\beta_0$ , if  $E_x[D(x, \beta_*)'V(x, \beta_*)^{-1}D(x, \beta_*)]$  has smallest eigenvalue bounded away from zero, then the  $\alpha$ -level critical region  $\mathcal{LM}_n > \chi_r^2(\alpha)$  defines a consistent test against parameter heterogeneity if*

$$E_x[D_\eta^c(x, \beta_*)V(x, \beta_*)^{-1}E[u(z, \beta_*)|x]] \neq 0.$$

## Appendix A: Unconditional Moments: Assumptions

This Appendix repeats NS Assumptions 1 and 2, p.226, and gives a revised NS Assumption 4, p.227.

Let  $G_\beta = E_z[\partial g(z, \beta_0)/\partial \beta']$  and  $\Omega = E_z[g(z, \beta_0)g(z, \beta_0)']$ .

**Assumption A.1** (a)  $\beta_0 \in \mathcal{B}$  is the unique solution to  $E_z[g(z, \beta)] = 0$ ; (b)  $\mathcal{B}$  is compact; (c)  $g(z, \beta)$  is continuous at each  $\beta \in \mathcal{B}$  with probability one; (d)  $E_z[\sup_{\beta \in \mathcal{B}} \|g(z, \beta)\|^\alpha] < \infty$  for some  $\alpha > 2$ ; (e)  $\Omega$  is nonsingular; (f)  $\rho(v)$  is twice continuously differentiable in a neighborhood of zero.

**Assumption A.2** (a)  $\beta_0 \in \text{int}(\mathcal{B})$ ; (b)  $g(z, \beta)$  is continuously differentiable in a neighborhood  $\mathcal{N}$  of  $\beta_0$  and  $E_z[\sup_{\beta \in \mathcal{N}} \|\partial g_i(\beta)/\partial \beta'\|] < \infty$ ; (c)  $\text{rank}(G_\beta) = p$ .

**Assumption A.3** The preliminary estimator  $\tilde{\beta}$  satisfies  $\tilde{\beta} = \beta_0 + O_p(1/\sqrt{n})$ .

## Appendix B: Conditional Moments: Assumptions

This Appendix collects together DIN Assumptions 1-6 for ease of reference.

**Assumption B.1** For all  $K$ ,  $E_x[q^K(x)'q^K(x)]$  is finite and for any  $a(x)$  with  $E_x[a(x)^2] < \infty$  there are  $K$ -vectors  $\gamma_K$  such that as  $K \rightarrow \infty$ ,

$$E_x[(a(x) - q^K(x)'\gamma_K)^2] \rightarrow 0.$$

Let  $\mathcal{X}$  denote the support of the random vector  $x$ .

**Assumption B.2** For each  $K$  there is a constant scalar  $\zeta(K)$  and matrix  $B_K$  such that  $\tilde{q}^K(x) = B_K q^K(x)$  for all  $x \in \mathcal{X}$ ,  $\sup_{x \in \mathcal{X}} \|\tilde{q}^K(x)\| \leq \zeta(K)$ ,  $E_x[\tilde{q}^K(x)\tilde{q}^K(x)']$  has smallest eigenvalue bounded away from zero uniformly in  $K$  and  $\sqrt{K} \leq \zeta(K)$ .

Next let  $u_\beta(z, \beta) = \partial u(z, \beta) / \partial \beta'$ ,  $D_\beta(x) = E_z[u_\beta(z, \beta_0) | x]$  and  $u_{\beta\beta}^j(z, \beta) = \partial^2 u^j(z, \beta) / \partial \beta \partial \beta'$ , ( $j = 1, \dots, s$ ). Also let  $\mathcal{N}$  denote a neighbourhood of  $\beta_0$ .

**Assumption B.3** The data are i.i.d. and (a) there exists a unique  $\beta_0 \in \mathcal{B}$  such that  $E_z[u(z, \beta) | x] = 0$ ; (b)  $\mathcal{B}$  is compact; (c)  $E_z[\sup_{\beta \in \mathcal{B}} \|u(z, \beta)\|^2 | x]$  is bounded; (d) for all  $\beta, \tilde{\beta} \in \mathcal{B}$ ,  $\|u(z, \beta) - u(z, \tilde{\beta})\| \leq \delta(z) \|\beta - \tilde{\beta}\|^\alpha$  for some  $\alpha > 0$  and  $\delta(z)$  such that  $E[\delta(z)^2 | x] < \infty$ .

**Assumption B.4** (a)  $\beta_0 \in \text{int}(\mathcal{B})$ ; (b)  $u(z, \beta)$  is twice differentiable in  $\mathcal{N}$ ,  $E_z[\sup_{\beta \in \mathcal{N}} \|u_\beta(z, \beta)\|^2 | x]$  and  $E_z[\|u_{\beta\beta}^j(z, \beta_0)\|^2 | x]$ , ( $j = 1, \dots, s$ ), are bounded; (c)  $E_x[D_\beta(x)'D_\beta(x)]$  is nonsingular.

**Assumption B.5** (a)  $\Sigma(x) = E_z[u(z, \beta_0)u(z, \beta_0)' | x]$  has smallest eigenvalue bounded away from 0; (b)  $E_z[\sup_{\beta \in \mathcal{N}} \|u(z, \beta)\|^4 | x]$  is bounded and, for all  $\beta \in \mathcal{N}$ ,  $\|u(z, \beta) - u(z, \beta_0)\| \leq \delta(z) \|\beta - \beta_0\|$  and  $E_z[\delta(z)^2 | x]$  is bounded.

**Assumption B.6** (a)  $\rho(\cdot)$  is twice continuously differentiable with Lipschitz second derivative in a neighbourhood of 0; (b)  $E_z[\sup_{\beta \in \mathcal{B}} \|u(z, \beta)\|^\gamma] < \infty$  and  $\zeta(K)^2 K / n^{1-2/\gamma} \rightarrow 0$  some  $\gamma > 2$ .

## Appendix C: Proofs of Results

### C.1 Neglected Heterogeneity Jacobian

Applying L'Hôpital's rule to the ratio

$$\frac{1}{2}\eta^{1/2}E\left[\frac{\partial g(z, \beta_0 + \eta^{1/2}w)}{\partial \beta'}w\right]/\eta,$$

and taking the limits  $\lim_{\eta \rightarrow 0_+}$  of numerator and denominator yields

$$\begin{aligned} G_\eta^j(\beta_0) &= \lim_{\eta \rightarrow 0_+} G_\eta^j(\beta_0, \eta) \\ &= \frac{1}{2} \lim_{\eta \rightarrow 0_+} \frac{1}{2\eta^{1/2}} E\left[\frac{\partial g(z, \beta_0 + \eta^{1/2}w)}{\partial \beta'}w\right] + \frac{1}{4} E_{z,w}[w' \lim_{\eta \rightarrow 0_+} \frac{\partial g^j(z, \beta_0 + \eta^{1/2}w)}{\partial \beta \partial \beta'}w] \\ &= \frac{1}{2} \lim_{\eta \rightarrow 0_+} G_\eta^j(\beta_0, \eta) + \frac{1}{4} \text{tr}(E_{z,w}[\frac{\partial^2 g^j(z, \beta_0)}{\partial \beta \partial \beta'}ww']). \end{aligned}$$

Therefore,

$$G_\eta^j(\beta_0) = \frac{1}{2} \text{tr}(E_{z,w}[\frac{\partial^2 g^j(z, \beta_0)}{\partial \beta \partial \beta'}ww']), (j = 1, \dots, m).$$

■

### C.2 GEL IM Test

The first order condition determining  $\hat{\lambda}(\beta)$  is  $\sum_{i=1}^n \rho_1(\hat{\lambda}(\beta)'g_i(\beta))g_i(\beta) = 0$ . Hence, by the implicit function theorem

$$\begin{aligned} \frac{\partial \hat{\lambda}(\beta)}{\partial \beta'} &= -\left[\sum_{i=1}^n \rho_2(\hat{\lambda}(\beta)'g_i(\beta))g_i(\beta)g_i(\beta)'/n\right]^{-1} \\ &\quad \times \sum_{i=1}^n [\rho_1(\hat{\lambda}(\beta)'g_i(\beta))G_{\beta i}(\beta) + \rho_2(\hat{\lambda}(\beta)'g_i(\beta))g_i(\beta)\hat{\lambda}(\beta)'G_{\beta i}(\beta)]/n. \end{aligned} \tag{C.1}$$

Recall that by Lemma A1 of NS

$$\rho_j(\hat{\lambda}'_0 g_i) \xrightarrow{p} -1, (j = 1, 2),$$

uniformly, ( $i = 1, \dots, n$ ). From the first order condition  $\sum_{i=1}^n \rho_1(\hat{\lambda}'_0 g_i)g_i = 0$ , w.p.a.1,

$$\sqrt{n}\hat{\lambda}_0 = \left[\sum_{i=1}^n \rho_{2i}g_i g_i'/n\right]^{-1} \sqrt{n}\hat{g}(\beta_0) + O_p(n^{-1/2}). \tag{C.2}$$



Thus, by a UWL,

$$\frac{\partial \hat{\lambda}(\beta_0)}{\partial \beta'} = -\Omega^{-1}[G + O_p(n^{-1/2})]$$

where the  $j$ th row of the  $O_p(n^{-1/2})$  term may be written as

$$\hat{\lambda}'_0 E_z[G_{\beta_i} g_i^j] + O_p(n^{-1}), (j = 1, \dots, m).$$

Hence, by a UWL,

$$\begin{aligned} & \sum_{i=1}^n \rho_{2i} G_{\beta_i}^j G_{\beta_i}^{k'} / n \xrightarrow{p} -E[G_{\beta_i}^j G_{\beta_i}^{k'}], \\ & \sum_{i=1}^n \rho_{2i} G_{\beta_i}^j g_i' / n \xrightarrow{p} -E[G_{\beta_i}^j g_i'], (j, k = 1, \dots, m), \\ & \sum_{i=1}^n \rho_{1i} G_{\beta_i} / n \xrightarrow{p} -G_{\beta}, \\ & \sum_{i=1}^n \rho_{2i} g_i g_i' / n \xrightarrow{p} -\Omega \end{aligned}$$

and

$$\sum_{i=1}^n \rho_{1i} \frac{\partial^2 g_i^j(\beta_0)}{\partial \beta \partial \beta'} / n \xrightarrow{p} -E\left[\frac{\partial^2 g_i^j(\beta_0)}{\partial \beta \partial \beta'}\right], (j = 1, \dots, m).$$

Therefore, evaluating the GEL Hessian (4.10) at  $\beta_0$ , the first term of the GEL Hessian is  $O_p(n^{-1})$  and both second and third terms are  $O_p(n^{-1/2})$ . The fourth term consists of two components: the  $O_p(1)$  component

$$-\sum_{i=1}^n \rho_{1i} G'_{\beta_i} \left[ \sum_{i=1}^n \rho_{2i} g_i g_i' / n \right]^{-1} \sum_{i=1}^n \rho_{1i} G_{\beta_i} / n,$$

which by a UWL is a consistent estimator for the asymptotic variance matrix  $G'_{\beta} \Omega^{-1} G_{\beta}$ ,

and the  $O_p(n^{-1/2})$  component

$$-\sum_{i=1}^n \rho_{1i} G'_{\beta_i} \left[ \sum_{i=1}^n \rho_{2i} g_i g_i' \right]^{-1} \sum_{i=1}^n \rho_{2i} g_i \hat{\lambda}'_0 G_{\beta_i} / n$$

Therefore in the GEL context the generalized information equality gives rise to the score

$$\begin{aligned} & \left[ \sum_{i=1}^n \rho_{1i} \frac{\partial^2 g_i'}{\partial \beta_k \partial \beta_l} \right] / n \\ & + \check{G}'_{\beta_k} \check{\Omega}^{-1} \sum_{i=1}^n \rho_{2i} g_i G'_{\beta_l} / n \\ & + \check{G}'_{\beta_l} \check{\Omega}^{-1} \sum_{i=1}^n \rho_{2i} g_i G'_{\beta_k} / n \hat{\lambda}_0, (k \leq l, l = 1, \dots, p), \end{aligned}$$

where  $\check{G}'_{\beta_k} = -\sum_{i=1}^n \rho_{1i} G_{\beta_k i} / n$ , ( $k = 1, \dots, p$ ), and  $\check{\Omega} = -\sum_{i=1}^n \rho_{2i} g_i g_i' / n$ . Asymptotically the implicit Jacobian is

$$\begin{aligned} & E\left[\frac{\partial^2 g(z, \beta_0)'}{\partial \beta_k \partial \beta_l}\right] \\ & + G'_{\beta_k} \Omega^{-1} E[g(z, \beta_0) G_{\beta_l}(z, \beta_0)'] \\ & + G'_{\beta_l} \Omega^{-1} E[g(z, \beta_0) G_{\beta_k}(z, \beta_0)'], (k \leq l, l = 1, \dots, p). \end{aligned}$$

■

### C.3 Empirical Likelihood

The relevant indicators for an EL-based test for the absence of parameter heterogeneity are the non-redundant elements of

$$\frac{1}{n} \sum_{i=1}^n \frac{1}{\hat{\pi}_i(\beta_0, \hat{\lambda}_0)} \frac{\partial^2 \hat{\pi}_i(\beta_0, \hat{\lambda}_0)}{\partial \beta_k \partial \beta_l}, (k \leq l, l = 1, \dots, p).$$

First

$$\frac{\partial \hat{\pi}_i(\beta_0, \hat{\lambda}_0)}{\partial \beta_k} = -\frac{1}{n(1 + \hat{\lambda}'_0 g_i)^2} [\hat{\lambda}'_0 G_{\beta_k i} + \frac{\partial \hat{\lambda}(\beta_0)'}{\partial \beta_k} g_i], (k = 1, \dots, p).$$

Secondly

$$\begin{aligned} \frac{\partial^2 \hat{\pi}_i(\beta_0, \hat{\lambda}_0)}{\partial \beta_k \partial \beta_l} &= \frac{2}{n(1 + \hat{\lambda}'_0 g_i)^3} [\hat{\lambda}'_0 G_{\beta_k i} + \frac{\partial \hat{\lambda}(\beta_0)'}{\partial \beta_k} g_i] [G'_{\beta_l i} \hat{\lambda}_0 + g_i' \frac{\partial \hat{\lambda}(\beta_0)}{\partial \beta_l}] \\ &\quad - \frac{1}{n(1 + \hat{\lambda}'_0 g_i)^2} [G'_{\beta_k i} \frac{\partial \hat{\lambda}(\beta_0)}{\partial \beta_l} + \frac{\partial^2 g_i(\beta_0)'}{\partial \beta_k \partial \beta_l} \hat{\lambda}_0 \\ &\quad + \frac{\partial \hat{\lambda}(\beta_0)'}{\partial \beta_k} G_{\beta_l i} + g_i' \frac{\partial^2 \hat{\lambda}(\beta_0)}{\partial \beta_k \partial \beta_l}]. \end{aligned}$$

Recall from the EL likelihood equations (4.11)

$$\sum_{i=1}^n \frac{1}{n(1 + \hat{\lambda}'_0 g_i)} g_i = 0.$$

Hence,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \frac{1}{\hat{\pi}_i(\beta_0, \hat{\lambda}_0)} \frac{\partial^2 \hat{\pi}_i(\beta_0, \hat{\lambda}_0)}{\partial \beta_k \partial \beta_l} &= \sum_{i=1}^n \frac{2}{n(1 + \hat{\lambda}'_0 g_i)^2} [\hat{\lambda}'_0 G_{\beta_k i} + \frac{\partial \hat{\lambda}(\beta_0)'}{\partial \beta_k} g_i] \\ &\quad \times [G'_{\beta_l i} \hat{\lambda}_0 + g_i' \frac{\partial \hat{\lambda}(\beta_0)}{\partial \beta_l}] \\ &\quad - \sum_{i=1}^n \frac{1}{n(1 + \hat{\lambda}'_0 g_i)} [G'_{\beta_k i} \frac{\partial \hat{\lambda}(\beta_0)}{\partial \beta_l} + \frac{\partial^2 g_i(\beta_0)'}{\partial \beta_k \partial \beta_l} \hat{\lambda}_0 + \frac{\partial \hat{\lambda}(\beta_0)'}{\partial \beta_k} G_{\beta_l i}]. \end{aligned}$$

From the implicit function theorem applied to the likelihood equations (4.11)

$$\frac{\partial \hat{\lambda}(\beta)}{\partial \beta_k} = \check{\Omega}(\beta)^{-1} \sum_{i=1}^n \left[ \frac{1}{n(1 + \hat{\lambda}(\beta)'g_i(\beta))} G_{\beta_{ki}}(\beta) - \frac{1}{n(1 + \hat{\lambda}(\beta)'g_i(\beta))^2} \hat{\lambda}(\beta)' G_{\beta_{ki}}(\beta) g_i(\beta) \right], \quad (k = 1, \dots, p), \quad (\text{C.3})$$

where

$$\check{\Omega}(\beta) = \sum_{i=1}^n \frac{1}{n(1 + \hat{\lambda}(\beta)'g_i(\beta))^2} g_i(\beta) g_i(\beta)'$$

Therefore, substituting for  $\partial \hat{\lambda}(\beta_0)/\partial \beta_k$ , ( $k = 1, \dots, p$ ), from (C.3), after cancelling terms,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \frac{1}{\hat{\pi}_i(\beta_0, \hat{\lambda}_0)} \frac{\partial^2 \hat{\pi}_i(\beta_0, \hat{\lambda}_0)}{\partial \beta_k \partial \beta_l} &= 2\hat{\lambda}'_0 \sum_{i=1}^n \frac{1}{n(1 + \hat{\lambda}'_0 g_i)^2} G_{\beta_{ki}} G'_{\beta_{li}} \hat{\lambda}_0 \\ &\quad - 2\hat{\lambda}'_0 \left[ \sum_{i=1}^n \frac{1}{n(1 + \hat{\lambda}'_0 g_i)^2} G_{\beta_{ki}} g'_i \right] \check{\Omega}(\beta_0)^{-1} \left[ \sum_{i=1}^n \frac{1}{n(1 + \hat{\lambda}'_0 g_i)^2} g_i G'_{\beta_{li}} \right] \hat{\lambda}_0 \\ &\quad + \left[ \sum_{i=1}^n \frac{1}{n(1 + \hat{\lambda}'_0 g_i)} G'_{\beta_{ki}} \right] \check{\Omega}(\beta_0)^{-1} \left[ \sum_{i=1}^n \frac{1}{n(1 + \hat{\lambda}'_0 g_i)^2} g_i G'_{\beta_{li}} \right] \hat{\lambda}_0 \\ &\quad + \left[ \sum_{i=1}^n \frac{1}{n(1 + \hat{\lambda}'_0 g_i)} G'_{\beta_{li}} \right] \check{\Omega}(\beta_0)^{-1} \left[ \sum_{i=1}^n \frac{1}{n(1 + \hat{\lambda}'_0 g_i)^2} g_i G'_{\beta_{ki}} \right] \hat{\lambda}_0 \\ &\quad - \left[ \sum_{i=1}^n \frac{1}{n(1 + \hat{\lambda}'_0 g_i)} \frac{\partial^2 g_i(\beta_0)'}{\partial \beta_k \partial \beta_l} \right] \hat{\lambda}_0 \\ &= \left[ \sum_{i=1}^n \frac{1}{n(1 + \hat{\lambda}'_0 g_i)} G'_{\beta_{ki}} \right] \check{\Omega}(\beta_0)^{-1} \left[ \sum_{i=1}^n \frac{1}{n(1 + \hat{\lambda}'_0 g_i)^2} g_i G'_{\beta_{li}} \right] \hat{\lambda}_0 \\ &\quad + \left[ \sum_{i=1}^n \frac{1}{n(1 + \hat{\lambda}'_0 g_i)} G'_{\beta_{li}} \right] \check{\Omega}(\beta_0)^{-1} \left[ \sum_{i=1}^n \frac{1}{n(1 + \hat{\lambda}'_0 g_i)^2} g_i G'_{\beta_{ki}} \right] \hat{\lambda}_0 \\ &\quad - \left[ \sum_{i=1}^n \frac{1}{n(1 + \hat{\lambda}'_0 g_i)} \frac{\partial^2 g_i(\beta_0)'}{\partial \beta_k \partial \beta_l} \right] \hat{\lambda}_0 + O_p(n^{-1}), \end{aligned}$$

( $k \leq l, l = 1, \dots, p$ ). Note that the first three terms are each  $O_p(n^{-1/2})$ .

## C.4 Proofs of Theorems

**Proof of Theorem 5.1:** First, note that, asymptotically, in the absence of parameter heterogeneity,  $\mathcal{LM}_n - n\hat{g}(\hat{\beta})'\hat{\Omega}(\hat{\beta})^{-1}\hat{G}_\eta(\hat{\beta})S\hat{\Sigma}(\hat{\beta})S'\hat{G}_\eta(\hat{\beta})'\hat{\Omega}(\hat{\beta})^{-1}\hat{g}(\hat{\beta}) \xrightarrow{p} 0$  where the  $r \times (p+r)$  matrix  $S$ , i.e.,  $S = (0, I_r)$ , selects out the components of  $\hat{G}(\beta)$  corresponding to the neglected heterogeneity hypothesis. Secondly,

$$\mathcal{LM}_n - n\hat{g}(\hat{\beta})'\hat{\Omega}(\hat{\beta})^{-1}\hat{G}_\eta(\hat{\beta})S(E_x[D(x)'V(x)^{-1}D(x)])^{-1}S'\hat{G}_\eta(\hat{\beta})'\hat{\Omega}(\hat{\beta})^{-1}\hat{g}(\hat{\beta}) \xrightarrow{p} 0,$$

since  $\hat{\Sigma}(\hat{\beta}) \xrightarrow{p} (E_x[D(x)'V(x)^{-1}D(x)])^{-1}$  by a similar argument to that used in the proof of Lemma A.3, pp.73-75, of DIN. Thirdly, since  $\|\hat{G}(\hat{\beta}) - \hat{G}(\beta_0)\| \xrightarrow{p} 0$ ,  $\|\hat{\Omega}(\hat{\beta}) - \hat{\Omega}(\beta_0)\| \xrightarrow{p} 0$ , by DIN Lemmata A.6, p.78, and A.7, p.79, and

$$\sqrt{n}(\hat{\beta} - \beta_0) - \hat{\Sigma}(\beta_0)\hat{G}_\beta(\beta_0)'\hat{\Omega}(\beta_0)^{-1}\sqrt{n}\hat{g}(\beta_0) \xrightarrow{p} 0,$$

cf. DIN Proofs of Theorems 5.4, pp.81-82, and 5.6, pp.86-87,  $\hat{G}_\eta^c(\hat{\beta})'\hat{\Omega}(\hat{\beta})^{-1}\sqrt{n}\hat{g}(\hat{\beta})$  is asymptotically equivalent to

$$\hat{G}_\eta^c(\beta_0)'\hat{\Omega}(\beta_0)^{-1} - \hat{\Omega}(\beta_0)^{-1}\hat{G}_\beta(\beta_0)\hat{\Sigma}(\beta_0)\hat{G}_\beta(\beta_0)'\hat{\Omega}(\beta_0)^{-1}\sqrt{n}\hat{g}(\beta_0).$$

Now, by Lemma A.3, p.73, of DIN,  $\hat{G}_\eta(\beta_0)'\hat{\Omega}(\beta_0)^{-1}\hat{G}_\beta(\beta_0) \xrightarrow{p} E_x[D_\eta^c(x)'V(x)D_\beta(x)]$  where  $D_\eta^c(x)$  comprises the selected vectors from  $[D_\eta(x, \beta_0)]_{kl} = E_z[\partial^2 u(z, \beta_0)/\partial\beta_k\partial\beta_l|x]$ , ( $k \leq l, l = 1, \dots, p$ ), and  $\hat{G}_\beta(\beta_0)'\hat{\Omega}(\beta_0)^{-1}\hat{G}_\beta(\beta_0) \xrightarrow{p} \mathcal{I}(\beta_0) = E_x[D_\beta(x)'V(x)D_\beta(x)]$ . Furthermore, by DIN, Lemma A.4, p.75,  $\hat{G}(\beta_0)'\hat{\Omega}(\beta_0)^{-1}\sqrt{n}\hat{g}(\beta_0) - \sum_{i=1}^n D(x_i)'V(x_i)^{-1}u_i(\beta_0)/\sqrt{n}$ , where  $D(x_i) = D_\beta(x_i, \beta_0)$ , ( $i = 1, \dots, n$ ). Therefore,  $\hat{G}_\eta^c(\hat{\beta})'\hat{\Omega}(\hat{\beta})^{-1}\sqrt{n}\hat{g}(\hat{\beta})$  is asymptotically equivalent to

$$(-E[D_\eta^c(x)'V(x)^{-1}D_\beta(x)]\mathcal{I}(\beta_0)^{-1}, I_r) \sum_{i=1}^n D(x_i)V(x_i)^{-1}u_i(\beta_0)/\sqrt{n}.$$

Hence, by an i.i.d. CLT,

$$\hat{G}_\eta^c(\hat{\beta})'\hat{\Omega}(\hat{\beta})^{-1}\sqrt{n}\hat{g}(\hat{\beta}) \xrightarrow{d} N(0, [S(E[D(x)'V(x)^{-1}D(x)])^{-1}S']^{-1}). \quad (\text{C.4})$$

The result then follows. ■

**Proof of Theorem 5.2:** By the mean value theorem applied to the first order conditions determining  $\hat{\lambda}$ , i.e.,  $\sum_{i=1}^n \rho_1(\hat{\lambda}'\hat{g}_i)\hat{g}_i = 0$ ,  $\sqrt{n}\hat{\lambda} - (-[\sum_{i=1}^n \rho_2(\hat{\lambda}'\hat{g}_i)\hat{g}_i\hat{g}_i'/n]^{-1}\sqrt{n}\hat{g}(\hat{\beta})) \xrightarrow{p} 0$  for some  $\hat{\lambda}$  on the line segment joining 0 and  $\hat{\lambda}$ . By a similar argument to that used in the proof of Theorem 5.1

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

Moreover,  $\hat{G}_\beta(\beta_0)'\hat{\Omega}(\beta_0)^{-1}\sqrt{n}\hat{g}(\hat{\beta}) \xrightarrow{p} 0$ . Hence, recalling  $S = (0, I_r)$ ,

$$\hat{G}(\hat{\beta})'\sqrt{n}\hat{\lambda} - S\hat{G}_\beta(\beta_0)'\hat{\Omega}(\beta_0)^{-1}\sqrt{n}\hat{g}(\hat{\beta}) \xrightarrow{p} 0.$$

Therefore, the result follows from (C.4). ■

**Proof of Theorem 5.3:** Application of Lemma A.3, p.73, of DIN, yields  $\hat{G}(\hat{\beta})'\hat{\Omega}(\hat{\beta})^{-1}\hat{g}(\hat{\beta}) \xrightarrow{p} E_x[D(x, \beta_*)V(x, \beta_*)^{-1}E_z[u(z, \beta_*)|x]]$  and  $\hat{G}(\hat{\beta})'\hat{\Omega}(\hat{\beta})^{-1}\hat{G}(\hat{\beta}) \xrightarrow{p} E_x[D(x, \beta_*)V(x, \beta_*)^{-1}D(x, \beta_*)]$ ,  
Therefore,

$$\begin{aligned} \mathcal{LM}_n/n &\xrightarrow{p} E_x[E_z[u(z, \beta_*)|x]'V(x, \beta_*)^{-1}D(x, \beta_*)] \\ &\times (E_x[D(x, \beta_*)'V(x, \beta_*)^{-1}D(x, \beta_*)])^{-1}E_x[D(x, \beta_*)V(x, \beta_*)^{-1}E_z[u(z, \beta_*)|x]]. \end{aligned}$$

Since  $E[D_\beta(x, \beta_*)V(x, \beta_*)^{-1}E[u(z, \beta_*)|x]] = 0$ , test consistency requires

$$E[D_\eta^c(x, \beta_*)V(x, \beta_*)^{-1}E[u(z, \beta_*)|x]] \neq 0.$$

■

## References

- BACK, K., AND D. BROWN (1993): “Implied Probabilities in GMM Estimators,” *Econometrica* 61, 971-976.
- BROWN, B.W. AND W.K. NEWAY (1992): “Bootstrapping for GMM”, *mimeo*, M.I.T.
- BROWN, B.W. AND W.K. NEWAY (1998): “Efficient Semiparametric Estimation of Expectations,” *Econometrica* 66, 453-464.
- CHAMBERLAIN, G. (1987): “Asymptotic Efficiency in Estimation with Conditional Moment Restrictions,” *Journal of Econometrics* 34, 305-334.
- CHESHER, A. (1984): “Testing for Neglected Heterogeneity”, *Econometrica* 52, 865-872.
- CRESSIE, N., AND T. READ (1984): “Multinomial Goodness-of-Fit Tests”, *Journal of the Royal Statistical Society Series B* 46, 440-464.
- DONALD, S.G., G.W. IMBENS, AND W.K. NEWAY (2003): “Empirical Likelihood Estimation and Consistent Tests With Conditional Moment Restrictions,” *Journal of Econometrics* 117, 55-93.

- HANSEN, L. P. (1982): “Large Sample Properties of Generalized Method of Moments Estimators”, *Econometrica* 50, 1029-1054.
- HANSEN, L.P., J. HEATON AND A. YARON (1996): “Finite-Sample Properties of Some Alternative GMM Estimators”, *Journal of Business and Economic Statistics* 14, 262-280.
- IMBENS, G.W. (1997): ”One-Step Estimators for Over-Identified Generalized Method of Moments Models,” *Review of Economic Studies* 64, 359-383.
- IMBENS, G.W., R.H. SPADY AND P. JOHNSON (1998): “Information Theoretic Approaches to Inference in Moment Condition Models,” *Econometrica* 66, 333-357.
- KELKER, D. (1970): “Distribution Theory of Spherical Distributions and a Location-Scale Parameter Generalization”, *Sankhyā, A*, 32, 419-430.
- KITAMURA, Y. (2007): “Empirical Likelihood Methods in Econometrics: Theory and Practice”. Chapter 8 in *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society*, Volume 3, eds. R.W. Blundell, W.K. Newey and T. Persson, *Econometric Society Monograph Series*, ESM 43 (2007), 174-237. Cambridge University Press: Cambridge.
- KITAMURA, Y., AND M. STUTZER (1997): “An Information-Theoretic Alternative to Generalized Method of Moments Estimation”, *Econometrica* 65, 861-874.
- LEE, L.-F. AND A. CHESHER (1986): “Specification Testing when Score Test Statistics are Identically Zero”, *Journal of Econometrics* 31, 121-149.
- NEWAY, W.K. (1985a): “Generalized Method of Moments Specification Testing”, *Journal of Econometrics* 29, 229-256.
- NEWAY, W.K. (1985b): “Maximum Likelihood Specification Testing and Conditional Moment Test”, *Econometrica* 53, 1047-1070.

- NEWKEY, W.K. (1997): "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics* 79, 147-168.
- NEWKEY, W.K. AND D. MCFADDEN (1994): "Large Sample Estimation and Hypothesis Testing," in Engle, R. and D. McFadden, eds., *Handbook of Econometrics, Vol. 4*, New York: North Holland.
- NEWKEY, W.K., AND R.J. SMITH (2004): "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica* 72, 219-255.
- NEWKEY, W.K., AND K.D. WEST (1987): "Hypothesis Testing with Efficient Method of Moments Estimation," *International Economic Review* 28, 777-787.
- OWEN, A. (1988): "Empirical Likelihood Ratio Confidence Intervals for a Single Functional," *Biometrika* 75, 237-249.
- OWEN, A. (2001): *Empirical Likelihood*, New York: Chapman and Hall.
- QIN, J. AND LAWLESS, J. (1994): "Empirical Likelihood and General Estimating Equations", *Annals of Statistics* 22, 300-325.
- SMITH, R. J. (1997): "Alternative Semi-Parametric Likelihood Approaches to Generalized Method of Moments Estimation", *Economic Journal* 107, 503-519.
- SMITH, R. J. (2010): "GEL Methods for Moment Condition Models." Forthcoming *Econometric Theory*. Earlier version CWP 19/04, cemmap, I.F.S. and U.C.L., available at <http://cemmap.ifs.org.uk/wps/cwp0419.pdf>.
- WHITE, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica* 48, 817-838.
- WHITE, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica* 50, 1-25.