ISSN 1755-5361

University of Essex

Department of Economics

Discussion Paper Series

No. 640 September 2007

GEL Estimation and Inference with Non-Smooth Moment Indicators and Dynamic Data

Gordon Kemp

GEL Estimation and Inference with Non-Smooth Moment Indicators and Dynamic Data

Gordon C.R. Kemp*

University of Essex

Revised Version: September 2007†

Abstract

In this paper we demonstrate consistency and asymptotic normality for Generalized Empirical Likelihood (GEL) estimation in dynamic models when the moment indicators being used are the non-differentiable functions of the parameters of interest.

KEYWORDS: Generalized Empirical Likelihood; non-differentiable sample moments; dynamic models.

---

*Address for correspondence: Dr. Gordon C.R. Kemp, Department of Economics, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, U.K. Tel: +44 (1206) 872766; Fax: +44 (1206) 872724; Email: kempgcr@essex.ac.uk.

†An earlier version of this paper was presented at the 2005 World Congress of the Econometric Society.

1 Introduction

In this paper we establish and asymptotic normality for Generalized Empirical Likelihood (GEL) estimators when population moment conditions are given by the expectations of non-differentiable functions of the parameters and the data are generated by a stationary ergodic time-series process. The GEL framework for estimation and inference was introduced by Smith (1997) as a common framework within which many quasi-likelihood alternatives to the Generalized Method of Moments (GMM) approach can be nested, including the Empirical Likelihood (EL) estimator introduced by Owen (1988) and further developed by Qin and Lawless (1994) and Imbens (1997), the continuous updating estimator (CUE) of Hansen, Heaton and Yaron (1996), and the exponential tilting estimator (ET) of Kitamura and Stutzer (1997) and Imbens, Spady and Johnson (1998). As such it has provided a useful framework for deriving the properties of such estimators and examining the relationships between them; see for example Newey and Smith (2004).

Smith (1997) provided heuristic demonstrations of consistency and asymptotic normality for GEL estimators when the data are generated by a stationary stochastic process and the population moment conditions are given by expectations of differentiable functions of the sample data. However, Smith (1997) provided neither formal statements of the assumptions under which consistency and asymptotic normality could be demonstrated nor formal proofs of such results. Newey and Smith (2004) rectified these deficiencies for the i.i.d. random sampling case in a paper primarily concerned with comparing and contrasting the higher-order asymptotic properties of the EL, CUE and ET estimators.

One feature of the existing literature on these quasi-likelihood alternatives is that it has dealt primarily with cases in which the population moment conditions are given by the expectations of differentiable functions of the parameters. In situations where these functions are non-differentiable the usual approach has been to smooth the functions as in Chen and Hall (1993) and Otsu (2003). The main exception appears to be Newey and Smith (2004) who demonstrate consistency, but not asymptotic normality, under assumptions which allow the population moment conditions to be given by the expectations of fairly general functions of the parameters which only need to be continuous with probability one at any fixed parameter value. However, as noted above, Newey and Smith (2004) limit attention to the i.i.d. random sampling case. The main contribution of the present paper is to demonstrate consistency and asymptotic normality for GEL estimators under an explicit set of assumptions when population moment conditions are given by the expectations of non-differentiable functions of the parameters and the data are generated by a stationary ergodic time-series process.

It is well-known that GMM and Maximum Likelihood (ML) estimators can be shown to be consistency and asymptotically normal in such circumstances under suitable conditions; see Newey and McFadden (1994). However, a distinguishing feature of the GEL class of estimators is that they are characterized as solutions to a saddlepoint problem rather than a minimization or maximization problem. Thus the theorems used by Newey and McFadden (1994) are not directly applicable as they do not cover saddlepoint problems as such. Instead, the strategy used in the present paper for demonstrating consistency is based on the saddlepoint approach of Newey and Smith (2004) while the strategy for demonstrating asymptotic normality is based on a combination of the saddlepoint approach of Newey and Smith (2004) and the approach of Newey and McFadden (1994). One point to note is that since the population moments are expectations which are not required to be continuous everywhere, it is possible that no solution to the GEL saddlepoint problem actually exists. The present paper considers estimators which asymptotically approximately solve the saddlepoint problem in a suitable fashion in order to address this issue.

In order to produce GEL estimators which are asymptotically as efficient as the standard two-stage GMM estimator in the context of dynamic data, the present paper like most of the literature on EL and related methods adopts the time-smoothing and blocking approach of Kitamura (1997) and Kitamura and Stutzer (1997). Like Kitamura and Stutzer (1997) we only consider formally the case where the estimator is generated using a uniform weighting function for smoothing the observations. However, it seems likely that the results can be generalized to other weighting functions albeit at the cost of modifying some of the assumptions and complicating some of the proofs.

The layout of the paper is as follows. Section 2 presents the model, objective function and the saddlepoint estimator. Section 3 demonstrates consistency for the saddlepoint estimator, using an approach similar to that of Newey and Smith (2004). Section 4 demonstrates asymptotic normality for the the saddlepoint estimator. Section 5 concludes the paper.

2 Model, Objective Function and Estimator

In the present paper, the parameter of interest, denoted β_0, which is an unknown element of the parameter space, denoted B_0 which is a known subset of p-dimensional Euclidean space, $\mathbb{R}^p$. This parameter of interest is characterized by means of a population moment condition:

$$E_0[g(Z,\beta_0)] = 0, \tag{1}$$

where Z is a $\mathcal{Z}$-valued random variable, where $\mathcal{Z} \subseteq \mathbb{R}^k$ for some $k < \infty$, $g(\cdot) : \mathbb{R}^k \times \mathbb{R}^p \to \mathbb{R}^q$ is a suitably measurable known function, and $E_0[\cdot]$ denotes the expectation operator with respect to the distribution of Z. The data on which the estimator will be based are obtained from a doubly-infinite stationary ergodic stochastic process $\{Z_i\}_{i=-\infty}^{+\infty}$ such that Z_i has the same marginal distribution as Z for each $i = 0, \pm 1, \pm 2, \ldots$.[1] We would like to define the estimator $\widehat{\beta}_n$ of β as the solution to the following saddlepoint optimization problem:

$$(P1) \qquad \min_{\beta \in B} \sup_{\lambda \in \widehat{\Lambda}_n(\beta)} \widehat{P}_n(\beta,\lambda), \tag{2}$$

[1] Notice that this framework is compatible with a situation in which there is an underlying doubly-infinite stationary ergodic stochastic process $\{X_i\}_{i=-\infty}^{+\infty}$ and $Z_i = (X_i, X_{i-1}, \ldots, X_{i-r})'$ for some fixed r for each $i = 0, \pm 1, \pm 2, \ldots$, as the resulting doubly-infinite stochastic process $\{Z_i\}_{i=-\infty}^{+\infty}$ will be stationary and ergodic.

where:

$$\widehat{P}_n(\beta, \lambda) = n^{-1} \sum_{i=1-m_n}^{n+m_n} \rho(\lambda' g_{i,n}^{\omega}(\beta)), \tag{3}$$

$$g_{i,n}^{\omega}(\beta) = (2m_n + 1)^{-1} \sum_{j=-m_n}^{m_n} g_{i-j,n}(\beta) \tag{4}$$

$$g_{i,n}(\beta) = 1_{\{1 \leq i \leq n\}} g_i(\beta), \tag{5}$$

$$g_i(\beta) = g(Z_i, \beta), \tag{6}$$

where $1_{\{\cdot\}}$ denotes the usual indicator function, $\rho(\cdot)$ is a twice continuously differentiable strictly concave function from $\mathcal{V}$, an open interval in $\mathbb{R}$ containing the point 0, to $\mathbb{R}$, such that the first derivative of $\rho(\cdot)$ at 0 is non-zero:

$$\widehat{\Lambda}_n(\beta) = \{\lambda \in \mathbb{R}^q : \lambda' g_{i,n}^{\omega}(\beta) \in \mathcal{V}, 1 - m_n \leq i \leq n + m_n\}, \tag{7}$$

and m_n is a lag truncation parameter which tends to infinity at a suitable rate as the sample size, n, tends to infinity. Without loss of generality, we will impose that $\rho(0) = 0$ and $\rho_1(0) = \rho_2(0) = -1$ where $\rho_1(u) = d\rho(u)/du$ and $\rho_2(u) = d^2\rho(u)/du^2$; see Newey and Smith (2004).

Unfortunately it is possible that there is no solution to the saddlepoint problem (P1) since $g(z, \beta)$ need not be continuous in β for all z and β. Therefore we will characterize the estimator $\widehat{\beta}_n$ of β as a mapping from $\{Z_i\}_{i=-\infty}^{\infty}$ to B such that:

$$\widehat{P}_n^*(\widehat{\beta}_n) \leq \inf_{\beta \in B} \widehat{P}_n^*(\beta) + o_p(n^{-\sigma}), \tag{8}$$

where:

$$\widehat{P}_n^*(\beta) = \sup_{\lambda \in \widehat{\Lambda}_n(\beta)} \widehat{P}_n(\beta, \lambda), \tag{9}$$

and σ is a suitable strictly positive constant. Since $0 \in \widehat{\Lambda}_n(\beta)$ for all $\beta \in B$ it follows $\widehat{P}_n^*(\beta) \geq 0$ for all $\beta \in B$ and hence $\inf_{\beta \in B} \widehat{P}_n^*(\beta)$ must exist and be non-negative. It follows that for any chosen $\sigma > 0$ there will exist a sequence of mappings $\{\widehat{\beta}_n\}_{n=1}^{\infty}$ satisfying Equation (8). This definition of the estimator does not guarantee that it is a random variable; however, this issue can easily be circumvented provided that we substitute outer measure for probability in the assumptions and the proofs.[2]

The objective function $\widehat{P}_n(\beta, \lambda)$ used here differs from that used by Smith (1997) in two respects. First, as noted in the Introduction, our objective function imposes a uniform weighting function in the construction of the smoothed contributions given by Equation (4), whereas Smith (1997) allows for a more general range of weighting functions. We have chosen this particular weighting function primarily for simplicity but it seems likely that the assumptions and proofs can be modified to enable us to demonstrate consistency and asymptotic normality for GEL estimators defined using a wider class of weighting functions. Suppose that $\{\omega_m(\cdot)\}_{m=0}^{\infty}$ is a sequence of weighting function such that for each $m = 0, 1, \ldots,$ $\omega_m(u) = 0$ for $|u| \geq (2m+1)$ and $\sum_{j=-m}^{m} \omega_m(j) = 1$. Now define $\sigma_{1m} = (2m+1)\sup_{-m \leq j \leq m} |\omega_m(j)|$, $\sigma_{2m} = \sum_{j=-m}^{m} |\omega_m(j)|$ and $\sigma_{2m} = (2m+1)\sum_{j=-m}^{m} \omega_m(j)^2$. The main conditions which we then need on $\{\omega_m(\cdot)\}_{m=0}^{\infty}$ seem to be that each σ_{jm} converges to a finite non-zero limit as $m \to \infty$ for $j = 1, 2, 3$. In the case of the uniform weighting function used here, $\sigma_{jm} = 1$ for all $m = 0, 1, 2, \ldots,$ and $j = 1, 2, 3$ so that these limiting conditions are automatically satisfied.

Second, we have treated the endpoints of the sample slightly differently than seems to be done by Smith (1997). In particular, we have smoothed in zeros at both ends of the sample. Doing so results in rather convenient expressions for the first and second derivatives of $\widehat{P}_n(\beta, \lambda)$ with respect to λ evaluated at $\lambda = 0$. In particular the first derivative of $\widehat{P}_n(\beta, \lambda)$ with respect to λ is given by:

$$\left[\frac{\partial \widehat{P}_n}{\partial \lambda}\right] = n^{-1} \sum_{i=1-m_n}^{n+m_n} \rho_1(\lambda' g_{i,n}^{\omega}(\beta)) g_{i,n}^{\omega}(\beta). \tag{10}$$

Evaluating this at $\lambda = 0$ gives:

$$\left[\left.\frac{\partial \widehat{P}_n}{\partial \lambda}\right|_{\lambda=0}\right] = -n^{-1} \sum_{s=1}^{n} g_s(\beta), \tag{11}$$

[2] If $(\Omega, \mathcal{F}, P)$ is a probability space then for any subset $A \subseteq \Omega$ we define the outer measure of A, denoted $P^*(A)$, as the infimum of $E(Y)$ over all random variables Y on $(\Omega, \mathcal{F}, P)$ such that $Y(\omega) \geq 1_A(\omega)$ for all $\omega \in \Omega$ where $1_A(\cdot) : \Omega \to \{0, 1\}$ denotes the indicator function for the set A.

where $S_n = (2m_n + 1)$, so the first derivative of $\widehat{P}_n(\beta, \lambda)$ with respect to λ evaluate at $\lambda = 0$ is equal to minus the sample moment condition used in GMM. In addition, the second derivative of $\widehat{P}_n(\beta, \lambda)$ with respect to λ is given by:

$$\left[\frac{\partial^2 \widehat{P}_n}{\partial \lambda \partial \lambda'}\right] = n^{-1} \sum_{i=1-m_n}^{n+m_n} \rho_2(\lambda' g^{\omega}_{i,n}(\beta)) g^{\omega}_{i,n}(\beta) g^{\omega}_{i,n}(\beta)' = \widehat{H}_n(\beta, \lambda). \tag{12}$$

Evaluating this at $\lambda = 0$ gives:

$$\widehat{H}_n(\beta, 0) = -S_n^{-1} \sum_{j=-2m_n}^{2m_n} \kappa_B \left(\frac{j}{2m_n + 1}\right) \widehat{\Gamma}_{s,n}(\beta), \tag{13}$$

where:

$$\widehat{\Gamma}_{j,n}(\beta) = n^{-1} \sum_{i=\max(1,1+j)}^{\min(n,n+j)} g_i(\beta) g_{i-j}(\beta)', \quad j = 0, \ldots, S_n, \tag{14}$$

and $\kappa_B(u) = \max(0, 1 - |u|)$ denotes the Bartlett kernel, so that the second derivative of $\widehat{P}_n(\beta, \lambda)$ with respect to λ evaluated at $\lambda = 0$ is proportional to the Bartlett kernel HAC estimator of the variance of the sample moment condition used in GMM. Note that this implicitly requires that $m_n < n/2$.

Designing the objective function so that it treats the endpoints in this fashion thus effectively ensures that the class of GEL estimators is as closely related as possible to the class of GMM estimators in which the weighting matrix is given by the inverse of a Bartlett kernel HAC estimator of the variance matrix of the sample moments. In particular, it guarantees that the class of GEL estimators as defined here does in fact include the CUE estimator based on the Bartlett kernel HAC estimator of the variance matrix of the sample moments.

In order to proceed further we make the following assumptions.

Assumptions

A1. *Data Generation Process (DGP)*

The doubly-infinite stochastic process $\{Z_i\}_{i=-\infty}^{+\infty}$ is strictly stationary and ergodic.

A2. *Parameter Space*

B is a compact subset of $\mathbb{R}^p$.

A3. *Moment Condition*

There exists a unique $\beta_0 \in B$ such that $E_0[g(Z_i, \beta_0)] = 0$ for $i = 0, \pm 1, \pm 2, \ldots$.

A4. *Continuity with Probability One*

For each $\bar{\beta} \in B$ and each $i = 0, \pm 1, \pm 2, \ldots$, $g(Z_i, \beta)$ is continuous with respect to β with probability one at $\beta = \bar{\beta}$.

A5. *Dominance*

There exists a measurable non-negative scalar function $d(z)$ and a finite scalar constant $\nu > 2$ such that:

(i) $\sup_{\beta \in B} \|g(z, \beta)\| \leq d(z)$ for all z; and

(ii) $E_0[d(Z_i)^\nu] < \infty$ for each $i = 0, \pm 1, \pm 2, \ldots$.

A6. *Limiting Variance Matrix*

There exists a non-singular matrix Σ_0 such that:

$$\lim_{n \to \infty} Var_0 \left[n^{-1/2} \sum_{i=1}^{n} g(Z_i, \beta_0) \right] = \Sigma_0.$$

A7. *Lag Truncation Sequence*

$\{m_n\}_{n=1}^{\infty}$ is a sequence of non-negative integers such that $0 \leq m_n < (n/2)$ for all n and such that $(2m_n + 1) = O(n^\delta)$ and $(2m_n + 1)^{-1} = O(n^{-\delta})$ for some $0 \leq \delta < \infty$.

A8. *Ideal HAC Estimator*

There exists a symmetric positive definite matrix Σ_1 such that:

$$\widehat{\Sigma}_n^0 \equiv \sum_{j=-2m_n}^{2m_n} \kappa_B \left(\frac{j}{2m_n + 1} \right) \widehat{\Gamma}_{s,n}(\beta_0) \xrightarrow{p} \Sigma_1,$$

where:

$$\widehat{\Gamma}_{j,n}(\beta_0) = n^{-1} \sum_{i=\max(1,1+j)}^{\min(n,n+j)} g_i(\beta_0) g_{i-j}(\beta_0)', \quad j = 0, 1, 2, \ldots,$$

and $\kappa_B(u) = \max(0, 1 - |u|)$ denotes the Bartlett kernel.

A9. *Carrier Function*

$\mathcal{V}$ is an open interval of $\mathbb{R}$ such that $0 \in \mathcal{V}$ and $\rho(\cdot) : \mathcal{V} \to \mathbb{R}$ is a twice continuously differentiable strictly concave function such that:

$$\rho(0) = 0, \quad \left[\frac{d\rho(u)}{du} \bigg|_{u=0} \right] = -1, \quad \left[\frac{d^2\rho(u)}{du^2} \bigg|_{u=0} \right] = -1.$$

A10. *Interior Parameter Value*

B has a non-empty interior, denoted $\text{int}(B)$, and $\beta_0 \in \text{int}(B)$.

A11. *Central Limit Theorem*

$n^{-1/2} \sum_{i=1}^{n} g(Z_i, \beta_0) \xrightarrow{D} N[0, \Sigma_0]$.

A12. *Stochastic Equicontinuity I*

The empirical process $Y_n(\beta)$ defined by:

$$Y_n(\beta) \equiv n^{-1/2} \sum_{i=1}^{n} \left[g_i(\beta) - E_0(g_i(\beta)) \right],$$

is stochastically equicontinuous on B.

A13. *Differentiable Population Moments*

$E_0[g(Z_i, \beta)]$ is continuously differentiable with respect to β on an open neighborhood of β_0 and its Jacobian matrix evaluated at $\beta = \beta_0$ has rank p.

A14. *Stochastic Equicontinuity II*

For each $l = 1, \ldots, q$, the empirical process $X_{n,l}(\beta)$ defined by:

$$\begin{aligned} X_{n,l}(\beta) &\equiv n^{-1/2} \sum_{i=1}^{n} \left[h_{i,l}(\beta) - E_0(h_{i,l}(\beta)) \right], \\ h_{i,l}(\beta) &\equiv |g_{i,l}(\beta) - g_{i,l}(\beta_0)|, \end{aligned}$$

is stochastically equicontinuous on B, where $g_{i,l}(\beta)$ denotes the l'th element of $g_i(\beta)$.

A15. *Lipschitz Continuity*

There exist an open neighborhood $B_0 \subset B$ of β_0 and a positive scalar $L_0 < \infty$ such that for each $i = 0, \pm 1, \pm 2, \ldots,$ and $l = 1, \ldots, q$:

$$\left| h_l^e(\bar{\beta}) - h_l^e(\beta_0) \right| \leq L_0 \cdot \|\bar{\beta} - \beta_0\|, \quad \forall \bar{\beta} \in B_0,$$

where $h_l^e(\beta) = E_0[h_{i,l}(\beta)]$.

Before proceeding further it is worthwhile making some comments on these assumptions, especially regarding how they relate to those made for GMM estimation with non-smooth objective functions. First, following the line of argument in Newey and McFadden (1994, pp. 2129, 2132) it is easy to show that Assumptions A1–A5 are sufficient to ensure the consistency of any simple GMM estimator $\tilde{\beta}_n$ characterized by:

$$(P2) \qquad \|\widehat{g}_{(n)}(\tilde{\beta}_n)\|^2 \leq \inf_{\beta \in B} \|\widehat{g}_{(n)}(\beta)\|^2 + o_p(n^{-\sigma}), \tag{15}$$

for some $\sigma \geq 0$. In fact $\tilde{\beta}_n$ can be shown to be consistent when the condition that $\nu > 2$ in Assumption A5 is replaced by the weaker condition that $\nu \geq 1$.

Second, Assumptions A3, A6 and A10–A13 in combination with the consistency of $\tilde{\beta}_n$, as implied by Assumptions A1–A5, are sufficient to enable us to invoke Theorem 7.2 of Newey and McFadden (1994) to show that if $\sigma \geq 1$ then $n^{1/2}(\tilde{\beta}_n - \beta_0)$ is asymptotically normal with mean zero and variance matrix $(G_0{}'G_0)^{-1}G_0{}'\Sigma_0 G_0(G_0{}'G_0)^{-1}$, where G_0 denotes the Jacobian matrix of $E_0[g(Z_i, \beta)]$ with respect to β evaluated at $\beta = \beta_0$. In fact, as noted by Newey and McFadden (1994, p. 2187), Assumption A12 is somewhat stronger than is necessary for invoking Theorem 7.2 of Newey and McFadden (1994).

Third, if we have a sequence of symmetric non-negative definite stochastic matrices $\{\widehat{C}_n\}_{n=1}^{\infty}$ such that $\widehat{C}_n$ converges in probability to Σ_0^{-1} then then the same arguments can be used to establish the existence, consistency and asymptotic normality of an optimal GMM estimator $\dot{\beta}_n$ defined by solving:

$$(P3) \qquad \widehat{g}_{(n)}(\dot{\beta}_n)'\widehat{C}_n\widehat{g}_{(n)}(\dot{\beta}_n) \leq \inf_{\beta \in B} \widehat{g}_{(n)}(\beta)'\widehat{C}_n\widehat{g}_{(n)}(\beta) + o_p(n^{-\sigma}), \tag{16}$$

for some $\sigma \geq 1$. The limiting distribution of $\dot{\beta}_n$ then has a variance matrix given by $(G_0{}'\Sigma_0^{-1}G_0)^{-1}$. Of course, as is well-known, the conditions on the limiting behavior of $\widehat{C}_n$ needed for an optimal GMM estimator are weaker than this; indeed, they do not require that Σ_0 be non-singular.

There are many possible choices for $\widehat{C}_n$ but one of the most commonly used is the feasible Bartlett kernel HAC estimator $\widehat{\Sigma}_n(\tilde{\beta}_n)$ of Σ_0, where:

$$\widehat{\Sigma}_n(\beta) = \sum_{j=-2m_n}^{2m_n} \kappa_B \left(\frac{j}{2m_n+1} \right) \widehat{\Gamma}_{j,n}(\beta), \tag{17}$$

and $\widehat{\Gamma}_{j,n}(\beta)$ is defined as in Equation (14). The properties of this estimator have been investigated by a number of authors, most notably Andrews (1991), in the case where $g(z,\beta)$ is differentiable with respect to β. In the present context, where $g(z,\beta)$ might be non-differentiable with respect to β, we show in Lemma 4 below that if $\bar{\beta}_n = \beta_0 + O_p(n^{-1/2})$ then $\widehat{\Sigma}_n(\bar{\beta}_n) \xrightarrow{p} \Sigma_1$ under Assumptions A1–A15.

It is clear that several of the assumptions made here, in particular Assumptions A6, A8, A11, A12 and A14, are fairly high level assumptions. Clearly, these assumptions all impose implicit conditions on the moments and time-series dependence in the stochastic process $\{Z_i\}_{i=-\infty}^{\infty}$ which are almost certainly stronger the conditions specified by Assumptions A1 and A5. However, Assumptions A6, A8 and A11 all hold provided Assumptions A1, A5 and A7 hold such that $\nu > 4$ and $\{Z_i\}_{i=-\infty}^{\infty}$ has strong mixing coefficients $\{\alpha_s\}_{s=1}^{\infty}$ which satisfy $\sum_{s=1}^{\infty} s^2 \alpha_s^{(b-1)/b} < \infty$ for some $1 < b \leq (\nu/4)$.[3][4][5]

Assumptions A12 and A14 are stochastic equicontinuity assumptions. Suppose that $\{W_{in} : i \leq n,\ n \geq 1\}$ is a triangular array of $\mathcal{W}$-valued random variables for $\mathcal{W} \subseteq \mathbb{R}^k$ and that $\Pi \subset \mathbb{R}^d$ is a bounded parameter space. Then let $\mathcal{M} = \{m(\cdot,\pi) : \pi \in \Pi\}$ denote a class of scalar real-valued functions and define the associated empirical process $\eta_n(\cdot)$ by:

$$\eta_n(\pi) = n^{-1/2} \sum_{i=1}^{n} [m(W_{in},\pi) - E(m(W_{in},\pi))].$$

[3] With these stronger conditions we can invoke Lemma 1 from Andrews (1991) to establish that $\Sigma_0 = \lim_{n\to\infty} Var_0 \left[n^{-1/2} \sum_{i=1}^{n} g(Z_i,\beta_0) \right]$ exists, though this does not ensure that Σ_0 is non-singular.

[4] With these stronger conditions we can invoke Proposition 1 from Andrews (1991) to establish that $\widehat{\Sigma}_n^0 \xrightarrow{p} \Sigma_0$.

[5] With these stronger conditions together with the Assumption A6 then we can invoke Corollary 24.7 of Theorem 24.6 from Davidson (1994) and establish that $n^{-1/2} \sum_{i=1}^{n} g(Z_i,\beta_0) \xrightarrow{D} N[0,\Sigma_0]$.

This process is said to be stochastically equicontinuous if for every sequence of positive scalar constants $\{\delta_n : n \geq 1\}$ that converges to zero:

$$\sup_{\pi_1,\pi_2 \in \Pi : \|\pi_1 - \pi_2\| \leq \delta_n} \|\eta_n(\pi_1) - \eta_n(\pi_2)\| \xrightarrow{p} 0.$$

Andrews (1993) provides a variety of conditions under which stochastic equicontinuity can be demonstrated for empirical processes based on dependent random variables. These conditions usually consist of three components. First, there are usually some restrictions on the mixing properties of the $\{W_{in} : i \leq n,\ n \geq 1\}$ array. Second, let the real-valued function $\bar{M}(\cdot)$ on $\mathcal{W}$ be an envelope of $\mathcal{M}$ if:

$$\sup_{\pi \in \Pi} |m(w, \pi)| \leq \bar{M}(\omega), \forall w \in \mathcal{W}.$$

Then there are usually requirements that $\mathcal{M}$ possesses an envelope function $\bar{M}(\cdot)$ satisfying certain restrictions. Third, there are usually restrictions on the size or complexity of the class $\mathcal{M}$, typically expressed in terms of so-called cover numbers. One particularly useful set of cover numbers are the so-called L^p-bracketing cover numbers defined as follows. Suppose $\mathcal{M}$ is a class of real-valued functions. For $\epsilon > 0$ and $p \in [1, \infty]$, the L^p-bracketing cover number $N_p^B(\epsilon, \mathcal{M})$ is the smallest value of n for which there exist real functions $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ on $\mathcal{W}$ such that for all $m \in \mathcal{M}$:

$$|m(w) - a_j(w)| \leq b_j(w) \quad \forall w \in \mathcal{W} \text{ for some } j \leq n$$

and:

$$\max_{j \leq n} \sup_{i \leq n, n \geq 1} \left(E\left[b_j(w_{in})^p\right]\right)^{1/p} \leq \epsilon.$$

The restrictions then take the form of bounds on how rapidly $N_p^B(\epsilon, \mathcal{M})$ can grow as $\epsilon \to 0$. One useful source of such bounds are provided by so-called L^p-continuity conditions. The class of functions $\mathcal{M}$ satisfies an L^p-continuity condition if there exist constants $p, C_0, \psi, r_0 > 0$ such that for every $\pi \in \Pi$ and $0 < \delta < r_0$:

$$\sup_{i\ \leq n, n \geq 1} \left(E \sup_{\pi_1 \in \Pi : \|\pi_1 - \pi\| \leq \delta} |m(W_{in}, \pi_1) - m(W_{in}, \pi)|^p\right)^{1/p} \leq C_0 \delta^{\psi}. \tag{18}$$

Then provided that $\Pi \subset \mathbb{R}^d$ is bounded it follows that $N_p^B(\epsilon, \mathcal{M}) \leq C_1 \epsilon^{-d/\psi}$ for some $C_1 < \infty$; see Andrews (1993, p. 201).

A vector real-valued process is then stochastically equicontinuous if each of its elements is stochastically equicontinuous. We can thus see that Assumptions A12 and A14 state that the processes Y_n and X_n are stochastically equicontinuous, where $W_i = Z_i$, $\pi = \beta$ and $\Pi = B$, and $m(\cdot,\cdot) = g(\cdot,\cdot)$ or $m(\cdot,\cdot) = h(\cdot,\cdot)$ respectively. In the present context, L^p-continuity conditions on the classes of functions are also potential useful because they help to provide a motivation for Assumption A15. Suppose that the elements of the class $\mathcal{M}$ are all non-negative and that furthermore there exists $\pi_0 \in \Pi$ such that $m(w, \pi_0) = 0$ for all $w \in \mathcal{W}$. Observe that if $\mathcal{M}$ satisfies the L^p-continuity condition with respect to the stochastic process $\{W_{in} : i \leq n,\ n \geq 1\}$ given in Equation (18) and $p \geq 1$ then it follows that for all $\pi \in \Pi$ and $0 < \delta < r_0$:

$$\begin{aligned} C_0 \delta^{\psi} &\geq E\left(\sup_{\pi_1 \in \Pi : \|\pi_1 - \pi\| \leq \delta} |m(W_{in}, \pi_1) - m(W_{in}, \pi)| \right) \\ &\geq \sup_{\pi_1 \in \Pi : \|\pi_1 - \pi\| \leq \delta} E\left(|m(W_{in}, \pi_1) - m(W_{in}, \pi)|\right) \\ &\geq \sup_{\pi_1 \in \Pi : \|\pi_1 - \pi\| \leq \delta} \left|E[m(W_{in}, \pi_1)] - E[m(W_{in}, \pi)]\right|, \end{aligned} \tag{19}$$

generating a property which closely resembles that Assumption A15.

Finally, it might seem simpler to replace Assumption A15 with an assumption that $E_0[h(Z_i, \beta)]$ is differentiable at $\beta = \beta_0$, thus paralleling Assumption A13 for the $g(\cdot,\cdot)$ function. Unfortunately, that would be a very restrictive assumption which would not be satisfied in certain leading cases of interest. In particular, suppose that $\{Z_i\}_{i=-\infty}^{+\infty}$ is a univariate stochastic process such that Z_i is a continuous random variable with a continuous and strictly positive density everywhere in $\mathbb{R}$, and that we wish to estimate the median of Z_i. A natural choice for the $g(\cdot)$ function is to set $g(z, \beta) = (1_{z<\beta} - 1_{z>\beta})/2$ which clearly satisfies Assumptions A4 and A5 given these conditions. Taking expectations we see that:

$$E_0[g(Z_i, \beta)] = (\Pr\{Z_i < \beta\} - \Pr\{Z_i > \beta\})/2 = \Pr\{Z_i \leq \beta\} - (1/2), \tag{20}$$

since Z_i has a continuous distribution. Furthermore, since Z_i has a strictly positive density it follows that $E_0[g(Z_i, \beta)] = 0$ has a unique solution at $\beta = \beta_0$ where β_0 is the median of Z_i characterized by $\Pr\{Z_i \leq \beta_0\} = \Pr\{Z_i \geq \beta_0\} = (1/2)$ thus satisfying Assumption A3.

If we define $F_Z(z) = \Pr\{Z_i \leq z\}$ then differentiating $E_0[g(Z_i, \beta)]$ gives:

$$\frac{dE_0[g(Z_i, \beta)]}{d\beta} = \frac{dF_Z(\beta)}{d\beta} = f_Z(\beta), \tag{21}$$

where $f_Z(z)$ denotes the density function of Z_i evaluated at z. By assumption, this is positive for all z and, in particular, it is positive at $z = \beta_0$ thus implying that $E_0[g(Z_i, \beta)]$ satisfies Assumption A13. However if we now define $h(Z_i, \beta) = |g(Z_i, \beta) - g(Z_i, \beta_0)|$ then taking expectations gives:

$$E_0[h(Z_i, \beta)] = \Pr\{\min(\beta, \beta_0) \leq Z_i \leq \min(\beta, \beta_0)\}. \tag{22}$$

It thus follows that $E_0[h(Z_i, \beta)]$ has a kink at $\beta = \beta_0$ such that its right derivative at $\beta = \beta_0$ is equal to $-f_Z(\beta_0)$ while its left derivative is equal to $f_Z(\beta_0)$. Hence $E_0[h(Z_i, \beta)]$ it is not differentiable at $\beta = \beta_0$ since the density of Z_i is assumed to be strictly positive everywhere. However, $E_0[h(Z_i, \beta)]$ is differentiable at every other value of β and if it is assumed that $f_Z(\cdot)$ is uniformly bounded above by $L_0 < \infty$ then it is easy to verify that Assumption A15.

3 Consistency

As stated in the Introduction, our strategy for demonstrating consisting of the GEL estimator is based on the approach used by Newey and Smith (2004) to demonstrate consistency in the i.i.d. sampling case. However, it is modified somewhat by the need to take the dependence in the $\{Z_i\}$ process into account.

The first step in demonstrating consistency is to establish the following lemma which directly parallels Lemma A1 from Newey and Smith (2004).

Lemma 1 *Under Assumptions A1, A2, A5, A7 and A9, for any fixed* ζ *such that* $(1/\nu) < \zeta < \infty$, *define* $\Lambda_n^\zeta = \{\lambda \in \mathbb{R}^q : \|\lambda\| \leq n^{-\zeta}\}$; *then:*

(i) $\sup_{\beta\in B,\lambda\in\Lambda_n^\zeta,1-m_n\le i\le n+m_n}|\lambda' g_{i,n}^\omega(\beta)| \xrightarrow{p} 0$*; and*

(ii) $\Lambda_n^\zeta \subseteq \cap_{\beta\in B}\widehat{\Lambda}_n(\beta)$*, with probability asymptotically one* (w. p. a. 1)*, where:*

$$\widehat{\Lambda}_n(\beta) = \{\lambda \in \mathbb{R}^q : \lambda' g_{i,n}^\omega(\beta) \in \mathcal{V}; 1 - m_n \le i \le n + m_n\}. \tag{23}$$

Proof. See Appendix. ■

Note that the assumptions we have made for Lemma 1 are stronger than are needed for the proof. in particular, the proof does not require: (a) the ergodicitity of $\{Z_i\}_{i=-\infty}^{+\infty}$ in Assumption A1; (b) the compactness of B in Assumption A2; (c) the requirement that $\nu > 2$ in Assumption A5, as the proof works for any $\nu \ge 1$; (d) the rate of growth restrictions on m_n in Assumption A7; or (e) any of the specific requirements on $\rho(\cdot)$ in Assumption A9. Indeed, if we replaced part (b) of Assumption A5 with the weaker condition that there exists $M_0 < \infty$ such that $E_0[d(Z_i)^\nu] \le M_0$ for all $i = 0, \pm1, \pm2, \ldots$, then we would not require $\{Z_i\}_{i=-\infty}^{\infty}$ to be stationary.

The second step is to establish the following lemma, based on part of Lemma A2 from Newey and Smith (2004), which controls the asymptotic behavior of $\widehat{P}_n^*(\beta_0)$.

Lemma 2 *Under Assumptions A1, A3 and A5–A9, suppose that* $\delta \le (1/2) - (1/\nu)$*; then:*

$$\widehat{P}_n^*(\beta_0) \equiv \sup_{\lambda\in\widehat{\Lambda}_n(\beta_0)} \widehat{P}_n(\beta_0, \lambda_0) = O_p(n^{\delta-1}). \tag{24}$$

Proof. See Appendix. ■

The third step is given by the following lemma which is a somewhat weakened version of Lemma A3 from Newey and Smith (2004).

Lemma 3 *Under Assumptions A1–A3 and A5–A9, suppose that* $\delta \le (1/2) - (1/\nu)$, $\sigma > (1\nu)$ *and* $\{\bar{\beta}_n\}_n^\infty$ *is a sequence of estimators of* β_0 *such that:*

$$\widehat{P}_n^*(\bar{\beta}_n) \le \widehat{P}_n^*(\beta_0) + o_p(n^{-\sigma}). \tag{25}$$

Then:

$$\|\widehat{g}_{(n)}(\bar{\beta}_n)\| = o_p(1). \tag{26}$$

Proof. See Appendix. ■

Note that the proof of Lemma 3 does not require the compactness of B in Assumption A2 and it is not entirely clear as to how important is the requirement that $\nu > 2$ in Assumption A5 for the proof. The proof uses Lemma 1 which, as noted above, only requires that $\nu \geq 1$. However, Assumption A8 almost certainly imposes stronger requirements on ν as discussed in Section 2 above.

We are now in a position to establish the consistency of $\widehat{\beta}_n$.

Theorem 1 *Under Assumptions A1–A9, suppose that* $\delta \leq (1/2) - (1/\nu)$ *and* $\{\bar{\beta}_n\}_n^{\infty}$ *is a sequence of estimators of* β_0 *such that* $\widehat{P}_n^*(\bar{\beta}_n) \leq \widehat{P}_n^*(\beta_0) + o_p(n^{-\sigma})$ *where* $\sigma > (1/\nu)$*; then* $\bar{\beta}_n \xrightarrow{p} \beta_0$.

Proof. See Appendix. ■

Corollary 1 *Under Assumptions A1–A9, if* $\delta \leq (1/2) - (1/\nu)$ *and* $\sigma > (1/\nu)$ *then* $\widehat{\beta}_n$ *is a consistent estimator of* β_0.

Proof. By construction $\inf_{\beta \in B} P_n^*(\beta) \leq P_n^*(\beta_0)$ and hence it follows that $\widehat{\beta}_n$ satisfies the conditions of Theorem 1. ■

Note that Theorem 1 does need stronger assumptions than do Lemmas 1–3. In particular, unlike Lemmas 1–3, Theorem 1 does require the compactness of B in Assumption $A2$ and the continuity with probability one of $g(\cdot)$ in Assumption A4.

Finally, it is worth comparing the assumptions under which consistency is demonstrated with the assumptions under which Newey and Smith (2004) are able to demonstrate consistency for the static GEL estimator, i.e. the estimator when $m_n = 0$, in the i.i.d. sampling case. Clearly, the assumption that the $\{Z_i\}$ are i.i.d replaces Assumption A1. In addition, the assumption that $m_n = 0$ clearly replaces Assumption A7. The other assumptions made by Newey and Smith (2004) for establishing consistency are then identical to those made here for establishing consistency.[6]

4 Asymptotic Normality

The first step in the proof of asymptotic normality of $\widehat{\beta}_n$ under suitable conditions on τ is to establish the following lemma.

Lemma 4 *Under Assumptions A1–A2, A5–A8 and A14–A15, suppose that* $\delta < (1/2) - (1/\nu)$ *and that* $\bar{\beta}_n = \beta_0 + 0_p(n^{-1/2})$ *then:*

$$\widehat{\Sigma}_n(\bar{\beta}_n) = \Sigma_1 + o_p(1). \tag{27}$$

Proof. See Appendix. ■

The second step is to to establish that $\widehat{\beta}_n$ is root-n consistent under suitable conditions on α, γ, δ, and σ.

Lemma 5 *Under Assumptions A1–A15, suppose that* $\delta < (1/2) - (1/\nu)$ *and* $\{\bar{\beta}_n\}_n^\infty$ *is a sequence of estimators of* β_0 *such that* $\widehat{P}_n^*(\bar{\beta}_n) \leq \widehat{P}_n^*(\beta_0) + o_p(n^{-\sigma})$ *where* $\sigma \geq (1-\delta)$ *; then* $\bar{\beta}_n = \beta_0 + O_p(n^{-1/2})$.

Proof. See Appendix. ■

Note that since Lemma 4 holds under a subset of the assumptions of Lemma 5 it follows that $\widehat{\Sigma}_n(\widehat{\beta}_n)$ is a consistent estimator of Σ_1.

The third step is to demonstrate that a rescaled version of $P_n^*(\beta)$ behaves approximately like $n\widehat{g}_{(n)}(\beta)'\Sigma_0\widehat{g}_{(n)}(\beta)$ in any sequence of neighborhoods of β_0 which shrink at rate $n^{-1/2}$.

[6] Newey and Smith (2004) assume that $Var_0[g(Z_i, \beta_0)]$ is non-singular but this is equivalent to Assumption A6 under i.i.d. sampling. In addition, they do not impose that $\rho(0) = 0$ in Assumption A9 but, as noted in Section 2 above, this condition is simply a normalization which is made without loss of generality. Furthermore, in the i.i.d. sampling case it is trivial to establish that with $m_n = 0$ then Assumption A8 is satisfied by the Law of Large Numbers under Assumptions A3 and A5.

Lemma 6 *Under Assumptions A1–A15, suppose that* $\delta < (1/2) - (1/\nu)$*,* $\sigma \geq (1 - \delta)$*, and* $\sigma_1 = \Sigma_0$*; then for any fixed* $0 < L < \infty$*:*

$$\sup_{\beta \in B_n(L)} \left| (2n/S_n)\widehat{P}_n^*(\beta) - 2n\widehat{Q}_n^*(\beta; \Sigma_0) \right| = o_p(1), \tag{28}$$

where $B_n(L) = \{\beta \in B : \|\beta - \beta_0\| \leq n^{-1/2}L\}$ *and:*

$$\widehat{Q}_n^*(\beta; A) = (1/2)\widehat{g}_{(n)}(\beta)' A^{-1} \widehat{g}_{(n)}(\beta), \tag{29}$$

for any $(q \times q)$ *symmetric positive definite matrix* A*.*

Proof. See Appendix. ■

It is immediately obvious that $\widehat{Q}_n^*(\beta; \Sigma_0)$ is an ideal optimal GMM objective function. We can now easily establish the main result of the paper.

Theorem 2 *Under Assumptions A1–A15, suppose that* $\delta < (1/2) - (1/\nu)$*,* $\sigma \geq (1 - \delta)$*, and* $\Sigma_1 = \Sigma_0$*; then:*

$$n^{1/2}(\widehat{\beta}_n - \beta_0) \xrightarrow{D} N[0, (G_0 \Sigma_0^{-1} G_0)^{-1}]. \tag{30}$$

Proof. See Appendix. ■

Thus $\widehat{\beta}_n$ is asymptotically efficient in that has the same asymptotic distribution as an optimal GMM estimator based on the same moment conditions. Indeed, the proof of Theorem 2 proceeds by demonstrating that $\widehat{\beta}_n$ is an optimal GMM estimator.

There are then three issues of interest. First, in order to ensure consistency we need to impose an upper bound on δ, namely that $\delta < (1/2) - (1/\nu)$. Since $\nu > 2$ by assumption, it is clear that we can select δ to satisfy this upper bound while still being non-negative. It is not clear whether this upper bound can be relaxed by using alternative smoothing windows; this remains a topic for further research.

Second, there is still an issue of estimating the asymptotic covariance matrix of $\widehat{\beta}_n$. Under the assumptions we can consistently estimate Σ_0 by $\widehat{\Sigma}_n(\widehat{\beta}_n)$. Furthermore, we can then apply theorem 7.4 from Newey and McFadden (1994) to provide a consistent estimator of G_0 under the assumptions made here. Putting these together then we can easily construct a consistent estimator of $(G_0\Sigma_0^{-1}G_0)^{-1}$.

Third, it is clear that we can implement asymptotically chi-square tests of hypotheses of interest either by means of Wald tests or by a comparison of evaluated objective functions. Thus let $H_0 : \phi(\beta) = 0$ be as set of $r \leq k$ restrictions forming a hypothesis of interest such that $\phi(\beta)$ is continuously differentiable with respect to β with rank r and suppose that $\widehat{\beta}_{R,n}$ satisfies $\phi(\widehat{\beta}_{R,n}) = 0$ and:

$$\widehat{P}_n^*(\widehat{\beta}_{R,n}) \leq \inf_{\beta \ inB:\phi(\beta)=0} \widehat{P}_n^*(\beta) + o_p(n^{-\sigma}). \tag{31}$$

Then following the same logic as used to prove Theorem 2 we can establish that $\widehat{\beta}_{R,n}$ is a restricted optimal GMM estimator and that $\widehat{P}_n^*(\widehat{\beta}_n)$ and $\widehat{P}_n^*(\widehat{\beta}_{R,n})$ are asymptotically equivalent to rescaled versions of an optimal GMM objective function evaluated at $\widehat{\beta}_n$ and $\widehat{\beta}_{R,n}$ respectively. This leads to the test statistic :

$$2nS_n^{-1}(\widehat{P}_n^*(\widehat{\beta}_{R,n}) - \widehat{P}_n^*(\widehat{\beta}_n)) \tag{32}$$

which should be asymptotically chi-square with r degrees of freedom under the null hypothesis. In a similar fashion we should find that $2nS_n^{-1}\widehat{P}_n^*(\widehat{\beta}_n)$ is asymptotically chi-square with $(q - k)$ degrees of freedom if the implicit overidentifying restrictions present in the moment conditions when $q > k$ are in fact valid.

5 Conclusions

In this paper we have demonstrated consistency and asymptotic normality for GEL estimators using non-smooth moment conditions with dynamic data. Our approach works directly with the relevant non-differentiable functions without any requiring us to smooth the non-differentiabilities.

Most of the assumptions made are fairly standard. However, in order to handle the non-differentiability we need to make fairly high-level assumptions to justify the use of empirical process methods. These high-level assumptions take the form of stochastic equicontinuity and Lipschitz-type conditions and would need to be verified using more primitive conditions in any specific case of interest.

Appendix

Proof of Lemma 1

Define:

$$\phi_n \equiv \sup_{\beta\in B,\lambda\in\Lambda_n^\zeta,1-m_n\leq i\leq n+m_n} |\lambda' g_{i,n}^{\omega}(\beta)|, \tag{A.1}$$

and observe that:

$$0 \;\leq\; \phi_n \;\leq\; n^{-\zeta} \sup_{\beta\in B,1-m_n\leq i\leq n+m_n} \|g_{i,n}^{\omega}(\beta)\|, \tag{A.2}$$

by application of the Cauchy-Schwartz inequality and the definition of Λ_n^ζ. Then observe that:

$$0 \;\leq\; \|g_{i,n}^{\omega}(\beta)\| \;\leq\; \sup_{1-m_n\leq i\leq n+m_n} \left\|(2m_n+1)^{-1} \sum_{j=-m_n}^{+m_n} g_{i-j,n}(\beta)\right\| \;\leq\; \max_{1\leq s\leq n} \|g_s(\beta)\|, \tag{A.3}$$

where $g_s(\beta) = g(Z_s, \beta)$, and hence that:

$$0 \;\leq\; \phi_n \;\leq\; n^{-\zeta} \sup_{\beta\in B,1\leq s\leq n} \|g_s(\beta)\| \;\leq n^{-\zeta} \max_{1\leq s\leq n} d(Z_s). \tag{A.4}$$

Application of the Markov inequality and Jensen's inequality then implies that $\max_{1\leq s\leq n} d(Z_s) = O_p(n^{1/\nu})$, since $\nu \geq 1$, and hence that $\phi_n \;=\; n^{-\zeta} O_p(n^{(1/\nu)}) = o_p(1)$, since $\zeta > (1/\nu)$, which establishes the first desired result.

Furthermore, Equation (A.4) immediately implies that:

$$\text{w. p. a. } 1: \quad \lambda' g_{i,n}^{\omega}(\beta) \in \mathcal{V} \quad \forall\; \lambda \in \Lambda_n^\zeta,\; \beta \in B,\; 1-m_n \leq i \leq n+m_n, \tag{A.5}$$

as $\mathcal{V}$ is an open interval containing 0. The second desired result then follows directly. $\square$

Proof of Lemma 2

Define $\widehat{g}_{(n)}(\beta) = n^{-1}\sum_{i=1}^{n} g(Z_i, \beta)$ and then observe that since $\rho(\cdot)$ is twice continuously differentiable, by Assumption A9, a Taylor series expansion of $\widehat{P}_n(\beta, \lambda)$ in λ around $\lambda = 0$ gives:

$$\begin{aligned}\widehat{P}_n(\beta,\lambda) &= -\lambda'\widehat{g}_{(n)}(\beta) + (1/2)\lambda'\widehat{H}_n(\beta,\pi\lambda)\lambda \\ &= -\lambda'\widehat{g}_{(n)}(\beta) + (1/2)\lambda'\left[n^{-1}\sum_{i=1-m_n}^{n+m_n}\rho_2(\pi\lambda' g^{\omega}_{i,n}(\beta))g^{\omega}_{i,n}(\beta)g^{\omega}_{i,n}(\beta)'\right]\lambda, \quad \text{(A.6)}\end{aligned}$$

for some $0 \le \pi \le 1$ which may depend on β, $\{Z_i\}_{i=1}^{n}$ and n. Now fix ζ such that $(1/\nu) < \zeta < (1/2) - \delta$ and define Λ_n^{ζ} as in Lemma 1; note that this is feasible since $0 < \delta < (1/2) - (1/\nu)$ by Assumption A7. Since $\rho(\cdot)$ is concave on the open set $\mathcal{V}$ by assumption it follows that $\rho(u)$ is continuous in $u \in \mathcal{V}$ and hence $\widehat{P}_n(\beta, \lambda)$ is continuous in $\lambda \in \widehat{\Lambda}_n(\beta)$ for given β. It then follows from Lemma 1 that:

$$\text{w.p.a.}\,1: \ \exists\tilde{\lambda}_0 \ \text{ s.t. } \ \tilde{\lambda}_0 = \arg\max_{\lambda\in\Lambda_n^{\zeta}} \widehat{P}_n(\beta_0, \lambda). \quad \text{(A.7)}$$

Furthermore, since Λ_n^{ζ} is convex and since $0 \in \Lambda_n^{\zeta}$, it also follows from Lemma 1 and Assumption A9 that:

$$\text{w.p.a.}\,1: \ \max_{1-m_n\le i\le n+m_n}\ \sup_{0\le\pi\le 1} \rho_2(\pi\tilde{\lambda}_0{}' g^{\omega}_{i,n}(\beta) < -(1/2) = (1/2)\rho_2(0), \quad \text{(A.8)}$$

since $\rho_2(0) = -1$. Combined with the Taylor series expansion above this implies that:

$$0 = \widehat{P}_n(\beta_0, 0) \le \widehat{P}_n(\beta_0, \tilde{\lambda}_0) \le -\tilde{\lambda}_0'\widehat{g}_{(n)}(\beta_0) + (1/4)\tilde{\lambda}_0'\widehat{H}_n(\beta_0, 0)\tilde{\lambda}_0. \quad \text{(A.9)}$$

But from Assumptions A6 and A8 it follows that $-S_n\widehat{H}_n(\beta_0, 0) = \widehat{\Sigma}_n(\beta) \xrightarrow{p} \Sigma_1$ under a subset of the assumptions made here. Furthermore, since Σ_1 is symmetric positive definite by assumption it follows that there exists $0 < C_0 < \infty$ such that $\lambda'\Sigma_0\lambda \ge C_0 \cdot \|\lambda\|^2$ for all $\lambda \in \mathbb{R}^q$ combined with the Cauchy-Schwartz inequality this implies that:

$$\text{w.p.a.}\,1: \ 0 \le \widehat{P}_n(\beta_0, \tilde{\lambda}_0) \le \|\tilde{\lambda}_0\| \cdot \|\widehat{g}_{(n)}(\beta_0)\| - (1/8)C_0(2m_n+1)^{-1} \cdot \|\tilde{\lambda}_0\|^2 \quad \text{(A.10)}$$

and thus that:

$$\Longrightarrow \ \text{w.p.a.}\,1: \ \|\tilde{\lambda}_0\| \le 8C_0^{-1}(2m_n+1) \cdot \|\widehat{g}_{(n)}(\beta_0)\|. \quad \text{(A.11)}$$

Assumption A6 also implies that $\widehat{g}_{(n)}(\beta_0) = O_p(n^{-1/2})$ since $E_0[n^{1/2}\widehat{g}_{(n)}(\beta_0)] = 0$ and $Var_0[n^{1/2}\widehat{g}_{(n)}(\beta_0)] = O(1)$. Since $m_n = O_p(n^{\delta})$ it thus follows from Equation (A.11) that $\|\tilde{\lambda}_0\| = O_p(n^{\delta-(1/2)})$ and since $\zeta < (1/2) - \delta$ it follows that $\|\tilde{\lambda}_0\| = o_p(n^{-\zeta})$ which in turn implies that $\|\tilde{\lambda}_0\| \in \text{int}(\Lambda_n^{\zeta})$, w. p. a. 1. But Lemma 1 implies that $\text{int}(\Lambda_n^{\zeta}) \subseteq \text{int}(\widehat{\Lambda}_n(\beta_0))$, w. p. a. 1, under a subset of the assumptions made here and since, by assumption, $\rho(\cdot)$ is concave on $\mathcal{V}$ and hence $\widehat{P}_n(\beta_0, \lambda)$ is concave in $\lambda \in \widehat{\Lambda}_n(\beta_0))$, it thus follows that:

$$\text{w. p. a. 1}: \ \|\tilde{\lambda}_0\| = \arg \max_{\lambda \in \widehat{\Lambda}_n(\beta_0))} \widehat{P}_n(\beta_0, \lambda) \tag{A.12}$$

$$\Longrightarrow \quad \text{w. p. a. 1}: \ 0 \leq \widehat{P}_n^*(\beta_0) = \widehat{P}_n(\beta_0, \tilde{\lambda}_0) \leq \|\tilde{\lambda}_0\| \cdot \|\widehat{g}_{(n)}(\beta_0)\|, \tag{A.13}$$

which then implies that $\widehat{P}_n^*(\beta_0) = O_p(n^{\delta-1})$ as desired. □

Proof of Lemma 3

The first step in the proof of this lemma is to establish that there exists $0 < C_1 < \infty$ such that:

$$\text{w. p. a. 1}: \sup_{\beta \in B, \lambda: \lambda'\lambda = 1} \left[-\lambda' \widehat{H}_n(\beta, 0)\lambda\right] < C_1. \tag{A.14}$$

Observe that:

$$-\lambda' \widehat{H}_n(\beta, 0)\lambda = n^{-1} \sum_{i=1-m_n}^{n+m_n} [\lambda' g_{i,n}^{\omega}(\beta)]^2 \leq n^{-1} \sum_{i=1-m_n}^{n+m_n} \|\lambda\|^2 \cdot \|g_{i,n}^{\omega}(\beta)\|^2 \tag{A.15}$$

which implies that:

$$\begin{aligned}
\sup_{\lambda: \lambda'\lambda = 1} \left[-\lambda' \widehat{H}_n(\beta, 0)\lambda\right] &\leq n^{-1} \sum_{i=1-m_n}^{n+m_n} \|g_{i,n}^{\omega}(\beta)\|^2 \\
&\leq n^{-1} \sum_{i=1-m_n}^{n+m_n} \left[(2m_n + 1)^{-1} \sum_{j=-m_n}^{m_n} \|g_{i,n}(\beta)\|^2\right] \\
&\leq n^{-1} \sum_{i=1}^{n} \|g(Z_i, \beta)\|^2. \qquad \text{(A.16)}
\end{aligned}$$

It then follows that:

$$\sup_{\beta \in B, \lambda: \lambda'\lambda = 1} \left[-\lambda' \widehat{H}_n(\beta, 0)\lambda\right] \leq n^{-1} \sum_{i=1}^{n} d(Z_i)^2 \xrightarrow{p} E[d(Z_i)^2] < \infty, \tag{A.17}$$

by the law of large numbers since $\{d(Z_i)^2\}_{i=-\infty}^{+\infty}$ is stationary and ergodic with a finite expectation. Setting $C_1 = E[d(Z_i)^2] + \epsilon$ for any $\epsilon > 0$ establishes Equation (A.14).

Second, fix ζ such that $(1/\nu) < \zeta < \min\{\sigma, (1-\delta)\}$, which is feasible by assumption since $\sigma > (1/nu)$ and $\delta < (1/2) - (1/\nu) < 1 - (1/\nu)$. Then define Λ_n^ζ, as in Lemma 1, and:

$$\dot{\lambda}_n = \begin{cases} -n^{-\zeta}\widehat{g}_{(n)}(\bar{\beta}_n)/\|\widehat{g}_{(n)}(\bar{\beta}_n)\|, & \text{if } \widehat{g}_{(n)}(\bar{\beta}_n) \neq 0, \\ -n^{-\zeta}(1,0,\ldots,0)', & \text{otherwise.} \end{cases} \tag{A.18}$$

By construction $\dot{\lambda}_n \in \Lambda_n^\zeta$ and hence it follows from Lemma 1 that $\dot{\lambda}_n \in \widehat{\Lambda}_n(\bar{\beta}_n)$, w. p. a. 1, under a subset of the assumptions made here. A Taylor series expansion then gives:

$$\text{w. p. a. 1}: \ \widehat{P}_n(\bar{\beta}_n, \dot{\lambda}_n) = -\dot{\lambda}_n{}'\widehat{g}_{(n)}(\bar{\beta}_n) + (1/2)\dot{\lambda}_n{}'\widehat{H}_n(\bar{\beta}_n, \pi\dot{\lambda}_n)\dot{\lambda}_n, \tag{A.19}$$

for some $0 \leq \pi \leq 1$ which may depend on $\dot{\lambda}_n$, $\widehat{\beta}_n$ and the $\{Z_i\}_{i=i-m_n}^{n+m_n}$. But Lemma 1 then implies that:

$$\sup_{0\leq\pi\leq1, 1-m_n\leq i\leq n+m_n} |\pi\dot{\lambda}_n{}'\widehat{g}_{(n)}(\bar{\beta}_n)| \xrightarrow{p} 0, \tag{A.20}$$

and hence that:

$$\text{w. p. a. 1}: \sup_{0\leq\pi\leq1, 1-m_n\leq i\leq n+m_n} |1 + \rho_2(\pi\dot{\lambda}_n{}'\widehat{g}_{(n)}(\bar{\beta}_n))| < (1/2), \tag{A.21}$$

which in turn implies that:

$$\text{w. p. a. 1}: \ \widehat{P}_n(\bar{\beta}_n, \dot{\lambda}_n) \geq -\dot{\lambda}_n{}'\widehat{g}_{(n)}(\bar{\beta}_n) + (3/4)\dot{\lambda}_n{}'\widehat{H}_n(\bar{\beta}_n, 0)\dot{\lambda}_n. \tag{A.22}$$

But then it follows from Equation (A.14) that:

$$\begin{aligned} \text{w. p. a. 1}: \ \widehat{P}_n(\bar{\beta}_n, \dot{\lambda}_n) &\geq -\dot{\lambda}_n{}'\widehat{g}_{(n)}(\bar{\beta}_n) - (3/4)C_1 \cdot \|\dot{\lambda}_n\|^2 \\ &= n^{-\zeta}\|\widehat{g}_{(n)}(\bar{\beta}_n)\| - (3/4)C_1 n^{-2\zeta}, \end{aligned} \tag{A.23}$$

by virtue of the definition of $\dot{\lambda}_n$. Clearly, $\widehat{P}_n(\bar{\beta}_n, \dot{\lambda}_n) \leq \bar{P}_n^*(\bar{\beta}_n)$, w. p. a. 1, and from the characterization of $\bar{\beta}_n$ together with the result of Lemma 2, which holds under a subset of the assumptions made here, it then follows that:

$$\widehat{P}_n^*(\bar{\beta}_n) \leq \widehat{P}_n^*(\beta_0) + o_p(n^{-\sigma}) = O_p(n^{-1+\delta}) + o_p(n^{-\sigma}). \tag{A.24}$$

Combined, these imply that:

$$\begin{aligned} & \text{w. p. a. 1}: \; n^{-\zeta}\|\widehat{g}_{(n)}(\bar{\beta}_n)\| - (3/4)C_1 n^{-2\zeta} \leq O_p(n^{-1+\delta}) + o_p(n^{-\sigma}) \\ \Longrightarrow \quad & \|\widehat{g}_{(n)}(\bar{\beta}_n)\| = O_p(n^{-\zeta}) + O_p(n^{-1+\delta+\zeta}) + o_p(n^{-\sigma+\zeta}). \qquad \text{(A.25)} \end{aligned}$$

But, by assumption, $\zeta > (1/\nu)$ so $-\zeta < 0$, $\zeta < 1-\delta$ so $-1+\delta+\zeta < 0$, and $\zeta < \sigma$ so $-\sigma+\zeta < 0$. Hence it follows that $O_p(n^{-\zeta}) + O_p(n^{-1+\delta+\zeta}) + o_p(n^{-\sigma+\zeta}) = o_p(1)$ and thus that $\|\widehat{g}_{(n)}(\bar{\beta}_n)\| = o_p(1)$ as desired. □

Proof of Theorem 1

For convenience in what follows, define $g^e(\beta) = E_0[g(Z_i,\beta)]$, which is valid in view of Assumptions A1 and A5. As noted in Section 2 above, it follows from Assumptions A1–A2 and A4–A5 that $g^e(\beta)$ is continuous on B and that $\widehat{g}_{(n)}(\beta)$ converges in probability to $g^e(\beta)$ uniformly on B by application of Lemma 2.4 from from Newey and McFadden (1994) and the discussion in the succeeding paragraphs. From Lemmma 3 if follows that $\|\bar{g}_{(n)}(\bar{\beta}_n)\| \xrightarrow{p} 0$, under a subset of the assumptions made here, so by the triangle inequality it follows that $g^e(\bar{\beta}_n) \xrightarrow{p} 0$. But then observe that Assumptions A2 and A3 combined with the continuity of $g^e(\beta)$ demonstrated above then implies that $\|g^e(\beta)\|$ has an identifiably unique minimum on B at $\beta = \beta_0$. The desired result follows immediately. □

Proof of Lemma 4

First, observe that:

$$\|\widehat{\Sigma}_n(\bar{\beta}_n)-\Sigma_1\| \leq \|\widehat{\Sigma}_n(\bar{\beta}_n)-\widehat{\Sigma}_n(\beta_0)\|+\|\widehat{\Sigma}_n(\beta_0)-\Sigma_1\| \leq \|\widehat{\Sigma}_n(\bar{\beta}_n)-\widehat{\Sigma}_n(\beta_0)\|+o_p(1), \quad \text{(A.26)}$$

by Assumption A8. Second, observe that:

$$\begin{aligned} \|\widehat{\Sigma}_n(\bar{\beta}_n) - \widehat{\Sigma}_n(\beta_0)\| & \leq \sum_{j=-2m_n}^{2m_n} \kappa_B(j/S_n)\cdot\|\widehat{\Gamma}_{j,n}(\bar{\beta}_n) - \widehat{\Gamma}_{j,n}(\beta_0)\| \\ & \leq 2S_n \sup_{|j|\leq 2m_n} \|\widehat{\Gamma}_{j,n}(\bar{\beta}_n) - \widehat{\Gamma}_{j,n}(\beta_0)\|, \qquad \text{(A.27)} \end{aligned}$$

using Equation (17) and noting that $0 \le \kappa_B(u) \le 1$ for all $u \in \mathbb{R}$. Third, observe that for each $|j| \le 2m_n$:

$$\begin{aligned}\widehat{\Gamma}_{j,n}(\beta) - \widehat{\Gamma}_{j,n}(\beta_0) &= n^{-1} \sum_{i=\max(1,1+j)}^{\min(n,n+j)} [g_i(\beta)g_{i-j}(\beta)' - g_i(\beta_0)g_{i-j}(\beta_0)'] \\ &= n^{-1} \sum_{i=\max(1,1+j)}^{\min(n,n+j)} [g_i(\beta) - g_i(\beta_0)]g_{i-j}(\beta)' \\ &\quad + n^{-1} \sum_{i=\max(1,1+j)}^{\min(n,n+j)} g_i(\beta_0)[g_{i-j}(\beta) - g_{i-j}(\beta_0)]', \end{aligned} \tag{A.28}$$

and hence that:

$$\begin{aligned}\|\widehat{\Gamma}_{j,n}(\beta) - \widehat{\Gamma}_{j,n}(\beta_0)\| &\le \{\sup_{1\le i\le n} \|g_i(\beta)\| + \sup_{1\le i\le n} \|g_i(\beta_0)\|\} \cdot\ n^{-1} \sum_{i=1}^{n} \|g_i(\beta) - g_i(\beta_0)\| \\ &\le 2 \sup_{1\le i\le n} d(Z_i) \cdot\ n^{-1} \sum_{i=1}^{n}\sum_{l=1}^{q} h_{i,l}(\beta). \end{aligned} \tag{A.29}$$

Since for given n, the right-hand side of Equation (A.29) is the same for all $j = -2m_n, \dots, 2m_n$, then substituting Equation (A.29) into Equation (A.27) implies that:

$$\|\widehat{\Sigma}_n(\bar{\beta}_n) - \widehat{\Sigma}_n(\beta_0)\| \le 4S_n \sup_{1\le i\le n} d(Z_i) \cdot\ n^{-1} \sum_{i=1}^{n}\sum_{l=1}^{q} h_{i,l}(\bar{\beta}_n). \tag{A.30}$$

Now Assumptions A1 and A5 imply that $n^{-1}\sum_{i=1}^{n} d(Z_i)^{\nu} = O_p(1)$ which by the Markov Inequality implies that $\sup_{1\le i\le n} d(Z_i) = O_p(n^{1/\nu})$. In conjunction with Equation (A.30) and Assumption A7 this then implies that:

$$\|\widehat{\Sigma}_n(\bar{\beta}_n) - \widehat{\Sigma}_n(\beta_0)\| \le O_p(n^{\delta+(1/\nu)}) \left\{ n^{-1} \sum_{i=1}^{n}\sum_{l=1}^{q} h_{i,l}(\bar{\beta}_n) \right\}. \tag{A.31}$$

Fourth, note that $\|g_i(\beta) - g_i(\beta_0)\| = h(Z_i, \beta) \ge 0$ for all $\beta \in B$ with equality if $\beta = \beta_0$. Since $\bar{\beta}_n = \beta_0 + O_p(n^{-1/2}) = \beta_0 + o_p(1)$, it follows from Assumption A14 that:

$$0 \le n^{-1} \sum_{i=1}^{n}\sum_{l=1}^{q} h_{i,l}(\bar{\beta}_n) \le \sum_{l=1}^{q} h_l^e(\bar{\beta}_n) + o_p(n^{-1/2}), \tag{A.32}$$

where $h_l^e(\beta) = E_0[h_{i,l}(\beta)]$. But since $h_{i,l}(\beta) \ge 0$ for all $\beta \in B$ with equality if $\beta = \beta_0$ it follows that that $h_{i,l}(\beta) = |h_{i,l}(\beta) - h_{i,l}(\beta_0)|$ for all $\beta \in B$ which combined with Assumption

A15 and the property that $\bar{\beta}_n = \beta_0 + O_p(n^{-1/2}) = \beta_0 + o_p(1)$ implies that:

$$0 \leq n^{-1} \sum_{i=1}^{n} \sum_{l=1}^{q} h_{i,l}(\bar{\beta}_n) \leq q L_0 \cdot \|\bar{\beta}_n - \beta_0\| + o_p(n^{-1/2}). \tag{A.33}$$

This in conjunction with Equation (A.31) then implies that:

$$\|\widehat{\Sigma}_n(\bar{\beta}_n) - \widehat{\Sigma}_n(\beta_0)\| \leq O_p(n^{\delta+(1/\nu)}) \left[\|\bar{\beta}_n - \beta_0\| + o_p(n^{-1/2})\right] \tag{A.34}$$

But by assumption $\delta < (1/2)-(1/\nu)$ so $\delta+(1/\nu)-(1/2) < 0$ and hence $\|\widehat{\Sigma}_n(\bar{\beta}_n)-\widehat{\Sigma}_n(\beta_0)\| = o_p(1)$ and hence $\|\widehat{\Sigma}_n(\bar{\beta}_n) - \Sigma_1\| = o_p(1)$ as desired. □

Proof of Lemma 5

First, observe from Assumptions A11 and A12 it follows that:

$$g(\bar{\beta}_n) = \widehat{g}_{(n)}(\bar{\beta}_n) - \widehat{g}_{(n)}(\beta_0) + o_p(n^{-1/2}) = \widehat{g}_{(n)}(\bar{\beta}_n) + O_p(n^{-1/2}), \tag{A.35}$$

which together with Assumption A13 then implies that:

$$G_0 \cdot (\bar{\beta}_n - \beta_0) + o_p(\|\bar{\beta}_n - \beta_0\|) = \widehat{g}_{(n)}(\bar{\beta}_n) + O_p(n^{-1/2}), \tag{A.36}$$

by application of the Mean Value Theorem, where denotes the Jacobian matrix of $g(\beta)$ with respect to β evaluate at $\beta = \beta_0$. Since G_0 has rank of p, by Assumption A13, it follows that:

$$\|\bar{\beta}_n - \beta_0\| \leq \|{G_0}'G_0\|^{-1/2} \cdot \|\widehat{g}_{(n)}(\bar{\beta}_n)\| + o_p(\|\bar{\beta}_n - \beta_0\|) + O_p(n^{-1/2}), \tag{A.37}$$

where for any matrix A, $\|A\| = [tr(A'A)]^{1/2}$, and hence that:

$$\|\bar{\beta}_n - \beta_0\| \leq O_p(1) \cdot \|\widehat{g}_{(n)}(\bar{\beta}_n)\| + O_p(n^{-1/2}). \tag{A.38}$$

Second, fix $\zeta = (1/2) - \delta$ and then define:

$$\dot{\lambda}_n = \begin{cases} -n^{-\zeta}\widehat{g}_{(n)}(\bar{\beta}_n)/\|\widehat{g}_{(n)}(\bar{\beta}_n)\|, & \text{if } \widehat{g}_{(n)}(\bar{\beta}_n) \neq 0, \\ -n^{-\zeta}(1, 0, \ldots, 0)', & \text{otherwise,} \end{cases} \tag{A.39}$$

as in the proof of Lemma 3. Following the line of argument in the proof of Lemma 3 we can then show that, under a subset of the assumptions made here:

$$\text{w.p.a.}\,1: \ -\dot{\lambda}_n{}'\widehat{g}_{(n)}(\bar{\beta}_n) + (3/4)\dot{\lambda}_n{}'\widehat{H}_n(\bar{\beta}_n, 0)\dot{\lambda}_n \leq \widehat{P}_n(\bar{\beta}_n, \dot{\lambda}_n), \tag{A.40}$$

and also that:

$$\text{w.p.a.}\,1: \ \widehat{P}_n(\bar{\beta}_n, \dot{\lambda}_n) \leq \widehat{P}^*_n(\bar{\beta}_n) \leq \widehat{P}^*_n(\beta_0) + o_p(n^{-\sigma}) = O_p(n^{-1+\delta}) + o_p(n^{-\sigma}). \tag{A.41}$$

Together with the results that $\widehat{H}_n(\bar{\beta}_n, 0) = -S_n^{-1}\widehat{\Sigma}_n(\bar{\beta}_n)$ and $-\dot{\lambda}_n{}'\widehat{g}_{(n)}(\bar{\beta}_n) = n^{-\zeta}\|\widehat{g}_{(n)}(\bar{\beta}_n)\|$ these imply that:

$$\text{w.p.a.}\,1: \ \|\widehat{g}_{(n)}(\bar{\beta}_n)\| \leq O_p(n^{-\delta-\zeta}) \cdot \|\widehat{\Sigma}_n(\bar{\beta}_n)\| + O_p(n^{-1+\delta+\zeta}) + o_p(n^{-\sigma+\zeta}). \tag{A.42}$$

Third, observe that:

$$\|\widehat{\Sigma}_n(\bar{\beta}_n)\| \leq \|\widehat{\Sigma}_n(\bar{\beta}_n) - \widehat{\Sigma}_n(\beta_0)\| + \|\widehat{\Sigma}_n(\beta_0) - \Sigma_1\| + \|\Sigma_1\|, \tag{A.43}$$

which in conjunction with Assumption A8 and Equation (A.34) from the proof of Lemma 4 then implies that:

$$\text{w.p.a.}\,1: \ \|\widehat{g}_{(n)}(\bar{\beta}_n)\| \leq O_p(n^{(1/\nu)-\zeta}) \cdot \|\bar{\beta}_n - \beta\| + O_p(n^{-1/2}). \tag{A.44}$$

In conjunction with Equation (A.38), this implies that:

$$\|\bar{\beta}_n - \beta\| \leq O_p(n^{(1/\nu)+\delta-(1/2)}) \cdot \|\bar{\beta}_n - \beta\| + O_p(n^{-1/2}). \tag{A.45}$$

Fourth, since $\delta < (1/2) - (1/\nu)$ it then follows that

$$\|\bar{\beta}_n - \beta\| \leq o_p(1) \cdot \|\bar{\beta}_n - \beta\| + O_p(n^{-1/2}). \tag{A.46}$$

from which the desired result follows immediately. □

Proof of Lemma 6

First, fix $0 < L < \infty$n and define $B_n(L) = \{\beta \in B : \|\beta - \beta_0\| \leq n^{-1/2}L\}$. Then, for any $(1/\nu) < \zeta < \infty$ define Λ_n^ζ as in Lemma 1 and observe that by application of Lemma 1 it follows that there exist $0 \leq \eta_{1,n} \leq 1 \leq \eta_{2,n}$ such that:

$$\text{w. p. a. } 1: \ \eta_{1,n} = \inf_{\beta \in B_n(L), \lambda \in \Lambda_n^\zeta, 1-m_n \leq i \leq n+m_n} [-\rho_2(\lambda' g_{i,n}^\omega(\beta)], \tag{A.47}$$

$$\text{w. p. a. } 1: \ \eta_{2,n} = \sup_{\beta \in B_n(L), \lambda \in \Lambda_n^\zeta, 1-m_n \leq i \leq n+m_n} [-\rho_2(\lambda' g_{i,n}^\omega(\beta)], \tag{A.48}$$

and $\eta_{j,n} = 1 + o_p(1)$ for $j = 1, 2$. By a Taylor series expansion it follows that for all $\beta \in B$ and $\lambda \in \widehat{\Lambda}_n(\beta)$:

$$\widehat{P}_n(\beta, \lambda) = -\lambda' \widehat{g}_{(n)}(\beta) + (1/2)\lambda' \widehat{H}_n(\beta, \pi\lambda)\lambda, \tag{A.49}$$

for some $0 \leq \pi \leq 1$ which may depend on β, λ, $\{Z_i\}_{i=-\infty}^{\infty}$ and n. By application of Lemma 1 it then follows that:

$$\text{w. p. a. } 1: \ 0 \leq \eta_{1,n}\widehat{\Sigma}_n(\beta) \leq [-S_n \widehat{H}_n(\beta, \lambda)] \leq \eta_{2,n}\widehat{\Sigma}_n(\beta), \quad \forall \beta \in B_n(L) \ \& \ \lambda \in \Lambda_n^\zeta, \tag{A.50}$$

where $\leq$ is used in the non-negative definite ordering sense for matrices so that $A \leq B$ means that $(B - A)$ is non-negative definite. Now since Λ_n^ζ is compact it also follows by application of Lemma 1 that there exists $\tilde{\lambda}_n(\beta)$ for each $\beta \in B_n(L)$ such that:

$$\text{w. p. a. } 1: \ \tilde{\lambda}_n(\beta) = \arg\max_{\lambda \in \Lambda_n^\zeta} \widehat{P}_n(\beta, \lambda), \quad \forall \beta \in B_n(L), \tag{A.51}$$

and hence that:

$$\text{w. p. a. } 1: \ 0 = \widehat{P}_n(\beta, 0) \leq \widehat{P}_n(\beta, \tilde{\lambda}_n(\beta)), \quad \forall \beta \in B_n(L). \tag{A.52}$$

Combined with the Taylor series expansion from above this implies that:

$$\text{w. p. a. } 1: \ 0 \leq -\tilde{\lambda}_n(\beta)' \widehat{g}_{(n)}(\beta) + (1/2)\tilde{\lambda}_n(\beta)' \widehat{H}_n(\beta, \pi\tilde{\lambda}_n(\beta))\tilde{\lambda}_n(\beta), \quad \forall \beta \in B_n(L), \tag{A.53}$$

for some $0 \leq \pi \leq 1$ which may depend on β, $\{Z_i\}_{i=-\infty}^{\infty}$ and n, and hence that:

$$\text{w. p. a. } 1: \ \tilde{\lambda}_n(\beta)' \widehat{\Sigma}_n(\beta)\tilde{\lambda}_n(\beta) \leq 2\eta_{1,n}^{-1} S_n \cdot \|\tilde{\lambda}_n(\beta)\| \cdot \|\widehat{g}_{(n)}(\beta)\|, \quad \forall \beta \in B_n(L). \tag{A.54}$$

Now let $\widehat{\mu}_n(\beta)$ denote the smallest eigenvalue of $\widehat{\Sigma}_n(\beta)$ and μ_0 denote the smallest eigenvalue of Σ_0. It follows by application of Lemmas 4 and 5 that:

$$\text{w. p. a. } 1: \widehat{\mu}_n(\beta) \geq (1/2)\mu_0 \geq 0, \quad \forall \beta \in B_n(L), \tag{A.55}$$

But Σ_0 is non-singular by assumption and hence $\mu_0 > 0$ which then implies that:

$$\text{w. p. a. } 1: \|\tilde{\lambda}_n(\beta)\| \leq 4(\mu_0 \eta_{1,n})^{-1} S_n \cdot \|\widehat{g}_{(n)}(\beta)\|, \quad \forall \beta \in B_n(L). \tag{A.56}$$

But then observe from by Assumptions A11–A13 it follows that:

$$\sup_{\beta \in B_n(L)} \|\widehat{g}_{(n)}(\beta)\| = O_p(n^{-1/2}), \tag{A.57}$$

and hence that:

$$\sup_{\beta \in B_n(L)} \|\tilde{\lambda}_n(\beta)\| = O_p(n^{-(1/2)+\delta}) = o_p(n^{-\zeta}), \tag{A.58}$$

which then implies that:

$$\text{w. p. a. } 1: \ \tilde{\lambda}_n(\beta) \in \text{int}(\Lambda_n^{\zeta}), \quad \forall \beta \in B_n(L), \tag{A.59}$$

and hence that:

$$\text{w. p. a. } 1: \ \tilde{\lambda}_n(\beta) \in \text{int}(\widehat{\Lambda}_n(\beta)), \quad \forall \beta \in B_n(L), \tag{A.60}$$

so that:

$$\text{w. p. a. } 1: \ \tilde{\lambda}_n(\beta) = \arg \max_{\lambda \in \widehat{\Lambda}_n(\beta)} \widehat{P}_n(\beta, \lambda), \quad \forall \beta \in B_n(L). \tag{A.61}$$

Now observe that by application of Lemma 4 it follows that there exists $0 \leq c_{1,n} \leq 1 \leq c_{2,n}$ such that:

$$c_{1,n}\widehat{\Sigma}_n(\beta) \leq \Sigma_0 \leq c_{2,n}\widehat{\Sigma}_n(\beta), \quad \forall \beta \in B_n(L), \tag{A.62}$$

(where $\leq$ is again used in the non-negative definite sense) and $c_{j,n} = 1 + o_p(1)$ for $j = 1, 2$. Hence it follows that:

$$\text{w. p. a. } 1: \ \widehat{Q}_n(\beta, \lambda; c_{1,n}^{-1}\eta_{2,n}S_n^{-1}\Sigma_0) \leq \widehat{P}_n(\beta, \lambda) \leq \widehat{Q}_n(\beta, \lambda; c_{2,n}^{-1}\eta_{1,n}S_n^{-1}\Sigma_0), \quad \forall \beta \in B_n(L) \& \lambda \in \Lambda_n^{\zeta}, \tag{A.63}$$

where:

$$\widehat{Q}_n(\beta, \lambda; A) = -\lambda' \widehat{g}_{(n)}(\beta) - (1/2)\lambda' A \lambda, \forall \beta \in B \& \lambda \in \mathbb{R}^q. \tag{A.64}$$

For any $\beta \in B$ and symmetic positive-definite matrix A define:

$$\bar{\lambda}_n(\beta; A) \equiv \arg\max_{\lambda \in \mathbb{R}^q} \widehat{Q}_n(\beta, \lambda; A) = -A^{-1}\widehat{g}_{(n)}(\beta), \tag{A.65}$$

$$\widehat{Q}_n^*(\beta; A) \equiv \max_{\lambda \in \mathbb{R}^q} \widehat{Q}_n(\beta, \lambda; A) = (1/2)\widehat{g}_{(n)}(\beta)' A^{-1} \widehat{g}_{(n)}(\beta). \tag{A.66}$$

It follows that:

$$\sup_{\beta \in B_n(L)} \|\bar{\lambda}_n(\beta; c_{1,n}^{-1}\eta_{2,n} S_n^{-1}\Sigma_0)\| = O_p(n^{-(1/2)+\delta}) = o_p(n^{-\zeta}), \tag{A.67}$$

$$\sup_{\beta \in B_n(L)} \|\bar{\lambda}_n(\beta; c_{2,n}^{-1}\eta_{1,n} S_n^{-1}\Sigma_0)\| = O_p(n^{-(1/2)+\delta}) = o_p(n^{-\zeta}), \tag{A.68}$$

and hence that:

$$\text{w. p. a. } 1: \; \widehat{Q}_n^*(\beta; c_{1,n}^{-1}\eta_{2,n} S_n^{-1}\Sigma_0) \le \widehat{P}_n^*(\beta) \le \widehat{Q}_n(\beta; c_{2,n}^{-1}\eta_{1,n} S_n^{-1}\Sigma_0), \quad \forall \beta \in B_n(L). \tag{A.69}$$

But this then implies that:

$$\text{w. p. a. } 1: \; 2n(c_{1,n}/\eta_{2,n})\widehat{Q}_n^*(\beta; \Sigma_0) \le (2n/S_n)\widehat{P}_n^*(\beta) \le 2n(c_{2,n}/\eta_{1,n})\widehat{Q}_n^*(\beta; \Sigma_0), \quad \forall \beta \in B_n(L). \tag{A.70}$$

Since $\sup_{\beta \in B_n(L)} \widehat{g}_n^*(\beta) = O_p(n^{-1/2})$ then $0 \le n\widehat{Q}_n^*(\beta; \Sigma_0) = O_p(1)$ and then since $(c_{1,n}/\eta_{2,n}) = 1 + o_p(1)$ and $(c_{2,n}/\eta_{1,n}) = 1 + o_p(1)$ it follows that:

$$\sup_{\beta \in B_n(L)} \left| (2n/S_n)\widehat{P}_n^*(\beta) - 2n\widehat{Q}_n^*(\beta; \Sigma_0) \right| = o_p(1), \tag{A.71}$$

as desired. □

Proof of Theorem 2

The first step is to observe Lemma 6 implies that there exists a sequence of strictly positive constants $\{L_n\}_{n=1}^{\infty}$ such that $L_n \to \infty$ and:

$$\sup_{\beta \in B_n(L_n)} \left| (2n/S_n)\widehat{P}_n^*(\beta) - 2n\widehat{Q}_n^*(\beta; \Sigma_0) \right| = o_p(1). \tag{A.72}$$

Next, there exists a sequence of mappings $\ddot{\beta}_n$ such that:

$$n\widehat{Q}_n^*(\ddot{\beta}_n;\Sigma_0) \leq \inf_{\beta\in B} n\widehat{Q}_n^*(\beta;\Sigma_0) + o_p(n^{-1}). \tag{A.73}$$

Then it follows from Theorems 2.6 and 7.2 of Newey and McFadden (1994) together with the discussion in Newey and McFadden (1994, p. 2133) that:

$$n^{1/2}(\ddot{\beta}_n - \beta_0) \xrightarrow{D} N[0,(G_0\Sigma_0^{-1}G_0)^{-1}]. \tag{A.74}$$

Note that if $\ddot{\beta}_n$ are not measurable then these statements should be interpreted in terms of outer measure rather than probability. It then follows from Lemma 5 that:

$$\Pr\left\{\widehat{\beta}_n \in B_n(L_n) \ \&\ \ddot{\beta}_n \in B_n(L_n)\right\} \to 1, \tag{A.75}$$

and thus:

$$\begin{aligned} 2n\widehat{Q}_n^*(\widehat{\beta}_n;\Sigma_0) &\leq 2nS_n^{-1}\widehat{P}_n^*(\widehat{\beta}_n) + o_p(1) \leq 2nS_n^{-1}\widehat{P}_n^*(\ddot{\beta}_n) + o_p(1) \\ &\leq 2n\widehat{Q}_n^*(\ddot{\beta}_n;\Sigma_0) + o_p(1) \leq 2n\inf_{\beta\in B}\widehat{Q}_n^*(\beta;\Sigma_0) + o_p(1). \end{aligned} \tag{A.76}$$

But this implies that $\widehat{\beta}_n$ satsifies the requirements of Theorem 7.2 of Newey and McFadden (1994) and hence that:

$$n^{1/2}(\widehat{\beta}_n - \beta_0) \xrightarrow{D} N[0,(G_0\Sigma_0^{-1}G_0)^{-1}]. \tag{A.77}$$

as desired. □

References

Andrews, D. W. K. (1991) ‘Heteroskedasticity and autocorrelation consistent covariance matrix estimation.’ *Econometrica* 59, 817–858

—— (1993) ‘An introduction to econometric applications of empirical process theory for dependent random variables.’ *Econometric Reviews* 12(2), 193–216

Chen, S., and P. Hall (1993) 'Smoothed empirical likelihood confidence intervals for quantiles.' *annals of Statistics* 21, 1166–1181

Davidson, J. (1994) *Stochastic Limit Theory: An Introduction for Econometricians* (Oxford: Oxford University Press)

Hansen, L. P., J. Heaton, and A. Yaron (1996) 'Finite-sample properties of some alternative GMM estimators.' *Journal of Business and Economic Statistics* 14, 262–280

Imbens, G. W. (1997) 'One-step estimators for over-identified generalized method of moments models.' *Review of Economic Studies* 64, 359–383

Imbens, G. W., R. H. Spady, and P. Johnson (1998) 'Information theoretic approaches to inference in moment condition models.' *Econometrica* 66, 333–357

Kitamura, Y. (1997) 'Empirical likelihood methods with weakly dependent processes.' *Annals of Statistics* 25, 2084–2102

Kitamura, Y., and M. Stutzer (1997) 'An information-theoretic alternative to generalized method of moments estimation.' *Econometrica* 65, 861–874

Newey, W. K., and D. L. McFadden (1994) 'Large sample estimation and hypothesis testing.' In *Handbook of Econometrics, Volume IV,* ed. R. F. Engle and D. L. McFadden (Amsterdam: North Holland) chapter 36, pp. 2111–2245

Newey, W. K., and R. J. Smith (2004) 'Higher order properties of GMM and generalized empirical likelihood estimators.' *Econometrica* 72, 219–256

Otsu, Taisuke (2003) 'Empirical likelihood for quantile regression.' Mimeo, University of Wisconsin-Madison

Owen, A. B. (1988) 'Empirical likelihood ratio confidence intervals for a single functional.' *Biometrika* 75, 237–249

Qin, J., and J. Lawless (1994) 'Empirical likelihood and general estimating equations.' *Annals of Statistics* 22, 300–325

Smith, R. J. (1997) 'Alternative semi-parametric likelihood approaches to generalized method of moments estimation.' *Economic Journal* 107, 503–519