Projection-based GEL score test for subsets of parameters with possible weak identification

Saraswata Chaudhuri*

April 7, 2009

Abstract

It has been recently shown that generalized empirical likelihood (GEL) methods can be used to design score tests for subsets of parameters such that the asymptotic size of the tests is equal to their nominal size when the nuisance parameters in the model are (strongly) identified. However, this does not necessarily hold if the nuisance parameters are not identified; and in such cases there is no guarantee that the standard (first-order) asymptotic χ^2 approximation of the score statistic will not result in over-rejection of the true value of the parameters of interest. In this paper we address this problem by proposing a new method of projection-based score test that guards against the uncontrolled over-rejection of the true value of the parameters of interest even when the nuisance parameters are not identified, while achieving asymptotic equivalence with the GEL score tests otherwise.

JEL Classification: C12; C13; C30
Keywords: Generalized empirical likelihood; Weak identification; Efficient score statistic; Nuisance parameters; Confidence intervals

*Department of Economics, CB 3305, University of North Carolina, Chapel Hill, NC 27519. Telephone: 919-966-3962. Fax: 919-966-4986. Email: saraswata_chaudhuri@unc.edu. We gratefully acknowledge helpful comments from Patrik Guggenberger and Eric Renault.

1 Introduction

Consider a model with p_1 parameters of interest denoted by θ_1 and the remaining p_2 nuisance parameters by θ_2. Suppose that the goal is to test hypotheses of the form $H : \theta_1 = \theta_{*1}$. Recently Guggenberger and Smith (2005, 2008) showed that generalized empirical likelihood (GEL) methods can be used to test the hypothesis such that the asymptotic size of the tests is equal to their nominal size irrespective of the strength of identification of the parameters of interest θ_1. This is a significant contribution to the literature because it has been widely documented that this important result does not generally hold for the conventional Wald, likelihood ratio and score tests based on the extremum estimation framework [see Stock et al. (2002) for a survey].

The GEL framework is a broad and important subclass of general extremum estimation and has received lot of attention recently [see, for example, Kitamura and Stutzer (1997), Imbens (2002), Newey and Smith (2004)]. Guggenberger and Smith (2005), henceforth GS, considered a test, the $GELR_\rho$ test, based on the likelihood ratio principle, and two tests, the LM_ρ and the S_ρ tests, based on the score principle. They recommended the LM_ρ test for testing hypotheses of the form $H : \theta_1 = \theta_{*1}$.

However, the nice result of GS does not hold if the nuisance parameters θ_2 are (close to being) not identified. In particular, there is no guarantee that the GEL tests will not result in an uncontrolled over-rejection of the true value of θ_1. In this paper we focus on the LM_ρ test that rejects $H : \theta_1 = \theta_{*1}$ at the nominal level α if $LM_\rho(\theta_{*1}, \hat{\theta}_2(\theta_{*1})) > \chi^2_{p_1}(1-\alpha)$ where $\hat{\theta}_2(\theta_{*1})$ is the GEL estimator of θ_2, restricted by the null hypothesis, and $\chi^2_{p_1}(1-\alpha)$ is the $100(1-\alpha)$-th quantile of a (central) chi-squared distribution with p_1 degrees of freedom. When the nuisance parameters are (close to being) not identified, the problem arises because $\hat{\theta}_2(\theta_{*1})$ is no longer consistent for θ_2 even when the null hypothesis is true, and as a result nothing is known about the asymptotic properties of the LM_ρ test. Although Kleibergen and Mavroeidis (2008) recently showed that the asymptotic size of a special case of this test based on the continuous updating GMM does not exceed its nominal size; there is no theoretical result that justifies the use of the LM_ρ test for the broader GEL class if the identifiability condition for the nuisance parameters is not satisfied.[1]

The contribution of the present paper is to propose a projection-based version of the LM_ρ test such that – (i) when θ_2 is not identified, one can impose any pre-specified upper bound to the asymptotic size of the new test, and (ii) when θ_2 is identified, the new test is asymptotically equivalent to the LM_ρ test proposed by GS against a set of alternatives that contains the usual $\sqrt{n}$-local alternatives. As shown in Chaudhuri et al. (2008), Chaudhuri (2008) and Chaudhuri and Zivot (2008), the projection used in this new test greatly reduces the undesirable loss in power associated with the usual projection-based tests. In fact, the projection used here makes the new test "as good as the LM_ρ test"of GS whenever the latter is guaranteed to work. The new test is motivated by Robins (2004), and similar methods were also proposed by Dufour (1990), Berger and Boos (1994) and Silvapulle (1996). Our result naturally extends to the other form of score test, the S_ρ test, but not to the $GELR_\rho$ test.

[1]Notwithstanding, GS mentioned that Monte Carlo simulations (not reported in their paper) showed that the size properties of the LM_ρ test "are not much affected by the strength or weakness of identification of"θ_2. It will be interesting to establish this result analytically following the approach of Kleibergen and Mavroeidis. However, this is beyond the scope of the present short note which focuses on the projection-based methods.

We use the following notations throughout. If $A = [A_1, \ldots, A_{bc}]$ is an $a \times bc$ matrix then $vecA = [A_1', \ldots, A_{bc}']'$, $devec_c A' = \left[(A_1, \ldots, A_c)', \ldots, (A_{(b-1)c+1}, \ldots, A_{bc})'\right]$ and $\|A\|$ is the positive square root of the largest eigenvalue of $A'A$. If A is full column rank then $P(A) = A(A'A)^{-1}A'$ and $N(A) = I_a - P(A)$ where I_a is the $a \times a$ identity matrix. If A is a symmetric positive semi-definite matrix then $A^{\frac{1}{2}}$ is the lower-triangular Cholesky factor of A such that $A = A^{\frac{1}{2}} A^{\frac{1}{2}'}$. For any function $x(w_t)$ we usually suppress its dependence on the data and, unless confusing, denote it by x_t and its sample (simple) average by $\bar{x}$. Lastly, we use $\overset{P}{\rightarrow}$, $\overset{d}{\rightarrow}$ to denote convergence in probability and distribution respectively, $\overset{A}{\sim}$ to denote "asymptotically follows"and the acronym w.p.a. for "with probability approaching".

2 The GEL framework

Let $g : \mathcal{S} \times \Theta \mapsto \mathbb{R}^k$ be such that for the sample of observations $\{w_t \in \mathcal{S} \subset \mathbb{R}^l : t = 1, \ldots, n\}$,

$$\begin{aligned} E[g(w_t, \theta)] &= 0 \text{ if } \theta = \theta_0, \\ &\neq 0 \text{ if } \theta \neq \theta_0. \end{aligned} \tag{2.1}$$

Equation (2.1) gives $k \geq p > 0$ moment restrictions for inference on p unknown elements of θ where p and k are fixed and finite numbers. Partition $\theta = (\theta_1', \theta_2')'$ such that for $i = 1, 2$, the $p_i \times 1$ sub-vector $\theta_i = (\theta_{ia}', \theta_{ib}')'$ where $p_i = p_{ia} + p_{ib}$ and $p = p_1 + p_2$.

Let $\rho : \mathcal{O} \mapsto \mathbb{R}$ where $\mathcal{O} \subset \mathbb{R}$ is an open interval containing 0 and let $\Lambda_n(\theta) = \{\lambda \in \mathbb{R}^k : \lambda' g_t(\theta) \in \mathcal{O} \text{ for } t = 1, \ldots, n\}$. The GEL criterion function is defined by

$$\hat{Q}_{\rho,n}(\theta, \lambda) = n^{-1} \sum_{t=1}^{n} \rho(\lambda' g_t(\theta)) - \rho(0).$$

In the following, unless there is a chance of confusion, we omit the subscripts ρ and n that respectively denote the choice of the function ρ (leading to various types of GEL such as continuous updating GMM, empirical likelihood, exponential tilting) and the sample size involved. The GEL estimator $\widehat{\theta}$ is defined as

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \sup_{\lambda \in \Lambda_n(\theta)} \hat{Q}(\theta, \lambda).$$

We list below the assumptions made following GS. Some of our assumptions are slightly stronger than those in GS because of three reasons – (i) we explicitly consider the power of the LM_ρ test for subsets of parameters against $\sqrt{n}$-local alternatives for the (strongly) identified components and fixed alternatives for the rest, (ii) we allow for identification failure of the nuisance parameters θ_2 and (iii) it eases the exposition without obscuring the basic idea behind the new projection-based LM_ρ test.

Assumption Θ*:* $\Theta = \Theta_1 \times \Theta_2$ where for $i = 1, 2$, let $\Theta_i = \Theta_{ia} \times \Theta_{ib}$ and for $j = a, b$, let $\theta_{0ij} \in$ interior(Θ_{ij}) where $\Theta_{ij} \subset \mathbb{R}^{p_{ij}}$ is compact.

More notations: For $j = a, b$, let $p_j = p_{1j} + p_{2j}$, $\theta_j = (\theta_{1j}', \theta_{2j}')'$ and $\Theta_j = \Theta_{1j} \times \Theta_{2j}$. This notation is used to regroup the weakly identified parameters as θ_a and the (strongly) identified parameters as θ_b (as defined in Assumption *ID*). $\mathcal{N}_i \in \Theta_i$ is used generically to denote some open neighborhood containing θ_{0i} for $i = 1, 2, 1a, 1b, 2a, 2b, a, b$. Statements such as *"A holds for $\theta_i \in \mathcal{N}_i$ "* should be read as *"there exists an open neighborhood $\mathcal{N}_i \subset \Theta_i$ containing θ_{0i} such that A holds for $\theta_i \in \mathcal{N}_i$"*.

Assumption ID:

(i) $E[\bar{g}(\theta)] = n^{-1/2}\tilde{m}_n(\theta) + m(\theta_b)$ where $\tilde{m}_n(\theta) : \Theta \mapsto \mathbb{R}^k$ and $m(\theta_b) : \Theta_b \mapsto \mathbb{R}^k$ are continuous functions such that $\tilde{m}_n(\theta) \to \tilde{m}(\theta)$ uniformly on Θ, $\tilde{m}(\theta)$ is continuous with $\tilde{m}(\theta_0) = 0$ and $m(\theta_b) = 0$ if and only if $\theta_b = \theta_{0b}$.

(ii) $\tilde{M}_n(\theta) = [\tilde{M}_{n1}(\theta), \tilde{M}_{n2}(\theta)]$ is bounded and continuous, where for $i = 1, 2$, the matrix $\tilde{M}_{ni}(\theta) = \partial\tilde{m}_n(\theta)/\partial\theta_i = [\tilde{M}_{nia}(\theta), \tilde{M}_{nib}(\theta)]$ converges uniformly to a bounded and continuous matrix $[\tilde{M}_{ia}(\theta), \tilde{M}_{ib}(\theta)] = \tilde{M}_i(\theta)$ for $\theta \in \Theta_a \times \mathcal{N}_b$.

(iii) $m(\theta_b)$ is continuously differentiable in $\theta_b \in \mathcal{N}_b$. $M(\theta_b) = \partial m(\theta_b)/\partial\theta_b = [M_1(\theta_b), M_2(\theta_b)]$ where for $i = 1, 2$, $M_i(\theta_b) = \partial m(\theta_b)/\partial\theta_{ib} \in \mathbb{R}^{k\times p_{ib}}$. $M(\theta_b)$ is bounded, continuous and has full column rank at θ_{0b}.

Assumption ρ:

(i) ρ is concave on its domain $\mathcal{O}$.

(ii) ρ is twice continuously differentiable in its domain. Defining $\rho_j(v) = \partial^j\rho(v)/\partial v^j$ for $j = 1, 2$ and $\rho_j = \rho_j(0)$, let $\rho_1 = \rho_2 = -1$.

Assumption S: The following are assumed to hold for each $\theta_1 \in \Theta_{1a} \times \mathcal{N}_{1b}$ (wherever applicable):

(i) $max_{1\le t\le n} \sup_{\theta_2\in\Theta_2} \|g_t(\theta_1, \theta_2)\| = o_p(n^{1/2})$, $\bar{g}(\theta_1, \theta_2)$ is integrable for all $\theta_2 \in \Theta_2$ and $\bar{g}(\theta_1, \theta_2) - E[\bar{g}(\theta_1, \theta_2)] = o_p(1)$ uniformly in $\theta_2 \in \Theta_2$.

(ii.a) $g_t(\theta)$ is twice continuously differentiable in $\theta \in \Theta$.

(ii.b) Let $G_t(\theta) = \partial g_t(\theta)/\partial\theta = [G_{t1}(\theta), G_{t2}(\theta)]$ where for $i = 1, 2$, $G_{ti}(\theta) = \partial g_t(\theta)/\partial\theta_i = [G_{tia}(\theta), G_{tib}(\theta)]$. Assume that $\bar{G}(\theta_1, \theta_2) = E[\bar{G}(\theta_1, \theta_2)] + o_p(1)$ for $\theta_2 \in \Theta_{2a} \times \mathcal{N}_{2b}$ where $E[\bar{G}(\theta_1, \theta_2)] = \partial E[\bar{g}(\theta_1, \theta_2)]/\partial\theta = n^{-1/2}\tilde{M}_n(\theta) + [0, M_1(\theta_b), 0, M_2(\theta_b)]$ by imposing the interchangeability of the order of differentiation and integration (and from assumption ID).

(ii.c) For $\theta \in \Theta_a \times \mathcal{N}_b$, there exists $\mathcal{G}_{ia}(\theta)$ bounded and continuous in θ such that $n^{-1}\sum_{t=1}^n \partial vec(G_{tia}(\theta))/\partial\theta_b - \mathcal{G}_{ia}(\theta) = o_p(1)$.

(iii) $n^{-1/2}\sum_{t=1}^n vec\left(g_t(\theta_{ab_0}) - E[g_t(\theta_{ab_0})], G_{ta}(\theta_{ab_0}) - E[G_{ta}(\theta_{ab_0})]\right)' \xrightarrow{d} [\Psi_g'(\theta_{ab_0}), \Psi_a'(\theta_{ab_0})]'$ where $\theta_{ab_0} = (\theta_a', \theta_{0b}')'$,

$$[\Psi_g'(\theta_{ab_0}), \Psi_a'(\theta_{ab_0})]' \sim N\left(0, V(\theta_{ab_0}) = \begin{bmatrix} \underset{k\times k}{V_{gg}(\theta_{ab_0})} & \underset{k\times kp_a}{V_{ga}(\theta_{ab_0})} \\ \underset{kp_a\times k}{V_{ag}(\theta_{ab_0})} & \underset{kp_a\times kp_a}{V_{aa}(\theta_{ab_0})} \end{bmatrix}\right).$$

$\Psi_a(\theta_{ab_0}) = [\Psi_{1a}'(\theta_{ab_0}), \Psi_{2a}'(\theta_{ab_0})]'$, $V_{ga}(\theta_{ab_0}) = [V_{g1}(\theta_{ab_0}), V_{g2}(\theta_{ab_0})] = V_{ag}(\theta_{ab_0})'$ and $V_{aa}(\theta_{ab_0}) = (V_{ij}(\theta_{ab_0}))_{i,j=1,2}$ are partitioned following the partition of $\theta_a = (\theta_{1a}', \theta_{2a}')'$. $V_{ag}(\theta)$ is bounded and continuous and $V_{gg}(\theta)$ is positive definite, bounded and continuous in $\theta \in \Theta_a \times \{\theta_{0b}\}$.

(iv) For $\theta \in \Theta_a \times \mathcal{N}_b$, let $[\hat{V}_{gg}(\theta), \hat{V}_{ga}(\theta)] = n^{-1}\sum_{t=1}^{n} g_t(\theta)[g_t(\theta)', vec(G_{ta}(\theta))']$. Assume that $\hat{V}_{ga}(\theta) = V_{ga}(\theta) + o_p(1)$ uniformly in $\theta_{2b} \in \mathcal{N}_{2b}$ and is bounded in probability and continuous. $\hat{V}_{gg}(\theta) = \Delta_{gg}(\theta) + o_p(1)$ uniformly in $\theta_b \in \mathcal{N}_{1b} \times \Theta_{2b}$ for some $\Delta_{gg}(\theta)$ that is uniformly bounded on $\theta_{2b} \in \Theta_{2b}$, continuous at θ_{0b} and $\Delta_{gg}(\theta_a, \theta_{0b}) = V_{gg}(\theta_a, \theta_{0b})$. $\hat{V}_{gg}(\theta)$ is bounded in probability, continuous and positive definite.[2]

3 Score test and projection-based score test

Usual score test

The LM_ρ statistic (henceforth, LM) of GS is defined as

$$LM(\theta) = n\bar{g}(\theta)'\hat{V}_{gg}^{-\frac{1}{2}}(\theta)P\left(\hat{V}_{gg}^{-\frac{1}{2}'}(\theta)\hat{D}(\theta)\right)\hat{V}_{gg}^{-\frac{1}{2}'}(\theta)\bar{g}(\theta) \tag{3.1}$$

where $\hat{D}(\theta) = [\hat{D}_1(\theta), \hat{D}_2(\theta)]$ and $\hat{D}_i(\theta) = n^{-1}\sum_{t=1}^{n}\rho_1(\lambda(\theta)'g_t(\theta))\,[G_{tia}(\theta), G_{tib}(\theta)] = [\hat{D}_{ia}(\theta), \hat{D}_{ib}(\theta)]$ for $i = 1, 2$. For a fixed θ, $\lambda(\theta)$ is defined as $\hat{Q}(\theta, \lambda(\theta)) = \sup_{\lambda \in \Lambda_n(\theta)} \hat{Q}(\theta, \lambda)$ [see Lemma A.1].

The LM test of GS rejects $H : \theta_1 = \theta_{*1}$ at nominal level α if

$$LM_\rho^{sub}(\theta_{*1}) := LM(\theta_{*1}, \hat{\theta}_2(\theta_{*1})) > \chi^2_{p_1}(1-\alpha)$$

where the restricted GEL estimator $\hat{\theta}_2(\theta_{*1})$ is defined by $\hat{\theta}_2(\theta_{*1}) = \arg\min_{\theta_2 \in \Theta_2} \sup_{\lambda \in \Lambda_n(\theta_{*1}, \theta_2)} \hat{Q}(\theta_{*1}, \theta_2, \lambda)$, and under our assumptions solves the first order condition $\hat{D}_2(\theta_{*1}, \hat{\theta}_2(\theta_{*1})) = 0$.

GS showed that when $p_{2a} = 0$, i.e. when the nuisance parameters θ_2 are identified, the asymptotic size of this test is equal to the nominal size α. However, no such result was established for the case $p_{2a} > 0$. The problem arises in the latter case because without identification it is not possible to estimate the nuisance parameters consistently (even $\hat{\theta}_2(\theta_{01})$ does not converge in probability to the true value θ_{02}). Hence it is not clear how one can justify the use of this test without pre-testing the identifiability of the nuisance parameters. Pre-testing, while popular among applied researchers, can lead to significant upward distortion of size of the (final) test for $H : \theta_1 = \theta_{*1}$ unless proper (and often difficult) size-adjustments are made [see Hall et al. (1996)].

New projection-based score test

In the following we describe a new test based on the LM statistic to avoid this problem. In particular, one can use the new test without any pre-test for identification of the nuisance parameters, and still put any pre-specified upper bound to the asymptotic size of the test. Moreover, if the nuisance parameters are identified, the new test is asymptotically equivalent to the LM test of GS for the alternatives considered in this paper.

Note that for any θ, the LM statistic in (3.1) can be expressed as the sum of squares of two vectors that are orthogonal by construction, i.e.,

$$LM(\theta) = LM_2(\theta) + LM_{1.2}(\theta).$$

[2] Since the setup only allows for i.i.d data or martingale difference sequences, and given that for θ as defined in Assumption S(iv) we have $E[\bar{g}(\theta)] = O_p(n^{-1/2})$, the convergence results in assumption S(iv) are not unreasonable.

$LM_2(\theta)$ is a quadratic form in the gradient (score) for θ_2. This is the LM statistic (of GS) that can used to test hypotheses of the form $\theta_2 = \theta_{*2}$ when θ_1 is known and is given by

$$LM_2(\theta) = n\bar{g}(\theta)'\hat{V}_{gg}^{-\frac{1}{2}}(\theta)P\left(\hat{V}_{gg}^{-\frac{1}{2}'}(\theta)\hat{D}_2(\theta)\right)\hat{V}_{gg}^{-\frac{1}{2}'}(\theta)\bar{g}(\theta). \tag{3.2}$$

The statistic $LM_{1.2}(\theta)$ is what is typically known in the maximum likelihood literature as the efficient LM (score) statistic for θ_1 [see Chaudhuri (2008)]. It is a quadratic form in the (sample) efficient gradient for θ_1 and is given by

$$LM_{1.2}(\theta) = n\bar{g}(\theta)'\hat{V}_{gg}^{-\frac{1}{2}}(\theta)P\left(N\left(\hat{V}_{gg}^{-\frac{1}{2}'}(\theta)\hat{D}_2(\theta)\right)\hat{V}_{gg}^{-\frac{1}{2}'}(\theta)\hat{D}_1(\theta)\right)\hat{V}_{gg}^{-\frac{1}{2}}(\theta)\bar{g}(\theta). \tag{3.3}$$

The rejection rule of the new projection-based score test for $H : \theta_1 = \theta_{*1}$ is described by the indicator function $\phi_n(\theta_{*1}) = 1$ where

$$\begin{aligned}\phi_n(\theta_{*1}) &= 1 \text{ if } \inf_{\theta_{*2}\in C_2(1-\zeta,\theta_{*1})=\left\{\theta_{*2}:LM_2(\theta_{*1},\theta_{*2})\leq\chi^2_{p_2}(1-\zeta)\right\}} LM_{1.2}(\theta_{*1},\theta_{*2}) > \chi^2_{p_1}(1-\epsilon)\\ &= 0 \text{ otherwise.}\end{aligned} \tag{3.4}$$

In Assumption ID(i) we modeled the two components θ_{1a} and θ_{1b} of the parameters of interest θ_1 as weakly identified and strongly identified respectively. As a result, with large number of observations the power curve along the axes of the identified components converges to 1 rapidly as the distance between θ_{*1b} and θ_{01b} (i.e. hypothesized value and true value) increases, whereas the power curve remains relatively flat along the axes of the weakly identified components even for very large distance between θ_{*1a} and θ_{01a}. Hence our analytical discussion only considers $\sqrt{n}$-local alternatives for θ_{1b} and fixed alternatives for θ_{1a} to ensure non-trivial power along all axes. In other words, we allow the hypothesized value θ_{*1} to belong to the set Θ_{n1} as defined below.

Hypothesized Value of θ_1*:* $\theta_{*1} \in \Theta_{n1} = \Theta_{1a} \times \Theta_{n1b}$ where $\Theta_{n1b} = \{\theta_{n1b} = \theta_{01b} + n^{-1/2}d_{1b}$ for some fixed $d_{1b} \in \mathbb{R}^{p_{1b}}$ such that $\theta_{n1b} \in \Theta_{1b}\}$.

Similarly, define Θ_{n2} and in the following, let θ_{n1} and θ_{n2} denote any arbitrary element of Θ_{n1} and Θ_{n2} respectively. In Lemma 3.1 we describe the properties of the LM_2 and the $LM_{1.2}$ statistics, evaluated at $(\theta_{n1}, \theta_{n2})$, based on which we further describe the properties of the new projection-based score test in Theorem 3.2.

Lemma 3.1 *Under assumptions* Θ, ID, ρ *and* S, *the following results hold for* $\theta_{ab_n} = (\theta'_{n1}, \theta'_{n2})'$*:*

(i) $LM_2(\theta_{ab_n}) \xrightarrow{d} g^*(\theta_{ab_0})'P\left(D_2^*(\theta_{ab_0})\right)g^*(\theta_{ab_0})$

(ii) $LM_{1.2}(\theta_{ab_n}) \xrightarrow{d} g^*(\theta_{ab_0})'P\left(N\left(D_2^*(\theta_{ab_0})\right)D_1^*(\theta_{ab_0})\right)g^*(\theta_{ab_0})$

where $g^*(\theta_{ab_0}) = V_{gg}^{-1/2'}(\theta_{ab_0})(\Psi_g(\theta_{ab_0}) + [\tilde{m}(\theta_{ab_0}) + M(\theta_{0b})d_b])$, $D_i^*(\theta_{ab_0}) = [D_{ia}^*(\theta_{ab_0}), D_{ib}^*(\theta_{ab_0})]$ *with* $D_{ia}^*(\theta_{ab_0}) = -V_{gg}^{-1/2'}(\theta_{ab_0})devec_k\left(\Psi_{i.g}(\theta_{ab_0}) - \left[\mathcal{G}_{ia}(\theta_{ab_0})d_b - V_{ig}(\theta_{ab_0})V_{gg}^{-1}(\theta_{ab_0})[\tilde{m}(\theta_{ab_0}) + M(\theta_{0b})d_b]\right]\right)$, $D_{ib}^*(\theta_{ab_0}) = -V_{gg}^{-1/2'}(\theta_{ab_0})M_i(\theta_{0b})$ *and* $\Psi_{i.g}(\theta_{ab_0}) = \Psi_{ia}(\theta_{ab_0}) - V_{ig}(\theta_{ab_0})V_{gg}^{-1}(\theta_{ab_0})\Psi_g(\theta_{ab_0})$ *for* $i = 1, 2$.

Remarks: We make the following remarks on the LM_2 and the $LM_{1.2}$ statistics respectively:

(i) $LM_2(\theta_0) \overset{A}{\sim} \chi^2_{p_2}$ and hence under $H: \theta_1 = \theta_{*1}$, the region $C_2(1-\zeta, \theta_{*1})$ will contain the true value θ_{02} of the nuisance parameters θ_2 w.p.a. $(1-\zeta)$. Now suppose that $p_{2a} = 0$, i.e. all the nuisance parameters are (strongly) identified. Then, $LM_2(\theta_{ab_n}) \overset{A}{\sim} \chi^2_{p_2}(\text{ncp} = \delta_2(\theta_{ab_0})'\delta_2(\theta_{ab_0}))$ where $\delta_2(\theta_{ab_0}) = P\left(D^*_{2b}(\theta_{ab_0})\right) V_{gg}^{-1/2'}(\theta_{ab_0})\left[\tilde{m}(\theta_{ab_0}) + M(\theta_{0b})d_b\right]$. Hence $LM_2(\theta)$ diverges to infinity with probability 1 when evaluated outside the $\sqrt{n}$-neighborhood of θ_{02b}, the true value of the nuisance parameters. Therefore, if $p_{2a} = 0$, the region $C_2(1-\zeta, \theta_{*1})$ can only contain θ_{n2b} in the $\sqrt{n}$-neighborhood of θ_{02b} with positive probability.

(ii) $LM_{1.2}(\theta_0) \overset{A}{\sim} \chi^2_{p_1}$. Now suppose that $p_{2a} = 0$, i.e. all the nuisance parameters are (strongly) identified. Then, conditional on $\Psi_{1.g}(\theta_{ab_0})$,

$$LM_{1.2}(\theta_{ab_n}) \xrightarrow{d} LM_{1.2}(\theta_{ab_0}) \overset{A}{\sim} \chi^2_{p_1}(\text{ncp} = \delta_{1.2}(\theta_{ab_0})'\delta_{1.2}(\theta_{ab_0})) \tag{3.5}$$

where $\delta_{1.2}(\theta_{ab_0}) = P\left(N\left(D^*_{2b}(\theta_{ab_0})\right) D^*_1(\theta_{ab_0})\right) V_{gg}^{-\frac{1}{2}'}(\theta_{ab_0})\left[\tilde{m}(\theta_{ab_0}) + M(\theta_{0b})d_b\right]$. That is, the asymptotic distribution is the same for all θ_{n2b}, including θ_{02b} and $\hat{\theta}_{2b}(\theta_{*1})$ (which GS showed to belong to the $\sqrt{n}$-neighborhood of θ_{02b}). The result holds unconditionally as well (although in that case the limiting distribution is no longer $\chi^2_{p_1}$). Hence, by virtue of the form of the $LM_{1.2}$ statistic, we can achieve asymptotic equivalence irrespective of whether the unknown (strongly identified) nuisance parameters are replaced by their true value or by any $\sqrt{n}$-consistent estimator. This is not possible with the other forms of LM statistic that are typically used in the econometrics literature [see Chaudhuri (2008)].

Theorem 3.2 *Under assumptions Θ, ID, ρ and S, the following results hold for $\theta_{*1} \in \Theta_{n1}$:*

(i) if $\theta_{*1} = \theta_{01}$ then $\lim_{n\to\infty} E_{\theta_{01}}[\phi_n(\theta_{*1})] \leq \zeta + \epsilon$

(ii) if $p_{2a} = 0$ then $\lim_{n\to\infty}\left[E_{\theta_{01}}[\phi_n(\theta_{*1})] - Pr_{\theta_{01}}\left[LM_{1.2}(\theta_{*1}, \theta_{02}) > \chi^2_{p_1}(1-\epsilon)\right]\right] = 0$.

We discuss the results of Theorem 3.2 in reverse order (this reflects our perception of the importance of the results). When the nuisance parameters are identified, i.e. when $p_{2a} = 0$, Theorem 3.2(ii) shows that for the alternatives considered in this paper, the new test is asymptotically equivalent to an infeasible test that rejects $H: \theta_1 = \theta_{*1}$ if $LM_{1.2}(\theta_{*1}, \theta_{02}) > \chi^2_{p_1}(1-\epsilon)$. And this result holds irrespective of the choice of $\zeta \in (0, 1-\epsilon)$. The latter test is infeasible because it uses the unknown true value θ_{02} of the nuisance parameters. GS showed that $\hat{\theta}_2(\theta_{*1})$ belongs to the $\sqrt{n}$-neighborhood of θ_{02}. Hence from Lemma 3.1(ii) (and Remarks(ii)) it follows that the infeasible test is asymptotically equivalent to the test that rejects $H: \theta_1 = \theta_{*1}$ if $LM_{1.2}(\theta_{*1}, \hat{\theta}_2(\theta_{*1})) > \chi^2_{p_1}(1-\epsilon)$. Further, note that $\hat{D}_2(\theta_{*1}, \hat{\theta}_2(\theta_{*1})) = 0$ under our assumptions and therefore $LM_{1.2}(\theta_{*1}, \hat{\theta}_2(\theta_{*1})) = LM(\theta_{*1}, \hat{\theta}_2(\theta_{*1}))$, i.e. the LM^{sub}_{ρ} statistic of GS. The asymptotic equivalence between the new projection-based score and GS's LM test (with nominal level ϵ) follows from the above argument.

This asymptotic equivalence is an interesting result. By its definition $\hat{\theta}_2(\theta_{*1})$ is always contained in the region $C_2(1-\zeta, \theta_{*1})$ and hence $\inf_{\theta_{*2} \in C_2(1-\zeta,\theta_{*1}) = \left\{\theta_{*2}: LM_2(\theta_{*1},\theta_{*2}) \leq \chi^2_{p_2}(1-\zeta)\right\}} LM_{1.2}(\theta_{*1}, \theta_{*2}) \leq LM_{1.2}(\theta_{*1}, \hat{\theta}_2(\theta_{*1}))$. Therefore for any choice of ζ, the LM test of GS is more powerful than the new test. However, the difference in power vanishes asymptotically as sample size increases. Note that this result only holds because of the use of the efficient score statistic $LM_{1.2}$ and will not hold for other forms of the score statistic used in the literature [see Chaudhuri (2008)]. Simulation results in Chaudhuri and Zivot (2008) and Zivot and Chaudhuri (2008) with reasonably small sample size also corroborate this fact for the continuous updating GMM (a subclass of GEL).

Theorem 3.2(i) shows that the asymptotic size of the new projection-based score test cannot exceed $\zeta + \epsilon$. [3] Simulations by Chaudhuri and his co-authors indicate that $\zeta + \epsilon$ is a conservative upper bound for the asymptotic size even when the nuisance parameters are not identified. Hence, if α is the desired asymptotic size of the test, practitioners can safely choose $\epsilon = \alpha$ and set ζ to 1% or 5% or 10% (as desired) and thus achieve asymptotic equivalence with the (nominal) level-α score test of GS when the latter is guaranteed to work and at the same time avoiding an uncontrolled over-rejection of the true parameter values otherwise.

The result of this paper is new in the econometrics literature. While Chaudhuri et al. (2008) and Chaudhuri and Zivot (2008) discuss this method in the context of split-sample score test of Dufour and Jasiak (2001) in the linear instrumental variables regression and the K test of Kleibergen (2005) based on continuous updating GMM, it is now known that there exists other tests that achieve the same purpose in those contexts [see Kleibergen (2007) and Kleibergen and Mavroeidis (2008)]. However, to the best of our knowledge, no test (except the "usual, conservative" projection-based ones) has been proposed yet that can be reliably applied to test for subsets of parameters in the context of generalized empirical likelihood irrespective of the identification failure of any parameters in the model. Our work in this paper was an attempt to fill this gap by avoiding an uncontrolled over-rejection of the true value of the parameters of interest even when the nuisance parameters are not identified, while achieving asymptotic equivalence with the GS's GEL score test that (arguably) works the best otherwise.

A Appendix

Lemma A.1 *Suppose assumptions Θ, ID, ρ and S hold. Then for $\theta_{ab_n} = (\theta'_{n1}, \theta'_{n2})'$ there exists $\lambda(\theta_{ab_n}) \in \Lambda_n(\theta_{ab_n})$ w.p.a.1 such that $\lambda(\theta_{ab_n}) = -V_{gg}^{-1}(\theta_{ab_0})\bar{g}(\theta_{ab_n}) + o_p(1)$ where θ_{ab_0} is as defined in assumption S(iii).*

Proof: First note that for n large enough $\theta_{nib} \in \mathcal{N}_{ib}$ for $i = 1, 2$. Then we verify the conditions of Lemma 8 in GS with $\Theta_n = \{\theta_{ab_n}\}$ by noting that – (i) $\max_{1 \le t \le n} \|g_t(\theta_{ab_n})\| = o_p(n^{-1/2})$ using assumption S(i); (ii) the minimum eigen-value of $\hat{V}_{gg}(\theta_{ab_n})$ is bounded away from 0 w.p.a.1 using assumption S(iv); and (iii) since $n^{1/2}E[\bar{g}(\theta_{ab_0})] \to \tilde{m}(\theta_{ab_0})$, a mean value expansion gives

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} g_t(\theta_{ab_n}) &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} g_t(\theta_{ab_0}) + \frac{1}{n} \sum_{t=1}^{n} G_{tb}(\theta_{ab_0}) d_b + o_p(1) \\ &\xrightarrow{d} \Psi_g(\theta_{ab_0}) + [\tilde{m}(\theta_{ab_0}) + M(\theta_{0b}) d_b] = O_p(1) \end{aligned} \tag{A.1}$$

using assumptions ID, S(i)-(iii).

Hence, using the differentiability of $\hat{Q}(\theta, \lambda)$ with respect to λ, we conclude that there exists $\lambda(\theta_{ab_n}) \in \Lambda_n(\theta_{ab_n})$ satisfying the first order condition $n^{-1} \sum_{t=1}^{n} \rho_1(\lambda(\theta_{ab_n})' g_t(\theta_{ab_n})) g_t(\theta_{ab_n}) = 0$

[3] Actually, since $LM_2(\theta)$ and $LM_{1.2}(\theta)$ are quadratic forms of orthogonal vectors, the upper bound can be made tighter $\zeta + \epsilon - \zeta\epsilon$. However, in practice ζ and ϵ are likely to be small and hence the tighter bound does not have much practical significance.

w.p.a.1 and $\lambda(\theta_{ab_n}) = O_p(n^{-1/2})$. Therefore, using assumption ρ, a mean value expansion gives

$$0 = \bar{g}_n(\theta_{ab_n}) + \left[\frac{1}{n}\sum_{t=1}^{n}\rho_2(\tilde{\lambda}' g_t(\theta_{ab_n}))g_t(\theta_{ab_n})g_t(\theta_{ab_n})'\right]\lambda(\theta_{ab_n})$$

for some $\tilde{\lambda}$ between 0 and $\lambda(\theta_{ab_n}) = O_p(n^{-1/2})$. Hence, using assumption S(iv) (and (iii)) and the same technique as in derivation of equation (A.5) in GS, we get $\lambda(\theta_{ab_n}) = -V_{gg}^{-1}(\theta_{ab_0})\bar{g}(\theta_{ab_n}) + o_p(1)$. ■

Lemma A.2 *Suppose assumptions Θ, ID, ρ and S hold. Let $\theta_{ab_n} = (\theta'_{n1}, \theta'_{n2})'$ and let $D_n(\theta) = \hat{D}(\theta)T_n$ where $T_n = diag(n^{1/2}1_{1\times p_{1a}}, 1_{1\times p_{1a}}, n^{1/2}1_{1\times p_{1a}}, 1_{1\times p_{1a}})$. Then*

$$D_n(\theta_{ab_n}) \xrightarrow{d} D^\dagger(\theta_{ab_0}) = \left[D^\dagger_{1a}(\theta_{ab_0}), D^\dagger_{1b}(\theta_{ab_0}), D^\dagger_{2a}(\theta_{ab_0}), D^\dagger_{2b}(\theta_{ab_0})\right] \quad \text{(A.2)}$$

where θ_{ab_0} is as defined in assumption S(iii), and for $i = 1, 2$,

$$\begin{aligned} vec(D^\dagger_{ia}(\theta_{ab_0})) &= -\Psi_{i.g}(\theta_{ab_0}) - \left[\mathcal{G}_{ia}(\theta_{ab_0})d_b - V_{ig}(\theta_{ab_0})V_{gg}^{-1}(\theta_{ab_0})[\tilde{m}(\theta_{ab_0}) + M(\theta_{0b})d_b]\right] \\ D^\dagger_{ib}(\theta_{ab_0}) &= -M_i(\theta_{0b}) \end{aligned}$$

where $\Psi_{i.g}(\theta_{ab_0}) = \Psi_{ia}(\theta_{ab_0}) - V_{ig}(\theta_{ab_0})V_{gg}^{-1}(\theta_{ab_0})\Psi_g(\theta_{ab_0})$ and is independent of $\Psi_g(\theta_{ab_0})$.

Proof: Note that $D_n(\theta_{ab_n}) = \hat{D}(\theta_{ab_n})T_n = n^{-1}\sum_{t=1}^n \rho_1(\lambda(\theta_{ab_n})'g_t(\theta_{ab_n}))G_t(\theta_{ab_n})T_n$. A mean value expansion gives $\rho_1(\lambda(\theta_{ab_n})'g_t(\theta_{ab_n})) = \rho_1(0) + \rho_2(\tilde{v}_t)g_t(\theta_{ab_n})'\lambda(\theta_{ab_n})$ for some $\tilde{v}_t$ between 0 and $g_t(\theta_{ab_n})'\lambda(\theta_{ab_n}) = O_p(n^{-1/2})$. Therefore, using the same technique as in Lemma A.1 and its result, and using $\tilde{G}_t(\theta_{ab_n})$ to denote $G_t(\theta_{ab_n})T_n$ we get

$$\begin{aligned} vec(D_n(\theta_{ab_n})) &= -\frac{1}{n}\sum_{t=1}^{n} vec(\tilde{G}_t(\theta_{ab_n})) + \frac{1}{n}\sum_{t=1}^{n}\rho_2(\tilde{v}_t)vec(\tilde{G}_t(\theta_{ab_n}))g_t(\theta_{ab_n})'\lambda(\theta_{ab_n}) \\ &= -\frac{1}{n}\sum_{t=1}^{n} vec(\tilde{G}_t(\theta_{ab_n})) + \frac{1}{n}\sum_{t=1}^{n}\rho_2(\tilde{v}_t)vec(\tilde{G}_t(\theta_{ab_n}))g_t(\theta_{ab_n})'V_{gg}^{-1}(\theta_{ab_0})\bar{g}(\theta_{ab_n}) + o_p(1) \\ &= vec(D_{n1a}(\theta_{ab_n}), D_{n1b}(\theta_{ab_n}), D_{n2a}(\theta_{ab_n}), D_{n2b}(\theta_{ab_n})) + o_p(1) \end{aligned}$$

where for $i = 1, 2$,

$$\begin{aligned} vec(D_{nia}(\theta_{ab_n})) &= -\frac{1}{\sqrt{n}}\sum_{t=1}^{n} vec(G_{tia}(\theta_{ab_n})) + V_{ig}(\theta_{ab_0})V_{gg}^{-1}(\theta_{ab_0})\frac{1}{\sqrt{n}}\sum_{t=1}^{n} g_t(\theta_{ab_n}) \\ vec(D_{nib}(\theta_{ab_n})) &= -vec(\bar{G}_{ib}(\theta_{ab_n})) + \frac{1}{n}\sum_{t=1}^{n}\rho_2(\tilde{v}_t)vec(\tilde{G}_{tib}(\theta_{ab_n}))g_t(\theta_{ab_n})'V_{gg}^{-1}(\theta_{ab_0})\bar{g}(\theta_{ab_n}). \end{aligned}$$

It follows directly from assumptions ID and S(i),(ii) and (iv) that $D_{nib}(\theta_{ab_n}) + M_i(\theta_{0b}) = o_p(1)$. Now to find the limit of $D_{nia}(\theta_{ab_n})$, note that by a mean value expansion

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} vec(G_{tia}(\theta_{ab_n})) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} vec(G_{tia}(\theta_{ab_0})) + \frac{1}{n}\sum_{t=1}^{n}\frac{\partial vec(G_{tia}(\theta_{ab_0}))}{\partial\theta_b}d_b + o_p(1) \xrightarrow{d} \Psi_{ia}(\theta_{ab_0}) + \mathcal{G}_{ia}(\theta_{ab_0})d_b.$$

Thus, using (A.1) we get $vec(D_{nia}(\theta_{ab_n})) \xrightarrow{d} -[\Psi_{ia}(\theta_{ab_0})+\mathcal{G}_{ia}(\theta_{ab_0})d_b]+V_{ig}(\theta_{ab_0})V_{gg}^{-1}(\theta_{ab_0})[\Psi_g(\theta_{ab_0})+\tilde{m}(\theta_{ab_0})+M(\theta_{0b})d_b] = -\Psi_{i.g}(\theta_{ab_0}) - \left[\mathcal{G}_{ia}(\theta_{ab_0})d_b - V_{ig}(\theta_{ab_0})V_{gg}^{-1}(\theta_{ab_0})[\tilde{m}(\theta_{ab_0})+M(\theta_{0b})d_b]\right]$ for $i = 1, 2$. From assumption S(iii), we know that $\Psi_{i.g}(\theta_{ab_0}) = \Psi_{ia}(\theta_{ab_0}) - V_{ig}(\theta_{ab_0})V_{gg}^{-1}(\theta_{ab_0})\Psi_g(\theta_{ab_0})$ and $\Psi_g(\theta_{ab_0})$ are uncorrelated by construction, and hence they are independent because they are normally distributed. ■

Proof of Lemma 3.1: Let $D_n(\theta)$ be as defined in Lemma A.2. Then using the fact that T_n is nonsingular, it is easy to see from (3.2) and (3.3) that $LM_2(\theta)$ and $LM_{1.2}(\theta)$ are invariant to the post-multiplication of $\hat{D}(\theta)$ by T_n. Hence noting that $\hat{V}_{gg}^{-\frac{1}{2}}(\theta_{ab_n}) - V_{gg}^{-\frac{1}{2}}(\theta_{ab_0}) = o_p(1)$, the results follow from (A.1) and Lemma A.2. ■

Proof of Theorem 3.2: Under our assumptions there exists $\hat{\theta}_2(\theta_{*1})$ such that $\hat{D}_2(\theta_{*1}, \hat{\theta}_2(\theta_{*1})) = 0$ and hence $LM_2(\theta_{*1}, \hat{\theta}_2(\theta_{*1})) = 0$. Therefore, $C_2(1-\zeta, \theta_{*1})$ is always nonempty and hence $\inf_{\theta_{*2} \in C_2(1-\zeta,\theta_{*1}) = \left\{\theta_{*2}: LM_2(\theta_{*1},\theta_{*2}) \le \chi^2_{p_2}(1-\zeta)\right\}} LM_{1.2}(\theta_{*1}, \theta_{*2})$ exists.
(i) Moreover, $C_2(1-\zeta, \theta_{01})$ contains θ_{02} w.p.a. $(1-\zeta)$. Therefore,

$$\begin{aligned}
& \lim_{n\to\infty} E_{\theta_{01}} \phi_n(\theta_{01}) \\
= \; & \lim_{n\to\infty} Pr_{\theta_{01}} \left[\inf_{\theta_{*2} \in \mathcal{C}_2(1-\zeta,\theta_{01})} LM_{1.2}(\theta_{01}, \theta_{*2}) > \chi^2_{p_1}(1-\epsilon) \right] \\
\le \; & 1 - \lim_{n\to\infty} Pr_{\theta_{01}} \left[\{\theta_{02} \in \mathcal{C}_2(1-\zeta, \theta_{01})\} \cap \{ \inf_{\theta_{*2} \in \mathcal{C}_2(1-\zeta,\theta_{01})} LM_{1.2}(\theta_{01}, \theta_{*2}) \le \chi^2_{p_1}(1-\epsilon) \} \right] \\
= \; & 1 - \lim_{n\to\infty} Pr_{\theta_{01}} \left[\inf_{\theta_{*2} \in \mathcal{C}_2(1-\zeta,\theta_{01})} LM_{1.2}(\theta_{01}, \theta_{*2}) \le \chi^2_{p_1}(1-\epsilon) \mid \theta_{02} \in \mathcal{C}_2(1-\zeta, \theta_{01}) \right] \\
& \times \lim_{n\to\infty} Pr_{\theta_{01}} [\, \theta_{02} \in \mathcal{C}_2(1-\zeta, \theta_{01})] \\
\le \; & 1 - \lim_{n\to\infty} Pr_{\theta_{01}} \left[LM_{1.2}(\theta_{01}, \theta_{02}) \le \chi^2_{p_1}(1-\epsilon) \right] \lim_{n\to\infty} Pr_{\theta_{01}} [\, \theta_{02} \in \mathcal{C}_2(1-\zeta, \theta_{01})] \\
\le \; & 1 - (1-\epsilon)(1-\zeta) \\
\le \; & \zeta + \epsilon.
\end{aligned}$$

(ii) Lemma 3.1(i) also implies that when $p_{2a} = 0$, the set $\mathcal{C}_2(1-\epsilon, \theta_{*1})$ is contained in the $\sqrt{n}$-neighborhood of θ_{02} w.p.a.1 under the conditions of the Theorem. Hence $\theta_2^{\inf}(\theta_{*1})$, where the infimum $\inf_{\theta_{*2} \in \mathcal{C}_2(1-\zeta,\theta_{*1})} LM_{1.2}(\theta_{*1}, \theta_{*2})$ is attained, is also in the $\sqrt{n}$-neighborhood of θ_{02}. Hence Lemma 3.1(i) directly applies and gives the local asymptotic equivalence of the tests. ■

References

Berger, R. L. and Boos, D. D. (1994). P-Values Maximized Over a Confidence Set for the Nuisance Parameter. *Journal of the American Statistical Association*, 89: 1012–1016.

Chaudhuri, S. (2008). *Projection-Type Score Tests for Subsets of Paramaters*. PhD thesis, University of Washington.

Chaudhuri, S., Richardson, T., Robins, J., and Zivot, E. (2008). A New Projection-Type Split-Sample Score Tests in Linear Instrumental Variables Regression. University of Washington Working Paper.

Chaudhuri, S. and Zivot, E. (2008). A new method of projection-based inference in GMM with weakly identified nuisance parameters. University of North Carolina, Chapel Hill.

Dufour, J. M. (1990). Exact Tests and Confidence Sets in Linear Regressions with Autocorrelated Errors. *Econometrica*, 58: 475–494.

Dufour, J. M. and Jasiak, J. (2001). Finite Sample Limited Information Inference Methods for Structural Equations and Models with Generated Regressors. *International Economic Review*, 42: 815–843.

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

Guggenberger, P. and Smith, R. (2008). Generalized empirical likelihood tests in time series models with potential identification failure. *Journal of Econometrics*, 142: 134–161.

Hall, A. R., Rudecusch, G. D., and Wilcox, D. W. (1996). Judging Instrumental Relevance in Instrumental Variables Estimation. *International Economic Review*, 37: 283–298.

Imbens, G. (2002). Generalized method of moments and empirical likelihood. *Journal of Business and Economic Statistics*, 20: 493–506.

Kitamura, Y. and Stutzer, M. (1997). An Information-Theoretic Alternative to Generalized Method of Moments Estimation. *Econometrica*, 65: 861–874.

Kleibergen, F. (2005). Testing Parameters In GMM Without Assuming That They Are Identified. *Econometrica*, 73: 1103–1123.

Kleibergen, F. (2007). Subset statistic in the linear IV regression model. Technical report, Brown University Working Paper.

Kleibergen, F. and Mavroeidis, S. (2008). Inference on Subsets of Parameters in GMM without Assuming Identification. Technical report, Brown University Working Paper.

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

Robins, J. M. (2004). Optimal Structural Nested Models for Optimal Sequential Decisions. In Lin, D. Y. and Heagerty, P., editors, *Proceedings of the Second Seattle Symposium on Biostatistics*. New York: Springer.

Silvapulle, M. J. (1996). A Test in the Presence of Nuisance Parameters. *Journal of the American Statistical Association*, 91: 1690–1693.

Stock, J. H., Wright, J. H., and Yogo, M. (2002). A Survey of Weak Instruments and Weak Identification in GMM. *Journal of Business and Economic Statistics*, 20: 518–529.

Zivot, E. and Chaudhuri, S. (2008). A Comment on Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve. University of Washingon Working Paper.