CONTRIBUTI DI RICERCA CRENOS

UNIVERSITÀ DI CAGLIARI UNIVERSITÀ DI SASSARI
CENTRE FOR NORTH SOUTH ECONOMIC RESEARCH
CRENoS
CENTRO RICERCHE ECONOMICHE NORD SUD

GMM, Generalized empirical likelihood, and time series

Federico Crudu

WORKING PAPERS

2009/12

CUEC

Centro Ricerche Economiche Nord Sud
(CRENoS)
Università di Cagliari
Università di Sassari

Il CRENoS è un centro di ricerca istituito nel 1993 che fa capo alle Università di Cagliari e Sassari ed è attualmente diretto da Stefano Usai. Il CRENoS si propone di contribuire a migliorare le conoscenze sul divario economico tra aree integrate e di fornire utili indicazioni di intervento. Particolare attenzione è dedicata al ruolo svolto dalle istituzioni, dal progresso tecnologico e dalla diffusione dell'innovazione nel processo di convergenza o divergenza tra aree economiche. Il CRENoS si propone inoltre di studiare la compatibilità fra tali processi e la salvaguardia delle risorse ambientali, sia globali sia locali.
Per svolgere la sua attività di ricerca, il CRENoS collabora con centri di ricerca e università nazionali ed internazionali; è attivo nell'organizzare conferenze ad alto contenuto scientifico, seminari e altre attività di natura formativa; tiene aggiornate una serie di banche dati e ha una sua collana di pubblicazioni.

www.crenos.it
info@crenos.it

CRENoS – Cagliari
Via San Giorgio 12, I-09100 Cagliari, Italia
tel. +39-070-6756406; fax +39-070- 6756402

CRENoS - Sassari
Via Torre Tonda 34, I-07100 Sassari, Italia
tel. +39-079-2017301; fax +39-079-2017312

Titolo: GMM, Generalized empirical likelihood, and time series

ISBN: 978 88 84 67 557 6

Prima Edizione: Novembre 2009

Via Is Mirrionis, 1
09123 Cagliari
Tel./Fax 070291201
www.cuec.it

GMM, Generalized Empirical Likelihood, and Time Series

Federico Crudu[1]*

University of Groningen and CRENoS

Abstract

In this paper we extend the results of Kitamura (1997) for BEL to the more general class of GEL estimators. The resulting BGEL estimator is proved to be consistent and asymptotically normal and attains the semiparametric lower bound. In addition, we define the BGEL version of the classical trinity of tests, Wald, Lagrange Multiplier, and Likelihood Ratio tests. The resulting tests are as expected chi square distributed. We find via Monte Carlo experiments that the overidentification tests that stem from the BGEL estimator have generally better small sample properties than the J test.

JEL Classification: C12, C14, C22.

Keywords: GMM, Generalized Empirical Likelihood, blocking techniques, α-mixing, overidentifying restrictions.

---

[1] Corresponding address: Department of Economics and Econometrics, University of Groningen, Net- telbosje 2, 9747 AE, Groningen, The Netherlands. Email: f.crudu@rug.nl.

* This paper is based on Chapter 2 of my PhD thesis at the University of York. I am grateful to Francesco Bravo and Fabrizio Iacone for helpful comments on the previous drafts of the paper. I also wish to thank Patrick Marsh, Jan Podivinsky, Christoph Hanck, Bruce Hansen for some useful hints on the use of his GAUSS codes, Jerry Lawless for having provided me with one of his papers (Qin and Lawless, 1995), and all the partecipants of the seminars at the University of Groningen, the 4th PhD meeting at UCL, and the 33rd Simposio de Analisis Economico at the University of Saragoza. All remaining errors are my own.

1 Introduction

Since Hansen's (1982) seminal paper GMM estimation has become a standard tool among applied economists in a wide variety of fields, such as business cycles analysis, covariance structure models, capital asset pricing models, stochastic volatility models, just to name a few (see Hall 2005, pp.3-4 for a more comprehensive list). The GMM framework may be easily used both for estimation and testing. In addition, a number of well known estimators may be described as special cases of GMM (for example OLS and Maximum Likelihood).

However, a major drawback is that GMM inference relies only on asymptotic approximations. In addition to that, a considerable amount of Monte Carlo evidence has pointed at the fact that the finite sample properties of the GMM estimator and inference tend to be very poor. In particular, it is well known that the J test tends to reject the null hypothesis too often (see for example Altonji and Segal 1996, Clark, 1996, and the other papers published in the 1996 and 2002 special issues of the *Journal of Business and Economic Statistics*).

In order to avoid the problems that GMM estimation and inference imply, we may want to concentrate our attention on some alternative estimators. In recent years the literature has proposed a number of possible options, for example the Empirical Likelihood (EL) estimator (Owen, 1988; Qin and Lawless, 1994), the Continuous Updating Estimator CUE (Hansen, Heaton and Yaron, 1996) and the Exponential Tilting (ET) estimator (Kitamura and Stutzer, 1997; Imbens, Spady and Johnson, 1998). Smith (1997) shows that EL and ET belong to the more general class of GEL estimators, and Newey and Smith (2004, NS henceforth) show that also CUE is a member of that class. All these estimators share with the GMM estimator the desirable property of having the same asymptotic distribution, even though they have different higher order asymptotic properties (see NS, and Anatolyev, 2005).

The purpose of this paper is twofold. First of all, we aim at extending the results of Kitamura (1997) on blockwise EL to the wider class of blockwise GEL (BGEL henceforth). The BGEL estimator includes as a special case the blockwise ET estimator. We study the first order properties of the resulting estimator and work out the BGEL version of the classical Maximum-Likelihood-based trinity of tests (i.e. Wald, Lagrange Multiplier, and Likelihood Ratio) both for overidentifying restrictions and for general possibly nonlinear restrictions. For the latter case we study the asymptotic distribution of a minimum chi square statistic. As a further result we define a Kullback-Leibler-type of statistic for overidentifying restrictions based on the BET estimator.

Second, we compare the performance of the J test in terms of size via Monte Carlo experiments against a set of tests based on two BGEL estimators, namely the blockwise EL (BEL) and the blockwise ET (BET). This type of analysis is interesting for two further reasons: the first reason is that the J test has become a standard tool among practitioners, despite its well-known problems in finite samples; hence, the use of statistics that enjoy better (theoretical) finite sample properties seems to be desirable; the second reason is that a comparison between tests based on BEL and BET estimators is almost natural as they are the most used in practice.

The outline of the paper is the following: in Section 2 we describe the estimation framework; Section 3 introduces the blockwise estimator. Section 4 gives account of the asymptotic results; Section 5 describes the Monte Carlo experiments; Section 6 concludes. Proofs and tables are in the appendix.

2 Estimation Framework

The starting point of our work is the classical (overidentified) moment condition model

$$E\left(g\left(x_t, \beta_0\right)\right) = 0 \tag{1}$$

where $g : \mathbb{R}^{L_x} \times \mathcal{B} \rightarrow \mathbb{R}^{L_g}$, given that $\mathcal{B}$ is the parameter space and $\beta_0 \in \mathcal{B} \subset \mathbb{R}^{L_\beta}$, where $L_\beta \leq L_g$. The vector of functions $g\left(x_t, \beta\right)$ is an L_g-vector and it is assumed to be twice continuously differentiable and the L_β-vector β is some feature of the distribution of x_t; the process that we consider, $\{x_t\}$, is assumed to be an $\mathbb{R}^{L_x}$-valued stationary process.

2.1 Generalized Method of Moments

The GMM estimator is computed by minimizing a quadratic form that includes the sample analogue of the moment functions and a matrix of weights

$$\bar{\beta} = \arg\min_{\beta \in \mathcal{B}} Q_n\left(\beta\right)$$

where $Q_n\left(\beta\right)$ is defined as

$$Q_n\left(\beta\right) = \hat{g}\left(\beta\right)' W_n \hat{g}\left(\beta\right)$$

and

$$\hat{g}\left(\beta\right) = \frac{1}{n}\sum_{t=1}^{n} g_t\left(\beta\right), \tag{2}$$

and $g\left(x_t, \beta\right) = g_t\left(\beta\right)$, whereas the square matrix of weights W_n is of size L_g, it is positive definite, and converges in probability to a matrix of constants. Given some regularity conditions (Hansen, 1982; Newey and McFadden, 1994) a GMM estimator is consistent and asymptotically normal

$$\sqrt{n}\left(\bar{\beta} - \beta_0\right) \rightarrow_d N\left(0, V_W\right)$$

where $V_W = \left(G(\beta_0)' W G(\beta_0)\right)^{-1} \left(G(\beta_0)' W \Omega(\beta_0) W G(\beta_0)\right) \left(G(\beta_0)' W G(\beta_0)\right)^{-1}$, where

$$\Omega(\beta_0) = E \sum_{j=-\infty}^{+\infty} g_{t+j}(\beta_0) g_t(\beta_0)' \tag{3}$$

is the long-run covariance matrix and

$$G(\beta_0) = E \frac{\partial}{\partial \beta} g_t(\beta_0) \tag{4}$$

is the expected value of the first derivative of the moment function. An important issue in the context of the GMM estimation procedure is the choice of the weighting matrix. An appropriate choice of the weighting matrix, in fact, causes the GMM estimator to be asymptotically efficient. The optimal unfeasible weighting matrix is the inverse of the long-run covariance of $g(x_t, \beta_0)$, $\Omega(\beta_0)^{-1}$. The estimator for $\Omega(\beta)$, is typically a kernel weighted average of sample autocovariances It is possible to show that the optimal weighting matrix is an estimator (say a Newey-West estimator) of the long-run covariance matrix

$$\tilde{\Omega}(\beta) = \hat{\Gamma}_0(\beta) + \sum_{j=1}^{\kappa} w_{j\kappa} \left(\hat{\Gamma}_j(\beta) + \hat{\Gamma}_j(\beta)'\right)$$

where $w_{j\kappa}$ is a kernel function, while $\hat{\Gamma}_j(\beta) = \frac{1}{n} \sum_{i=1}^{n-j} g_t(\beta) g_{t+j}(\beta)'$. The early literature suggested to use as a weight $w_{j\kappa} = 1$, the so-called uniform or truncated kernel. Unfortunately, this kind of solution does not ensure $\tilde{\Omega}(\beta)$ to be positive definite. Other popular choices for $w_{j\kappa}$ are the Bartlett kernel, the Parzen kernel and the spectral quadratic kernel.

Thus, given a preliminary consistent estimator $\bar{\beta}$ the minimization problem is

$$\hat{\beta} = \arg\min_{\beta \in \mathcal{B}} Q_n\left(\beta, \bar{\beta}\right)$$

while the updated criterion function becomes

$$Q_n\left(\beta,\bar{\beta}\right)=\hat{g}\left(\beta\right)'\tilde{\Omega}\left(\bar{\beta}\right)^{-1}\hat{g}\left(\beta\right).$$

This estimator is sometimes called two-step GMM estimator. Such approach can be generalized by repeating the procedure above a number $k(>2)$ of times

$$\hat{\beta}_{(k)}=\arg\min_{\beta\in\mathcal{B}}Q_n\left(\beta,\hat{\beta}_{(k-1)}\right)$$

which is called iterated GMM estimator. Both the two-step and the iterated estimator are normally distributed:

$$\sqrt{n}\left(\hat{\beta}-\beta_0\right)\rightarrow_d N\left(0,\left(G\left(\beta_0\right)'\Omega\left(\beta_0\right)^{-1}G\left(\beta_0\right)\right)^{-1}\right). \tag{5}$$

Finally, the normalized objective function evaluated at the estimated parameter $J_n\left(\hat{\beta}\right)=nQ_n\left(\hat{\beta}\right)$ converges to a chi square with L_g-L_β degrees of freedom

$$J_n\left(\hat{\beta}\right)\rightarrow_d\chi^2_{L_g-L_\beta} \tag{6}$$

which is used to test the overidentifying restrictions. Notice that $\hat{\beta}$ in (6) is either the two-step or the iterated GMM estimator.

2.2 Generalized Empirical Likelihood

As proven by NS, the GEL estimation problem has a dual nature, and it may be interpreted as a Minimum Distance (MD) estimation problem, where the parameters of interest are computed by minimizing the distance between the empirical density and an artificial density, given a certain set of constraints (Corcoran, 1988). The GEL objective function is

$$\hat{P}(\beta,\lambda) = \frac{1}{n}\sum_{t=1}^{n}\rho\left(\lambda' g_t(\beta)\right) \tag{7}$$

where the carrier function $\rho(\nu)$ is concave in its domain, and it is normalized to be $\rho_1(0) = \rho_2(0) = -1$, given that $\rho_j(\nu)$, $j = 1,2$ is the jth derivative of $\rho(\nu)$ with respect to ν (Smith,1997). This method reduces the problem to finding the following saddle point:

$$\left(\hat{\beta}', \hat{\lambda}'\right)' = \arg\min_{\beta\in\mathcal{B}} \sup_{\lambda\in\hat{\Lambda}_n(\beta)} \sum_{t=1}^{n}\rho\left(\lambda' g_t(\beta)\right),$$

where $\hat{\Lambda}_n(\beta) = \{\lambda : \lambda' g_t(\beta) \in \mathcal{V}, t = 1, ..., n\}$ and $\mathcal{V}$ is an open set containing zero. The moment condition model may be framed into a MD constrained optimization problem, which essentially is a nonparametric maximum likelihood problem,

$$\mathcal{L}(\pi,\beta,\tau,\mu) = \sum_{t=1}^{n}\psi(\pi_t) + \mu\left(1 - \sum_{t=1}^{n}\pi_t\right) + \tau'\left(\sum_{t=1}^{n}\pi_t g_t(\beta)\right)$$

where the convex function $\psi(\cdot)$ is our (pseudo) likelihood function and π_t, $t = 1, ..., n$ is a set of probabilities such that $\pi_t \geq 0$ for all t and $\sum_{t=1}^{n}\pi_t = 1$. Whenever $\psi(\cdot)$ belongs to the Cressie-Read (Cressie and Read, 1984) family of discrepancies, i.e. $\psi(\pi_t) = \frac{1}{n}\frac{(n\pi_t)^{\gamma+1}-1}{\gamma(\gamma+1)}$, it admits a dual representation as a GEL estimator (see NS). From the limiting cases $\gamma = -1$ and $\gamma = 0$, we get the EL and the ET estimators respectively, while from $\gamma = 1$ we obtain the EU estimator, or its dual CUE (Hansen, Heaton, and Yaron, 1996). The GEL estimator is consistent and Normally distributed

$$n^{1/2}\begin{pmatrix}\hat{\beta}-\beta_0\\ \hat{\lambda}-0\end{pmatrix} \to_d N\left(0, \begin{pmatrix}V_\beta & 0\\ 0 & V_\lambda\end{pmatrix}\right)$$

where $V_\beta = \left(G(\beta_0)' \Omega(\beta_0)^{-1} G(\beta_0)\right)^{-1}$ and $V_\lambda = \Omega(\beta_0)^{-1}\left(I - G(\beta_0) V_\beta G(\beta_0)' \Omega(\beta_0)^{-1}\right)$, and $\Omega(\beta_0) = E\left(g_t(\beta_0) g_t(\beta_0)'\right)$, while $G(\beta_0)$ is as in (4).

3 A Blockwise Approach to GEL Estimation

In this section we provide a generalization of Kitamura's (1997) Blockwise EL estimator to the more general family of Blockwise GEL estimators. We assume that the following strong mixing conditions are satisfied

$$\alpha_x(k) \to 0, \;\; k \to \infty$$

where $\alpha_x(k) = \sup_{A,B} |\Pr(A \cap B) - \Pr(A)\Pr(B)|$, $A \in F_{-\infty}^0$, $B \in F_k^\infty$, and $F_{m'}^{m''} = \sigma(x_i : m' \leq i \leq m'')$. We also assume $\sum_{k=1}^\infty \alpha_x(k)^{1-\frac{1}{c}} < \infty$ for some constant $c > 1$. Let now M and L be two integers both dependent on the sample size n, where $M \to \infty$ as $n \to \infty$, $M = o\left(n^{\frac{1}{2}}\right)$, $L = O(M)$ and $L \leq M$. Furthermore, let z_i, $i \in \mathbb{N}$ be a row vector of M consecutive observations $\left(x_{(i-1)L+1}, ..., x_{(i-1)L+M}\right)$. The parameter M is the block length, while L is the distance between the first observation of z_i and z_{i+1}. Note that, if $M = L$, the blocks are adjacent, whereas in all the other cases we have a certain degree of overlapping.

In order to describe the features that derive from the time series properties of the data, we apply to the GEL framework the block structure of Kitamura (1997) and originally devised by Künsch (1989) (see also Carlstein, 1986). The estimating functions (2) are redefined as a mapping of the blocks. Hence,

$$h_i(\beta) = \frac{1}{M}\sum_{j=1}^{M} g\left(x_{(i-1)L+j}, \beta\right) \tag{8}$$

where the function h is $h : \mathbb{R}^{L_x M} \times \mathcal{B} \to \mathbb{R}^{L_g}$, $b = \left[\frac{n-M}{L}\right] + 1$ is the new sample size and $[\cdot]$

is the integer part of $\cdot$.

A general framework can be introduced in the same way of NS. The Blockwise GEL (BGEL) criterion function is then defined as

$$\hat{R}(\beta, \lambda) = \frac{1}{b} \sum_{i=1}^{b} \rho(\lambda' h_i(\beta)) \tag{9}$$

and the resulting BGEL estimator is the saddle point of (9)

$$\left(\hat{\beta}', \hat{\lambda}'\right)' = \arg\min_{\beta \in \mathcal{B}} \sup_{\lambda \in \hat{\Lambda}_b(\beta)} \sum_{i=1}^{b} \rho(\lambda' h_i(\beta)),$$

where $\hat{\Lambda}_b(\beta) = \{\lambda : \lambda' g_t(\beta) \in \mathcal{V}, i = 1, ..., b\}$ and $\mathcal{V}$ is an open set containing zero. It is also possible to define a formula for the probability of the observations associated to each (B)GEL estimator (see NS). The resulting general expression is

$$\pi_i = \frac{\rho_1(\lambda' h_i(\beta))}{\sum_{j=1}^{b} \rho_1(\lambda' h_i(\beta))}$$

for $i = 1, ..., b$.

4 Asymptotic Results

In this section we analyze more in detail the asymptotic properties of the BGEL estimator we briefly described above.They are analyzed through the results provided in their classical papers by Wald (1949) and Wolfowitz (1949), and then adapted by Kitamura (1997) to his blockwise empirical likelihood estimator (see also Kitamura and Stutzer, 1997). We also exploit the results of NS.

Assume that:

A1. $\{x_t\}_{t \in \mathbb{Z}}$ is a strictly stationary strong mixing sequence of size $-\alpha/(\alpha - 2)$ where

$\alpha > 2$.

A2. (i) $\mathcal{B}$ is the parameter space and is compact, (ii) $E(g_t(\beta_0)) = 0$ and β_0 is unique, (iii) for small enough $\delta > 0$ and $\eta > 0$, $E \sup_{\beta^* \in \mathcal{N}(\beta,\delta)} \|g_t(\beta^*)\|^{2(1+\eta)} < \infty$ for all $\beta \in \mathcal{B}$, where $\mathcal{N}(\beta,\delta)$ is an open sphere in $\mathcal{B}$ centred in β and of radius δ, (iv) $\{\beta_j\}_{j\in\mathbb{Z}}$ is a sequence that converges to a certain $\beta \in \mathcal{B}$ as $j \to \infty$, then $g_t(\beta_j)$ converges to $g_t(\beta)$ for all x_t except perhaps on a null set, which may depend on the limit point β, (v) $\Omega(\beta_0) = \lim_{n\to\infty} Var(\sqrt{n}\hat{g}(\beta_0))$ is positive definite.

A3. (i) $\beta_0 \in int(\mathcal{B})$ and $g_t(\beta)$ is twice continuously differentiable at the true value β_0, (ii) $E\|g_t(\beta_0)\|^{2c} < \infty$ for $c > 1$, $E \sup_{\beta^* \in \mathcal{N}(\beta_0,\delta)} \|g_t(\beta^*)\|^{2+\varepsilon} < \infty$, $M = o\left(n^{1/2-1/(2+\varepsilon)}\right)$ for some $\varepsilon > 0$, $E \sup_{\beta^* \in \mathcal{N}(\beta_0,\delta)} \|\partial g_t(\beta^*)/\partial\beta'\|^2 < \infty$, (iii) $E \sup_{\beta^* \in \mathcal{N}(\beta_0,\delta)} \|\partial^2 g_{tj}(\beta^*)/\partial\beta\partial\beta'\| < \infty$ for all j, where $g_{tj}(\beta)$ is the jth element of $g_t(\beta)$, (iv) $G(\beta_0) = E(\partial g_t(\beta_0)/\partial\beta')$ is full column rank.

A4. $\rho(\cdot)$ is twice continuously differentiable in an open neighborhood of 0, and $\rho_1(0) = \rho_2(0) = -1$.

The following theorem proves consistency and asymptotic Normality of the BGEL estimator. Such a result is new in the sense that it generalizes the result of Kitamura (1997).

Theorem 1 *If assumptions A1-A4 hold, then, the BGEL estimator* $\hat{\beta}$ *is consistent and asymptotically Normal*

$$\begin{pmatrix} n^{1/2}\left(\hat{\beta} - \beta_0\right) \\ M^{-1} n^{1/2}\left(\hat{\lambda} - 0\right) \end{pmatrix} \to_d N\left(0, \begin{pmatrix} V_\beta & 0 \\ 0 & V_\lambda \end{pmatrix}\right)$$

where $V_\beta = \left(G(\beta_0)' \Omega(\beta_0)^{-1} G(\beta_0)\right)^{-1}$ *corresponds to the optimally weighted covariance matrix of the GMM estimator, and* $V_\lambda = \Omega(\beta_0)^{-1}\left(I - G(\beta_0) V_\beta G(\beta_0)' \Omega(\beta_0)^{-1}\right)$.

Let us suppose we want to test the following hypothesis (i.e. overidentification):

$$H_0 : \beta_0 \in \mathcal{B} \text{ such that } E\left(g_t\left(\beta_0\right)\right) = 0.$$

The resulting testing functions are the following distance (D), blockwise J (BJ), and Lagrange multiplier (LM) statistics

$$\begin{aligned} D^{BGEL}\left(\hat{\beta}\right) &= 2M^{-1}n\left(\hat{R}\left(\hat{\beta},\hat{\lambda}\right)-\rho\left(0\right)\right) && (10)\\ BJ^{BGEL}\left(\hat{\beta}\right) &= n\hat{h}\left(\hat{\beta}\right)'\hat{\Omega}\left(\hat{\beta}\right)^{-1}\hat{h}\left(\hat{\beta}\right) && (11)\\ LM^{BGEL}\left(\hat{\beta}\right) &= M^{-2}n\hat{\lambda}'\hat{\Omega}\left(\hat{\beta}\right)\hat{\lambda}. && (12)\end{aligned}$$

The distance statistic D accepts as a special case the $BELLR$ of Kitamura (1997). The following theorem gives the asymptotic distribution of the above statistics.

Theorem 2 *If assumptions A1-A4 hold then*

$$D^{BGEL}, BJ^{BGEL}, LM^{BGEL} \to_d \chi^2_{L_g - L_\beta}.$$

Similarly to all the BGEL estimators, BET estimator has a dual nature, and it can be also defined as the solution of a Lagrangian problem:

$$\mathcal{L}\left(\pi,\mu,\tau,\beta\right) = \sum_{i=1}^{b}\pi_i \log\left(\frac{\pi_i}{1/b}\right) + \mu\left(1-\sum_{i=1}^{b}\pi_i\right) - n\tau'\left(\sum_{i=1}^{b}\pi_i h_i\left(\beta\right)\right).$$

The BET objective function

$$KL\left(\pi_i, 1/b\right) = \sum_{i=1}^{b}\pi_i \log\left(\frac{\pi_i}{1/b}\right) \tag{13}$$

is actually the sample counterpart of the Kullback-Leibler discrepancy:

$$KL(d\omega, d\nu) = \int \log\left(\frac{d\omega}{d\nu}\right) d\omega$$

where $d\omega$ and $d\nu$ are two density functions. The quantity $KL(d\omega, d\nu)$ is zero only when $d\omega = d\nu$. The following corollary analyzes the asymptotic behaviour of the optimal $KL(\hat{\pi}_i, 1/b)$ criterion.

Corollary 1 *If assumptions A1-A4 hold then*

$$2KL(\hat{\pi}_i, 1/b) \rightarrow_d \chi^2_{L_g - L_\beta}$$

where $KL(\hat{\pi}_i, 1/b) = \frac{n}{M}\sum_{i=1}^{b} \hat{\pi}_i \log\left(\frac{\hat{\pi}_i}{1/b}\right)$ *and* $\pi_i = \frac{\exp(\lambda' h_i(\beta))}{\sum_{j=1}^{b}\exp(\lambda' h_j(\beta))}$.

Let us suppose that there exists a subset of $\mathcal{B}$, $\mathcal{B}_a$, such that $a(\beta_0) = 0$ and that we are interested in testing the following hypothesis

$$H_0 : \beta \in \mathcal{B}_a \subseteq \mathcal{B} \textit{ such that } a(\beta) = 0 \tag{14}$$

where the subset of $\mathcal{B}$, $\mathcal{B}_a$, defines the collection of β such that $a(\beta) = 0$. Then, following Qin and Lawless (1995) and Kitamura (1997), we can work out the BGEL version of the classical trinity of test statistics (Wald, Lagrange Multiplier and Distance tests) and minimum chi square, MC:

$$D_a^{BGEL}\left(\tilde{\beta}, \hat{\beta}\right) = 2M^{-1}n\left(\hat{R}\left(\hat{\beta}, \hat{\lambda}\right) - \hat{R}\left(\tilde{\beta}, \tilde{\lambda}\right)\right) \tag{15}$$

$$Wald_a^{BGEL}\left(\hat{\beta}\right) = na\left(\hat{\beta}\right)'\left(\frac{\partial a\left(\hat{\beta}\right)}{\partial \beta'}\hat{V}_\beta \frac{\partial a\left(\hat{\beta}\right)}{\partial \beta}\right)^{-1} a\left(\hat{\beta}\right) \tag{16}$$

$$LM_a^{BGEL}\left(\tilde{\beta}\right) = M^{-2}n\tilde{\lambda}'\hat{G}\tilde{V}_\beta\hat{G}'\tilde{\lambda} \tag{17}$$

$$MC_a^{BGEL}\left(\tilde{\beta}, \hat{\beta}\right) = n\left(\tilde{\beta} - \hat{\beta}\right)'\hat{V}_\beta^{-1}\left(\tilde{\beta} - \hat{\beta}\right), \tag{18}$$

where $\left(\hat{\beta}, \hat{\lambda}\right)$ and $\left(\tilde{\beta}, \tilde{\lambda}\right)$ are the unconstrained and the constrained estimators respectively. The following theorem establishes the asymptotic distribution of the above statistics.

Theorem 3 *If assumptions A1-A4 hold and that there exists a $L_a \times 1$ vector of functions $a : \mathbb{R}^{L_\beta} \rightarrow \mathbb{R}^{L_a}$, such that $a(\beta)$ is continuous and differentiable in its argument, and the matrix of first derivatives is full column rank: rank $\left(\frac{\partial a(\beta)}{\partial \beta}\right) = L_a$. Then,*

$$D_a^{BGEL}, Wald_a^{BGEL}, LM_a^{BGEL}, MC_a^{BGEL} \rightarrow_d \chi^2_{L_a}.$$

Proofs are in the Appendix.

5 Experimental design

We carry out some Monte Carlo experiments considering a linear model and a nonlinear model, in order to verify the small sample properties of the BGEL and GMM estimators and of the corresponding overidentification tests. We focus our attention on the size of the tests because we know that the J test performs poorly in finite samples and we want to see if the size BGEL-based statistics get closer to the nominal levels. The null hypothesis of an overidentification test is that $E(g(x_t, \beta)) = 0$, where g is the vector of moment functions. The initial step for the GMM estimator utilizes an identity matrix, while for the subsequent steps we use a Newey-West matrix with Bartlett kernel. The associated bandwidth parameter is computed by means of an automatic procedure, as described in Newey and West (1994). The tests we implement are the J test for both 2-step (J_{2GMM}) and iterated GMM (J_{IGMM}) estimators, and the D test, the LM test, and the BJ test, computed both at the BEL and BET estimator. For the LM test we use two specifications based on two estimators of the matrix $\Omega(\beta)$:

$$\hat{\Omega}_{\pi}\left(\hat{\beta}\right) = M \sum_{i=1}^{b} \hat{\pi}_i h_i\left(\hat{\beta}\right) h_i\left(\hat{\beta}\right)'$$

for $\hat{\pi}_i = \frac{1/b}{1+\hat{\lambda}' h_j\left(\hat{\beta}\right)}$ and $\hat{\pi}_i = \frac{\exp\left(\hat{\lambda}' h_i\left(\hat{\beta}\right)\right)}{\sum_{j=1}^{b} \exp\left(\hat{\lambda}' h_j\left(\hat{\beta}\right)\right)}$, and

$$\hat{\Omega}_{BW}\left(\hat{\beta}\right) = \frac{M}{b} \sum_{i=1}^{b} h_i\left(\hat{\beta}\right) h_i\left(\hat{\beta}\right)'$$

The two statistics are then named LM_{π} and LM_{BW}. In order to distinguish the results from the two estimators, we use the superscripts $"^{EL}"$ and $"^{ET}"$. The KL test is computed at the BET estimator only, but for consistency of notation in the comments and the tables we refer to it as KL^{ET}. The tables in the appendix outline the results. They include the empirical size of the tests for two levels of (theoretical) significance i.e. 5% and 10%.

All the simulations are implemented in R and the algorithms we use are derived from Bruce Hansen's GAUSS code for EL.

5.1 Experiment I: the Linear Model

With the following experiment we aim at analyzing the behaviour of our set of statistics when we increase the degree of overidentification of the system. The linear model is the following

$$y_t = \theta_1 + \theta_2 x_t + u_t$$

(Inoue and Shintani, 2001; Allen, Gregory and Shimotsu, 2005). While the processes x_t and u_t are defined as

$$\begin{aligned} u_t &= \rho u_{t-1} + \varepsilon_{1t} \\ x_t &= \rho x_{t-1} + \varepsilon_{2t} \end{aligned}$$

where $\varepsilon_{it} \sim N(0, .16)$ $i = 1, 2$; the errors are, moreover, uncorrelated and the coefficient ρ is set to 0.4, while $\theta_1 = \theta_2 = 0$, where θ_1 is considered as given and θ_2 is the parameter to be estimated. The vector of instruments is $z_t = (x_t, x_{t-1})'$, $z_t = (x_t, x_{t-1,} x_{t-2})'$ and $z_t = (x_t, x_{t-1,} x_{t-2}, x_{t-3})'$, i.e. $L_g = 2, 3, 4$. That means that we end up having one, two, and three overidentifying restrictions respectively. Finally, the sample sizes and the block lengths are set to $n = 256, 512$ and $M = 4, 8$. The experiments are then repeated 5000 times. Results are reported in Table 1.

Table 1 approximately here

From our simulations we find that the sizes of the J_{IGMM} test and the J_{2GMM} test are identical. The size of the J_{IGMM} (and of the J_{IGMM}) test is accurate as long as there is only one overidentifying restriction, whereas for $L_g = 3, 4$ the empirical size worsens dramatically. Moreover, the effect of a larger sample turns out to have little effect on the size of the tests. The D^{EL} test and the D^{ET} test provide a good approximation of the correct sizes for any choice of L_g. The size distortion that exists for $n = 256$ tends to disappear when we increase the sample to $n = 512$. LM_{π}^{EL} test and BJ^{EL} test reproduce approximately the same results as the distance statistic for $L_g = 2, 3$, whereas for $L_g = 4$ the two tests overreject for both sample sizes. The uniformly weighted Lagrange multiplier test LM_{BW}^{EL} does not benefit of the effect of the EL probabilities and tends to have sightly worse empirical sizes than the efficiently weighted counterpart. The LM_{BW}^{ET} and the LM_{π}^{ET}

tests are similar to their EL counterpart; however, the distortion of the LM_{BW}^{ET} test is generally larger than that of the LM_{BW}^{EL} test. Hence, weights appear to have bigger impact on the reduction of the size distortion. Interestingly, the BJ^{ET} test is almost indifferent to the degree of overidentification (it actually tends to be undersized) and it is close to the asymptotic size for both $n = 256$ and $n = 512$. Finally, the KL^{ET} test is similar to the corresponding distance statistic.

5.2 Experiment II: the Nonlinear Model

The nonlinear model that we take into account is due to Hall and Horowitz (1996; see also Gregory, Lamarche and Smith, 2005, and Schennach, 2007). With this experiment we want to see how the size of our set of statistics behaves when we both increase the persistence of the data and the degree of overidentification of the system. The model consists of two parameters $(\mu, \beta)'$, given that their true value is $(-0.72, 3)'$. In the existent literature the first parameter is considered as given and the second is to be estimated. We consider an extension of such a model by estimating the whole parameter vector $(\mu, \beta)'$. In order to preserve overidentification, we add a further moment condition:

$$E \begin{pmatrix} \exp\left(\mu - \beta\left(x_t + y_t\right) + 3y_t\right) - 1 \\ y_t\left(\exp\left(\mu - \beta\left(x_t + y_t\right) + 3y_t\right) - 1\right) \\ \left(r_t - 1\right)\left(\exp\left(\mu - \beta\left(x_t + y_t\right) + 3y_t\right) - 1\right) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{19}$$

A second specification with five moment functions (i.e. with three overidentifying restrictions) is considered:

$$E\begin{pmatrix} \exp\left(\mu-\beta\left(x_t+y_t\right)+3y_t\right)-1 \\ y_t\left(\exp\left(\mu-\beta\left(x_t+y_t\right)+3y_t\right)-1\right) \\ \left(r_{1t}-1\right)\left(\exp\left(\mu-\beta\left(x_t+y_t\right)+3y_t\right)-1\right) \\ \left(r_{2t}-1\right)\left(\exp\left(\mu-\beta\left(x_t+y_t\right)+3y_t\right)-1\right) \\ \left(r_{3t}-1\right)\left(\exp\left(\mu-\beta\left(x_t+y_t\right)+3y_t\right)-1\right) \end{pmatrix}=\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{20}$$

The data generating mechanism develops along the following design

$$\begin{aligned} x_t &= \rho x_{t-1}+\sqrt{1-\rho^2}u_t \\ y_t &= \rho y_{t-1}+\sqrt{1-\rho^2}v_t \end{aligned}$$

where $\rho = .4, .6, .8$. Note that, when $\rho = 0$, we end up having *iid* data. The disturbances u_t and v_t are both set to be $N(0,1)$, while both r_t and r_{it} $i = 1,2,3$ are chi square distributed with one degree of freedom. Finally, sample sizes, block lengths and repetitions are chosen to be equal to the ones selected for the experiment with the linear specification, i.e. $n = 256, 512$, $M = 4, 8$, and 5000 repetitions. Results for the model described in (19) are reported in Table 2, whereas for (20) results are reported in Table 3. The performance of the overidentification tests for the nonlinear model is analyzed along two dimensions: the number of overidentifying restrictions (i.e. degrees of freedom) and the persistence of the $AR(1)$ processes that generate the data.

For all the cases we consider, the J_{IGMM} test dominates the J_{2GMM}. Therefore, only the J_{IGMM} test is taken into account in the comments.

Table 2 approximately here

If we consider $L_g = 3$ (one degree of freedom), for $\rho = .4, .6$, and $n = 512$ the J_{IGMM} test rejects the null at about 7% when the asymptotic size is 5%, while for $\rho = .8$ is about 9%[1]. The only BGEL-based test that improves the J_{IGMM} test and provides a good approximation of the nominal sizes is the LM_{π}^{ET} test. The D^{EL}, the KL^{ET} and the D^{ET} tests are similar but inferior with respect to the LM_{π}^{ET} test, while the LM_{BW}^{EL} and LM_{BW}^{ET} are the worst in the panel; interestingly, the LM_{BW}^{EL} test seems to converge faster to the nominal rejection rates than its ET analogue. BJ tests also improve the J_{IGMM} test. More specifically the BJ^{ET} test is similar to the KL^{ET} test, while the BJ^{EL} statistic is the best among the EL-based tests. In the majority of cases, overlapping blocks are better than the nonoverlapping ones, and we can better appreciate the difference when $\rho = .8$.

Table 3 approximately here

For $L_g = 5$ (three degrees of freedom), the J_{IGMM} test reveals to be very inadequate, since the rejection rate is never less than about 35%, for a 5% nominal size. Also in this case the LM_{π}^{ET} test is the best, and provides good approximations to the true rejection rates also in the more extreme cases. Also the KL^{ET} test provides good results: the corresponding sizes at 5% and for $n = 512$ range between 7.3% and 14.6% for $\rho = .4$, 6.7% and 6.9% for $\rho = .6$, and 6.4% and 9.6% for $\rho = .8$. The sizes of the D^{ET} test are similar but not as good as those provided by the KL^{ET} test. Moreover, it seems to be more sensitive than the D^{ET} to the increase in the data persistence. The rejection rates at 5% of the D^{EL}, LM_{BW}^{ET}, and LM_{π}^{ET} statistics are quite high, well above 10% (and up to 40% for the LM_{BW}^{ET} when $\rho = .8$). That suggests the EL-based tests being less robust than the LM_{π}^{ET}, KL^{ET}, and D^{ET} tests to an increase in degrees of freedom. Only the BJ^{EL}

[1]The J_{IGMM} test slightly improves the J_{2GMM}; therefore only the former is taken into account. Moreover, we find, as we expected, improvements as the sample size increases, and that generally there is consistency in the results between the two sample sizes. Thus, we constrain the comments on the results to the larger sample case ($n = 512$).

test yields results that are comparable to some ET-based tests (such as the D^{ET} test and the BJ^{ET} test), but its size at 5% is always about 10%. For $L_g = 5$, we can still express a preference for overlapping blocks, in particular when $\rho = .6, .8$ and when $n = 256$ and $\rho = .4$.

6 Conclusions

The GMM setup is a powerful framework for estimation and testing. It has the advantage of being easy to use, as it consists of minimizing a quadratic form in a set of moment functions and a possibly arbitrary positive definite matrix of weights, and allows the researcher to use more estimating equations than parameters; moreover, a number of well known estimators may be viewed as being GMM (obviously MM, but also Maximum Likelihood, and Least Squares). Such features have made the GMM estimator popular among practitioners, who have applied it in a very diverse range of contexts. The use of GMM must be, however, critical, as its finite sample properties have proven to be often poor.

In this paper we propose a generalization to Kitamura's (1997) BEL estimator as an alternative to GMM. The resulting BGEL estimator is consistent and asymptotically normal and attains the semiparametric lower bound of Chamberlain (1987). In addition, we define the BGEL version of the classical trinity of tests, Wald, Lagrange Multiplier, and Likelihood Ratio tests for overidentifying restrictions and for general possibly nonlinear restrictions. A Kullback-Leibler-type of statistic for overidentifying restrictions is also defined. The resulting tests are as expected chi square distributed.

There is limited Monte Carlo evidence on the performance of BGEL estimators in the context of time series (see for example Gregory, Lamarche and Smith, 2002), and what often comes out is that, although BGEL estimators enjoy better theoretical features than GMM, their finite sample properties are not always satisfactory and do not offer an acceptable alternative to the J test. Our Monte Carlo results to some extent provide a clearer point of view on the behaviour of the BGEL-based inference in a variety of settings. We compute distance, LM, and BJ statistics at the EL and ET estimator and we see that ET-based tests tend to yield very good sizes (maybe with the exception of the LM_{BW}^{ET} test) and often dominate their EL-based counterpart (and of course the standard J test). In addition, EL-based inference is sometimes unable to improve the J test. We also notice that efficient weighting has a considerable effect on the results particularly on the ET-based LM test. In fact, the LM_{π}^{ET} test turns out to be almost independent on the level of persistence of the data and on the degree of overidentification. Similar results are found for the BJ^{ET} test, but only in the context of the linear model. On the basis of our simulations, then, we strongly advise the use of BGEL estimators and inference, in particular the LM_{π}^{ET} test. The LM_{π}^{ET} test, based on the BET estimator and weights, is the only case where the empirical size of the test (almost) matches the asymptotic size regardless the specification of the model.

7 References

1. Altonji, J. G., L. M. Segal (1996): Small-Sample Bias in GMM Estimation of Covariance Structures, *Journal of Business and Economic Statistics*, **14**, 353-366.

2. Anatolyev, S., (2005): GMM, GEL, Serial Correlation, and Asymptotic Bias, *Econometrica*, **73**, 983-1002.

3. Andrews, D. W. A. (1997): A Stopping Rule for the Computation of Generalized Method of Moments Estimators, *Econometrica*, **65**, 913-931.

4. Bravo, F. (2005): Blockwise Empirical Entropy Tests for Time Series Regressions. *Journal of Time Series Analysis*, **26**, 185-210.

5. Bravo, F. (2004): Empirical Likelihood Based Inference with Applications to some Econometric Models. *Econometric Theory*, **20**, 231-264.

6. Carlstein, E. (1986): The Use of Subseries Methods for Estimating the Variance of a General Statistic from a Stationary Time Series, *The Annals of Statistics*, **14**, 1171-1179.

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

8. Christiano, L. J., W. J. den Haan (1996): Small-Sample Properties of GMM for Business-Cycle Analysis, *Journal of Business and Economic Statistics*, **14**, 309-327.

9. Clark, T. E. (1996): Small-Sample Properties of Estimators of Nonlinear Models of Covariance Structure, *Journal of Business and Economic Statistics*, **14**, 367-373.

10. Corcoran, S. A. (1998): Bartlett Adjustment of Empirical Discrepancy Statistics. *Biometrika*, **85**, 967-972.

11. Davidson, R., J. G. MacKinnon (2003): *Econometric Theory and Methods*, OUP.

12. Fitzenberger, B. (1997), The Moving Blocks Bootstrap and Robust Inference for Linear Least Squares and Quantile Regression, *Journal of Econometrics*, **82**, 235-287.

13. Greene, W. H. (2003): *Econometric Analysis*, Prentice-Hall.

14. Gregory, A. W., J.-F. Lamarche, G. W. Smith (2002): Information-theoretic Estimation of Preference Parameters: Macroeconomic Applications and Simulation Evidence. *Journal of Econometrics*, **107**, 213-233.

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

16. Hahn, J., A. Inoue (2002): A Monte Carlo Comparison of Various Asymptotic Approximation to the Distribution of Instrumental Variables Estimator, *Econometric Reviews*, **21**, 309-336.

17. Hall, A. R. (2005): *Generalized Method of Moments*. OUP.

18. Hall, P., J. L. Horowitz (1996): Bootstrap Critical Values for Tests Based on Generalized Method of Moment Estimators, *Econometrica*, **64**, 891-916.

19. Hansen, B. E. (2009): Econometrics, Manuscript, University of Wisconsin.

20. Hansen, L. P. (1982): Large Sample Properties of Generalized Method of Moments Estimator, *Econometrica*, **50**, 1029-1054.

21. Hansen, L. P., J. Heaton, A. Yaron (1996): Finite-Sample Properties of Some Alternative GMM Estimators, *Journal of Business and Economic Statistics*, **14**, 262-280.

22. Ibragimov, I. A., Y. V. Linnik (1971): *Independent and Stationary Sequences of Random Variables.* Wolters-Noordhoff, Groningen.

23. Imbens, G. W., R. H. Spady, P. Johnson (1998): Information Theoretic Approaches to Inference in Moment Condition, *Econometrica*, **66**, 333-357.

24. Kitamura, Y. (2006): Empirical Likelihood Methods in Econometrics: Theory and Practice, Cowles Foundation Discussion Paper.

25. Kitamura, Y. (1997): Empirical Likelihood Methods with Weakly Dependent Processes, *The Annals of Statistics*, **25**, 2084-2102.

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

27. Künsch, H. R. (1989): The Jackknife and the Bootstrap for General Stationary Observations, *The Annals of Statistics*, **17**, 1217-1241.

28. Newey, W. K., D. McFadden (1994): Large Sample Estimation and Hypothesis Testing, in *Handbook of Econometrics vol. IV*, ed. R. Engle and D. McFadden. North Holland.

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

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

31. Owen, A. B. (1990): Empirical Likelihood Ratio Confidence Regions, *The Annals of Statistics*, **18**, 90-120.

32. Owen, A. B. (2001): *Empirical Likelihood*, Chapman-Hall.

33. Otsu, T. (2006): Generalized Empirical Likelihood Inference for Nonlinear and Time Series Models Under Weak Identification, *Econometric Theory*, **22**, 513-527.

34. Qin, J., J. Lawless (1995): Estimating Equations, Empirical Likelihood and Constraints on Parameters, *Canadian Journal of Statistics*, **23**, 595-601.

35. Qin, J., J. Lawless (1994): Empirical Likelihood and General Estimating Equations,*The Annals of Statistics*, **22**, 300-325.

36. Rao, C. R. (1966): *Linear Statistical Inference and Its Applications*, Wiley.

37. Schennach, S. M (2007): Point Estimation with Exponentially Tilted Empirical Likelihood, *The Annals of Statistics*, **2**, 634-672.

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

39. Wald, A. (1949): Note on the Consistency of the Maximum Likelihood Estimate, *The Annals of Mathematical Statistics*, **20**, 595-601.

40. Wolfowitz, J. (1949): On the Wald's Proof of the Consistency of the Maximum Likelihood Estimate, *The Annals of Mathematical Statistics*, **20**, 601-603.

8 Appendix: Proofs and Tables

Some abbreviations are used throughout the appendix. ULLN stands for uniform law of large numbers, while CLT and CMT indicate central limit theorem and continuous mapping theorem respectively. In addition, we use the following notation: $\rightarrow_p$ and $\rightarrow_d$ denote convergence in probability and convergence in distribution; C is a generic positive constant; CS and T denote Cauchy-Schwarz inequality and triangular inequality respectively; $\|\cdot\|$ is the Euclidean norm of $\cdot$. $1\{\cdot\}$ is the indicator function, i.e., given a certain event A, $1\{A\} = 1$ and $1\{nonA\} = 0$.

8.1 Preliminary Lemmata

Lemma 1 *Given assumptions A1-A4 and* $\ell \in \{0, 1, 2\}$

$$\sup_{\beta \in \mathcal{B}} \left\| \frac{\partial^\ell}{\partial \beta_j^\ell} \hat{h}(\beta) - E \frac{\partial^\ell}{\partial \beta_j^\ell} g(\beta) \right\| = o_p(1)$$

Proof. By Fitzenberger (1997)

$$\frac{\partial^\ell}{\partial \beta_j^\ell} \hat{h}(\beta) = \frac{\partial^\ell}{\partial \beta_j^\ell} \hat{g}(\beta) + O_p\left(\frac{M}{n}\right)$$

uniformly in β. Thus, by means of triangular inequality,

$$\left\| \frac{\partial^\ell}{\partial \beta_j^\ell} \hat{h}(\beta) - \frac{\partial^\ell}{\partial \beta_j^\ell} \hat{g}(\beta) + \frac{\partial^\ell}{\partial \beta_j^\ell} \hat{g}(\beta) - E \frac{\partial^\ell}{\partial \beta_j^\ell} g(\beta) \right\| \leq o_p(1) + \left\| \frac{\partial^\ell}{\partial \beta_j^\ell} \hat{g}(\beta) - E \frac{\partial^\ell}{\partial \beta_j^\ell} g(\beta) \right\|.$$

The result follows from an application of the ULLN. ■

Lemma 2 *Given assumptions A1-A4*

$$\left\| \hat{\Omega}\left(\hat{\beta}\right) - \Omega \right\| = o_p(1)$$

Proof. By Fitzenberger (1997)

$$\begin{aligned}\frac{M}{b}\sum_{i=1}^{b} h_i\left(\hat{\beta}\right) h_i\left(\hat{\beta}\right)' &= \frac{1}{n}\sum_{t=1}^{n} g_t\left(\hat{\beta}\right) g_t\left(\hat{\beta}\right)' + O_p\left(\frac{M^2}{n}\right) \\ &= \hat{\Omega}\left(\hat{\beta}\right) + O_p\left(\frac{M^2}{n}\right)\end{aligned}$$

then, by assuming $\sqrt{n}$-consistency of $\hat{\beta}$, by a mean value expansion of $\hat{\Omega}\left(\hat{\beta}\right)$ about β_0 and CS, we get

$$\left\|\hat{\Omega}\left(\hat{\beta}\right) - \hat{\Omega}\left(\beta_0\right)\right\| \leq 2\left(\sum_{t=1}^{n} \sup_{\beta\in\mathcal{B}} \left\|G_t\left(\beta\right)\right\|^2\right)^{\frac{1}{2}} \left(\sum_{t=1}^{n} \sup_{\beta\in\mathcal{B}} \left\|g_t\left(\beta\right)\right\|^2\right)^{\frac{1}{2}} \left\|\hat{\beta} - \beta_0\right\|$$

and, by Lemma and $\sqrt{n}$-consistency of $\hat{\beta}$, we obtain

$$\left\|\hat{\Omega}\left(\hat{\beta}\right) - \hat{\Omega}\left(\beta_0\right)\right\| \leq o_p\left(1\right). \tag{21}$$

Consider now the following

$$\begin{aligned}\left\|\hat{\Omega}\left(\hat{\beta}\right) - \Omega\right\| &= \left\|\hat{\Omega}\left(\hat{\beta}\right) - \hat{\Omega}\left(\beta_0\right) + \hat{\Omega}\left(\beta_0\right) - \Omega\right\| \\ &\leq \left\|\hat{\Omega}\left(\beta_0\right) - \hat{\Omega}\left(\beta_0\right)\right\| + \left\|\hat{\Omega}\left(\beta_0\right) - \Omega\right\|\end{aligned}$$

by T; then by (21)

$$\begin{aligned}\left\|\hat{\Omega}\left(\hat{\beta}\right) - \Omega\right\| &\leq \left\|\hat{\Omega}\left(\beta_0\right) - \Omega\right\| + o_p\left(1\right) \\ &\leq \sup_{\beta\in\mathcal{B}} \left\|\hat{\Omega}\left(\beta_0\right) - \Omega\right\| + o_p\left(1\right).\end{aligned}$$

The result follows by ULLN. ■

8.2 Proof of the Main Theorems

Proof of Theorem 1. First, we want to show that $\hat{\lambda} = \hat{\lambda}(\beta_0) = O_p\left(Mn^{-1/2}\right)$. Let us define a neighborhood of zero such that $\mathcal{N}\left(0, n^{-1/(2+\eta)}\right)$, and $\lambda \in \mathcal{N}\left(0, n^{-1/(2+\eta)}\right)$. Moreover, since $\max_t \|g(x_t, \beta_0)\| = o_p\left(n^{1/2c}\right)$, $\max_t \sup_{\lambda \in \Lambda} |\lambda' g(x_t, \beta_0)| = o_p(1)$, for $c \geq 1 + \eta/2$. Let us now expand $\hat{R}\left(\beta_0, \hat{\lambda}\right)$ about $\lambda = 0$ with Lagrange remainder

$$\hat{R}\left(\beta_0, \hat{\lambda}\right) = \hat{R}(\beta_0, 0) + \hat{\lambda}' \frac{\partial \hat{R}(\beta_0, 0)}{\partial \lambda} + \frac{1}{2}\hat{\lambda}' \frac{\partial^2 \hat{R}\left(\beta_0, \dot{\lambda}\right)}{\partial \lambda \partial \lambda'} \hat{\lambda}$$

where $\frac{\hat{R}(\beta_0, 0)}{\partial \lambda} = -\frac{1}{b}\sum_i h_i(\beta_0) = -\hat{h}(\beta_0)$, and $\frac{\hat{R}\left(\beta_0, \dot{\lambda}\right)}{\partial \lambda \partial \lambda'} = \frac{1}{b}\sum_i \rho_2\left(\dot{\lambda}' h_i(\beta_0)\right) h_i(\beta_0) h_i(\beta_0)'$. Then, similarly to NS (Lemma A2),

$$\begin{aligned} \hat{R}\left(\beta_0, \hat{\lambda}\right) &= \hat{R}(\beta_0, 0) - \hat{\lambda}' \hat{h}(\beta_0) + \frac{1}{2}\hat{\lambda}' \left(\frac{1}{b}\sum\nolimits_i \rho_2\left(\dot{\lambda}' h_i(\beta_0)\right) h_i(\beta_0)'\right) \hat{\lambda} \\ &\leq \rho(0) + \left\|\hat{\lambda}\right\| \left\|\hat{h}(\beta_0)\right\| - \frac{C}{M}\left\|\hat{\lambda}\right\|^2. \end{aligned}$$

Recall that $\tilde{\lambda}$ is a maximizer for $\hat{R}(\beta_0, \lambda)$; therefore, $\rho(0) = \hat{R}(\beta_0, 0) \leq \hat{R}\left(\beta_0, \hat{\lambda}\right)$. Thus,

$$\rho(0) \leq \hat{R}\left(\beta_0, \hat{\lambda}\right) \leq \rho(0) + \left\|\hat{\lambda}\right\| \left\|\hat{h}(\beta_0)\right\| - \frac{C}{M}\left\|\hat{\lambda}\right\|^2$$

where C is a generic positive constant. Thus, for $\left\|\hat{h}(\beta_0)\right\| = O_p\left(n^{-1/2}\right)$,

$$\left\|\hat{\lambda}\right\| = O_p\left(Mn^{-1/2}\right).$$

Since $O_p\left(Mn^{-1/2}\right)$ goes to zero as the sample size increases, $\hat{\lambda}$ is consistent. Let us define now the following set $C_n = \left\{x_t : \|g(x_t, \beta)\| \leq n^{1/(2+2\eta)}, \forall \beta \in \mathcal{B}\right\}$ and $g_n(x_t, \beta) = g(x_t, \beta) \times 1\{C_n\}$. Also let $q_{n,\beta}(\ell) = E(\rho(\ell' g_n(x_t, \beta)))$. Then, by mean value theorem, we have that $\lim_{n \to \infty} \frac{\partial}{\partial \ell} q_{n,\beta}(\ell) = -E(g(x_t, \beta))$ uniformly in the neighborhood

$\ell \in \mathcal{N}\left(0, n^{-1/(2+\eta)}\right)$. Moreover, let us define the set $\mathcal{N}_n = \left\{\ell : \ell = n^{-1/(2+\eta)}u, \|u\| = 1\right\}$, and $\ell_n(\beta) = \arg\max_{\ell \in \mathcal{N}_n} E\left(\rho\left(\ell' g_n(x_t, \beta)\right)\right)$. By means of mean value theorem, we expand the maximand, and noting that $\lim_{n\to\infty} u_n = -E(g(x_t, \beta)) / \|E(g(x_t, \beta))\|$, we have,

$$\lim_{n\to\infty} E\left(n^{1/(2+\eta)}\left(\rho\left(\ell_n(\beta)' g_n(x_t, \beta)\right) - \rho(0)\right)\right) = \|E(g(x_t, \beta))\|$$

By A2(iv)

$$\lim_{n\to\infty} \lim_{\delta\downarrow 0} n^{1/(2+\eta)} E\left(\sup_{\beta^* \in \mathcal{N}(\beta,\delta)} \left(\rho\left(\ell(\beta^*)' g_n(x_t, \beta^*)\right) - \rho(0)\right)\right) = \|E(g(x_t, \beta))\|. \qquad (22)$$

By uniqueness of β_0 and (22), the parameter space $\mathcal{B}$ can be covered by a finite number of open spheres $\mathcal{N}\left(\beta_j, \delta_j\right)$,with their center in β_j and radius δ_j. The radius is chosen so that $n^{1/(2+\eta)} E\left(\sup_{\beta^* \in \mathcal{N}\left(\beta_j,\delta_j\right)} \left(\rho\left(\ell(\beta^*)' g_n(x_t, \beta^*)\right) - \rho(0)\right)\right) + o(1) = 2V_j$, where $j = 1...v$. Moreover, assumption A2(iii) implies that

$$\max_t \sup_{\beta^* \in \mathcal{N}\left(\beta_j,\delta_j\right)} \|g(x_t, \beta^*)\| = o\left(n^{1/(2+2\eta)}\right).$$

Hence, there exists a sufficiently large integer V_j such that for a sufficiently small $\varepsilon > 0$

$$P\left\{\frac{1}{n}\sum_t \sup_{\beta^* \in \mathcal{N}\left(\beta_j,\delta_j\right)} \left(\rho\left(\ell(\beta^*)' g_n(x_t, \beta^*)\right) - \rho(0)\right) < n^{-1/(2+\eta)} V_j\right\} < \frac{\varepsilon}{2v}$$

where $j = 1, ..., v$, and for all $n > n_j$. Thus, it follows

$$P\left\{\sup_{\beta^* \in \mathcal{B}(\delta)} \frac{1}{n}\sum_t \left(\rho\left(\ell(\beta^*)' g_n(x_t, \beta^*)\right) - \rho(0)\right) < n^{-1/(2+\eta)} V\right\} < \frac{\varepsilon}{2}$$

for $V = \min_j V_j$ and for all $n > \max_j n_j$ and $\mathcal{B}(\delta) = \mathcal{B} \backslash \mathcal{N}(\beta_0, \delta)$. Since $\lambda(\beta)$ is optimal

$$\frac{1}{b}\sum\nolimits_i \rho\left(\lambda(\beta)' h_i(\beta)\right) \geq \frac{1}{n}\sum\nolimits_t \rho\left(\ell(\beta)' g_t(\beta)\right) + o_p(1)$$

Therefore, there exists a large enough $n_A \in \mathbb{N}$, such that

$$P\left\{\sup_{\beta^* \in \mathcal{B}(\delta)} \frac{1}{b}\sum\nolimits_i \left(\rho\left(\hat{\lambda}' h_i(\beta)\right) - \rho(0)\right) < n^{-1/(2+\eta)} V\right\} < \frac{\varepsilon}{2} \tag{23}$$

for all $n > n_A$. Since $\hat{\lambda}$ is a maximizer and is consistent, by a mean value expansion, we get,

$$\rho(0) \leq \hat{R}\left(\beta_0, \hat{\lambda}\right) \leq \rho(0) - \hat{\lambda}' \hat{h}(\beta_0) + O_p\left(\frac{M}{n}\right).$$

Thus, noting that $\hat{\lambda}' \hat{h}(\beta_0)$ is $o_p\left(n^{-\frac{1}{2}}\right)$,

$$0 \leq \hat{R}\left(\beta_0, \hat{\lambda}\right) - \rho(0) \leq o_p\left(n^{-\frac{1}{2}}\right)$$

That implies that there exists a large enough $n_B \in \mathbb{N}$ such that

$$P\left\{\frac{1}{b}\sum\nolimits_i \left(\rho\left(\hat{\lambda}' h_i(\beta_0)\right) - \rho(0)\right) > n^{-1/2} V\right\} < \frac{\varepsilon}{2} \tag{24}$$

for all $n > n_B$. By (23) and (24), and for any δ small enough and $n = \max(n_A, n_B)$ we get $P\left(\left\|\hat{\beta} - \beta_0\right\| > \delta\right) < \varepsilon$, which implies $\hat{\beta} \to_p \beta_0$. To prove asymptotic normality, let us consider the first order condition from the BGEL criterion function, and let us expand them by means of mean value theorem about the couple $(\beta_0, 0)$, i.e. the true parameters for β and λ respectively:

$$\begin{aligned} 0 &= M^{-1} n^{1/2} R_\beta\left(\hat{\beta}, \hat{\lambda}\right) = M^{-1} n^{1/2} \hat{R}_\beta(\beta_0, 0) \\ &\quad + \hat{R}_{\beta\lambda}\left(\dot{\beta}, \dot{\lambda}\right) M^{-1} n^{1/2}\left(\hat{\lambda} - 0\right) + M^{-1} \hat{R}_{\beta\beta}\left(\dot{\beta}, \dot{\lambda}\right) n^{1/2}\left(\hat{\beta} - \beta_0\right) \end{aligned}$$

$$
\begin{aligned}
0 &= n^{1/2}\hat{R}_{\lambda}\left(\hat{\beta},\hat{\lambda}\right)=n^{1/2}R_{\lambda}\left(\beta_0,0\right)\\
&+M\hat{R}_{\lambda\lambda}\left(\dot{\beta},\dot{\lambda}\right)M^{-1}n^{1/2}\left(\hat{\lambda}-0\right)\\
&+\hat{R}_{\lambda\beta}\left(\dot{\beta},\dot{\lambda}\right)n^{1/2}\left(\hat{\beta}-\beta_0\right)
\end{aligned}
$$

Note that $\hat{R}_{\beta\lambda}(\beta,\lambda)=\hat{R}_{\lambda\beta}(\beta,\lambda)'$. The derivatives of the BGEL function are computed following NS, by keeping constant, according to convenience, one element of the carrier function. Moreover, the asymptotic behaviour of the components of the first order conditions is analyzed following the results in Lemma 1, Lemma 2, and consistency of $\hat{\lambda}$:

$$
M^{-1}\hat{R}_{\beta\beta}\left(\dot{\beta},\dot{\lambda}\right)=\frac{1}{Mb}\sum_{i=1}^{b}\rho_1\left(\dot{\lambda}'h_i\left(\dot{\beta}\right)\right)\frac{\partial^2 h_i\left(\dot{\beta}\right)}{\partial\beta\partial\beta'}\dot{\lambda}\rightarrow_p 0 \tag{25}
$$

$$
\hat{R}_{\beta\lambda}\left(\dot{\beta},\dot{\lambda}\right)'=\hat{R}_{\lambda\beta}\left(\dot{\beta},\dot{\lambda}\right)=\frac{1}{b}\sum_{i=1}^{b}\rho_1\left(\dot{\lambda}'h_i\left(\dot{\beta}\right)\right)\frac{\partial h_i\left(\dot{\beta}\right)}{\partial\beta}\rightarrow_p -G\left(\beta_0\right) \tag{26}
$$

$$
M\hat{R}_{\lambda\lambda}\left(\dot{\beta},\dot{\lambda}\right)=\frac{M}{b}\sum_{i=1}^{b}\rho_2\left(\dot{\lambda}'h_i\left(\dot{\beta}\right)\right)h_i\left(\dot{\beta}\right)h_i\left(\dot{\beta}\right)'\rightarrow_p -\Omega\left(\beta_0\right). \tag{27}
$$

By summarizing the FOCs in matrix form and rearranging,

$$
\begin{pmatrix} n^{1/2}\left(\hat{\beta}-\beta_0\right) \\ M^{-1}n^{1/2}\left(\hat{\lambda}-0\right)\end{pmatrix}=-\begin{pmatrix} -V_\beta & \Xi' \\ \Xi & V_\lambda\end{pmatrix}\begin{pmatrix} 0 \\ n^{1/2}\hat{h}\left(\beta_0\right)\end{pmatrix}+o_p\left(1\right) \tag{28}
$$

where $V_\beta=\left(G\left(\beta_0\right)'\Omega\left(\beta_0\right)^{-1}G\left(\beta_0\right)\right)^{-1}$, $V_\lambda=\Omega\left(\beta_0\right)^{-1}\left(I-G\left(\beta_0\right)V_\beta G\left(\beta_0\right)'\Omega\left(\beta_0\right)^{-1}\right)$, and $\Xi=\Omega\left(\beta_0\right)^{-1}G\left(\beta_0\right)V_\beta$. By CLT for strong mixing sequences (see e.g. Ibragimov and Linnik, 1971) $\sqrt{n}\hat{h}\left(\beta_0\right)$ is Normally distributed with zero mean and variance $\Omega\left(\beta_0\right)$. The thesis follows from an application of the CMT. ■

Proof of Theorem 2. Consider the optimized BGEL criterion function

$$\hat{R}\left(\hat{\beta},\hat{\lambda}\right)=\frac{1}{b}\sum_{i=1}^{b}\rho\left(\hat{\lambda}'h_i\left(\hat{\beta}\right)\right) \tag{29}$$

being $\left(\hat{\beta},\hat{\lambda}\right)'$ consistent. Let us mean value expand the above function with Lagrange reminder

$$\hat{R}\left(\hat{\beta},\hat{\lambda}\right)=\rho\left(0\right)-\hat{\lambda}'\hat{h}\left(\hat{\beta}\right)-\frac{1}{2}\hat{\lambda}'\left(\frac{\hat{\Omega}\left(\hat{\beta}\right)}{M}+o_p\left(\frac{1}{b}\right)\right)\hat{\lambda} \tag{30}$$

where $\hat{h}\left(\hat{\beta}\right)=\sum_{i=1}^{b}h_i\left(\hat{\beta}\right)$, and, $\hat{\Omega}\left(\hat{\beta}\right)=M\sum_{i=1}^{b}h_i\left(\hat{\beta}\right)h_i\left(\hat{\beta}\right)'/b$. Moreover, for a given $\dot{\lambda}$ lying between 0 and $\hat{\lambda}$, and being $\hat{\lambda}$ consistent, we have $\rho_2\left(\dot{\lambda}'h_i\left(\hat{\beta}\right)\right)=-1+o_p\left(1\right)$. From (28) we obtain

$$\sqrt{n}\left(\hat{\beta}-\beta_0\right)=-\Xi'\sqrt{n}\hat{h}\left(\beta_0\right)+o_p\left(1\right) \tag{31}$$

and

$$\frac{\sqrt{n}}{M}\hat{\lambda}=V_\lambda\sqrt{n}\hat{h}\left(\beta_0\right)+o_p\left(1\right). \tag{32}$$

Consider again a mean value expansion on $\hat{h}\left(\hat{\beta}\right)$:

$$\hat{h}\left(\hat{\beta}\right)=\hat{h}\left(\beta_0\right)+\hat{G}\left(\dot{\beta}\right)\left(\hat{\beta}-\beta_0\right).$$

Then, from (31) and rescaling we obtain

$$\sqrt{n}\hat{h}\left(\hat{\beta}\right)=-\Omega\left(\beta_0\right)\frac{\sqrt{n}}{M}\hat{\lambda}+o_p\left(1\right) \tag{33}$$

$$=\Omega\left(\beta_0\right)V_\lambda\sqrt{n}\hat{h}\left(\beta_0\right)+o_p\left(1\right). \tag{34}$$

Notice that V_λ is a projection matrix; therefore, $V_\lambda=V_\lambda\Omega\left(\beta_0\right)V_\lambda$. By substituting (31), (32), and (34) into (30) and by appropriate rescaling, we get

$$
\begin{aligned}
\frac{n}{M}\left(\hat{R}\left(\hat{\beta},\hat{\lambda}\right)-\rho\left(0\right)\right) &= -\left(-V_{\lambda}\sqrt{n}\hat{h}\left(\beta_{0}\right)+o_{p}\left(1\right)\right)'\left(\Omega V_{\lambda}\sqrt{n}\hat{h}\left(\beta_{0}\right)+o_{p}\left(1\right)\right) \\
&\quad -\frac{1}{2}\left(-V_{\lambda}\sqrt{n}\hat{h}\left(\beta_{0}\right)+o_{p}\left(1\right)\right)'\left(\hat{\Omega}\left(\hat{\beta}\right)+o_{p}\left(\frac{M}{b}\right)\right) \\
&\quad \times\left(-V_{\lambda}\sqrt{n}\hat{h}\left(\beta_{0}\right)+o_{p}\left(1\right)\right).
\end{aligned}
$$

Rearranging and exploiting the properties of projection matrices

$$
2\frac{n}{M}\left(\hat{R}\left(\hat{\beta},\hat{\lambda}\right)-\rho\left(0\right)\right)=\sqrt{n}\hat{h}\left(\beta_{0}\right)'V_{\lambda}\Omega\left(\beta_{0}\right)V_{\lambda}\sqrt{n}\hat{h}\left(\beta_{0}\right)+o_{p}\left(1\right). \tag{35}
$$

By CLT for strong mixing sequences (see e.g. Ibragimov and Linnik, 1971) $\hat{h}\left(\beta_{0}\right)$ is Normally distributed with zero mean and variance $\Omega\left(\beta_{0}\right)$. Furthermore, the quadratic form in (35) converges to a chi square (See Rao (1965) pp. 153-157; see also Greene (2003) Theorems B.8, B.10, B.11). The corresponding degrees of freedom are given by the trace of $V_{\lambda}\Omega\left(\beta_{0}\right)$. Thus,

$$
\begin{aligned}
tr\left(V_{\lambda}\Omega\left(\beta_{0}\right)\right) &= tr\left(\left(\Omega\left(\beta_{0}\right)^{-1}-\Omega\left(\beta_{0}\right)^{-1}G\left(\beta_{0}\right)V_{\beta}G\left(\beta_{0}\right)'\Omega\left(\beta_{0}\right)^{-1}\right)\Omega\left(\beta_{0}\right)\right) \\
&= tr\left(I_{L_{g}}\right)-tr\left(G\left(\beta_{0}\right)'\Omega\left(\beta_{0}\right)^{-1}G\left(\beta_{0}\right)V_{\beta}\right) \\
&= L_{g}-L_{\beta}.
\end{aligned}
$$

Hence, the distance statistic $2\frac{n}{M}\left(\hat{R}\left(\hat{\beta},\hat{\lambda}\right)-\rho\left(0\right)\right)$ is distributed as a $\chi^{2}_{L_{g}-L_{\beta}}$. By (34), and solving for $\hat{h}\left(\beta_{0}\right)$ we can redefine (35) as

$$
2\frac{n}{M}\left(\hat{R}\left(\hat{\beta},\hat{\lambda}\right)-\rho\left(0\right)\right)=n\hat{h}\left(\hat{\beta}\right)'\Omega\left(\beta_{0}\right)^{-1}\hat{h}\left(\hat{\beta}\right)+o_{p}\left(1\right). \tag{36}
$$

Analogously, by (33), we find

$$2\frac{n}{M}\left(\hat{R}\left(\hat{\beta},\hat{\lambda}\right)-\rho\left(0\right)\right)=\frac{n}{M^{2}}\hat{\lambda}'\Omega\left(\beta_{0}\right)\hat{\lambda}+o_{p}\left(1\right). \tag{37}$$

Hence. (35), (36). and (37) are asymptotically equivalent and chi square distributed with $L_g - L_\beta$ degrees of freedom. ■

Proof of Corollary 1. By substituting (**??**) into $2\frac{n}{M}\sum_{i=1}^{b}\hat{\pi}_i\log\left(\frac{\hat{\pi}_i}{1/b}\right)$ and after some tedious algebra

$$\begin{aligned}&2\frac{n}{M}\left(\sum_{i=1}^{b}\frac{\exp\left(\hat{\lambda}'h_i\left(\hat{\beta}\right)\right)}{\sum_{j=1}^{b}\exp\left(\hat{\lambda}'h_j\left(\hat{\beta}\right)\right)}\log\left(\frac{b\exp\left(\hat{\lambda}'h_i\left(\hat{\beta}\right)\right)}{\sum_{j=1}^{b}\exp\left(\hat{\lambda}'h_j\left(\hat{\beta}\right)\right)}\right)\right)\\ =\;&2\frac{n}{M}\left(\log\frac{1}{b}\sum_{i=1}^{b}\exp\left(\hat{\lambda}'h_i\left(\hat{\beta}\right)\right)\right).\end{aligned}$$

The result follows also by exploiting the fact that $\sum_{i=1}^{b}\hat{\pi}_i h_i\left(\hat{\beta}\right)=0$. Let us mean value expand the right hand side of the above expression about $\lambda=0$. By consistency of $\hat{\lambda}$ we get

$$\begin{aligned}2\frac{n}{M}\left(\log\frac{1}{b}\sum_{i=1}^{b}\exp\left(\hat{\lambda}'h_i\left(\hat{\beta}\right)\right)\right) &= 2\frac{n}{M}\left(-\hat{\lambda}'\hat{h}\left(\hat{\beta}\right)-\frac{1}{2}\hat{\lambda}'\left(\frac{1}{b}\sum_{i=1}^{b}h_i\left(\hat{\beta}\right)h_i\left(\hat{\beta}\right)'\right)\hat{\lambda}\right)\\ &\quad+2\frac{n}{M}\left(\frac{1}{2}\left(\hat{\lambda}'\hat{h}\left(\hat{\beta}\right)\right)\left(\hat{\lambda}'\hat{h}\left(\hat{\beta}\right)\right)'\right).\end{aligned}$$

Notice that since the order of magnitude of $\hat{\lambda}'\hat{h}\left(\hat{\beta}\right)$ is $O_p\left(M/n\right)$ the last term of the expansion is negligible. Hence, by (33)

$$2\frac{n}{M}\left(\log\frac{1}{b}\sum_{i=1}^{b}\exp\left(\hat{\lambda}'h_i\left(\hat{\beta}\right)\right)\right)=\frac{n}{M^{2}}\hat{\lambda}'\Omega\hat{\lambda}+o_p\left(1\right),$$

so from Theorem 2 the result follows. ■

Proof of Theorem 3. Let us define the first order conditions (FOCs) for the BGEL function. Consider the restriction

$$a(\beta_0) = 0,$$

such that $a(\cdot) \in \mathbb{R}^{L_a}$, and expand $a\left(\hat{\beta}\right)$ about β_0 by mean value theorem:

$$\begin{aligned} a\left(\hat{\beta}\right) &= a(\beta_0) + \frac{\partial a\left(\dot{\beta}\right)}{\partial \beta}\left(\hat{\beta} - \beta_0\right) \\ &= \frac{\partial a\left(\dot{\beta}\right)}{\partial \beta}\left(\hat{\beta} - \beta_0\right) \end{aligned}$$

Then

$$n^{1/2} a\left(\hat{\beta}\right) = \frac{\partial a\left(\dot{\beta}\right)}{\partial \beta} n^{1/2}\left(\hat{\beta} - \beta_0\right)$$

Since $\dot{\beta}$ lies in the line that joins β_0 and $\hat{\beta}$, and since $\hat{\beta}$ is a consistent estimator for β_0, $\dot{\beta}$ converges in probability to β_0, i.e. $\dot{\beta} \to_p \beta_0$. From Theorem 1

$$n^{1/2}\left(\hat{\beta} - \beta_0\right) \to_d N(0, V_\beta)$$

where $V_\beta = \left(G(\beta_0)' \Omega(\beta_0)^{-1} G(\beta_0)\right)^{-1}$. $G(\beta_0)$ and $\Omega(\beta_0)$ are described in assumptions A3(iv) and A2(v) respectively. Thus, by CLT and CMT,

$$n^{1/2} a\left(\hat{\beta}\right) \to_d N\left(0, \frac{\partial a(\beta_0)}{\partial \beta'} V_\beta \frac{\partial a(\beta_0)}{\partial \beta}\right) \tag{38}$$

or, similarly,

$$\left(\frac{\partial a(\beta_0)}{\partial \beta'} V_\beta \frac{\partial a(\beta_0)}{\partial \beta}\right)^{-1/2} n^{1/2} a\left(\hat{\beta}\right) \to_d N(0, I)$$

For the same arguments used in Theorem 2 for the convergence of quadratic forms (i.e.

Greene (2003) Theorems B.8, B.10, and B.11, and Rao (1965) pp. 153-157),

$$na\left(\hat{\beta}\right)'\left(\frac{\partial a\left(\beta_0\right)}{\partial\beta'}V_\beta\frac{\partial a\left(\beta_0\right)}{\partial\beta}\right)^{-1}a\left(\hat{\beta}\right)\rightarrow_d\chi^2_{L_a} \tag{39}$$

Let us define the Lagrangian for the constrained optimization under the constraints defined above

$$\begin{aligned}\mathcal{L}\left(\beta,\lambda,\zeta\right) &= \left(\hat{R}\left(\beta,\lambda\right)-\rho\left(0\right)\right)-\zeta' a\left(\beta\right)\\ &= \frac{1}{b}\sum_{i=1}^{b}\left(\rho\left(\lambda' h_i\left(\beta\right)\right)-\rho\left(0\right)\right)-\zeta' a\left(\beta\right)\end{aligned} \tag{40}$$

The resulting FOCs are

$$0=\hat{R}_\beta\left(\tilde{\beta},\tilde{\lambda},\tilde{\zeta}\right)=\frac{1}{b}\sum_{i=1}^{b}\rho_1\left(\tilde{\lambda}' h_i\left(\tilde{\beta}\right)\right)\frac{\partial h_i\left(\tilde{\beta}\right)}{\partial\beta'}\tilde{\lambda}-\frac{\partial a\left(\tilde{\beta}\right)}{\partial\beta'}\tilde{\zeta} \tag{41}$$

$$0=\hat{R}_\lambda\left(\tilde{\beta},\tilde{\lambda},\tilde{\zeta}\right)=\frac{1}{b}\sum_{i=1}^{b}\rho_1\left(\tilde{\lambda}' h_i\left(\tilde{\beta}\right)\right)h_i\left(\tilde{\beta}\right) \tag{42}$$

$$0=\hat{R}_\zeta\left(\tilde{\beta},\tilde{\lambda},\tilde{\zeta}\right)=a\left(\tilde{\beta}\right) \tag{43}$$

Consistency of the constrained estimator $\tilde{\beta}$ is ensured by the fact that the corresponding parameter space $\mathcal{B}_a$, defined as $\mathcal{B}\cap\{\beta : a(\beta)=0\}$, is compact. Let us now expand around $(\beta_0, 0)$ the FOCs of the unconstrained optimization by mean value theorem evaluated at both the constrained estimator and the unconstrained estimator

$$\begin{aligned}0 &= M^{-1}n^{1/2}\hat{R}_\beta\left(\hat{\beta},\hat{\lambda}\right)\\ &= M^{-1}n_\beta^{1/2}\hat{R}_\beta\left(\beta_0,0\right)+\hat{R}_{\beta\lambda}\left(\dot{\beta},\dot{\lambda}\right)M^{-1}n^{1/2}\left(\hat{\lambda}-0\right)\\ &\quad +M^{-1}\hat{R}_{\beta\beta}\left(\dot{\beta},\dot{\lambda}\right)n^{1/2}\left(\hat{\beta}-\beta_0\right)\end{aligned}$$

$$\begin{aligned} 0 &= n^{1/2}\hat{R}_{\lambda}\left(\hat{\beta},\hat{\lambda}\right) \\ &= n^{1/2}\hat{R}_{\lambda}\left(\beta_0,0\right)+M\hat{R}_{\lambda\lambda}\left(\dot{\beta},\dot{\lambda}\right)M^{-1}n^{1/2}\left(\hat{\lambda}-0\right) \\ &\quad +\hat{R}_{\lambda\beta}\left(\dot{\beta},\dot{\lambda}\right)n^{1/2}\left(\hat{\beta}-\beta_0\right) \end{aligned}$$

and

$$\begin{aligned} & M^{-1}n^{1/2}\hat{R}_{\beta}\left(\tilde{\beta},\tilde{\lambda}\right) \\ =\ & M^{-1}n^{1/2}\hat{R}_{\beta}\left(\beta_0,0\right)+\hat{R}_{\beta\lambda}\left(\dot{\beta},\dot{\lambda}\right)M^{-1}n^{1/2}\left(\tilde{\lambda}-0\right) \\ & +M^{-1}\hat{R}_{\beta\beta}\left(\dot{\beta},\dot{\lambda}\right)n^{1/2}\left(\tilde{\beta}-\beta_0\right) \end{aligned}$$

$$\begin{aligned} 0 &= n^{1/2}\hat{R}_{\lambda}\left(\tilde{\beta},\tilde{\lambda}\right) \\ &= n^{1/2}\hat{R}_{\lambda}\left(\beta_0,0\right)+M\hat{R}_{\lambda\lambda}\left(\dot{\beta},\dot{\lambda}\right)M^{-1}n^{1/2}\left(\tilde{\lambda}-0\right) \\ &\quad +\hat{R}_{\lambda\beta}\left(\dot{\beta},\dot{\lambda}\right)n^{1/2}\left(\tilde{\beta}-\beta_0\right) \end{aligned}$$

Note that $0=\hat{R}_{\lambda}\left(\tilde{\beta},\tilde{\lambda},\tilde{\zeta}\right)=\hat{R}_{\lambda}\left(\tilde{\beta},\tilde{\lambda}\right)$ and $\hat{R}_{\beta\lambda}\left(\beta,\lambda\right)=\hat{R}_{\lambda\beta}\left(\beta,\lambda\right)'$. Again, the asymptotic behaviour of the components of the first order conditions is analyzed following the results in Lemma 1 Lemma 2, and consistency of $\tilde{\lambda}$:

$$M^{-1}\hat{R}_{\beta\beta}\left(\dot{\beta},\dot{\lambda}\right)=\frac{1}{Mb}\sum_{i=1}^{b}\rho_1\left(\dot{\lambda}'h_i\left(\dot{\beta}\right)\right)\frac{\partial^2 h_i\left(\dot{\beta}\right)}{\partial\beta\partial\beta'}\dot{\lambda}\rightarrow_p 0 \tag{44}$$

$$\hat{R}_{\beta\lambda}\left(\dot{\beta},\dot{\lambda}\right)'=\hat{R}_{\lambda\beta}\left(\dot{\beta},\dot{\lambda}\right)=\frac{1}{b}\sum_{i=1}^{b}\rho_1\left(\dot{\lambda}'h_i\left(\dot{\beta}\right)\right)\frac{\partial h_i\left(\dot{\beta}\right)}{\partial\beta}\rightarrow_p -G \tag{45}$$

$$M\hat{R}_{\lambda\lambda}\left(\dot{\beta},\dot{\lambda}\right)=\frac{M}{b}\sum_{i=1}^{b}\rho_2\left(\dot{\lambda}'h_i\left(\dot{\beta}\right)\right)h_i\left(\dot{\beta}\right)h_i\left(\dot{\beta}\right)'\rightarrow_p -\Omega \tag{46}$$

By taking the difference of the first order conditions with respect to the constrained and the unconstrained estimator and reformulating the equations in matrix form, we get

$$\begin{pmatrix} 0 \\ M^{-1}n^{1/2}\hat{R}_{\beta}\left(\tilde{\beta},\tilde{\lambda}\right) \end{pmatrix} = \begin{pmatrix} M\hat{R}_{\lambda\lambda}\left(\dot{\beta},\dot{\lambda}\right) & \hat{R}_{\lambda\beta}\left(\dot{\beta},\dot{\lambda}\right) \\ \hat{R}_{\beta\lambda}\left(\dot{\beta},\dot{\lambda}\right) & M^{-1}\hat{R}_{\beta\beta}\left(\dot{\beta},\dot{\lambda}\right) \end{pmatrix} \begin{pmatrix} M^{-1}n^{1/2}\left(\tilde{\lambda}-\hat{\lambda}\right) \\ n^{1/2}\left(\tilde{\beta}-\hat{\beta}\right) \end{pmatrix}$$

The 2×2 block matrix converges in probability to the quantities in (44)-(46). Hence, by solving for $\left(\tilde{\lambda}-\hat{\lambda}\right)$ and $\left(\tilde{\beta}-\hat{\beta}\right)$

$$\begin{pmatrix} M^{-1}n^{1/2}\left(\tilde{\lambda}-\hat{\lambda}\right) \\ n^{1/2}\left(\tilde{\beta}-\hat{\beta}\right) \end{pmatrix} = -\begin{pmatrix} V_{\lambda} & \Xi \\ \Xi' & -V_{\beta} \end{pmatrix} \begin{pmatrix} 0 \\ M^{-1}n^{1/2}\hat{R}_{\beta}\left(\tilde{\beta},\tilde{\lambda}\right) \end{pmatrix} + o_p(1). \tag{47}$$

Note now that from the constraints we obtain

$$n^{1/2}a\left(\hat{\beta}\right) = \frac{\partial a\left(\dot{\beta}\right)}{\partial\beta} n^{1/2}\left(\tilde{\beta}-\hat{\beta}\right). \tag{48}$$

Then, by substituting the equation for $n^{1/2}\left(\tilde{\beta}-\hat{\beta}\right)$ in (47) in (48), we find

$$\begin{aligned} n^{1/2}a\left(\hat{\beta}\right) &= \frac{\partial a\left(\dot{\beta}\right)}{\partial\beta} V_{\beta} M^{-1} n^{1/2}\hat{R}_{\beta}\left(\tilde{\beta},\tilde{\lambda}\right) + o_p(1) \\ &= M^{-1}n^{1/2}\frac{\partial a\left(\dot{\beta}\right)}{\partial\beta} V_{\beta} \frac{\partial a\left(\tilde{\beta}\right)}{\partial\beta'}\tilde{\zeta} + o_p(1) \end{aligned} \tag{49}$$

Note that $\hat{R}_{\beta}\left(\tilde{\beta},\tilde{\lambda},\tilde{\zeta}\right) = \hat{R}_{\beta}\left(\tilde{\beta},\tilde{\lambda}\right) - \frac{\partial a(\tilde{\beta})}{\partial\beta'}\tilde{\zeta} = 0$. Therefore, by (38),

$$\begin{aligned} M^{-1}n^{1/2}\tilde{\zeta} &= \left(\frac{\partial a\left(\dot{\beta}\right)}{\partial\beta} V_{\beta} \frac{\partial a\left(\tilde{\beta}\right)}{\partial\beta'}\right)^{-1} n^{1/2}a\left(\hat{\beta}\right) \\ &\to {}_d N\left(0, \left(\frac{\partial a\left(\beta_0\right)}{\partial\beta'} V_{\beta} \frac{\partial a\left(\beta_0\right)}{\partial\beta}\right)^{-1}\right) \end{aligned}$$

and

$$\left(\frac{\partial a(\beta_0)}{\partial \beta'} V_\beta \frac{\partial a(\beta_0)}{\partial \beta}\right)^{1/2} M^{-1} n^{1/2} \tilde{\zeta} \to_d N(0, I).$$

Finally, by the properties of the quadratic forms invoked in Theorem 2

$$M^{-2} n \tilde{\zeta}' \left(\frac{\partial a(\beta_0)}{\partial \beta'} V_\beta \frac{\partial a(\beta_0)}{\partial \beta}\right) \tilde{\zeta} \to_d \chi^2_{L_a} \tag{50}$$

Let us now expand again by mean value theorem the BGEL function, evaluated at the constrained $\left(\tilde{\beta}, \tilde{\lambda}\right)$ estimator, around the unconstrained estimator $\left(\hat{\beta}, \hat{\lambda}\right)$:

$$\begin{aligned}
& \hat{R}\left(\tilde{\beta}, \tilde{\lambda}\right) \\
= \; & \hat{R}\left(\hat{\beta}, \hat{\lambda}\right) + R_\beta\left(\dot{\beta}, \dot{\lambda}\right)\left(\tilde{\beta} - \hat{\beta}\right) \\
& + \hat{R}_\lambda\left(\dot{\beta}, \dot{\lambda}\right)\left(\tilde{\lambda} - \hat{\lambda}\right) + \frac{1}{2}\left(\tilde{\beta} - \hat{\beta}\right)' \hat{R}_{\beta\beta}\left(\dot{\beta}, \dot{\lambda}\right)\left(\tilde{\beta} - \hat{\beta}\right) \\
& + \frac{1}{2}\left(\tilde{\lambda} - \hat{\lambda}\right)' \hat{R}_{\beta\lambda}\left(\dot{\beta}, \dot{\lambda}\right)\left(\tilde{\lambda} - \hat{\lambda}\right) + \frac{1}{2}\left(\tilde{\beta} - \hat{\beta}\right)' \hat{R}_{\lambda\beta}\left(\dot{\beta}, \dot{\lambda}\right)\left(\tilde{\beta} - \hat{\beta}\right) \\
& + \frac{1}{2}\left(\tilde{\lambda} - \hat{\lambda}\right)' \hat{R}_{\lambda\lambda}\left(\dot{\beta}, \dot{\lambda}\right)\left(\tilde{\lambda} - \hat{\lambda}\right)
\end{aligned}$$

As $\left\|\hat{\beta} - \beta_0\right\| = \tau \left\|\dot{\beta} - \beta_0\right\|$ and $\left\|\hat{\lambda}\right\| = \tau \left\|\dot{\lambda}\right\|$ for $0 \leq \tau < 1$, we have that $\left(\dot{\beta}, \dot{\lambda}\right) \to_p (\beta_0, 0)$. Rearranging, we get

$$\begin{aligned}
& 2nM^{-1}\left(\hat{R}\left(\tilde{\beta}, \tilde{\lambda}\right) - \hat{R}\left(\hat{\beta}, \hat{\lambda}\right)\right) \\
= \; & n^{1/2} \begin{pmatrix} M^{-1}\left(\tilde{\lambda} - \hat{\lambda}\right) \\ \tilde{\beta} - \hat{\beta} \end{pmatrix}' \begin{pmatrix} M\hat{R}_{\lambda\lambda}(\beta_0, 0) & \hat{R}_{\lambda\beta}(\beta_0, 0) \\ \hat{R}_{\beta\lambda}(\beta_0, 0) & 0 \end{pmatrix} n^{1/2} \begin{pmatrix} M^{-1}\left(\tilde{\lambda} - \hat{\lambda}\right) \\ \tilde{\beta} - \hat{\beta} \end{pmatrix} + o_p(1)
\end{aligned}$$

Moreover, from (47) and rearranging

$$2nM^{-1}\left(\hat{R}\left(\tilde{\beta}, \tilde{\lambda}\right) - \hat{R}\left(\hat{\beta}, \hat{\lambda}\right)\right) = M^{-2} n \tilde{\lambda}' G V_\beta G' \tilde{\lambda} + o_p(1).$$

Which implies,

$$2nM^{-1}\left(\hat{R}\left(\tilde{\beta},\tilde{\lambda}\right)-\hat{R}\left(\hat{\beta},\hat{\lambda}\right)\right)\rightarrow_{d}\chi^{2}_{L_a} \tag{51}$$

Consider now the Minimum Chi Square criterion, the result in (48), and (49), then,

$$n\left(\tilde{\beta}-\hat{\beta}\right)'V_{\beta}^{-1}\left(\tilde{\beta}-\hat{\beta}\right)=\frac{n}{M^2}\zeta'\frac{\partial a\left(\tilde{\beta}\right)}{\partial\beta'}V_{\beta}\frac{\partial a\left(\tilde{\beta}\right)}{\partial\beta}\zeta+o_p\left(1\right)$$

for $V_{\beta}^{-1}=\hat{V}_{\beta}^{-1}+o_p\left(1\right)$. Thus, because of (50) the above quadratic form is chi square distributed with L_a degrees of freedom. ∎

	J_{2GMM}	J_{IGMM}		D^{EL}	LM_{BW}^{EL}	LM_{π}^{EL}	BJ^{EL}	D^{ET}	LM_{BW}^{ET}	LM_{π}^{ET}	BJ^{ET}	KL^{ET}
$n = 256$			M, L									
$L_g = 2$	.054 .104	.054 .104	8, 1	.073 .120	.080 .130	.068 .117	.054 .105	.068 .119	.094 .144	.068 .120	.053 .105	.073 .122
			8, 8	.073 .124	.091 .143	.073 .125	.054 .108	.068 .122	.094 .146	.066 .120	.054 .107	.072 .124
			4, 1	.060 .115	.068 .119	.059 .119	.052 .104	.060 .114	.078 .129	.060 .114	.052 .104	.063 .116
			4, 4	.061 .115	.072 .121	.061 .114	.054 .105	.061 .115	.076 .132	.060 .113	.054 .105	.062 .117
$L_g = 3$	.158 .263	.158 .263	8, 1	.074 .128	.068 .121	.068 .121	.041 .094	.072 .127	.123 .181	.069 .127	.045 .101	.078 .132
			8, 8	.089 .148	.136 .188	.087 .146	.045 .099	.070 .134	.117 .169	.053 .109	.052 .112	.066 .118
			4, 1	.065 .113	.073 .124	.059 .107	.042 .095	.062 .115	.094 .142	.059 .113	.050 .106	.066 .118
			4, 4	.066 .123	.088 .145	.065 .119	.046 .097	.063 .116	.088 .141	.055 .103	.059 .111	.060 .109
$L_g = 4$	.284 .432	.284 .432	8, 1	.081 .142	.231 .305	.219 .294	.096 .179	.077 .135	.158 .223	.071 .127	.032 .086	.085 .143
			8, 8	.113 .173	.312 .381	.301 .372	.112 .205	.090 .153	.175 .238	.074 .134	.038 .091	.095 .153
			4, 1	.066 .118	.172 .244	.165 .238	.103 .183	.065 .117	.112 .164	.058 .113	.043 .096	.068 .120
			4, 4	.078 .134	.224 .295	.215 .286	.113 .197	.072 .127	.121 .177	.062 .119	.047 .105	.073 .130
$n = 512$												
$L_g = 2$	.047 .096	.047 .096	8, 1	.052 .100	.054 .102	.049 .098	.044 .095	.053 .102	.065 .115	.055 .102	.045 .095	.054 .103
			8, 8	.054 .103	.060 .106	.053 .106	.048 .097	.053 .104	.066 .115	.053 .103	.052 .097	.055 .104
			4, 1	.050 .099	.049 .099	.048 .097	.046 .097	.050 .101	.057 .105	.050 .101	.050 .097	.050 .101
			4, 4	.052 .100	.052 .104	.050 .099	.048 .095	.052 .099	.058 .109	.051 .099	.059 .095	.053 .100
$L_g = 3$	.146 .254	.146 .254	8, 1	.059 .106	.063 .107	.055 .100	.047 .093	.059 .107	.081 .130	.057 .105	.053 .104	.060 .107
			8, 8	.064 .117	.078 .129	.061 .114	.050 .095	.061 .112	.076 .125	.053 .095	.056 .110	.055 .100
			4, 1	.051 .100	.053 .099	.050 .099	.047 .093	.045 .102	.069 .118	.054 .102	.058 .109	.055 .104
			4, 4	.057 .104	.059 .111	.055 .102	.049 .094	.057 .103	.063 .114	.048 .094	.059 .111	.050 .098
$L_g = 4$	.270 .423	.270 .423	8, 1	.056 .108	.173 .240	.168 .237	.111 .194	.058 .109	.093 .147	.052 .104	.041 .095	.058 .110
			8, 8	.063 .121	.218 .284	.215 .280	.119 .206	.058 .114	.099 .158	.055 .104	.045 .100	.060 .112
			4, 1	.046 .095	.130 .196	.127 .196	.113 .191	.048 .095	.070 .120	.045 .092	.045 .099	.047 .096
			4, 4	.053 .101	.151 .218	.146 .216	.117 .195	.052 .099	.075 .125	.048 .094	.049 .099	.051 .099

Table 1: Overidentification tests for the linear model

L_g= 3	J_{2GMM}	J_{IGMM}	M,L	D^{EL}	LM^{EL}_{BW}	LM^{EL}_{π}	BJ^{EL}	D^{ET}	LM^{ET}_{BW}	LM^{ET}_{π}	BJ^{ET}	KL^{ET}
$n = 256$												
$\rho = .4$	.098 .155	.076 .132	8, 1	.115 .175	.146 .196	.112 .163	.077 .133	.137 .191	.191 .250	.098 .145	.139 .192	.116 .175
			8, 8	.120 .182	.159 .218	.125 .178	.079 .136	.095 .151	.168 .213	.058 .106	.095 .153	.084 .139
			4, 1	.107 .167	.107 .153	.107 .153	.078 .129	.092 .144	.153 .207	.054 .096	.093 .145	.077 .129
			4, 4	.114 .171	.143 .193	.115 .162	.080 .133	.088 .143	.159 .208	.055 .099	.091 .145	.079 .133
$\rho = .6$	.099 .158	.072 .126	8, 1	.111 .170	.151 .200	.118 .163	.067 .123	.091 .147	.162 .216	.062 .096	.090 .146	.081 .132
			8, 8	.130 .192	.177 .219	.139 .186	.067 .128	.101 .155	.164 .220	.061 .100	.098 .164	.083 .135
			4, 1	.102 .161	.140 .187	.107 .153	.068 .116	.090 .143	.152 .202	.056 .092	.093 .148	.076 .128
			4, 4	.121 .182	.160 .211	.127 .175	.072 .142	.097 .155	.157 .208	.062 .102	.095 .167	.086 .139
$\rho = .8$	.139 .197	.104 .159	8, 1	.151 .207	.208 .230	.160 .208	.078 .134	.105 .153	.178 .227	.068 .100	.092 .157	.085 .132
			8, 8	.219 .275	.289 .325	.234 .283	.126 .186	.138 .187	.201 .252	.089 .125	.143 .209	.113 .159
			4, 1	.171 .230	.189 .231	.137 .177	.078 .131	.114 .164	.183 .233	.073 .118	.104 .167	.097 .148
			4, 4	.196 .258	.245 .290	.205 .258	.122 .183	.131 .180	.202 .255	.084 .122	.135 .197	.113 .161
$n = 512$												
$\rho = .4$	.086 .140	.075 .129	8, 1	.087 .152	.100 .150	.074 .128	.075 .131	.083 .137	.131 .182	.056 .100	.085 .143	.078 .126
			8, 8	.086 .134	.110 .160	.086 .134	.076 .130	.096 .159	.135 .186	.069 .124	.102 .172	.085 .145
			4, 1	.086 .146	.094 .147	.075 .123	.075 .129	.079 .133	.126 .171	.052 .095	.083 .137	.073 .119
			4, 4	.089 .149	.102 .154	.079 .130	.073 .128	.081 .134	.125 .175	.051 .098	.083 .138	.072 .125
$\rho = .6$	.087 .140	.071 .123	8, 1	.084 .144	.105 .153	.076 .124	.065 .120	.083 .132	.144 .197	.060 .092	.087 .139	.078 .123
			8, 8	.103 .162	.126 .180	.101 .147	.071 .127	.084 .138	.150 .203	.057 .095	.091 .149	.082 .126
			4, 1	.084 .139	.094 .148	.074 .119	.065 .116	.084 .138	.150 .203	.057 .095	.082 .140	.082 .126
			4, 4	.095 .153	.110 .165	.089 .135	.070 .125	.080 .136	.150 .191	.054 .096	.087 .141	.080 .127
$\rho = .8$	.112 .169	.089 .144	8, 1	.106 .161	.145 .187	.105 .151	.065 .118	.086 .137	.169 .216	.059 .094	.084 .143	.079 .124
			8, 8	.110 .167	.202 .246	.156 .220	.097 .151	.108 .159	.176 .229	.074 .108	.121 .186	.093 .159
			4, 1	.149 .217	.143 .183	.107 .153	.065 .115	.110 .162	.174 .230	.077 .119	.117 .173	.097 .147
			4, 4	.104 .155	.192 .221	.139 .183	.088 .134	.199 .254	.259 .316	.144 .198	.226 .284	.173 .241

Table 2: Overidentification tests for the nonlinear model with three moment functions

L_g= 5	J_{2GMM}	J_{IGMM}		D^{EL}	LM^{EL}_{BW}	LM^{EL}_{π}	BJ^{EL}	D^{ET}	LM^{ET}_{BW}	LM^{ET}_{π}	BJ^{ET}	KL^{ET}
$n = 256$			M, L									
$\rho = .4$	.425 .568	.366 .531	8, 1	.265 .327	.350 .403	.300 .369	.131 .212	.127 .192	.303 .361	.062 .095	.100 .174	.093 .145
			8, 8	.236 .309	.361 .413	.253 .321	.089 .166	.141 .201	.305 .365	.063 .102	.118 .200	.098 .149
			4, 1	.238 .312	.343 .384	.261 .321	.135 .210	.100 .161	.270 .331	.043 .070	.090 .159	.073 .121
			4, 4	.236 .307	.345 .390	.257 .317	.126 .198	.126 .183	.271 .340	.053 .085	.118 .192	.086 .136
$\rho = .6$	.433 .582	.368 .535	8, 1	.200 .273	.337 .392	.211 .281	.067 .132	.131 .187	.296 .359	.058 .091	.100 .173	.090 .133
			8, 8	.254 .336	.411 .460	.287 .361	.089 .163	.148 .206	.301 .358	.066 .098	.123 .208	.099 .145
			4, 1	.185 .258	.324 .366	.217 .280	.073 .134	.105 .162	.279 .339	.044 .072	.096 .159	.073 .117
			4, 4	.222 .294	.359 .405	.258 .322	.089 .162	.131 .190	.279 .342	.052 .085	.125 .203	.087 .129
$\rho = .8$	.459 .599	.395 .556	8, 1	.233 .308	.406 .451	.266 .332	.072 .129	.149 .202	.297 .352	.056 .086	.095 .168	.086 .124
			8, 8	.327 .406	.487 .536	.366 .429	.142 .219	.198 .256	.320 .371	.094 .126	.175 .265	.126 .166
			4, 1	.223 .295	.379 .420	.256 .327	.084 .141	.146 .203	.303 .360	.063 .094	.112 .183	.094 .142
			4, 4	.278 .355	.427 .473	.315 .383	.135 .209	.174 .234	.324 .377	.080 .109	.169 .248	.118 .162
$n = 512$												
$\rho = .4$	.380 .536	.350 .513	8, 1	.161 .230	.205 .256	.162 .213	.105 .176	.137 .218	.304 .378	.109 .181	.110 .173	.146 .225
			8, 8	.169 .239	.221 .274	.172 .232	.106 .177	.085 .139	.248 .308	.048 .082	.101 .171	.084 .127
			4, 1	.132 .202	.173 .217	.136 .186	.098 .161	.082 .134	.233 .300	.044 .077	.097 .163	.076 .127
			4, 4	.142 .213	.186 .236	.143 .199	.103 .163	.079 .134	.233 .293	.041 .073	.098 .162	.073 .118
$\rho = .6$	.388 .539	.348 .512	8, 1	.153 .217	.222 .260	.163 .211	.093 .155	.090 .141	.241 .310	.040 .068	.092 .160	.067 .114
			8, 8	.184 .254	.267 .314	.200 .260	.102 .171	.104 .152	.247 .313	.043 .069	.115 .189	.069 .117
			4, 1	.137 .199	.196 .234	.147 .195	.084 .147	.084 .133	.242 .304	.040 .068	.097 .158	.069 .116
			4, 4	.158 .222	.219 .266	.165 .220	.100 .164	.085 .136	.242 .300	.040 .063	.102 .171	.069 .114
$\rho = .8$	.412 .565	.369 .539	8, 1	.183 .256	.208 .343	.208 .268	.094 .157	.104 .152	.250 .310	.048 .068	.097 .163	.064 .096
			8, 8	.259 .329	.373 .416	.283 .351	.142 .207	.138 .186	.259 .319	.066 .088	.153 .226	.089 .122
			4, 1	.365 .424	.407 .442	.351 .399	.243 .313	.110 .156	.275 .331	.055 .080	.110 .173	.080 .121
			4, 4	.281 .350	.355 .397	.287 .343	.186 .253	.134 .185	.286 .342	.068 .094	.158 .228	.096 .140

Table 3: Overidentification tests for the nonlinear model with five moment functions

Ultimi Contributi di Ricerca CRENoS

I Paper sono disponibili in: http://www.crenos.it

09/11 *Francesca Mameli, Gerardo Marletto,* "Can national survey data be used to select a core set of indicators for monitoring the sustainability of urban mobility policies?"

09/10 *Emanuela Marrocu, Raffaele Paci,* "They arrive with new information. Tourism flows and production efficiency in the European regions"

09/09 *Oliviero A. Carboni,* "An Empirical investigation of the Determinants of R&D Cooperation"

09/08 *Fabiano Schivardi, Eliana Viviano,* "Entry Barriers in Retail Trade"

09/07 *Rinaldo Brau, Carlo Carraro,* "The Design of Voluntary Agreements in Oligopolistic Markets"

09/06 *Francesca Mameli, Gerardo Marletto,* "A participative procedure to select indicators of sustainable urban mobility policies"

09/05 *Claudio Detotto, Manuela Pulina,* "Does more crime mean fewer jobs? An ARDL model"

09/04 *Francesco Pigliaru,* "Persistent Regional Gaps and the Role of Social Capital: Hints from the Italian Mezzogiorno's case"

09/03 *Giovanni Sulis,* "Wage Returns to Experience and Tenure for Young Men in Italy"

09/02 *Guido Ferrari, Giorgio Garau, Patrizio Lecca,* "Constructing a Social Accounting Matrix for Sardinia"

09/01 *Giorgio Garau, Patrizio Lecca, Lucia Schirru,* "Does Deflation Method Matter for Productivity Measures?"

08/23 *Barbara Dettori, Emanuela Marrocu, Raffaele Paci,* "Total factor productivity, intangible assets and spatial dependence in the European regions"

08/22 *Fabio Cerina, Sauveur Giannoni,* "Pollution Adverse Tourists and Growth"

08/21 *Carina Hirsch,Giovanni Sulis,* "Schooling, Production Structure and Growth: An Empirical Analysis on Italian Regions"

08/20 *Fabio Cerina, Francesco Mureddu,* "Agglomeration and Growth with Endogenous Expediture Shares"

08/19 *Dimitri Paolini,* "Screening and short – term contracts"

08/18 *Massimo Del Gatto, Adriana Di Liberto, Carmelo Petraglia,* "Measuring Productivity"

08/17 *Edoardo Otranto,* "Identifying Financial Time Series with Similar Dynamic Conditional Correlation"

08/16 *Rinaldo Brau, Raffaele Doronzo, Carlo V. Fiorio, Massimo Florio,* "Gas Industry Reforms and Consumers' Prices in the European Union: an Empirical Analysis"

08/15 *Oliviero A. Carboni,* "The Effect of R&D Subsidies on Private R&D:Evidence from Italian Manufacturing Data"

08/14 *Gerardo Marletto,* "Getting out of the car. An institutional/evolutionary approach to sustainable transport policies"

08/13 *Francesco Lisi, Edoardo Otranto,* "Clustering Mutual Funds by Return and Risk Levels"

08/12 *Adriana Di Liberto, Francesco Pigliaru, Piergiorgio Chelucci,* "International TFP Dynamics and Human Capital Stocks: a Panel Data Analysis, 1960-2003"

08/11 *Giorgio Garau, Patrizio Lecca,* "Impact Analysis of Regional Knowledge Subsidy: a CGE Approach"

Finito di stampare nel mese di Novembre 2009
Presso **studiografico&stampadigitale Copy Right**
Via Torre Tonda 8 – Tel. 079.200395 – Fax 079.4360444
07100 Sassari

www.crenos.it

ISBN 978-88-84-67-557-6

9 788884 675576