Neglected Heterogeneity in Moment Condition Models*

Jinyong Hahn
Department of Economics
U.C.L.A.
8283 Bunche Hall
Mail Stop: 147703
Los Angeles, CA 90095
U.S.A.

Whitney K. Newey
Department of Economics
M.I.T.
50 Memorial Drive
E52-391
Cambridge, MA 02142
U.S.A.

Richard J. Smith†
Centre for Microdata Methods and Practice
Institute for Fiscal Studies
7 Ridgmount Street
London WC1E 7AE
U.K.
Faculty of Economics
University of Cambridge
Austin Robinson Building
Sidgwick Avenue
Cambridge CB3 9DD
U.K.
and
Economics and Finance Department
University of Canterbury
Private Bag 4800
Christchurch 8140
New Zealand

This Draft: June 2013

Abstract

The central concern of this paper is parameter heterogeneity in models specified by a number of unconditional or conditional moment conditions and thereby the provision of a framework for the development of apposite optimal m-tests against its potential presence. We initially consider the unconditional moment restrictions framework. Optimal m-tests against moment condition parameter heterogeneity are derived with the relevant Jacobian matrix obtained in terms of the second order own derivatives 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 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 and the provision of a limited set of simulation experiments to illustrate the efficacy of the proposed tests.

JEL Classification: C13, C30

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

*We are grateful for the constructive comments from the Editor and two referees that have helped to improve this paper. Earlier versions were presented at Nuffield College, Oxford, the Universities of Cambridge and Nottingham, Victoria University of Wellington, the Conference in Honour of P.C.B. Phillips, University of York, 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. Part of the research for the preparation of the initial version of the paper 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. The authors would also like to thank C.E. Hernandez and J. Kang who provided excellent research assistance.

†Corresponding author at: Faculty of Economics, University of Cambridge, Austin Robinson Building, Sidgwick Avenue, Cambridge CB3 9DD, UK. *E-mail address*: rjs27@econ.cam.ac.uk (R.J. Smith)

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 parameter heterogeneity in models specified by moment conditions and thereby the provision of a framework for the development of apposite optimal m-tests against its potential presence. We consider both unconditional and conditional model settings. 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 order own derivatives 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 efficient two-step (2S) generalized method of moments (GMM) [Hansen (1982)] and generalized empirical likelihood (GEL) [Newey and Smith (2004), henceforth NS, and Smith (1997, 2011)] estimation.

The 2SGMM-based version of the generalized IM test statistic employs all second derivatives including cross-derivatives of the moments. The GEL form is associated with a more general form of parameter heterogeneity test involving additional components that may be interpreted in terms of a particular correlation structure linking the sample Jacobian and the random variates 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 against parameter heterogeneity in Chesher (1984). We then consider the effect of parameter heterogeneity on 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 asymptotic local power. Section 4 of the paper provides moment specification tests obtained by consideration of generalized forms of the IM equality appropriate for efficient 2SGMM and GEL estimation. These statistics are then compared with those against moment condition parameter heterogeneity developed in section 3. Components of the 2SGMM form coincide with those of section 3 whereas the GEL statistic incorporates additional terms that implicitly allow for a particular form of correlation between the sample Jacobian and the random variates 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. Section 6 provides a set of simulation experiments to illustrate the potential efficacy for empirical research of the tests proposed in the paper. The paper is concluded in Section 7. The Appendices contain relevant assumptions and proofs of results and assertions made in the main text.

Throughout the text (x_i, z_i), $(i = 1, ..., n)$, will denote i.i.d. observations on the observable d_x-dimensional covariate or instrument vector x and the d_z-dimensional vector z that may include a sub-vector of x. The vector β 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 jth 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 β 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 β_0 with limiting representation

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

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

$$n^{1/2}(\hat{\beta}_{ML} - \beta_0) \xrightarrow{d} N(0, \mathcal{I}(\beta_0)^{-1}).$$

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 β demonstrates that the density function $f(z, \cdot)$ obeys the familiar IM *equality*

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

Therefore, under correct specification, i.e., z distributed with density function $f(z, \beta_0)$, and given the consistency of $\hat{\beta}_{ML}$ for β_0, by an i.i.d. UWL, the contrast with zero

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

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[1,2]

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

[1] 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.

[2] Chesher and Smith (1997) provides a likelihood ratio form of (conditional) moment specification test. An attractive feature of this test is that it admits a "Bartlett correction" by division by a scale factor that creates a statistic with higher order accuracy as compared to conventional moment-based tests.

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 β as a random vector and the density $f(z, \beta)$ as the *conditional* density of z given β. Absence of parameter heterogeneity corresponds to $\beta = \beta_0$ almost surely.

Suppose that the marginal density of β 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 β, $E[\beta] = \beta_0$, where $E[\cdot]$ here denotes expectation with respect to the density of β. Equivalently, writing $\beta = \beta_0 + \eta^{1/2} w$, w has the symmetric continuous density $h(w'w)$. Thus, 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 η 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[3]

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

[3] Alternatively specifying the marginal density of β as $\kappa^{-p} h((\beta - \beta_0)/\kappa)$ with $h(\cdot)$ symmetric and κ 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 respec to κ

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

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

Taking the limit $\lim_{\eta \to 0_+}$ gives the score for η as

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

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

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

i.e., the indicators on which the IM test statistic is based; cf. (2.2). See Chesher (1984, p.686). Apart from the scalar parameter case, however, the classical IM test is sub-optimal examining a wider alternative hypothesis than neglected parameter heterogeneity, an optimal test being based on the diagonal elements of (2.4); see the discussion preceding Theorem 3.1 below.

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, β a $p \times 1$ parameter vector with $m \geq p$. In the absence of parameter heterogeneity, we assume there is a true parameter value β_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 efficient 2SGMM and GEL estimators for β_0 obtained under the moment constraint (3.1) where there is no possibility of confusion; see sections 4.1 and 4.2 below where 2SGMM and GEL are briefly described.

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 α 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 $\alpha_0 = 0$ with the two sided alternative hypothesis $\alpha_0 \neq 0$; we then write the vector of moment functions under $\alpha_0 = 0$, cf. (3.1), as $g(z, \beta) = g(z; 0, \beta)$. Let $g_i(\beta) = g(z_i, \beta)$, $(i = 1, ..., n)$, 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 $\alpha_{0n} = \delta/n^{1/2}$ 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.,

$$n\hat{g}(\hat{\beta})'\hat{\Omega}^{-1}\hat{G}(\hat{G}'\hat{\Omega}^{-1}\hat{G})^{-1}\hat{G}'\hat{\Omega}^{-1}\hat{g}(\hat{\beta}), \tag{3.2}$$

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

3.2 Neglected Heterogeneity

The approach adopted here is similar to that of Chesher (1984) for the classical likelihood context described above in section 2.3. As there, for ease of exposition, we centre β at β_0 and write

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

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

Assumption 3.1 *(Parameter Heterogeneity.) The parameter vector β is a random vector with (unconditional) mean β_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. fn.3 above 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 β, i.e., w,

$$E_z[g(z, \beta)|w] = 0, \tag{3.3}$$

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 κ is then given by

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

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

$$\begin{aligned} G_\kappa &= G_\kappa(\beta_0, 0) \\ &= E_{z,w}[\frac{\partial g(z, \beta_0)}{\partial \beta'} w] \\ &= E_w[E_z[\frac{\partial g(z, \beta_0)}{\partial \beta'} | w] w]. \end{aligned} \tag{3.4}$$

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_κ is not identically zero unless w and $\partial g(z, \beta_0)/\partial \beta'$ are uncorrelated. However, the Jacobian expression (3.4) 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 the leading case in which the null hypothesis Jacobian G_κ is identically zero, i.e., conditions under which w and $\partial g(z, \beta_0)/\partial \beta'$ are uncorrelated are examined. We return to the general case in section 4 when we consider generalized IM statistics appropriate for the moment condition context and where this leading case implicitly features centrally in the analysis; see, in particular, sections 4.1 and 4.2.

First, G_κ is identically zero if the derivative matrix $\partial g(z, \beta_0)/\partial \beta'$ is conditionally mean independent of w since from (3.4) then

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

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 κ 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 κ 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; see section 5. Consequently, we re-interpret the moment condition under parameter heterogeneity (3.3) 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 β is a random vector with conditional mean β_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 (3.4) with respect to κ is then

$$\begin{aligned} G_\kappa(\beta_0, \kappa) &= E_x[E_{z,w}[\frac{\partial g(z, \beta_0 + \kappa w)}{\partial \kappa}|x]] \qquad (3.5) \\ &= E_x[E_{z,w}[\frac{\partial g(z, \beta_0 + \kappa w)}{\partial \beta'} w|x]]. \end{aligned}$$

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 κ 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 (3.5) 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 κ 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. fn.3 in 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 κ 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 η is

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

Evaluation at $\eta = 0$ results in the indeterminate ratio $0/0$. Define $G^j_\eta(\beta_0, \eta) = E_{z,w}[\partial g^j(z, \beta)/\partial \eta]$, $(j = 1, ..., m)$. Applying L'Hôpital's rule to the ratio

$$\frac{1}{2}\eta^{1/2} E_{z,w}[\frac{\partial g(z, \beta_0 + \eta^{1/2} w)}{\partial \beta'} w]/\eta, \tag{3.6}$$

and taking the limits $\lim_{\eta \to 0_+}$ of numerator and denominator in (3.6), results in the following expression for the Jacobian with respect to η at $\eta = 0$

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

See Appendix C.1.

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

$$\begin{aligned} G^j_\eta &= \frac{1}{2} tr(E_w[E_z[\frac{\partial^2 g^j(z, \beta_0)}{\partial \beta \partial \beta'} | w] ww']) \\ &= \frac{1}{2} tr(E_z[\frac{\partial^2 g^j(z, \beta_0)}{\partial \beta \partial \beta'}] var_w[w]), (j = 1, ..., m). \end{aligned} \tag{3.7}$$

Cf. Corollary 3.1.

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

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

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

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

Cf. Corollary 3.2. Moreover, if the random variation in β, i.e., w, is also second moment independent of x, $var_w[w|x] = var_w[w]$, the resultant Jacobian is identical to (3.7), i.e.,

$$G^j_\eta = \frac{1}{2} tr(E_z[\frac{\partial^2 g^j(z, \beta_0)}{\partial \beta \partial \beta'}] var_w[w]), (j = 1, ..., m).$$

When β is scalar, from (3.7) the Jacobian with respect to η is

$$G_\eta^j = \frac{1}{2} var_w[w] E_z[\frac{\partial^2 g^j(z,\beta_0)}{\partial\beta^2}], (j = 1, ..., m);$$

cf. Chesher (1984, p. 868). In this case and under the above conditions the Jacobian of an optimal m-test against neglected parameter heterogeneity equals the vector $E_z[\partial^2 g(z,\beta_0)/\partial\beta^2]$ since the LM-type statistic (3.2) is invariant to the scalar heterogeneity variance $var_w[w]$; cf. section 3.1. More generally for vector β, by analogy with the discussion in Chesher (1984, p.868) and in section 2.3, the expression (3.7) suggests a form for the Jacobian given by the non-redundant elements of

$$E_z[\frac{\partial^2 g(z,\beta_0)}{\partial\beta_k\partial\beta_l}], (k \leq l, l = 1, ..., p). \tag{3.8}$$

Although the LM-type statistic (3.2) based on the Jacobian definition (3.8) may offer an efficacious test against parameter heterogeneity it examines a wider alternative hypothesis than that of parameter heterogeneity and as such is likely to be sub-optimal in the sense of section 3.1. To see this, consider a re-specification of the elements of the heterogeneous parameter vector β as $\beta_k = \beta_{0k} + \eta_k^{1/2} w_k$, $(k = 1, ..., p)$, that allows for individual parameter heterogeneity, i.e., the absence of parameter heterogeneity now corresponds to the hypothesis $\eta_k = 0$, $(k = 1, ..., p)$. Let $G_{\eta_k}(\beta_0, \eta) = E_{z,w}[\partial g(z,\beta)/\partial\eta_k]$, $(k = 1, ..., p)$, where $\eta = (\eta_1, ..., \eta_p)'$. A similar analysis to that following Corollary 3.2 leading to (3.7) yields the Jacobian with respect to η_k

$$\begin{aligned} G_{\eta_k} &= \lim_{\eta\to 0_+} G_{\eta_k}(\beta_0, \eta) \\ &= \frac{1}{2} E_{z,w}[\frac{\partial^2 g(z,\beta_0)}{\partial\beta_k^2} w_k^2] \\ &= \frac{1}{2} var_w[w_k] E_z[\frac{\partial^2 g(z,\beta_0)}{\partial\beta_k^2}], (k = 1, ..., p). \end{aligned}$$

The disparity between the Jacobians formed from (3.8) and (3.9) and the resultant difference in statistic degrees of freedom may be accounted for by the former essentially failing to account for the covariances between the elements $\eta_k^{1/2} w_k$, $(k = 1, ..., p)$, taking the value zero in the absence of parameter heterogeneity, $\eta_k = 0$, $(k = 1, ..., p)$, i.e., the off-diagonal components $\partial^2 g(z,\beta_0)/\partial\beta_k\partial\beta_l$, $(k<l, l=1,...,p)$, in the Jacobian (3.8) are irrelevant to the construction of an optimal m-statistic against parameter heterogeneity. As in the scalar case the optimal LM-type statistic (3.2) based on (3.9) is invariant to the heterogeneity variances $var_w[w_k]$, $(k=1,...,p)$.[4]

We summarise the above development in the following result.

Theorem 3.1 *Either (a) under Assumption 3.1, if $\partial g(z,\beta_0)/\partial\beta'$ and $\partial^2 g(z,\beta_0)/\partial\beta_k^2$, $(k=1,...,p)$, are conditionally mean independent of w, or (b) under Assumption 3.2, if $\partial g(z,\beta_0)/\partial\beta'$ and $\partial^2 g(z,\beta_0)/\partial\beta_k^2$, $(k=1,...,p)$, 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 linearly independent vectors comprising*

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

3.3 Test Statistics

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

$$G_{\beta i}(\beta)=\frac{\partial g_i(\beta)}{\partial\beta'}, [G_{\eta i}(\beta)]_k=\frac{\partial^2 g_i(\beta)}{\partial\beta_k^2}, (k=1,...,p).$$

We stack the vectors $[G_{\eta i}(\beta)]_k$, $(k=1,...,p)$, as columns of the $m\times p$ matrix $G_{\eta i}(\beta)$, $(i=1,...,n)$.

Let $G_\beta=E_z[\partial g(z,\beta_0)/\partial\beta']$ and $[G_\eta]_k=E_z[\partial^2 g(z,\beta_0)/\partial\beta_k^2]$, $(k=1,...,p)$, stacked similarly to $[G_{\eta i}(\beta)]_k$, $(k=1,...,p)$, as the columns of the $m\times p$ matrix G_η. As in the classical case, there may be a linear dependence among the columns of the population matrix G_η taken together with G_β. Moreover, for economic theoretic reasons, parameter heterogeneity may only be suspected in a subset of the elements of β. Therefore, we

[4] The asymptotic local optimality of the LM-type statistic (3.2) concerns tests against a two-sided local alternative of the form $\alpha_{0n}=\delta/n^{1/2}$ where $\delta\neq 0$. Correspondingly, as in section 3.1, the alternative hypothesis for individual parameter heterogeneity is $\eta\neq 0$, i.e., $\eta_k\neq 0$ for at least one k, $(k=1,...,p)$. Strictly speaking, the relevant alternative hypothesis is the one-sided hypothesis $\eta\geq 0$ with $\eta_k>0$ for at least one k, $(k=1,...,p)$. Tests incorporating these inequality restrictions, see *inter alia* Andrews (2001), would display greater power than the LM-type statistic (3.2) but at the expense of increased computational complexity.

adopt the notation G^c_η for those non-redundant r columns chosen from G_η with $G^c_{\eta i}(\beta)$, $(i = 1, ..., n)$, their sample counterparts..

To define the requisite 2SGMM and GEL statistics, define the sample Jacobian estimators $\hat{G}_\beta(\beta) = \sum_{i=1}^n G_{\beta i}(\beta)/n$, $\hat{G}^c_\eta(\beta) = \sum_{i=1}^n G^c_{\eta i}(\beta)/n$ and write $\hat{G}(\beta) = (\hat{G}_\beta(\beta), \hat{G}^c_\eta(\beta))$ with the consequent $\hat{\Sigma}(\beta) = (\hat{G}(\beta)'\hat{\Omega}(\beta)^{-1}\hat{G}(\beta))^{-1}$.

The optimal 2SGMM or GEL LM-type statistic for neglected heterogeneity is

$$\mathcal{LM}_n = n\hat{g}(\hat{\beta})'\hat{\Omega}(\hat{\beta})^{-1}\hat{G}(\hat{\beta})\hat{\Sigma}(\hat{\beta})\hat{G}(\hat{\beta})'\hat{\Omega}(\hat{\beta})^{-1}\hat{g}(\hat{\beta}); \tag{3.9}$$

see Newey and West (1987) and Smith (2011). Given the optimal 2SGMM 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 (3.9) may also be defined as

$$\mathcal{S}_n = n\hat{\lambda}'\hat{G}(\hat{\beta})\hat{\Sigma}(\hat{\beta})\hat{G}(\hat{\beta})'\hat{\lambda}. \tag{3.10}$$

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

The limiting distributions of the statistics $\mathcal{LM}_n$ (3.9) and $\mathcal{S}_n$ (3.10) in the absence of parameter heterogeneity may then be described. Let $\mathcal{N}$ denote a neighbourhood of β_0. Let $G = (G_\beta, G^c_\eta)$.[5]

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

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

Assumptions A.1 and A.2 repeat NS Assumptions 1 and 2, p.226, respectively, with Assumption A.3 the consistency of the preliminary estimator $\tilde{\beta}$ required for efficient

[5] Test statistics $\mathcal{LM}_n$ (3.9) and $\mathcal{S}_n$ (3.10) based on the Jacobian (3.8) require the definitions $[G_{\eta i}(\beta)]_{kl} = \partial^2 g_i(\beta)/\partial\beta_k\partial\beta_l$, $(k \leq l, l = 1, ..., p)$, with the vectors $[G_{\eta i}(\beta)]_{kl}$, $(k \leq l, l = 1, ..., p)$, as columns of the $m \times p(p+1)/2$ matrix $G_{\eta i}(\beta)$, $(i = 1, ..., n)$. Similarly $[G_\eta]_{kl} = E_z[\partial^2 g(z, \beta)/\partial\beta_k\partial\beta_l]$, $(k \leq l, l = 1, ..., p)$, stacked similarly to $[G_{\eta i}(\beta)]_{kl}$, $(k \leq l, l = 1, ..., p)$, form the columns of the $m \times p(p+1)/2$ matrix G_η. Theorem 3.2 requires the condition $E[\sup_{\beta \in \mathcal{N}} \|\partial^2 g(z, \beta)/\partial\beta_k\partial\beta_l\|] < \infty$, $(k \leq l, l = 1, ..., p)$, in place of $E_z[\sup_{\beta \in \mathcal{N}} \|\partial^2 g(z, \beta)/\partial\beta_k^2\|] < \infty$, $(k = 1, ..., p)$.

2SGMM, cf. NS Assumption A.4, p.227. As noted in sections 4.1 and 4.2 below Assumptions A.1-A.3 are sufficient for the consistency for β_0 and asymptotic normality of 2SGMM and GEL. Moreover, by a UWL, $\hat{\Omega}(\hat{\beta}) \xrightarrow{p} \Omega$. The consistency of $\hat{G}(\hat{\beta})$ for G also follows by a UWL from $E_z[\sup_{\beta\in\mathcal{N}} \|\partial^2 g(z,\beta)/\partial\beta_k^2\|] < \infty$, $(k = 1, ..., p)$, with the additional hypothesis $rank(G) = p + r$, $p + r \leq m$, of Theorem 3.2 ensuring $G'\Omega^{-1}G$ is p.d. Hence, $\hat{\Sigma}(\hat{\beta}) \xrightarrow{p} (G'\Omega^{-1}G)^{-1}$. Since $n^{1/2}\hat{g}(\hat{\beta}) \xrightarrow{d} N(0, \Omega - G_\beta(G'_\beta\Omega^{-1}G_\beta)^{-1}G'_\beta)$ by a CLT the limiting distribution of $\mathcal{LM}_n$ (3.9) and, likewise, that of $\mathcal{S}_n$ (3.10) stated in Theorem 3.2 follow directly; see, e.g., Newey and West (1987) and Smith (2011).

Note that the Jacobian estimator $\hat{G}(\beta)$ may equivalently be replaced by its GEL counterpart $\tilde{G}(\beta) = (\tilde{G}_\beta(\beta)', \tilde{G}_\eta(\beta)) = \sum_{i=1}^n \hat{\pi}_i(\beta, \hat{\lambda}(\beta))(G_{\beta i}(\beta)', G^c_{\eta i}(\beta)')$, where the implied probabilities $\hat{\pi}_i(\beta, \hat{\lambda}(\beta))$, $(i = 1, ..., n)$, 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.,

$$E_{z,w}[\frac{\partial g(z, \beta_0)}{\partial\beta'} w] = 0, \tag{3.11}$$

necessitating the use of L'Hôpital's rule to obtain the Jacobian with respect to η 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 mimics the classical IM test of White (1982) discussed in section 2.2. Indeed, the GEL-based test defined in section 4.2 incorporates a component that corresponds to an implicit particular correlation structure between w and $\partial g(z, \beta_0)/\partial\beta'$; see the discussion following (4.10) below.

4 Generalized Information Matrix Tests

Optimal 2SGMM or GEL tests for neglected heterogeneity based on the moment indicator own second derivatives $\partial^2 g(z, \beta_0)/\partial\beta_k^2$, $(k = 1, ..., p)$, described in section 3.3, may also be interpreted in terms of 2SGMM and GEL versions of a generalized IM equality. As is well known, see, e.g., Tauchen (1985), the 2SGMM 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 β is the efficient 2SGMM estimator due to Hansen (1982). Suppose $\tilde{\beta}$ is a preliminary consistent estimator for β_0. The 2SGMM estimator is defined as

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

Under (3.1) and, in particular, Assumptions A.1(a)-(e) and A.3 of the Appendix it is straightforward to show that $\hat{\beta}_{2S} \xrightarrow{p} \beta_0$ and, in addition if Assumption A.2 holds, $\hat{\beta}_{2S}$ is asymptotically normally distributed, i.e., $n^{1/2}(\hat{\beta}_{2S} - \beta_0) \xrightarrow{d} N(0, \Sigma_{\beta\beta})$ where $\Sigma_{\beta\beta} = (G_\beta'\Omega^{-1}G_\beta)^{-1}$. See, e.g., Newey and McFadden (1994, Theorems 2.6, p.2132, and 3.4, p.2148). 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 $\Sigma_{\beta\beta}$, i.e., the asymptotic variance of the 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 β in $\hat{\Omega}(\beta)$, i.e., the CUE is given by

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

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

Now consider the rescaled 2SGMM objective function

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

To describe a generalized IM equality similar to (2.2) for the 2SGMM criterion $\hat{Q}(\beta)$, first, under Assumptions A.1-A.3, by a UWL and a CLT, the limiting normal distribution associated with the score $\partial\hat{Q}(\beta_0)/\partial\beta = \hat{G}_\beta(\beta_0)'\hat{\Omega}(\tilde{\beta})^{-1}\hat{g}(\beta_0)$ obtained from (4.3) may be stated as

$$n^{1/2}\frac{\partial\hat{Q}_n(\beta_0)}{\partial\beta} \xrightarrow{d} N(0, G'_\beta\Omega^{-1}G_\beta).$$

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.*

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

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 2SGMM-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 $n^{1/2}\partial\hat{Q}(\beta_0)/\partial\beta$, i.e., the generalized IM $G'_\beta\Omega^{-1}G_\beta$, with the Hessian evaluated at the 2SGMM estimator $\hat{\beta}_{2S}$,

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

A standard estimator for the generalized IM $G'_\beta\Omega^{-1}G_\beta$ is $\hat{G}_\beta(\hat{\beta}_{2S})'\hat{\Omega}(\tilde{\beta})^{-1}\hat{G}_\beta(\hat{\beta}_{2S})$. This estimator also has an interpretation as an outer product form of estimator based on the "scores" $\hat{G}_\beta(\hat{\beta}_{2S})'\hat{\Omega}(\tilde{\beta})^{-1}g_i(\tilde{\beta})$, $(i = 1, ..., n)$; 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 2SGMM IM specification test statistic is therefore based on the non-redundant "scores" from

$$\frac{\partial^2\hat{g}(\hat{\beta}_{2S})'}{\partial\beta_k\partial\beta_l}\hat{\Omega}(\tilde{\beta})^{-1}\hat{g}(\hat{\beta}_{2S}), (k \le l, l = 1, ..., p).$$

The Jacobian underpinning this generalized 2SGMM IM specification test statistic is thereby constructed from

$$E_z[\frac{\partial^2 g(z,\beta_0)}{\partial\beta_k\partial\beta_l}], (k \le l, l = 1, ..., p), \tag{4.4}$$

which is identical to the Jacobian (3.8) defined analogously to the classical IM statistic against neglected heterogeneity of Chesher (1984) given in section 2.3. As discussed in section 3.2 a generalized 2SGMM IM specification statistic $\mathcal{LM}_n$ (3.9) or $\mathcal{S}_n$ (3.10) based on the above "scores" is likely to offer a less powerful test against parameter heterogeneity than those proposed in section 3.3 since the diagonal components $\partial^2 g(z,\beta_0)/\partial\beta_k^2$, $(k = 1,...,p)$, only in the Jacobian (4.4) are relevant for the construction of an optimal *m*-statistic against parameter heterogeneity; see Theorems 3.1 and 3.2. Recall from section 3.2 that the formulation (4.4) 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.

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, 2011). Indeed CUE (4.2) is included as a special case of GEL; see fn.6 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 $\rho_j = \rho_j(0)$, $(j = 0,1,2,...)$. Let $\hat{\Lambda}_n(\beta) = \{\lambda : \lambda' g_i(\beta) \in \mathcal{V}, i = 1,...,n\}$. The GEL criterion is defined as

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

with the GEL estimator $\hat{\beta}$ of β given as the solution to a saddle point problem; *viz.*[6]

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

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 A.1 of the Appendix, by NS Theorem 3.1, p.226, $\hat{\beta} \xrightarrow{p} \beta_0$ and $\hat{\lambda} \xrightarrow{p} 0$ and, with the additional

[6] 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].

Assumption A.2, $n^{1/2}(\hat{\beta}-\beta_0) \xrightarrow{d} N(0, \Sigma_{\beta\beta})$ and $n^{1/2}\hat{\lambda} \xrightarrow{d} N(0, \Omega^{-1} - \Omega^{-1} G_\beta \Sigma_{\beta\beta} G'_\beta \Omega^{-1})$ by 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

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

cf. NS and Brown and Newey (1992, 2002).[7]

A similar analysis to that described above in section 4.1 for 2SGMM 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 λ with $\hat{\lambda}(\beta)$, i.e.,

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

Hence, by the envelope theorem, the score with respect to β is

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

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

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

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 $G_{\beta_k i}(\beta) = \partial g_i(\beta)/\partial \beta_k$, $(k = 1, ..., p)$, $(i = 1, ..., n)$.

[7] The GEL empirical probabilities $\hat{\pi}_i(\beta, \hat{\lambda}(\beta))$, $(i = 1, ..., n)$, 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 $\hat{\lambda}(\hat{\beta})' \hat{g}_i(\hat{\beta})$ 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 $E_z[a(z, \beta_0)]$.

Evaluating the Hessian (4.10) at β_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

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

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

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

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 $\hat{G}_{\beta_k i} = G_{\beta_k i}(\hat{\beta})$, $(k = 1,...,p)$, $(i = 1,...,n)$. Similarly to 2SGMM, 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

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

this estimator has the approximate interpretation as an outer product form of estimator based on the "scores" $-[\sum_{i=1}^{n}\hat{\rho}_{1i}\hat{G}'_{\beta i}/n][\sum_{i=1}^{n}\hat{\rho}_{2i}\hat{g}_i\hat{g}_i'/n]^{-1}\sqrt{-\hat{\rho}_{2i}}\hat{g}_i$, $(i = 1,...,n)$; 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

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

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

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

where $G_{\beta_k} = E_z[\partial g(z, \beta_0)/\partial \beta_k]$, $(k = 1, ..., p)$.

The first term in (4.10) is identical to the 2SGMM Jacobian (4.4) but, interestingly, the second and third terms are absent for 2SGMM. This occurs because of the use of the preliminary consistent estimator $\tilde{\beta}$ to estimate Ω in 2SGMM whereas GEL implicitly also optimises a variance component over β, 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" $G'_{\beta_k}\Omega^{-1}g(z, \beta_0)$, $(k \leq l, l = 1, ..., p)$. Cf. G_κ (3.4). These terms thereby implicitly allow for a particular form of correlation between the heterogeneity variate w and the sample Jacobian $\partial g(z, \beta_0)/\partial \beta'$. To see how this arises rewrite the implicit Jacobian (4.10) as the non-redundant terms of

$$E_z[\frac{\partial^2 g_0^j}{\partial \beta \partial \beta'}] + G'_\beta \Omega^{-1} E_z[g_0 \frac{\partial g_0^j}{\partial \beta'}] + E_z[\frac{\partial g_0^j}{\partial \beta} g_0']\Omega^{-1} G_\beta, (j = 1, ..., m),$$

where $g_0 = g(z, \beta_0)$, $\partial g_0^j/\partial \beta = \partial g^j(z, \beta_0)/\partial \beta$ and $\partial^2 g_0^j/\partial \beta \partial \beta' = \partial^2 g^j(z, \beta_0)/\partial \beta \partial \beta'$, $(j = 1, ..., m)$. Suppose that the heterogeneity random vector w, see section 3.2, may be decomposed as $w = \eta^{1/2} var_v[v] G'_\beta \Omega^{-1} g_0 + v$. The random vector v is assumed here to satisfy hypotheses (a) and (b) on w in Theorem 3.1. Hence, the covariance between the heterogeneity variate w and the sample Jacobian $\partial g(z, \beta_0)/\partial \beta'$ is $\eta^{1/2} var_v[v] G'_\beta \Omega^{-1} E_z[g_0 \partial g_0^j/\partial \beta']$, $(j = 1, ..., m)$. Then, a similar analysis to that in section 3.2 yields,

$$\begin{aligned} G_\eta^j &= \frac{1}{2} tr(E_z[\frac{\partial^2 g_0^j}{\partial \beta \partial \beta'}] var_v[v]) + tr(E_z[\frac{\partial g_0^j}{\partial \beta} g_0']\Omega^{-1} G_\beta) var_v[v]) \\ &= \frac{1}{2} tr((E_z[\frac{\partial^2 g_0^j}{\partial \beta \partial \beta'}] + E_z[\frac{\partial g_0^j}{\partial \beta} g_0']\Omega^{-1} G_\beta + G'_\beta \Omega^{-1} E_z[g_0 \frac{\partial g_0^j}{\partial \beta'}]) var_v[v]), \end{aligned}$$

$(j = 1, ..., m)$; cf. (3.8). Therefore, the implicit Jacobian (4.10) corresponds to one associated with an LM-type statistic (3.2) defined analogously to the classical IM statistic against neglected parameter heterogeneity; see section 2.3.

The discussion in section 3.2 preceding Theorem 3.1 suggests that an optimal m-statistic against parameter heterogeneity that incorporates the above form of implicit correlation between the heterogeneity variate w and the sample Jacobian $\partial g(z,\beta_0)/\partial\beta'$ would be based on the linearly independent vectors comprising

$$E_z[\frac{\partial^2 g_0}{\partial\beta_k^2}]+G'_{\beta_k}\Omega^{-1}E_z[g_0\frac{\partial g'_0}{\partial\beta_k}]+E_z[\frac{\partial g_0}{\partial\beta_k}g'_0]\Omega^{-1}G_{\beta_k},(k=1,...,p). \tag{4.11}$$

In this case, the heterogeneity random vector w is decomposed as $w_k=\eta_k^{1/2}var_v[v_k]G'_{\beta_k}\Omega^{-1}g_0+v_k$, $(k=1,...,p)$.

Note that the additional terms in (4.10) and (4.11) 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 β_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 $-\breve\Omega(\hat\beta)^{-1}n^{1/2}\hat g(\hat\beta)$ for $n^{1/2}\hat\lambda(\hat\beta)$ since $n^{1/2}\hat\lambda(\hat\beta)=-\Omega^{-1}n^{1/2}\hat g(\hat\beta)+O_p(n^{-1/2})$ in the absence of parameter heterogeneity.

4.3 An Example: Empirical Likelihood[8]

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 propbabilities, cf. (4.7), are

$$\hat\pi_i(\beta,\lambda)=\frac{1}{n(1+\lambda' g_i(\beta))},(i=1,...,n),$$

[8]We are grateful to Y. Kitamura for suggesting this example.

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

$$\begin{aligned}0 &= -\sum_{i=1}^{n}\hat{\pi}_i(\beta,\hat{\lambda}(\beta))g(z_i,\beta) \qquad (4.12)\\ &= -\sum_{i=1}^{n}\frac{1}{n(1+\hat{\lambda}(\beta)'g_i(\beta))}g_i(\beta),\end{aligned}$$

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

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

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

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

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

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

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

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

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 $\rho_{2i} = -1/n(1-\hat{\lambda}_0' g_i)^2$, $(i = 1, ..., n)$.

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 2SGMM 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 β. 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 $\beta_0 \in int(\mathcal{B})$. 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 $q^K(x) = (q_{1K}(x), ..., q_{KK}(x))'$ 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 γ_K

such that as $K \to \infty$, $E[(a(x) - q^K(x)'\gamma_K)^2] \to 0$. 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 $\sup_{x\in\mathcal{X}} \left\|\tilde{q}^K(x)\right\| \leq \zeta(K)$ and $\sqrt{K} \leq \zeta(K)$.

Efficient 2SGMM 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.,[9]

$$n^{1/2}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, \mathcal{I}(\beta_0)^{-1}).$$

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

[9]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 $\zeta(K)^2K/n \to 0$ and $\zeta(K)^2K^2/n \to 0$. See DIN, Theorems 5.4, p.66, and 5.6, p.67, respectively.

5.1 Neglected Heterogeneity

The relevant Jacobian terms follow directly from the analysis for the unconditional moment case. Thus, if w is conditionally uncorrelated with $\partial g(z,\beta_0)'/\partial\beta$ given x, see section 3.2, cf. Theorem 3.1 and the 2SGMM IM test of section 4.1, define[10]

$$\begin{aligned} [G_{\eta i}(\beta)]_k &= \frac{\partial^2 g_i(\beta)}{\partial\beta_k^2} \\ &= \frac{\partial^2 u_i(\beta)}{\partial\beta_k^2}\otimes q_i, (k=1,...,p). \end{aligned} \tag{5.2}$$

Incorporating an implicit (conditional) correlation similar to that described above for the GEL IM statistic of section 4.2 in the unconditional case, cf. (4.11), define[11]

$$\begin{aligned} [G_{\eta i}(\beta)]_k &= \frac{\partial^2 g_i(\beta)}{\partial\beta_k^2} \\ &\quad +G_{\beta_k i}(\beta)g_i(\beta)'\Omega^{-1}G_{\beta_k} \\ &\quad +G_{\beta_k i}(\beta)g_i(\beta)'\Omega^{-1}G_{\beta_k} \\ &= \frac{\partial^2 u_i(\beta)}{\partial\beta_k^2}\otimes q_i \\ &\quad +[\frac{\partial u_i(\beta)}{\partial\beta_k}u_i(\beta)'\otimes q_i q_i']\Omega^{-1}G_{\beta_k} \\ &\quad +[\frac{\partial u_i(\beta)}{\partial\beta_k}u_i(\beta)'\otimes q_i q_i']\Omega^{-1}G_{\beta_k}, (k=1,...,p). \end{aligned} \tag{5.3}$$

5.2 Test Statistics

As previously, define the $m\times p$ matrix $G_{\eta i}(\beta)$ with columns $[G_{\eta i}(\beta)]_k$, $(k=1,...,p)$, defined in (5.2) or (5.3), $(i=1,...,n)$. Since there may be linear dependencies among

[10]The terms corresponding to the 2SGMM Jacobian (4.4), cf. (3.8), are $[G_{\eta i}(\beta)]_{kl}=\partial^2 g_i(\beta)/\partial\beta_k\partial\beta_l=\partial^2 u_i(\beta)/\partial\beta_k\partial\beta_l\otimes q_i$, $(k\le l, l=1,...,p)$.

[11]The terms for the GEL Jacobian (4.10) are

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

the population counterparts of these columns taken together with those of $G_{\beta i}(\beta)$, $(i = 1, ..., n)$, let $G^c_{\eta i}(\beta)$ denote those r non-redundant columns chosen from $G_{\eta i}(\beta)$, $(i = 1, ..., n)$.

For the requisite 2SGMM and GEL statistics, let the sample Jacobian estimator $\hat{G}^c_\eta(\beta) = \sum_{i=1}^n G^c_{\eta i}(\beta)/n$ and write $\hat{G}(\beta) = (\hat{G}_\beta(\beta), \hat{G}^c_\eta(\beta))$ with the consequent $\hat{\Sigma}(\beta) = (\hat{G}(\beta)'\hat{\Omega}(\beta)^{-1}\hat{G}(\beta))^{-1}$.

The respective optimal 2SGMM or GEL score and LM statistics for neglected heterogeneity are then 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 $\mathcal{LM}_n$ and $\mathcal{S}_n$ in the absence of parameter heterogeneity. This requires the strengthening of the assumptions therein, in particular, Assumption B.5 of Appendix B.

Let $u_{\beta_k\beta_k}(z, \beta) = \partial^2 u(z, \beta)/\partial\beta_k^2$, $(k = 1, ..., p)$, and $u_{\beta_k\beta_k\beta}(z, \beta) = \partial^3 u(z, \beta)/\partial\beta_k^2\partial\beta'$, $(k = 1, ..., p)$. Also let $\mathcal{N}$ denote a neighbourhood of β_0 and $D_\beta(x) = E_z[u_\beta(z, \beta_0)|x]$. We write $D^c_\eta(x, \beta)$ as the linearly independent components selected from either $E_z[u_{\beta_k\beta_k}(z, \beta)|x]$ or $E_z[u_{\beta_k\beta_k}(z, \beta)|x] + E_z[u_{\beta_k}(z, \beta)u(z, \beta)'|x]V(x, \beta)^{-1}D_{\beta_k}(x, \beta) + E_z[u_{\beta_k}(z, \beta)u(z, \beta)'|x] \times V(x, \beta)^{-1}D_{\beta_k}(x, \beta)$, $(k = 1, ..., p)$, with $D(x, \beta) = (D_\beta(x, \beta), D^c_\eta(x, \beta))$ and $D(x) = D(x, \beta_0)$.[12]

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

[12]Jacobians (4.4) and (4.10) require the additional conditions: (b') $E_z[\sup_{\beta\in\mathcal{N}} \left\|u^j_{\beta\beta}(z, \beta)\right\|^2 |x]$ and $E_z[\sup_{\beta\in\mathcal{N}} \left\|u^j(z, \beta)u_\beta(z, \beta)\right\|^2 |x]$, $(j = 1, ..., s)$, are bounded and (c') $E_z[\left\|u^j_{\beta\beta\beta_k}(z, \beta_0)\right\|^2 |x]$, $(k = 1, ..., p)$, $E_z[\|u_\beta(z, \beta_0)\|^4 |x]$ and $E_z[\left\|u^j(z, \beta_0)u^k_{\beta\beta}(z, \beta_0)\right\|^2 |x]$, $(j, k = 1, ..., s)$, are bounded. Also $D^c_\eta(x, \beta)$ is now defined as the non-redundant components selected from either $E_z[u_{\beta_k\beta_l}(z, \beta)|x]$ or $E_z[u_{\beta_k\beta_l}(z, \beta)|x] + E_z[u_{\beta_k}(z, \beta)u(z, \beta)'|x]V(x, \beta)^{-1}D_{\beta_l}(x, \beta) + E_z[u_{\beta_l}(z, \beta)u(z, \beta)'|x]V(x, \beta)^{-1}D_{\beta_k}(x, \beta)$, $(k \leq l, l = 1, ..., p)$.

$E_z[\left\|u_{\beta_k}(z,\beta_0)u^j_\beta(z,\beta_0)\right\|^2|x]$, *and* $E_z[\|u_{\beta_k\beta}(z,\beta_0)u^j(z,\beta_0)\|^2|x]$, $(j=1,...,s)$, $(k=1,...,p)$, *are bounded; (d)* $E_x[D(x)'D(x)]$ *is nonsingular.*

Consequently, similar results to those in the unconditional case may be stated for the LM and score statistics $\mathcal{LM}_n$ and $\mathcal{S}_n$.

Theorem 5.1 *Let Assumptions B.1-B.5,* $\tilde{\beta}=\beta_0+O_p(n^{-1/2})$ *and* $\zeta(K)^2K/n\to 0$ *be satisfied for GMM or, for GEL, let Assumptions B.1-B.6 and* $\zeta(K)^2K^2/n\to 0$ *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,*

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

Indeed, as the proof of this theorem attests, $\mathcal{LM}_n$ 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 5.2 *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* β_* *replacing* β_0, *if* $E_x[D(x,\beta_*)'V(x,\beta_*)^{-1}D(x,\beta_*)]$ *has smallest eigenvalue bounded away from zero, then the* α*-level critical region* $\mathcal{LM}_n>\chi^2_r(\alpha)$ *defines a consistent test against parameter heterogeneity if*

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

The conclusion of Theorem 5.3 critically depends on the point-wise consistency of $\hat{\beta}$ for some parameter vector β_*. In general this condition may be violated under parameter heterogeneity, in particular, no such β_* may exist. To gain some insight into the purport of this condition, suppose that $\lim_{n\to\infty}\mathcal{P}\{\hat{\beta}\in\mathcal{B}_*\}=1$ for some set $\mathcal{B}_*\subseteq\mathcal{B}$. Although we do not attempt to establish results formally here, it might be hazarded that if the conditions $E_z[u(z,\beta_*)|x] \neq 0$ and $E_x[D^c_\eta(x,\beta_*)V(x,\beta_*)^{-1}E[u(z,\beta_*)|x]] \neq 0$ hold for all $\beta_* \in \mathcal{B}_*$ then the conclusion of Theorem 5.3 could continue to be valid. If, however, there exists some subset $\mathcal{B}_{**} \subseteq \mathcal{B}_*$ such that either condition fails for all $\beta_* \in \mathcal{B}_{**}$ then Theorem 5.3 is unlikely to hold.

6 Simulation Evidence

This section presents the results from a limited set of simulation experiments designed to elicit the potential efficacy for empirical research of the tests introduced above. The experiments concern the LM statistic $\mathcal{LM}_n$ (3.9).

6.1 Experimental Design

The simulation study is based on a nonlinear regression specification inspired by a panel Poisson regression model and considers a simple design based on the panel exponential regression model. More precisely, the analysis examines the conditional moment conditions

$$E_z[y_t - \exp(\beta x_t)|\{x_s\}_{s=1}^2] = 0, (t = 1, 2),$$

where β is a scalar parameter, $p = 1$. Thus, the neglected heterogeneity dimension and the number of restrictions under test is unity, $r = 1$.

The conditional moment conditions are reformulated as $m = 4$ unconditional moment restrictions defined by

$$E_z[x_s(y_t - \exp(\beta x_t))] = 0, (s, t = 1, 2).$$

Therefore, the moment function m-vector $g(z, \beta)$ has elements

$$x_s(y_t - \exp(\beta x_t)), (s, t = 1, 2).$$

The respective derivative vectors $\partial g(z, \beta)/\partial\beta$ and $\partial^2 g(z, \beta)/\partial\beta^2$ have elements

$$-x_s x_t \exp(\beta x_t), (s, t = 1, 2),$$

and

$$-x_s x_t^2 \exp(\beta x_t), (s, t = 1, 2)$$

The covariates x_s, $(s = 1, 2)$, are independently distributed as $N(0, 1)$ and $N(1, 1)$ variates respectively and distributed independently of the regression errors $y_t - \exp(\beta x_t)$, $(t = 1, 2)$, which are themselves independently and standard normally distributed $N(0, 1)$.[13] The true value $\beta_0 = 0$ with the heterogeneity error term w distributed $N(0, 1)$ independently of the covariates $\{x_s\}_{s=1}^2$. Under this design only the moment function second derivative $\partial^2 g(z, \beta)/\partial\beta^2$ is relevant since the additional terms in the implicit GEL IM test Jacobian (4.11) are null. The initial estimator $\tilde{\beta}$ required for implementation of 2SGMM is obtained by non-linear least squares.

The range of values for the heterogeneity parameter $\kappa = \eta^{1/2}$ is 0.00, 0.25, 0.50, 0.75, 1.00, 1.50 and 2.00. Random sample sizes $n = 100$, 200, 500 and 1000 were considered. Each experiment comprised 5000 replications.

6.2 Results

Table 1 presents the results under parameter homogeneity, i.e., $\beta = \beta_0$. The columns titled $E_n[\hat{\beta}]$ and $E_n[\mathcal{LM}_n]$ detail the simulation means of the 2SGMM estimator $\hat{\beta}$ (4.1) and the LM-type statistic $\mathcal{LM}_n$ (3.9) respectively. Columns $c10$ and $c05$ are the respective empirical rejection probabilities (ERPs) associated with the 0.10 and 0.05 critical values of the nominal chi-square distribution with one degree of freedom.

Table 1 about here

As expected, the 2SGMM estimator $\hat{\beta}$ (4.1) is approximately unbiased, the bias decreasing as the sample size increases, and the simulation mean of the LM-type statistic $\mathcal{LM}_n$ (3.9) likewise very closely approximates that of the nominal χ_1^2 variate. Similarly empirical and nominal sizes are very similar at both nominal 0.10 and 0.05 levels.

[13] It is necessary to impose a non-zero mean for at least one of the covariates x_s, $(s = 1, 2)$. Let $G = (G_\beta, G_\eta)$, where $G_\beta = E_z[\partial g(z, \beta_0)/\partial\beta]$ and $G_\eta = E_z[\partial^2 g(z, \beta_0)/\partial\beta^2]$, cf. section 3.3. Then, if the the covariates $x_s \sim N(\mu_s, 1)$, $(s = 1, 2)$, are independently distributed, the matrix G is full rank if $\mu_1, \mu_2 \neq 0$ but if $\mu_1 = \mu_2 = 0$ G has rank 1.

Tables 2 and 3 about here

Since empirical and nominal sizes are approximately equal, Tables 2 and 3 present ERPs that are not size-adjusted. As the heterogeneity parameter κ increases from 0.25 to 0.50 ERPs increase dramatically at all sample sizes for both 0.10 and 0.05 nominal critical values as does the simulation mean $E_n[\mathcal{LM}_n]$ of the LM-type statistic $\mathcal{LM}_n$ (3.9) . However, for $\kappa = 0.75$, although ERPs increase for sample sizes $n = 100$ and 200 they decline for $n = 500$ and 1000 relative to $\kappa = 0.50$, which is reflected in the relative increase in the simulation mean $E_n[\mathcal{LM}_n]$. The phenomena of non-monotonic power and moderate increase in $E_n[\mathcal{LM}_n]$ are repeated for further increases in the heterogeneity parameter κ; indeed ERPs for $\kappa = 2.00$ are less than those for $\kappa = 0.50$ except at $n = 100$. Such results are unsurprising since the LM-type statistic $\mathcal{LM}_n$ (3.9) (and likewise the score-type statistic $\mathcal{S}_n$ (3.10)) is designed to yield an asymptotically locally most powerful test; see section 3.1. Overall ERPs increase with sample size n although the relative increase in ERPs declines and is very moderate for higher values of κ.

7 Concluding Remarks

This paper considers parameter heterogeneity in models specified by moment conditions and provides a framework for the development of optimal m-tests against its possible presence. Both unconditional and conditional model settings are examined. In a leading case, the optimal m-test statistic is expressed in terms of the second order own derivatives of the moments with respect to the potentially random parameters. We also consider generalized IM equalities associated with efficient 2SGMM and GEL estimation. The GMM-based version of the generalized IM test statistic corresponds to an m-statistic employing all second order derivatives including cross-derivatives of the moments whereas the GEL form involves additional components. These additional terms are associated with a more general form of parameter heterogeneity test that may be interpreted in terms of a particular form of correlation structure between the sample Jacobian and the random variables driving potential parameter heterogeneity.

The paper also provides the results from a limited set of simulation experiments. These experiments indicate that although the LM-type test proposed in this paper displays non-monotic power it may prove to be efficacious against moderate parameter heterogeneity.

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)** *Ω is nonsingular;* **(f)** *$\rho(v)$ is twice continuously differentiable in a neighborhood of zero.*

Assumption A.2 **(a)** *$\beta_0 \in int(\mathcal{B})$;* **(b)** *$g(z,\beta)$ is continuously differentiable in a neighborhood $\mathcal{N}$ of β_0 and $E_z[\sup_{\beta\in\mathcal{N}} \|\partial g(z,\beta)/\partial\beta'\|] < \infty$;* **(c)** *$rank(G_\beta) = p$.*

Assumption A.3 *The preliminary estimator $\tilde{\beta}$ satisfies $\tilde{\beta} = \beta_0 + O_p(n^{-1/2})$.*

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 γ_K such that as $K \to \infty$,*

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

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 $x \in \mathcal{X}$, $\sup_{x\in\mathcal{X}} \left\|\tilde{q}^K(x)\right\| \leq \zeta(K)$, $E_x[\tilde{q}^K(x)\tilde{q}^K(x)']$ 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 $u^j_{\beta\beta}(z, \beta) = \partial^2 u^j(z, \beta)/\partial\beta\partial\beta'$, $(j = 1, ..., s)$. Also let $\mathcal{N}$ denote a neighbourhood of β_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 $\beta, \tilde{\beta} \in \mathcal{B}$, $\left\|u(z, \beta) - u(z, \tilde{\beta})\right\| \leq \delta(z) \left\|\beta - \tilde{\beta}\right\|^\alpha$ for some $\alpha > 0$ and $\delta(z)$ such that $E[\delta(z)^2|x] < \infty$.*

Assumption B.4 *(a) $\beta_0 \in int(\mathcal{B})$; (b) $u(z, \beta)$ is twice differentiable in $\mathcal{N}$, $E_z[\sup_{\beta\in\mathcal{N}} \|u_\beta(z, \beta)\|^2 |x]$ and $E_z[\left\|u^j_{\beta\beta}(z, \beta_0)\right\|^2 |x]$, $(j = 1, ..., s)$, 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 $\zeta(K)^2 K/n^{1-2/\gamma} \to 0$ some $\gamma > 2$.*

Appendix C: Proofs of Results

C.1 Neglected Heterogeneity Jacobian

Applying L'Hôpital's rule to the ratio

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

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

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

Therefore,

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

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

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

Recall that by Lemma A1 of NS

$$\rho_j(\hat\lambda_0'g_i) \xrightarrow{p} -1, (j=1,2),$$

uniformly, $(i=1,...,n)$. From the first order condition $\sum_{i=1}^n \rho_1(\hat\lambda_0'g_i)g_i = 0$, w.p.a.1,

$$n^{1/2}\hat\lambda_0 = [\sum_{i=1}^n \rho_{2i}g_ig_i'/n]^{-1}n^{1/2}\hat g(\beta_0) + O_p(n^{-1/2}). \tag{C.2}$$

Thus, by a UWL,

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

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

$$\hat\lambda_0'E_z[G_{\beta i}g_i^j] + O_p(n^{-1}), (j=1,...,m).$$

Hence, by a UWL,

$$\sum_{i=1}^{n} \rho_{2i} G^{j}_{\beta i} G^{k\prime}_{\beta i}/n \xrightarrow{p} -E_z[G^{j}_{\beta i} G^{k\prime}_{\beta i}],$$
$$\sum_{i=1}^{n} \rho_{2i} G^{j}_{\beta i} g_i'/n \xrightarrow{p} -E_z[G^{j}_{\beta i} g_i'], (j, k = 1, ..., m),$$
$$\sum_{i=1}^{n} \rho_{1i} G_{\beta i}/n \xrightarrow{p} -G_\beta,$$
$$\sum_{i=1}^{n} \rho_{2i} g_i g_i'/n \xrightarrow{p} -\Omega$$

and

$$\sum_{i=1}^{n} \rho_{1i} \frac{\partial^2 g^j_i(\beta_0)}{\partial\beta\partial\beta'}/n \xrightarrow{p} -E_z[\frac{\partial^2 g^j(\beta_0)}{\partial\beta\partial\beta'}], (j = 1, ..., m).$$

Therefore, evaluating the GEL Hessian (4.10) at β_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

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

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

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

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

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

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

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

C.3 Empirical Likelihood

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

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

First

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

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 (4.12)

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

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 (4.12)

$$\frac{\partial\hat{\lambda}(\beta)}{\partial\beta_k}=\breve{\Omega}(\beta)^{-1}\sum_{i=1}^{n}[\frac{1}{n(1+\hat{\lambda}(\beta)'g_i(\beta))}G_{\beta_k i}(\beta)-\frac{1}{n(1+\hat{\lambda}(\beta)'g_i(\beta))^2}\hat{\lambda}(\beta)'G_{\beta_k i}(\beta)g_i(\beta)],(k=1,...,p), \tag{C.3}$$

where

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

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

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

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

C.4 Proofs of Theorems

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

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

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 $\left\|\hat{G}(\hat{\beta})-\hat{G}(\beta_0)\right\|\xrightarrow{p}0$, $\left\|\hat{\Omega}(\hat{\beta})-\hat{\Omega}(\beta_0)\right\|\xrightarrow{p}0$, by DIN Lemmata A.6, p.78, and A.7, p.79, and

$$n^{1/2}(\hat{\beta}-\beta_0)+\hat{\Sigma}(\beta_0)\hat{G}_\beta(\beta_0)'\hat{\Omega}(\beta_0)^{-1}n^{1/2}\hat{g}(\beta_0)\xrightarrow{p}0,$$

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

$$\hat{G}^c_\eta(\beta_0)'(\hat\Omega(\beta_0)^{-1} - \hat\Omega(\beta_0)^{-1}\hat G_\beta(\beta_0)\hat\Sigma(\beta_0)\hat G_\beta(\beta_0)'\hat\Omega(\beta_0)^{-1})n^{1/2}\hat g(\beta_0).$$

Now, by Lemma A.3, p.73, of DIN, $\hat G^c_\eta(\beta_0)'\hat\Omega(\beta_0)^{-1}\hat G_\beta(\beta_0) \xrightarrow{p} E_x[D^c_\eta(x)'V(x)D_\beta(x)]$ where $D^c_\eta(x)$ comprises the selected vectors from $[D_\eta(x,\beta_0)]_k = E_z[\partial^2 u(z,\beta_0)/\partial\beta_k^2|x]$, $(k = 1,...,p)$, 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, $\hat G(\beta_0)'\hat\Omega(\beta_0)^{-1}\sqrt{n}\hat g(\beta_0) - \sum_{i=1}^n D(x_i)'V(x_i)^{-1}u_i(\beta_0)/n^{1/2} \xrightarrow{p} 0$, where $D(x_i) = (D_\beta(x_i), D^c_\eta(x_i))$, $(i = 1,...,n)$. Therefore, $\hat G^c_\eta(\hat\beta)'\hat\Omega(\hat\beta)^{-1}\sqrt{n}\hat g(\hat\beta)$ is asymptotically equivalent to

$$(-E_x[D^c_\eta(x)'V(x)^{-1}D_\beta(x)]\mathcal{I}(\beta_0)^{-1}, I_r)\sum_{i=1}^n D(x_i)'V(x_i)^{-1}u_i(\beta_0)/n^{1/2}.$$

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

$$\hat G^c_\eta(\hat\beta)'\hat\Omega(\hat\beta)^{-1}n^{1/2}\hat g(\hat\beta) \xrightarrow{d} N(0, [S(E_x[D(x)'V(x)^{-1}D(x)])^{-1}S']^{-1}). \qquad \text{(C.4)}$$

The result $\mathcal{LM}_n \xrightarrow{d} \chi^2_r$ then follows.

For the score statistic $\mathcal{S}_n$, by the mean value value theorem applied to the first order conditions determining $\hat\lambda$, i.e., $\sum_{i=1}^n \rho_1(\hat\lambda'\hat g_i)\hat g_i = 0$, $n^{1/2}\hat\lambda - (-[\sum_{i=1}^n \rho_2(\dot\lambda'\hat g_i)\hat g_i\hat g_i'/n]^{-1}n^{1/2}\hat g(\hat\beta)) \xrightarrow{p} 0$ for some $\dot\lambda$ on the line segment joining 0 and $\hat\lambda$. By a similar argument to that used above for $\mathcal{LM}_n$

$$\hat G(\hat\beta)'n^{1/2}\hat\lambda - \hat G(\beta_0)'\hat\Omega(\beta_0)^{-1}n^{1/2}\hat g(\hat\beta) \xrightarrow{p} 0.$$

Moreover, $\hat G_\beta(\beta_0)'\hat\Omega(\beta_0)^{-1}n^{1/2}\hat g(\hat\beta) \xrightarrow{p} 0$. Hence, recalling $S = (0, I_r)$,

$$\hat G(\hat\beta)'n^{1/2}\hat\lambda - S'\hat G^c_\eta(\beta_0)'\hat\Omega(\beta_0)^{-1}n^{1/2}\hat g(\hat\beta) \xrightarrow{p} 0.$$

Therefore, the result $\mathcal{S}_n \xrightarrow{d} \chi^2_r$ follows from (C.4). ∎

Proof of Theorem 5.2: 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 $E_x[D_\beta(x,\beta_*)V(x,\beta_*)^{-1}E[u(z,\beta_*)|x]] = 0$, test consistency requires

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

$\blacksquare$

References

Andrews, D.W.K., 2001, Testing when a parameter is on the boundary of the maintained hypothesis. Econometrica 69, 683-734.

Back, K. and D. Brown, 1993, Implied probabilities in GMM estimators. Econometrica 61, 971-976.

Brown, B.W. and W.K. Newey, 1992, Bootstrapping for GMM. Mimeo, M.I.T.

Brown, B.W. and W.K. Newey, 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.

Chesher, A. and R.J. Smith, 1997, Likelihood ratio specification tests. Econometrica 65, 627-646.

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. Newey, 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, in: R.W. Blundell, W.K. Newey and T. Persson, (Eds.), Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society, Vol. 3, Econometric Society Monograph Series, ESM 43. Cambridge University Press, Cambridge, pp. 174-237.

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.

Newey, W.K., 1985a, Generalized method of moments specification testing. Journal of Econometrics 29, 229-256.

Newey, W.K., 1985b, Maximum likelihood specification testing and conditional moment test. Econometrica 53, 1047-1070.

Newey, W.K., 1997, Convergence rates and asymptotic normality for series estimators. Journal of Econometrics 79, 147-168.

Newey, W.K. and D. McFadden, 1994, Large sample estimation and hypothesis testing, in: R. Engle and D. McFadden, (Eds.), Handbook of Econometrics, Vol. 4. New York: North Holland, pp. 2111-2245.

Newey, W.K. and R.J. Smith, 2004, Higher order properties of GMM and generalized empirical likelihood estimators. Econometrica 72, 219-255.

Newey, 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 J. Lawless, 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., 2011, GEL methods for moment condition models. Econometric Theory 27, 1192-1235.

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.

Table 1. Empirical Size

n	κ	$E_n[\hat{\beta}]$	$E_n[\mathcal{LM}_n]$	$c10$	$c05$
100	0.00	-0.0018	1.0125	0.1030	0.0502
200	0.00	-0.0008	1.0137	0.1024	0.0520
500	0.00	-0.0007	1.0092	0.1050	0.0532
1000	0.00	-0.0003	1.0081	0.0986	0.0492

Table 2. Empirical Power: $\kappa = 0.25$, 0.50 and 0.75

n	κ	$E_n[\hat{\beta}]$	$E_n[\mathcal{LM}_n]$	$c10$	$c05$
100	0.25	0.0382	1.1829	0.1306	0.0700
200	0.25	0.0404	1.4446	0.1752	0.0990
500	0.25	0.0420	2.2367	0.3052	0.2012
1000	0.25	0.0424	3.5904	0.4936	0.3698
100	0.50	0.1325	2.1457	0.3080	0.1886
200	0.50	0.1432	3.6945	0.5394	0.3978
500	0.50	0.1523	7.0720	0.8560	0.7730
1000	0.50	0.1583	14.1523	0.9746	0.9548
100	0.75	0.2411	2.9231	0.4360	0.2936
200	0.75	0.2714	4.7618	0.6402	0.5182
500	0.75	0.3045	8.6731	0.8200	0.7546
1000	0.75	0.3239	13.8112	0.8660	0.8298

Table 3. Empirical Power: $\kappa = 1.00$, 1.50 and 2.00

n	κ	$E_n[\hat{\beta}]$	$E_n[\mathcal{LM}_n]$	$c10$	$c05$
100	1.00	0.3052	3.1092	0.4412	0.3160
200	1.00	0.3356	4.3417	0.5304	0.4298
500	1.00	0.3617	6.0517	0.5810	0.5096
1000	1.00	0.3915	7.6099	0.6012	0.5354
100	1.50	0.3424	2.8626	0.3914	0.2804
200	1.50	0.3845	3.1605	0.4064	0.3020
500	1.50	0.4428	3.5256	0.4136	0.3136
1000	1.50	0.4637	3.8138	0.4172	0.3258
100	2.00	0.3822	2.5593	0.3646	0.2434
200	2.00	0.4330	2.6515	0.3616	0.2486
500	2.00	0.4780	2.8244	0.3764	0.2656
1000	2.00	0.5293	2.9611	0.3786	0.2762