Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19103-6297
pier@ssc.upenn.edu
http://www.econ.upenn.edu/pier

PIER Working Paper 97-028

"A LIL for m-Estimators and Applications to Hypothesis Testing with Nuisance Parameters"

by

Filippo Altissimo and Valentina Corradi

A LIL for m-Estimators and Applications to Hypothesis Testing with Nuisance Parameters

Filippo Altissimo(*)(**)[1]

Valentina Corradi(*)

(*) University of Pennsylvania
Department of Economics
3718 Locust Walk
Philadelphia, PA 19104-6297
corradi@econ.sas.upenn.edu, tel. (215) 898-1505

(**) Bank of Italy
Research Department
via Nazionale 91
Rome-00184, Italy
altissi@mbox.vol.it , tel: (0039-6) 4792-3441

First Draft: June 1996
This version, March 1997

[1]We are grateful to Herman Bierens, Jin Hahn, Ingmar Prucha, Shinichi Sakata, Giovanni Violante, as well as to the seminar participants at the University of Pennsylvania, the University of Maryland and the 1996 Summer Meetings of the North American Econometric Society for helpful comments and suggestions.

Abstract

The purpose of this paper is twofold: on one side we aim to provide easily computable almost sure bounds for inference in presence of nuisance parameters unidentified under the null. On the other hand, more generally, we aim to provide a flexible completely consistent procedure for inference in, possibly mispecified, parametric models, in the case of dependent and heterogeneous observations. With the term completely consistent we mean that the asymptotic size is zero and the asymptotic power is one. The small sample behavior of our procedure is analyzed via few Monte Carlo simulations; in particular we consider (i) conditional moment tests, (ii) testing for nonlinearities in the SETAR model. Overall the size approaches zero relatively slowly, while the power approaches one very quickly.

1 Introduction

The purpose of this paper is twofold: on one hand we aim to provide easily computable almost sure bounds for inference in presence of nuisance parameters. On the other hand more generally, we provide a flexible, completely consistent procedure for inference in, possibly mispecified, parametric models, in the case of dependent and heterogeneous observations. With the term completely consistent we mean that both type I and type II errors approach zero asymptotically, or equivalently the size approaches zero and the power approaches 1.

To fulfill this task we begin by providing a law of the iterated logarithm (LIL) for m-estimators (or extremum estimators). Basically, under slightly stronger moment requirements than in Gallant and White (1988), we show that $\limsup_{n\to\infty} d_n \|\hat{V}_n^{-1/2}\hat{A}_n(\hat{\theta}_n - \theta_n^*)\| \leq 1$, almost surely, where $d_n = \frac{\sqrt{n}}{\sqrt{2\log\log n}}$, $\|\cdot\|$ denotes the Euclidean norm, $\hat{V}_n$ and $\hat{A}_n$ are, respectively, estimators for the variance of the score and for the Hessian. In order to have almost sure, nuisance parameter free bounds, we also provide conditions under which heteroskedasticity-autocorrelation (HAC) robust estimators are strongly consistent for the "true" variance or for "something larger" than the true variance, in the case of heterogeneous mispecified models. This extends lemma 2 in Corradi (1996) to a general class of nonlinear estimators.

The almost sure bound above is the basis for our inference analysis. We then construct Wald and Lagrange Multiplier statistics, say W_n and LM_n, for possibly nonlinear parametric restrictions, and show that under the null, almost surely, $\limsup_{n\to\infty} d_n\|W_n(LM_n)\| \leq 1$, while under the alternative $d_n\|W_n\|, d_n\|LM_n\|$ diverge at rate $\sqrt{n}/\sqrt{2\log\log n}$. Hereafter, unless otherwise specified, for alternative hypothesis, we mean the global alternative. Thus we decide in favor of H_0 if we get a number smaller than or equal to one and we decide to reject H_0 if we get a number larger than one. As the probability of rejecting (accepting) H_0 when it is true (H_A is true) is asymptotically 0, the procedure is completely consistent.

We then move to hypothesis testing when nuisance parameters are present only under the alternative. In particular we are interested in testing $H_0 : h(\theta^*(\pi)) = 0, \forall\pi \in \Pi$ versus $H_A : h(\theta^*(\pi)) \neq 0$, for all $\pi \in \Pi$ or for $\pi \in \Pi^1$, with Π^1 having positive Lebesgue measure. Several economic problems involve hyphotesis testing with nuisance parameters present only under the alternative; as an example we may think of (consistent) tests for the functional form of the conditional mean (e.g. Bierens (1982, 1990), Bierens and Ploberger (1997), Lee , White and Granger (1993), Stinchcombe and White (1995), Hansen (1996)), of test for (the absence of) structural breaks (Andrews (1993), Andrews and Ploberger (1994), Bai and Perron (1995)) and of test for the presence of regime switching in models with unobserved regimes (e.g. Hamilton (1989), Hansen (1992), Garcia (1995)). It is well known that the asymptotic distribution in presence of nuisance parameters unidentified under the null is non standard; although the critical values may be well approximated via bootstrapping tecniques or via the p-value transformation approach of Hansen (1996), the computational burden is non indifferent. The computational problem of performing inference in presence of nuisance parameters can easily be overcome using a LIL based inference procedure which does not require approximation of the distribution of the statistics via simulation. In fact we just need to compute the statistics for different values of the nuisance parameters and then take the sup; under the null it has a well defined almost sure bound, under the alternative it diverges.

Stinchcombe and White (1995) were the first who derive almost sure bounds for hypothesis testing with nuisance parameters present only under the alternative. They obtain their result by appealing to the LIL for Banach-space valued iid random variables; we believe that their approach can be extended to the case of weak dependence, as LIL for stationary mixing Banach-space valued random variables are available in the probability literature (e.g. Berger (1990)). However, at least to the best of our knowledge, there are not Banach space LIL for the dependent and heterogeneous case.

Here instead we follow a different approach that allows for both dependence and heterogeneity. We show that, under mild conditions, under null, the properly rescaled score, say $\frac{1}{\sqrt{2n\log\log(n)}}\sum\nabla_\theta q_t(\theta^*(\pi),\pi)$, is a.s. asymptotically uniform equicontinuous (a.s. AUEC in Pötscher and Prucha (1994), terminology), and that this implies a LIL for $\sup_{\pi\in\Pi} d_n\|W_n(\pi)\|$ and $\sup_{\pi\in\Pi} d_n\|LM_n(\pi)\|$. We also allow for dynamic mispecification; thus the score is no longer a martingale difference sequence and we need to use a HAC-robust type estimator for its variance, say $\hat{V}_n(\pi)$; we shall show that $\sup_{\pi\in\Pi}|\hat{V}_n^{ij}(\pi) - V^{*ij}(\pi)| = o_{as}(1), \forall i,j$. The almost sure bound we derive in the dependent and heterogeneous case requires the differentiability of the score with respect to the nuisance parameters; while this allows to consider several interesting cases, e.g. conditional moment tests, it nevertheless rules out some other relevant cases, such as testing for structural breaks, thresholds type nonlinearities. Such a smoothness requirement is relaxed in the stationary case, where we derive almost sure bounds under the same assumptions as in Hansen (1996), except for the fact that we also allow for dynamic mispecification[1].

We finally provide some Monte Carlo simulations in order to assess the finite sample performance of our procedure. More precisely we consider (i) tests for the functional form of the conditional mean as in Bierens (1982, 1990) and (ii) testing for nonlinearities in the SETAR model. Overall we find that the size is approaching zero relatively slowly, but the power goes to one very quickly. This finding is peculiar to the LIL based inference procedure given that other procedures for testing in presence of nuisance paramenter based on simulation methods share to opposite property of good size but poor power in small sample. In this sense a sensible strategy is to start by using the LIL-based procedure; if we get a value less or equal to one we can safely decide in favor of the null, otherwise we may wish to use also other testing procedures.

The paper is organized as follows. Section 2 provides a LIL for m-estimators in the case of dependent and heterogeneous observations. Section 3 and 4 give almost sure bounds for Wald and Lagrange multiplier statistics, respectively in absence and in presence of nuisance parameter unidentified under the null. Section 5 applies the testing procedure developed above and gives some simulation results. Appendix 1 contains all the assumptions, Appendix 2 contains all the proofs.

2 A LIL for m-Estimators

To avoid notational burden, the assumptions used in all the sections and the proofs of the theorems are collected in Appendix 1 and Appendix 2, respectively. Starting, let us define the

[1] As Andrew and Ploberger (1994) and Hansen (1996) our approach cannot handle the case of test for regime switching in the Hamilton's model due to the fact that in that case the problem of the nuisance parameter is combined with the difficulty of the singularity of the information matrix under the null of the test.

m-estimator as

$$\hat{\theta}_n = argmin_{\theta \in \Theta} \frac{1}{n} \sum_{t=1}^{n} q_t(X^t, \theta)$$

where $X^t = (X_1, X_2, \ldots X_t)$ and similarly

$$\theta_n^* = argmin_{\theta \in \Theta} \frac{1}{n} \sum_{t=1}^{n} E(q_t(X^t, \theta)).$$

We have that

THEOREM 1

Let A1-A4, A6, A8, A9(i)-(iii), A10-A13 hold, then

$$\sqrt{\frac{n}{2 \log \log(n)}} ||\hat{\theta}_n - \theta_n^*|| = O_{as}(1).$$

Theorem 1 simply states an almost sure rate of convergence for m-estimators. A1-A9(i)-(iii) are standard memory, moment, near-epoch dependent, smoothness and a.s. Lipschitz assumptions which are usually used in order to derive asymptotic normality for nonlinear estimators. The price we have to pay in order to have an almost sure rate, A10-13, are conditions imposing some constraints on the degree of heterogeneity we can allow for; so we do not need A10-13 in the strictly starionary case. In particular, A10 requires $A_n^* = n^{-1} \sum E(\nabla_\theta^2 q_t(\theta_n^*))$ converges to A^* rather than just being $O(1)$. A11 requires that $\theta_{n,m}^* = argmin_{\theta \in \Theta} n^{-1} \sum_{t=m+1}^{m+n} E(\nabla_\theta q_t(X^t, \theta))$ converges to θ^*, uniformly in m. Finally A12 is a sort of linear growth conditions on the variance of the score and A13 imposes a constraint on the heterogeneity of the conditional variance.

As we mentioned in the introduction, in order to have an almost sure bound useful in practice, we need to provide a strong consistent estimator for the variance of the score. Let us define the covariance matrix of the score as

$$\hat{V}_n = n^{-1} \sum_{t=1}^{n} \nabla q_t(\hat{\theta}_n) \nabla q_t(\hat{\theta}_n)' \tag{1}$$

$$+ n^{-1} \sum_{\tau=1}^{l_n} w_\tau \sum_{t=\tau+1}^{n} \left(\nabla q_t(\hat{\theta}_n) \nabla q_{t-\tau}(\hat{\theta}_n)' + \nabla q_{t-\tau}(\hat{\theta}_n) \nabla q_t(\hat{\theta}_n)' \right),$$

where $w_\tau = \left(1 - \frac{\tau}{l_n+1}\right)$ and $l_n = o\left(\sqrt{\frac{n}{2 \log \log(n)}}\right)$.

We have then

THEOREM 2

Let A1-A3, A5(ii)(iii), A7(ii)-(iii), A8-A9, A11-A20 hold. Then

$$\hat{V}_n - (V^* + U^*) = o_{a.s.}(1)$$

where V^* and U^* are as defined in the statement of A12 and A20 respectively, and $\hat{V}_n$ is as defined in (1).

Theorem 2 provides an almost sure consistency result for autocorrelation-heteroskedasticity robust estimators of the variance. Basically we use Theorem 1 and A14-A19 in order to show that

$$\hat{V}_n = n^{-1} \sum \nabla_\theta q_t(\theta_n^*) \nabla_\theta q_t(\theta_n^*)'$$

$$+n^{-1} \sum_{\tau=1}^{l_n} w_\tau \sum_{t=\tau+1}^{n} \left(\nabla_\theta q_t(\theta_n^*) \nabla_\theta q_{t-\tau}(\theta_n^*)' + \nabla_\theta q_{t-\tau}(\theta_n^*) \nabla_\theta q_t(\theta_n^*)' \right) + o_{as}(1)$$

and we then apply Lemma 2 in Corradi (1996). By somewhat restricting the degree of heterogeneity we allow for (we indeed require via A11-A20 asymptotic covariance stationarity and asymptotic homokurtosis), and by requiring an a.s. Lipschitz condition, A9(iv), on the elements of $(\nabla_\theta^2 q_t(\theta_n^*))^2$, we strenghten the weak consistency result of Gallant and White (ch.6, 1988) into an almost sure result. Needless to say we obtain an almost sure consistent estimator for the variance of the score in the homogeneneous and/or correct specification case and an overestimate of the variance in the case of heterogeneity and incorrect specification of the conditional mean, as in the latter case, although $n^{-1} \sum E(\nabla_\theta q_t(\theta_n^*)) = 0$, in general $E(\nabla_\theta q_t(\theta_n^*)) \neq 0$. In the independent case or in the case of dynamic correct specification, the score is martingale difference, so we do not need to take into account the cross-terms in the estimation of the variance. Thus A14-A19 and A9(iv) can be relaxed and the almost sure consistency of $\hat{V}_n$ comes straightforwardly from the uniform strong law of large numbers. We are now in the position of providing a law of the iterated logarithm (LIL) for the m-estimator,

THEOREM 3

Let A1-A3, A5, A7-A20 hold. Then

$$(i) \limsup_{n \to \infty} \sqrt{\frac{n}{2loglog(n)}} ||\hat{V}_n^{-1/2} \hat{A}_n (\hat{\theta}_n - \theta_n^*)|| \leq 1$$

almost surely. And if U^*, as defined in A20 is equal 0, then

$$(ii) \limsup_{n\to\infty} \sqrt{\frac{n}{2loglog(n)}} ||\hat{V}_n^{-1/2} \hat{A}_n(\hat{\theta}_n - \theta_n^*)|| = 1$$

almost surely.

Theorem 3, providing an almost sure nuisance parameter free bound for $d_n(\hat{\theta}_n - \theta_n^*)$, forms the basis of our inference procedure. The strong consistency of $\hat{V}_n$ and $\hat{A}_n$ for $V^* + U^*$ and for A^*, respectively, preserves the almost sure nature of the bound. In fact if $\hat{A}_n$ and $\hat{V}_n$ were only weakly consistent, the result above would hold in probability rather than almost surely.

The class of m-estimators we consider includes ordinary least squares, nonlinear least squares, maximum likelihood, quasi maximum likelihood. However from the proofs of theorem 1-3 we see that our result can be straighforwardly generalized to (nonlinear) instrumental variable estimators. Also, at least in the homogeneous case and provided the estimator of the weighting matrix converges almost surely to a symmetric, positive definite matrix, analogous bounds also hold for the generalized method of moments estimators.

The result stated in Theorem 3 forms the basis of the hypothesis testing procedure developed in the next two sessions, respectively, in absence and in presence of nuisance parameters unidentified under the null.

3 Hypothesis Testing

We start by developing the inference procedure through LIL in the case of absence of nuisance parameters. Let H_L and H_A denote respectively the local and the global alternatives. We consider

$$H_0 : d_n h_n(\theta_n^*) \to 0$$

$$H_A : \forall n > n_0, \exists \epsilon > 0, s.t., |h(\theta_n^*)| > \epsilon$$

$$H_L : \{d_n h(\theta_n^*)\} = O(1)$$

H_0, H_A, H_L are defined following the framework for hypothesis testing for heterogeneous, possibly mispecified models suggested by White (1994, Ch.8), replacing $\sqrt{n}$ with d_n. As for the null, we require that the hypothized restrictions hold in the limit and converge sufficiently fast.

Clearly, if we rule out the possible joint presence of mispecification and heterogeneity, we can simply state the null as $H_0 : h(\theta^*) = 0$. As for the alternative we require $h(\theta_n^*)$ to be bounded away from zero $\forall n$ sufficiently large; if we rule out the joint presence of heterogeneity and mispecification, we could simply state $H_A : h(\theta^*) \neq 0$. As for the local alternative, we require $d_n h(\theta_n^*)$ to be asymptotically bounded, that is we require the set of restriction to approach zero sufficiently slow. Again if we rule out the joint presence of heterogeneity and mispecification, we could have defined the local in the usual way, $H_{L_n} = c/d_n$, i.e. by slightly perturbating the data generating process (DGP) and considering a sequence of alternatives.
Define $\widetilde{\theta}_n$ as

$$\begin{aligned} \widetilde{\theta}_n &= arg\ \min_{\theta} \frac{1}{n}\sum_{t=1}^{n} q_t(X^t, \theta) \\ with\ \theta &\in \Theta_0 = \{\theta \in \Theta : \ h_n(\theta) = 0\} \end{aligned} \tag{2}$$

and θ_n^0 as

$$\begin{aligned} \theta_n^0 &= arg\ \min_{\theta} \frac{1}{n}\sum_{t=1}^{n} E(q_t(X^t, \theta)) \\ with\ \theta &\in \Theta_0 = \{\theta \in \Theta : \ h_n(\theta) = 0\} \end{aligned}$$

while $\widehat{\theta}_n$ and θ_n^* are the unrestricted counterparts.
Let define the Lagrange Multiplier statistic as

$$LM_n = \widetilde{B}_n^{-\frac{1}{2}} H_n(\widetilde{\theta}_n)' \left(\nabla^2 Q_n(\widetilde{\theta}_n)\right)^{-1} \nabla Q_n(\widetilde{\theta}_n) \tag{3}$$

where $Q_n(\theta) = n^{-1}\sum q_t(X^t, \theta)$, $H_n(\theta)$ is the Jabobian of h_n and

$$\widetilde{B}_n = H_n(\widetilde{\theta}_n)' \left(\nabla^2 Q_n(\widetilde{\theta}_n)\right)^{-1} \widetilde{V}_n \left(\nabla^2 Q_n(\widetilde{\theta}_n)\right)^{-1} H_n(\widetilde{\theta}_n)$$

with

$$\widetilde{V}_n = n^{-1}\sum_{t=1}^{n} \nabla q_t(\widetilde{\theta}_n) \nabla q_t(\widetilde{\theta}_n)'$$

$$+ n^{-1}\sum_{\tau=1}^{l_n} w_\tau \sum_{t=\tau+1}^{n} \left(\nabla q_t(\widetilde{\theta}_n) \nabla q_{t-\tau}(\widetilde{\theta}_n)' + \nabla q_{t-\tau}(\widetilde{\theta}_n) \nabla q_t(\widetilde{\theta}_n)'\right).$$

Similarly we define the Wald statistics as

$$W_n = \widehat{B}_n^{-1/2} h_n(\widehat{\theta}_n)$$

and

$$\widehat{B}_n = H_n(\widehat{\theta}_n)' \left(\nabla^2 Q_n(\widehat{\theta}_n)\right)^{-1} \widehat{V}_n \left(\nabla^2 Q_n(\widehat{\theta}_n)\right)^{-1} H_n(\widehat{\theta}_n)$$

$$\widehat{V}_n = n^{-1} \sum_{t=1}^{n} \nabla q_t(\hat{\theta}_n) \nabla q_t(\hat{\theta}_n)'$$

$$+n^{-1} \sum_{\tau=1}^{l_n} w_\tau \sum_{t=\tau+1}^{n} \left(\nabla q_t(\hat{\theta}_n) \nabla q_{t-\tau}(\hat{\theta}_n)' + \nabla q_{t-\tau}(\hat{\theta}_n) \nabla q_t(\hat{\theta}_n)'\right).$$

We have,

THEOREM T1

Let A1-A3, A5, A7-A20 and AT1 and AT2 hold. Define $d_n = \sqrt{\frac{n}{2loglog(n)}}$;

(i) Under H_0,

$$(a) \limsup_{n\to\infty} d_n \|W_n\| \leq 1$$
$$(b) \limsup_{n\to\infty} d_n \|LM_n\| \leq 1$$

almost surely. And if U^*, as defined in A20 is equal 0, then (a) and (b) hold with equalities.

(ii) Under H_A,

$$(a) P\{\lim_{n\to\infty} d_n \|W_n\| > c_n\} = 1$$

$$(b) \lim_{n\to\infty} P[d_n \|LM_n\| > c_n] = 1$$

where $c_n \times (\frac{n}{2 \ln \ln n})^{-(\frac{1}{2}-\delta)} \to \infty$ as n→ ∞, with $\delta \in (0, 1/2)$.

(iii) Under H_L,

$$(a) \limsup_{n\to\infty} d_n \|W_n\| \leq 1 + \Delta$$

almost surely, with $0 < \Delta < \infty$,

$$(b) \lim_{n\to\infty} P[d_n \|LM_n\| \leq 1 + \Delta] = 1.$$

Under the global alternative $d_n\|W_n\|$ diverges with probability one, while $d_n\|LM_n\|$ diverges with probability approaching one. This is due to the fact that we did not impose that, under H_A, $\theta^0_{n,m} = argmin_{\theta\in\Theta:h(\theta)=0} n^{-1}\sum_{t=m+1}^{m+n} q_t(X^t,\theta)$ converge to some limit θ^0 uniformly in m and that A12-A19 hold with $\theta^*_{n,m}$ replaced by $\theta^0_{n,m}$. Under these further assumptions also $d_n\|LM_n\|$ would have diverged with probability one. Under the local alternative, $d_n\|W_n\|$ has an almost sure bound which is looser than one and $d_n\|LM_n\|$ has a weak bound which is looser than one.

From Theorem T1, we see that a testing procedure can be based on rejecting (accepting) H_0 according whether $d_n\|LM_n\|$ or $d_n\|W_n\|$ is greater than one (smaller or equal to one). This delivers a flexible, completely consistent procedure for inference in, possibly mispecified, parametric models. With the term completely consistent we mean that both type I and type II errors approach zero asymptotically, or equivalently the size approaches zero and the power approaches one.

For example, Sin and White (1996) consider penalized likelihood criteria that guarantee the choice, with probability one, of the model attaining the lower Kullback Liebler Information Criterion (KLIC) or, if both models reach the same KLIC, of the more parsimonious model. In particular they suggest to pick model A (B) according whether their information criterion is negative (positive). Sin and White derive their result from the LIL for weakly dependent processes, in this sense their framework is close to the framework considered here. Phillips and Ploberger (1996) show that, for various classes of likelihood functions and prior density, the Bayesian data density is asymptotically of the same exponential form and propose a model selection approach based on picking the model that maximize the (log of) exponential data density. Such a criterion is called PIC as it maximizes the posterior information criterion. By deciding in favor of model A (B) according to whether the PIC is smaller or larger than a certain number, we choose with probability approaching one the model maximizing the posterior information criterion. Thus model selection via information criteria leads to choose with probability (approaching) one the model maximizing the posterior information criterion or minimizing the KLIC, and so to discard with probability (approaching) one the other competing model. Our approach instead leads to accept with probability one a specific hypothesis, H_0, when it is true and to reject it with probability one when instead H_A is true. In this sense our approach is complementary to the model selection approach; simply our procedure is preferable when we are interested in testing a specific hypothesis, but we want to be sure to choose (reject) H_0 with probability one when it's correct (false). Nevertheless there are cases, such as lag order selection, in which inference via $d_n||W_n||, d_n||LM_n||$ and via model selection leads, asymptotically, to the same answer.

Second, in classical hyphotesis testing, the joint effect of heterogeneity and mispecification is to lead to a conservative inference, in that $U^* > 0$ and so the statistic is "smaller" than it should be and the actual size of the test is smaller than the nominal size. Here instead when $U^* > 0$ instead of having a exact almost sure bound, we have an almost sure upper bound; however the asympotic size is zero regardless $U^* = 0$ or $U^* > 0$ and so we do not have a size problem connected with the joint effect of heterogeneity and mispecification.

Third, the zero size property of ther LIL based inference could be particularly useful in a sequential testing framework (e.g. lag order selection), in that situation classical tests suffer of severe size distorsion in both finite and large samples[2].

Finally, although we have considered only hypothesis testing for nested models, this framework is easily generalizable to test non-nested hypotheses and to Hausman type tests. As an example we may consider testing for the weakly exogeneity of X_t in $y_t = W_t'\beta_0 + \varepsilon_t$ where $W_t = (X_t, Z_t)$. Let $\hat{\beta}_n$ and $\tilde{\beta}_n$ be respectively the OLS and the IV estimator (using Z_t as instruments). Under H_0 both $d_n||\hat{\beta}_n - \beta_0||$ and $d_n||\tilde{\beta}_n - \beta_0||$ satisfy a LIL, so we can find an almost sure bound for $d_n||\hat{\beta}_n - \tilde{\beta}_n||$; on the other way under H_A, $d_n||\hat{\beta}_n - \beta_0||$ diverges while $d_n||\tilde{\beta}_n - \beta_0||$ satisfy a LIL, so we expect $d_n||\hat{\beta}_n - \tilde{\beta}_n||$ to diverge at rate d_n.

4 Hypothesis Testing in Presence of Nuisance Parameters

The computational burden associated with the calculation of the asymptotic critical values for testing in presence of unidentified nuisance parameters has been recently overcome by Stinch-

[2]Some montecarlo experimentes have been performed on the use of a LIL testing procedure in the chioce of order lag of an AR process in small sample. The result are rather mixed with the LIL approach performing better in term of size than PIC, BIC and classical sequantial tests in the case of low values of the roots of the characteristic polynomial of the underlying process, and the converse in the case of high values of the roots.

combe and White (1995) who obtain a LIL-based almost sure bound, in the iid framework, for conditional moment type test. Their result is derived from the LIL for Banach-valued iid random variables.

Here we derive almost sure bounds for $d_n\|W_n\|$ and $d_n\|LM_n\|$ via simpler tools that allow for both dependence and heterogeneity. In order to define the null and the alternative in a simple way, we rule out the joint presence of heterogeneity and mispecification, we can allow for either, but not simultaneously. Nevertheless we always allow for dynamic mispecification; broadly speaking in the heterogeneous case we require the model to be correctly specified for $E(y_t|X_t)$, but we do not require the model to be correctly specified for $E(y_t|F_{t-1})$, $F_t = \sigma(X_1, X_2, \ldots X_{t+1}, y_1, \ldots y_t)$. By allowing for dynamic mispecification, we do not require the score to be martingale difference, thus we need to use a HAC robust estimator for the covariance matrix, say $\hat{V}_n(\pi)$ and we shall show that $\sup_{\pi\in\Pi} |\hat{V}_n^{ij}(\pi) - V_n^{*ij}(\pi)| = o_{as}(1), \forall i, j$. This generalizes the result of Theorem 2 to the case in which the score depends on nuisance parameters.

Define the null as,

$$H_0 : h(\theta^*(\pi)) = h(\theta^*) = 0, \ \forall\pi \in \Pi$$

where $\Pi \in R^k$ and π are the nuisance parameters that are unidentified under the null. Define the alternative as

$$H_A : h(\theta^*(\pi)) \neq 0 \ \forall\pi \in \Pi^1$$

where $\mu(\Pi^1) > 0$, with μ being the Lebesgue measure on R^k; alternatively we can specify the alternative as

$$H_{A'} : h(\theta^*(\pi)) \neq 0 \ \forall\pi \in \Pi$$

except for a set of Lebesgue measure zero.

Thus we consider the case in which nuisance parameters are present only under the alternative. Let's define

$$\theta^*(\pi) = argmin_{\theta\in\Theta} E(q_t(X^t, \theta, \pi)),$$

and the sample analog

$$\hat{\theta}_n(\pi) = argmin_{\theta\in\Theta} \frac{1}{n} \sum_{t=1}^{n} q_t(X^t, \theta, \pi)$$

The restricted counterparts θ_n^0 and $\tilde{\theta}_n$ are defined as in the previous section. It is worthy to point out again that, under H_0, q_t and θ_n^0 do not depend on π, also under the null, $\theta_n^0 = \theta_0, \forall n$. Let's define

$$W_n(\pi) = \widehat{B}_n^{-1/2}(\pi)\, h(\widehat{\theta}_n(\pi))$$

where

$$\widehat{B}_n(\pi) = H_n(\widehat{\theta}_n(\pi))' \left(\nabla^2 Q_n(\widehat{\theta}_n(\pi), \pi)\right)^{-1} \widehat{V}_n \left(\nabla^2 Q_n(\widehat{\theta}_n(\pi), \pi)\right)^{-1} H_n(\widehat{\theta}_n(\pi))$$

and

$$\widehat{V}_n = n^{-1} \sum_{t=1}^{n} \nabla q_t(\hat{\theta}_n(\pi), \pi) \nabla q_t(\hat{\theta}_n(\pi), \pi)'$$

$$+n^{-1} \sum_{\tau=1}^{l_n} w_\tau \sum_{t=\tau+1}^{n} \left(\nabla q_t(\hat{\theta}_n(\pi), \pi) \nabla q_{t-\tau}(\hat{\theta}_n(\pi), \pi)' + \nabla q_{t-\tau}(\hat{\theta}_n(\pi), \pi) \nabla q_t(\hat{\theta}_n(\pi), \pi)'\right),$$

also let define

$$LM_n(\pi) = \widetilde{B}_n^{-1/2}(\pi) H(\widetilde{\theta}_n)' \left(\nabla^2 Q_n(\widetilde{\theta}_n, \pi)\right)^{-1} \nabla Q_n(\widetilde{\theta}_n, \pi)$$

with $\widetilde{B}_n$ and $\tilde{V}_n$ defined as $\widehat{B}_n$ and $\hat{V}_n$, replacing $\hat{\theta}_n(\pi)$ with $\tilde{\theta}_n$.

Before stating the main theorem, we need,

LEMMA TN1

Let A1,A2, ATN1-ATN3(a), ATN6 hold then

$$(i) \sup_{\pi \in \Pi} |\hat{\theta}_n(\pi) - \theta^*(\pi)| = o_{as}(1)$$

$$(ii) \sup_{\pi \in \Pi} |\hat{A}_n(\pi) - A_n^*(\pi)| = o_{as}(1)$$

where $\hat{A}_n(\pi) = n^{-1} \sum \nabla_\theta^2 q_t(\hat{\theta}_n(\pi), \pi)$ and $A_n^*(\pi) = n^{-1} \sum E(\nabla_\theta^2 q_t(\theta^*(\pi), \pi))$.

THEOREM TN1

(i) Let A1-A2, and ATN1-ATN11 hold. Then, under H_0,

$$\begin{aligned} (a)\ \limsup_{n\to\infty} \sup_{\pi\in\Pi} d_n \|W_n(\pi)\| &= 1 \\ (b)\ \limsup_{n\to\infty} \sup_{\pi\in\Pi} d_n \|LM_n(\pi)\| &= 1 \end{aligned}$$

almost surely.

(ii)(a) Let A1-A2, ATN1-ATN2, ATN3(a), ATN4, ATN6, ATN9-ATN11 hold. Then under H_A or under $H_{A'}$,

$$P[\lim_{n\to\infty} \sup_{\pi\in\Pi} d_n\|W_n(\pi)\| > c_n] = 1$$

(ii)(b) Let A1, A2, ATN1-ATN2, ATN3(b), ATN4, ATN11B hold,

$$P[\lim_{n\to\infty} \sup_{\pi\in\Pi} d_n\|LM_n(\pi)\| > c_n] = 1$$

with $c_n \times (\frac{n}{2\ln\ln n})^{-(\frac{1}{2}-\delta)} \to \infty$ as n$\to \infty$, with $\delta \in (0, 1/2)$.

About the assumptions of the theorem above, ATN1 is the usual regularity condition on the parameter space. Assumption ATN2 and ANT3 are the previous regularity conditions extended to the domain of π. ATN4 is a further smoothness requirement for the score, we infact require that the score be differentiable on Π. This assumption actually rules out some interesting applications, as testing for structural breaks or for threshold type nonlinearities. We discuss at which price ATN4 can be relaxed at the end of this session. ATN5 is satisfied whenever the score, evaluated at $\theta^*(\pi) = \theta^*$, under H_0, can be written in the form $\epsilon_t g(X^t, \theta^*, \pi))$, with $\epsilon_t = y_t - E(y_t|X_t)$. For example this is the case of conditional moment tests. We use ATN8 in order to show that $b_n^{-1} \sum \nabla_\theta q_t(\theta^*(\pi), \pi)$, with $b_n = \frac{1}{\sqrt{2n\ln\ln n}}$, is a.s. asymptotically uniform equicontinuous under H_0. ATN7, ATN9-ATN11 are required in order to show that, under both H_0 and H_A, $\sup_{\pi\in\Pi} |\hat{V}_n^{ij}(\pi) - V^{*ij}(\pi)| = o_{as}(1), \forall i, j$. Finally ATN11B guarantees the almost sure consistency of the variance of the score evaluated at $\tilde{\theta}_n$, under the alternative; we are not sure whether, by dropping ATN11B, we can still get uniform weak convergence, as in the non-nuisance parameter case.

Thus we reject (accept) H_0 according to whether $\sup_{\pi\in\Pi} \|W_n(\pi)\|, \sup_{\pi\in\Pi} \|LM_n(\pi)\|$ is greater (equal or smaller) than one. As in the nuisance parameter free case, this delivers a completely consistent procedure.

We mentioned above that ATN4 rules out important cases as for example testing for structural breaks, testing versus threshold alternatives. We now relax it. We can provide an almost sure bound for the same class of problems considered by Hansen (1996) and under analogous assumptions, but we allow for dynamic mispecification and so we do not require the score to be a martingale difference sequence. This can be seen as an easily and quickly computable alternative to the conditional p-value trasformation approach.

Note that ATN17 is the analogous of Assumption 2 in Hansen 1996, applied to the score evalatuated at $\theta^*(\pi)$.

THEOREM TN2

Let A2, ATN1-ATN4, ATN12-ATN19 hold.

(i) Under H_0,

$$\begin{aligned}(a)\ \limsup_{n\to\infty}\sup_{\pi\in\Pi} d_n\|W_n(\pi)\| &= 1\\ (b)\ \limsup_{n\to\infty}\sup_{\pi\in\Pi} d_n\|LM_n(\pi)\| &= 1\end{aligned}$$

almost surely.

(ii) Under H_A or under $H_{A'}$,

$$P[\lim_{n\to\infty}\sup_{\pi\in\Pi} d_n\|W_n(\pi)\| > c_n] = 1$$

$$P[\lim_{n\to\infty}\sup_{\pi\in\Pi} d_n\|LM_n(\pi)\| > c_n] = 1$$

with $c_n \times (\frac{n}{2\ln\ln n})^{-(\frac{1}{2}-\delta)} \to \infty$ as n$\to \infty$, with $\delta \in (0, 1/2)$.

The proof of Theorem TN2 above is very close to the proof of theorem 1 in Hansen 1996. In fact Hansen shows that the conditions for $n^{-1/2}\sum s(\gamma)\epsilon_t$ (in his notation) being asymptotically uniformly stochastic equicontinuous on Γ are satisfied; here instead we show that the condition for $b_n^{-1}\sum \nabla_\theta q(\theta^*(\pi), \pi)$ be almost surely asymptotically uniformly stochastic equicontinuous on Π are satisfied.

5 Applications and Small Sample Properties

Next we show by two limited monte carlo exercises how the testing procedure proposed in the previous sections behaves in finite sample. In the first one we shall consider the conditional moment test of functional form proposed by Bierens (1982, 1990). After that, in the line of Hansen (1996), we will look at the case of testing for the significance of the nonlinear term in a self-exciting threshold model (SETAR).

5.1 Testing of functional form of nonlinear regression model

Conditional moment specification tests are based on the property that the conditional expectation of certain functions (e.g. weighting functions) of the innovations and of the observations of the nonlinear regression model should be almost surely equal to zero if the model is correctly specified. The choice of the weighting function becomes crucial in term of power of the specification test, given that with a finite set of moment conditions imposed the test cannot be consistent in all the directions. Bierens (1982, 1990) overcame this point proposing a test which uses an infinite set of moment conditions and which is based on the following lemma:

Lemma 1 (Bierens 1990)

Let v be a random variable or vector satisfying $E|v| < \infty$ and let x be a bounded random vector in $\Re^k$ such that $P[E(v|x) = 0] < 1$. Then the set $S = \left\{\pi \in \Re^k : E[v \exp(\pi' x)] = 0\right\}$ has measure zero.

In the context of parameter regression analysis, it is assumed that the regression function $E(y|x)$ belongs to the parametric form of function $f(x, \phi)$ where $\phi \in \Phi$, the parameters space. The model is defined to be correctly specified if the data generating process is such that it exists a $\phi^0 \in \Phi$

$$P\left[E(y|x) = f(x, \phi^0)\right] = 1,$$

and on the contrary the model is mispecified respect to the underlying data generating process if

$$P[E(y|x) = f(x, \phi)] < 1$$

for every $\phi \in \Phi$. From Bierens' lemma and the definition of mispecification it follows that if the model is mispecified the set

$$S = \left\{\pi \in \Re^k : E[(y - f(x, \phi)) \exp(\pi' g(x))] = 0\right\}$$

has measure zero, where g is a measurable bounded one to one function.

To perform a conditional moment test we use the following statistic

$$\widehat{M}_n(\pi) = \frac{1}{n} \sum \left(y_t - f(x_t, \widehat{\phi})\right) \exp(\pi' g(x_t))$$

where $\widehat{\phi}$ is the $\arg\min_{\phi} \frac{1}{n} \sum (y - f(x,\phi))^2$. Under the null $\hat{M}_n(\pi)$, for every π converges to a normal with zero mean and variance $s(\pi)$.

In Bierens (1990), it was proposed the following specification of the test

$$\sup_{\pi} \left(n \frac{\widehat{M}_n(\pi)^2}{s(\pi)} \right)$$

which under the null converges to $sup_{\pi}(z(\pi)^2)$ where $z(\pi)$ is a gaussian element with variance covariance $\Gamma(\pi)$ which depends on the model and on the distribution of (y, x). Inference based on the statistic above requires the computation of the critical values. To this purpose, Bierens provided also an approximation procedure that alleviates the computational burden implied by the exact computation of the critical values.

Thus conditional moment specification test casts exactly in our framework given that the Bierens' conditional moment test is equivalent of testing for the null of $\gamma = 0$ in the following nonlinear regression model

$$y_t = f(x_t, \phi) + \gamma \exp\left(\pi' g\left(x_t\right)\right) + \varepsilon_t$$

and so, once the nuisance parameters π has been assumed to belong an hypercube $\Pi \subset \Re^k$, the conditional moment can be performed in the line of the Lagrangian Multipliers test as it has been proposed in the previous section with a clear gain in term of computational requirement. Following the notation of section 4, it is easy to see that $\theta = \{\phi, \gamma\}$ and the restricted estimator is $\widehat{\theta} = \left\{\widehat{\phi}, 0\right\}$, while the score evaluated at the restricted estimator reduces to

$$\nabla_{\theta} q_t(\widehat{\theta}, \pi) = \left\{ \begin{array}{c} 0 \\ \left(y_t - f(x_t, \widehat{\phi})\right) \exp\left(\pi' g\left(x_t\right)\right) \end{array} \right. .$$

In the following, we will look at the performance of the LIL based conditional moment test in some numerical example; to this purpose we will use a variation of the example of Bierens (1982, 1990). The data generating process we designed, is the following. Let the regressors x_{1t} and x_{2t} be autoregressive of order one with coefficient α_1 and α_2 respectively and innovation u_{1t} and u_{1t}, which are independent draws from the standard normal distribution. The dependent variable is generated according to $y_t = 1 + x_{1t} + x_{2t} + v_t$ where v_t is either $v_t = u_{1t} u_{1t} + \sigma \varepsilon_t$, with ε_t standard normal , or $v_t = \sqrt{1+\sigma^2}\varepsilon_t$, we call them as DGP1 and DGP2 respectively. The regression model fitted is

$$y_t = \phi_0 + \phi x_{1t} + \phi_2 x_{2t} + v_t$$

and we test the hypothesis that it is correctly specified.

Clearly the model is mispecified under DGP1, while it is correctly specified under DGP2; note in any case that under the two data generating processes the probability limit of the regression coefficient is the same, e.g. (1,1,1), and that the variance of the innovation v_t is equal to $(1+\sigma^2)$ in both cases. The weighting function g has been specified according to Bierens (1990) as

$$\exp(\pi g\,(x_t)) = \prod_{i=1}^{2} \exp(\pi_i \times \operatorname{atan}\left(\frac{(x_{it} - \overline{x}_i)}{2s_i}\right))$$

where $\overline{x}_i$ and s_i are the mean and the standard deviation of x_{it}, $i = 1,2$. The nuisance parameter π belongs to the hypercube $[0,4] \times [0,4]$ and the Lagrange Multiplier statistics has been evaluated on a grid of 225, (15×15), points.

We performed the monte carlo exercise for different specifications of (α_1, α_2)=$\{(.4, .4)$, $(.4, -.4)$, $(.4, .8)\}$ and for different values of the variance $\sigma = \{.5, 1, 2\}$. In particular, it is interesting to see how in the DGP2 the change in the relative variance of the two component of v can affect the power of the test. The monte carlo simulation has been conducted on a sample of size 100, 200 and 400, performing 500 replication using common number generator. The small sample size of the LIL based testing procedure in the case of correctly specified model are in table 1, while the small sample power, when the true model is a DGP1, is in table 2.

Table 1 - Small Sample Size of the LM statistic

α_1, α_2	σ	$T = 100$	$T = 200$	$T = 400$
(.4, .4)	.5	0.28	0.10	0.06
(.4, .4)	1	0.28	0.09	0.06
(.4, .4)	2	0.28	0.11	0.07
(.4, −.4)	.5	0.30	0.09	0.07
(.4, −.4)	1	0.31	0.10	0.08
(.4, −.4)	2	0.30	0.09	0.07
(.4, .8)	.5	0.20	0.12	0.10
(.4, .8)	1	0.20	0.12	0.10
(.4, .8)	2	0.20	0.12	0.09

Table 2 - Small Sample Power of the LM statistic

α_1, α_2	σ	$T = 100$	$T = 200$	$T = 400$
(.4, .4)	.5	0.98	0.99	1.00
(.4, .4)	1	0.97	0.99	1.00
(.4, .4)	2	0.96	0.98	1.00
(.4, −.4)	.5	0.99	1.00	1.00
(.4, −.4)	1	0.99	1.00	1.00
(.4, −.4)	2	0.97	0.99	1.00
(.4, .8)	.5	0.92	0.98	1.00
(.4, .8)	1	0.85	0.95	0.98
(.4, .8)	2	0.84	0.93	0.97

The results of the exercise show that the size approaches zero rather slowly, for example a sample of two hundred observations the size around ten percent. On the contrary the power goes to one very quickly; it is also evident that the change in the ratio between the two component of the innovation of DGP1 has an effect on the small sample power. In fact the variance σ seems to be inversely related with the power of the test, given that the relative importance of the source of mispecification is "reduced" for large σ. More recently Bierens and Ploberger (1997) proposed the following integrated conditional moment test $\int \left(\frac{n\hat{M}_n^2(\pi)}{s(\pi)} \right) \mu(d\pi)$, that has particularly appealing properties in terms of power towards sequence of local alternatives; while we conjecture that an almost sure bound can be provided also for that statistic, Theorem TN1 does not straightforwardly apply.

5.2 Testing nonlinearity in SETAR model

Threshold autoregressive models have the appealing feature of capturing the possible different dynamics of macro time series in different phases of the economic cycle, see Potter (1995) and Altissimo and Violante (1996). As the threshold parameters is present only under the alternative, testing for the significance of the nonlinear term falls into the class of hypothesis testing with nuisance parameters unidentified under the null. Two possible alternatives have been proposed in the literature: the first is to bootstrap the distribution of the statistics under the null and the second is to apply the p-value transformation approach of Hansen (1996). Here we show by a numerical example how the LIL inference procedure, based on the result of theorem TN2, can handle this case.

Consider the following self-exciting threshold model (SETAR) model

$$\begin{aligned} y_t &= \alpha_0 + \alpha_1 y_{t-1} + \ldots + \alpha_p y_{t-p} + \\ &\quad + (\beta_0 + \beta_1 y_{t-1} + \ldots + \beta_p y_{t-p}) \, 1\{y_{t-d} \leq \gamma\} + \varepsilon_t. \end{aligned} \tag{4}$$

where γ is the threshold parameters and d is the lag order at which the switch operates. We are going to test the significance of the nonlinear term. The null is $H_0 : \beta_0 = \ldots = \beta_p = 0$ and so the parameter $\{\gamma\}$ is not identified.

Under the null, if the innovation is iid, all the roots of the characteristic equation are strictly inside the unit circle, $E\left|\varepsilon_t\right|^{4p} < \infty$ for some $p > 1$ and the density of the innovation is bounded and continuous then the conditions for the theorem (3.1) of Pham and Tram (1985) are satisfied and the $\{y_t\}$ generated by (4) is strictly stationary process with $\beta-$mixing coefficient exponentially declining and satisfying ATN13. So under the null the assumption for theorem TN2 (i) are satisfied. Nothing we can say about TN2 (ii) given that it is not clear if the SETAR process is an absolutely regular process. In spite of that we will perform the monte carlo exercise both for the size and for the power of the LIL inference procedure[3].

The monte carlo exercise has been structured in the following way. The process in (4) has been specified to be of order one, $p = 1$, and the nonlinear factor operates at one lag, $d = 1$. The innovation ε_t has been assumed to be independent drawing from the standard normal distribution and for the autoregressive coefficient two different specifications have been chosen, $\alpha_1 = \{.4, \ .8\}$.

The LIL inference procedure consists in the evaluation of the statistics, $d_n\|W_n(\pi)\|$ and $d_n\|LM_n(\pi)\|$ described in section 4, over a grid of points in the space of the nuisance parameters and then consider the sup of the statistics over the grid of point. The interval of the nuisance parameter has been set to $[-1, 1]$ for the first specification of α_1 and to $[-1.5, 1.5]$ in the second one, and a grid of forty points has been used.

We consider samples of 100, 200 and 400 observations and in each monte carlo exercise we use 500 replications. The following table shows the small sample size of the LIL based tests, *supW* and *supLM*, for the absence of threshold effect in the SETAR model.

Table 3 - Small Sample Size

α_1		T=100	T=200	T=400
.5	*supW*	0.32	0.16	0.10
.5	*supLM*	0.18	0.12	0.09
.8	*supW*	0.25	0.12	0.10
.8	*supLM*	0.15	0.10	0.07

As in the conditional moment test the size of the test is approaching zero rather slowly; we can also note that the increase of the persistence of the series "reduces" the finite size.

The power property of the test has been assessed by looking to various possible modifications of the linear dynamics due to the threshold effect both in term of intercept modification and

[3]It is the case to note that in the case of SETAR of order one with two state Petrucelli and Woodford (1984) have proved that iff the roots in the two states are strictly less than one and their product is also strictly less than one, then the model is geometrically ergodic. In our exercise the specifications of the alternative satisfy those conditions.

in term of slope. We consider the following parameterizations: ($\beta_0 = 1, 1.5, \beta_1 = 0$), and ($\beta_0 = 0, \beta_1 = -0.5, -1$); in all the cases the threshold parameter γ has been set equal to zero.

In the following tables we reported the results for the small sample power properties of the two statistics, for the case of low and high persistence respectively.

Table 4 - Small Sample Power for $\alpha_1 = 0.5$

		T=100	T=200	T=400
$\beta_0 = 1$	$supW$	0.87	1.00	1.00
$\beta_0 = 1$	$\sup LM$	0.82	0.99	1.00
$\beta_0 = 1.5$	$supW$	0.97	1.00	1.00
$\beta_0 = 1.5$	$\sup LM$	0.97	1.00	1.00
$\beta_1 = -.5$	$supW$	0.68	0.93	0.98
$\beta_1 = -.5$	$\sup LM$	0.65	0.92	0.95
$\beta_1 = -1$	$supW$	0.97	1.00	1.00
$\beta_1 = -1$	$\sup LM$	0.97	1.00	1.00

Table 5 - Small Sample Power for $\alpha_1 = 0.8$

		T=100	T=200	T=400
$\beta_0 = 1$	$\sup W$	0.67	0.95	1.00
$\beta_0 = 1$	$\sup LM$	0.65	0.91	1.00
$\beta_0 = 1.5$	$\sup W$	0.85	0.97	1.00
$\beta_0 = 1.5$	$\sup LM$	0.79	0.96	1.00
$\beta_1 = -.5$	$\sup W$	0.70	0.94	0.98
$\beta_1 = -.5$	$\sup LM$	0.69	0.92	0.96
$\beta_1 = -1$	$\sup W$	0.98	1.00	1.00
$\beta_1 = -1$	$\sup LM$	0.98	1.00	1.00

Similarly to the case of conditional moment specification tests the small sample power of the LIL based inference procedure is quite satisfactory and, as expected, it is increasing with the "distance" of the alternative and with the sample size.

On this ground the proposed procedure presents the opposite feature of the Hansen's (1996) one, which has a good small sample size but a rather low power. This fact makes the two procedures complementary and the choice between them has to be done not only in function of the different computational burden but also in term of the choice of being conservative towards the null of the test. In practice if we obtain a value below or equal to one, we can "safely" decide in favor of the null; if instead we get a value above one, we may want to use Hansen's approach too.

6 Appendix 1

A1 - X_t, $t = 1, 2, \ldots n$ is a R^d-valued α-mixing process of size $-2r/(r-2)$, with $r > 2$.

A2 - $\theta \in \Theta$ compact in R^p.

A3 - $q_t(\theta)$ is twice continuously differentiable in the interior of Θ.

A4 - (i) $q_t(\theta)$, is r-dominated on Θ uniformly in t, (ii) $\nabla_\theta q_t(\theta)$ is r-dominated on Θ uniformly in t, (iii) $\nabla_\theta^2 q_t(\theta)$ is r-dominated on θ uniformly in t; $r > 2$.

A5 - (i), (ii), (iii) same as in A4, with r-domination replaced by $2r$-domination.

Hereafter NED is defined as in Definition 3.13 in Gallant White (1988, p.32, hereafter GW).

A6 - (i) $q_t(\theta)$ is L_2-NED on X_t with size -1, uniformly in (Θ, ρ),(ii) $\nabla q_t(\theta)$ is L_2-NED on X_t with size -1, uniformly in (Θ, ρ), (iii) $\nabla^2 q_t(\theta)$ is L_2- NED on X_t with size -1, uniformly in (Θ, ρ), and ρ denotes a proper metric on Θ.

A7 - (i)-(iii) same as in A6, but L_2-NED on X_t with size $-2(r-1)/(r-2)$,$r > 2$.

Let

$$\theta_n^* = argmin_{\theta \in \Theta} \frac{1}{n} \sum E(q_t(X^t, \theta))$$

where $X^t = (X_0, X_1, \ldots X_t)$.

A8 - The sequence θ_n^* is uniquely identifiable on Θ as in definition 3.2 of Gallant and White (1988), uniformly in n. Also θ_n^* is an interior of Θ, $\forall n$.

Hereafter the a.s. Lipschitz condition is defined as in Andrews (1992, p.248, assumption S-LIP)

A9 - (i) $q_t(\theta)$ is a.s. Lipschitz on Θ, (ii) $\nabla_\theta q_t(\theta)$ is a.s. Lipschitz on Θ, (iii) $\nabla^2_\theta q_t(\theta)$ is a.s. Lipschitz on Θ, (iv) the elements of $(\nabla^2_\theta q_t(\theta))^2$ are a.s. Lipschitz on Θ.

A10 - $n^{-1}\sum E(\nabla^2 q_t(\theta)) = A^*_n(\theta)$ is uniformly positive definite in Θ and there exists

$$A^*(\theta) = \lim_{n\to\infty} A^*_n(\theta), \forall\theta \in \Theta$$

A11 - There exist $\theta^* \in R^p$, such that as $n \to \infty$

$$\theta^*_{n,m} \to \theta^*$$

uniformly in m, where

$$\theta^*_{n,m} = argmin_{\theta\in\Theta} n^{-1} \sum_{t=m+1}^{m+n} E(q_t(X^t, \theta))$$

and $\theta^*_n = \theta_{n,0}$.

Hereafter, for notational convenience, we suppress X^t from the argument of q_t. Let

$$S_n(m) = \sum_{t=m+1}^{m+n} (\nabla q_t(\theta^*_{n,m}) - E(\nabla q_t(\theta^*_{n,m})) = \sum_{t=m+1}^{m+n} (\nabla \bar{q}_t(\theta^*_{n,m}))$$

and let $S^i_n(m)$ be the $i-th$ element of the R^p-valued vector $S_n(m)$.

A12 - There exists a positive definite matrix $V^* = [v^*_{ij}]$, $i, j = 1, 2, \ldots p$ with $v_{ji} \leq \Delta, \forall i, j$ and such that

$$|n^{-1}E(S^i_n(m)S^j_n(m)) - v^*_{i,j}| \leq \Delta_1 n^{-\rho}$$

for $\rho > 0$, and Δ_1 independent of m.

Let $F_m = \sigma(X_1, X_2, \ldots X_m)$,

A13 -

$$E|n^{-1}E(S^i_n(m)S^j_n(m)|F_m) - n^{-1}E(S^i_n(m)S^j_n(m))| \leq \Delta_2 n^{-\eta}$$

with $\eta > 0$, and Δ_2 independent of m.

Define

$$Z_{t,\tau}(\theta^*_{n,m}) = \left[\nabla q_t(\theta^*_{n,m})\nabla q_{t+\tau}(\theta^*_{n,m})' - E\left(\nabla q_t(\theta^*_{n,m})\nabla q_{t+\tau}(\theta_{n,m})'\right)\right]$$
$$Y_{t,\tau}(\theta^*_{n,m}) = \left[\nabla^2 q_t(\theta^*_{n,m}) \otimes \nabla q_{t+\tau}(\theta^*_{n,m})' - E\left(\nabla^2 q_t(\theta^*_{n,m}) \otimes \nabla q_{t+\tau}(\theta_{n,m})'\right)\right]$$
$$W_{t,\tau}(\theta^*_{n,m}) = \left[\nabla^2 q_t(\theta^*_{n,m}) \otimes \nabla^2 q_{t+\tau}(\theta^*_{n,m})' - E\left(\nabla^2 q_t(\theta^*_{n,m}) \otimes \nabla^2 q_{t+\tau}(\theta_{n,m})'\right)\right].$$

Let $Z^{i,j}_{t,\tau}(\theta^*_{n,m})$ the the i,j-th element of $Z_{t,\tau}(\theta^*_{n,m})$, for $\tau = 1, 2, \dots l_n$ and $i, j = 1, 2, \dots p$.

A14 - There exists $0 < \delta < \omega^{\tau}_{1,ij} < \delta^{-1}$, $\forall i, j$, and δ independent of τ, such that

$$|n^{-1}E\left(\sum_{t=m+1}^{m+n} Z^{i,j}_{t,\tau}(\theta_{n,m})\right)^2 - \omega^{\tau}_{1,ij}| \leq C_1 n^{-\rho}$$

with $\rho > 0$ and C_1 independent of m and τ.

A15 - $\forall i, j = 1, 2, \dots p$,

$$E|n^{-1}E\left(\left(\sum_{t=m+1}^{m+n} Z^{i,j}_{t,\tau}(\theta^*_{n,m})\right)^2 |F_m\right) - n^{-1}E\left(\sum_{t=m+1}^{m+n} Z^{i,j}_{t,\tau}(\theta^*_{n,m})\right)^2 | \leq C_2 n^{-\eta}$$

with $\eta > 0$ and C_2 independent of m and τ.

Let $Y^{i,j}_{t,\tau}(\theta^*_{n,m})$ the i,j-th element of $Y_{t,\tau}(\theta^*_{n,m})$, for $\tau = 1, 2, \dots l_n$ and $i = 1, 2, \dots p, j = 1, 2 \dots p^2$.

A16 - There exists $0 < \delta < \omega^{\tau}_{2,ij} < \delta^{-1}$ $\forall i, j$, and δ independent of τ such that

$$|n^{-1}E\left(\sum_{t=m+1}^{m+n} Y^{i,j}_{t,\tau}(\theta_{n,m})\right)^2 - \omega^{\tau}_{2,ij}| \leq C_3 n^{-\rho}$$

with $\rho > 0$ and C_3 independent of m and τ.

A17 - For every $i, j = 1, 2, \dots p^2$,

$$E|n^{-1}E\left(\left(\sum_{t=m+1}^{m+n} Y_{t,\tau}^{i,j}(\theta_{n,m}^*)\right)^2 |F_m\right) - n^{-1}E\left(\sum_{t=m+1}^{m+n} Y_{t,\tau}^{i,j}(\theta_{n,m}^*)\right)^2 | \leq C_4 n^{-\eta}$$

with $\eta > 0$ and C_4 independent of m and τ.

Let $W_{t,\tau}^{i,j}(\theta_{n,m}^*)$ the i,j-th element of $W_{t,\tau}(\theta_{n,m}^*)$, for $\tau = 1, 2, \dots l_n$ and $i, j = 1, 2, \dots p^2$.

A18 - There exists $0 < \delta < \omega_{3,ij}^{\tau} < \delta^{-1}$, $\forall i, j$ and δ independent of τ, such that

$$|n^{-1}\left(\sum_{t=m+1}^{m+n} W_{t,\tau}^{i,j}(\theta_{n,m})\right)^2 - \omega_{3,ij}^{\tau}| \leq C_5 n^{-\rho}$$

with $\rho > 0$ and C_5 independent of m and τ.

A19 -

$$|n^{-1}E\left(\left(\sum_{t=m+1}^{m+n} W_{t,\tau}^{i,j}(\theta_{n,m}^*)\right)^2 |F_m\right) - n^{-1}E\left(\sum_{t=m+1}^{m+n} W_{t,\tau}^{i,j}(\theta_{n,m}^*)\right)^2 | \leq C_6 n^{-\eta}$$

with $\eta > 0$ and C_6 independent of m and τ.

A20 - Let $w_\tau = 1 - \frac{\tau}{l_n+\tau}$ and $l_n = o(\sqrt{\frac{n}{2\log\log(n)}})$. Define

$$U_n^* = n^{-1}\sum_t^n E(\nabla q_t(\theta_n^*))E(\nabla q_t(\theta_n^*)')+$$

$$n^{-1}\sum_\tau^{l_n} w_\tau \sum_{t=\tau+1}^{n} E(\nabla q_t(\theta_n^*))E(\nabla q_{t-\tau}(\theta_n^*)\prime) + E(\nabla q_{t-\tau}(\theta_n^*))E(\nabla q_t(\theta_n^*)')$$

U_n^* is positive definite and $\lim_{n\to\infty} U_n^* = U^*$, where U^* has bounded trace.

AT1 - $h : \Theta \to \Re^q$ is twice continuosly differentiable in the interior of Θ, for all n, and its Jacobian $H_n(\cdot)$ has full column rank, uniformly in n.

AT2 - A8 is satisfied under both the null and the alternative. θ_n^0 is in the interior of Θ_0, and A20 holds with θ_n^* replaced by θ_n^0.

ATN1 - $\pi \in \Pi$ compact in $\Re^k$.

ATN2 - A3-A7 and A9 hold on $\Theta \times \Pi$. Also $h : \Theta \times \Pi \rightarrow R^q$ is twice continuosly differentiable in the interior of Θ, and its Jacobian $H_n(\cdot)$ has full column rank uniformly in n and π.

ATN3 - (a) $\theta^*(\pi)$ is uniquely identifiable in the sense that

$$\inf_{\pi \in \Pi} \inf_{\theta \in \eta^c(\theta^*(\pi))} |n^{-1} \sum E(q_t(\theta, \pi)) - n^{-1} \sum E(q_t(\theta^*(\pi), \pi)| > 0$$

under both H_A and H_0, where

$$\eta_{\theta_\pi} = [\theta : |\theta - \theta^*(\pi)| < \delta]$$

with δ independent of π.
(b) θ_n^0 as defined in section 3 is uniquely identifiable, under both H_A and H_0.

ATN4 - For all t, $\nabla_\theta q_t(\theta^*(\pi), \pi)$ is a continuously differentiable function of π in the interior of Π.

ATN5 - Under H_0 the elements of $E\left(\nabla_\pi \left(\nabla_\theta q_t(\theta^*, \pi)\right)\right)$ are equal to zero, $\forall \pi \in \Pi$.

ATN6 - $n^{-1} \sum E(\nabla_\theta^2 q_t(\theta, \pi)) = A_n^*(\theta, \pi)$ is uniformly positive definite over $\Theta \times \Pi$; and $\lim_{n \rightarrow \infty} A_n(\theta, \pi) = A^*(\theta, \pi)$

In the definition of $S_n(m), Z_{t,\tau}(\theta_{n,m}^*), Y_{t,\tau}(\theta_{n,m}^*), W_{t,\tau}(\theta_{n,m}^*)$, given before A14, replace $q_t(\theta_{n,m}^*)$ with $q_t(\theta^*(\pi), \pi)$, and correspondingly define $S_n(m, \pi), Z_{t,\tau}(\theta^*(\pi), \pi), Y_{t,\tau}(\theta^*(\pi), \pi), W_{t,\tau}(\theta^*(\pi), \pi)$.

ATN7 - The elements of $Z_{t,\tau}(\theta^*(\pi), \pi), Y_{t,\tau}(\theta^*(\pi), \pi), W_{t,\tau}(\theta^*(\pi), \pi)$ satisfy, $\forall \pi \in \Pi$, respectively, A14-A15, A16-A17, A18-A19, with $\omega_{h,\tau}^{ij}$ replaced by $0 < \delta < \omega_{h,\tau}^{ij}(\pi) < \delta^{-1}$, $h = 1, 2, 3$, δ

independent of π, and $C_i, i = 1, \ldots 6$ independent of π. Also $S_n(m, \pi)$, $\forall \pi \in \Pi$ satisfy A12-A13 with v_{ij}^* replaced by $0 < \delta < v^*(\pi) < \delta^{-1}$, with δ and $\Delta_i, i = 1, 2$ independent of π.

ATN8 - The elements of $\nabla_\pi S_n(m, \pi)$ satisfy A12-A13, $\forall \pi \in \Pi$, with v_{ij}^* replaced by $0 < \delta < v_{ij}^{**}(\pi) < \delta^{-1}$, with δ and $\Delta_i, i = 1, 2$ independent of π.

ATN9 - $Z_{t,\tau}^{i,j}(\theta^*(\pi), \pi)$, $Y_{t,\tau}^{i,j}((\theta^*(\pi), \pi), W_{t,\tau}^{i,j}(\theta^*(\pi), \pi)$, where i, j denotes the $ij - th$ elements of the corresponding matrix, are continuously differentiable in the interior of Π, $\forall i, j$. Let's call the respective gradients $DZ_{t,\tau}^{i,j}(\theta^*(\pi), \pi)$, $DY_{t,\tau}^{i,j}(\theta^*(\pi), \pi)$, $DW_{t,\tau}^{i,j}(\theta^*(\pi), \pi)$.

ATN10 - For every i, j the elements of $DZ_{t,\tau}^{i,j}(\theta^*(\pi), \pi), DY_{t,\tau}^{ij}(\theta^*(\pi), \pi), DW_{t,\tau}^{i,j}(\theta^*(\pi), \pi)$ satisfy respectively A14-15, A16-A17, A18-A19, $\forall \pi \in \Pi$, with $\omega_{h,\tau}$ replaced by $0 < \delta < \omega_{h,\tau}^*(\pi) < \delta^{-1}$, $h = 1, 2, 3$, with δ and $C_i, i = 1, \ldots 6$ independent of π.

ATN11 - $\forall i, j$, the element of the R^k-valued vectors $DZ_{t,\tau}^{ij}(\theta^*(\pi), \pi)$, $DY_{t,\tau}^{i,j}(\theta^*(\pi), \pi)$ and $DW_{t,\tau}^{i,j}(\theta^*(\pi), \pi)$ are $2r$-dominated $\forall \pi \in \Pi$.

ATN11B - ATN7, ATN9-ATN11 hold with θ^* replaced by

$$\theta_{n,m}^0 = \underset{\theta \in \Theta s.t. h(\theta)=0}{\arg\min} \quad n^{-1} \sum_{t=m+1}^{m+n} Eq_t(\theta).$$

ATN12 - X_t is a strictly stationary β-mixing sequence, with coefficient β_k defined as

$$\beta_k = \frac{1}{2} \sup \left[\sum_{i=1}^{I} \sum_{j=1}^{J} |P(A_i \cap B_j) - P(A_i)P(B_j)| \right]$$

where $[A_i]_{i=1}^I$ is a partition of $\sigma(X_1, \ldots X_l)$ and $[B]_{j=1}^J$ is a partition of $\sigma(X_{k+l}, X_{k+l+1} \ldots)$.

ATN13 - $\sum_{k=0}^{\infty} \beta_k k^{2/(p-2)} (\log \log k) < \infty$, with $p > 2$.

ATN14 - $q(X^t, \theta, \pi) = q(X_t, X_{t-1}, \ldots X_{t-q}, \theta, \pi)$, with q finite.

ATN15 - A3-A5 hold over $\Theta \times \Pi$, with Θ and π compact in R^p and R^k respectively.

ATN16 - $\sup_{\pi \in \Pi} \nabla_\theta q(X^t, \theta^*(\pi)) \in L^p$, $p > 2$.

For a generic matrix A, let $\|A\|_p = (E(X^p))^{1/p}$, and $|A| = (tr(A\prime A))^{1/2}$

ATN17 - $\|\nabla_\theta q^j(\theta^*(\pi), \pi) - \nabla_\theta q^j(\theta^*(\pi'), \pi')\|_l \leq B|\pi - \pi\prime|^\nu$, $\forall j = 1, 2 \ldots p$, $l > 2$, $\nu > 0$, $B < \infty$, $\forall \pi, \pi' \in \Pi$.

ATN18 - Let $Z^{ij}_{t,\tau}(\pi)$ be defined as in ATN7, $\forall i, j, \forall t, \tau$,

$$\left(E \left(\sup_{\pi \in \Pi} |Z^{i,j}_{t,\tau}(\pi)| \right)^p \right)^{1/2} \leq \Delta$$

with $p > 2$, $0 < \Delta < \infty$ independent of π and τ.

ATN19 - Under the null, $\forall i, j, \forall t, \tau$,

$$\|Z^{i,j}_{t,\tau}(\theta^*, \pi) - Z^{i,j}_{t,\tau}(\theta^*, \pi')\|_l \leq B|\pi - \pi\prime|^\nu$$

with $l > 2$, $\nu > 0$.

7 Appendix 2

PROOF OF THEOREM 1 : Unless otherwise specified, all summations are from 1 to n.

$$n^{-1}\sum \nabla_\theta q_t(\hat{\theta}_n) = n^{-1}\sum \nabla_\theta q_t(\theta_n^*) + n^{-1}\sum \nabla_\theta^2 q_t(\bar{\theta}_n)(\hat{\theta}_n - \theta_n^*) \quad (5)$$

where $\bar{\theta}_n \in (\hat{\theta}_n, \theta_n^*)$, and $q_t(\theta) = q_t(\theta, X^t)$, with $X^t = (X_1, X_2, \ldots X_t)$. From (5) we have

$$(\hat{\theta}_n - \theta_n^*) = -(n^{-1}\sum \nabla_\theta^2 q_t(\bar{\theta}_n))^{-1}(n^{-1}\sum \nabla_\theta q_t(\theta_n^*))$$

Let $d_n = \sqrt{\frac{n}{2loglog(n)}}$, and $b_n = \sqrt{2n\ loglog(n)}$, we have

$$d_n(\hat{\theta}_n - \theta_n^*) = -(n^{-1}\sum E(\nabla_\theta^2 q_t(\theta_n^*)))^{-1}(\frac{1}{b_n}\sum \nabla_\theta q_t(\theta_n^*))$$

$$-\left((n^{-1}\sum \nabla_\theta^2 q_t(\bar{\theta}))^{-1} - (n^{-1}\sum E(\nabla_\theta^2 q_t(\theta_n^*)))^{-1}\right)\frac{1}{b_n}\sum \nabla_\theta q_t(\theta_n^*)$$

Let

$$A_n^*(\bar{\theta}_n) = n^{-1}\sum E(\nabla_\theta^2 q_t(\bar{\theta}_n))$$

$$\hat{A}_n(\bar{\theta}_n) = n^{-1}\sum \nabla_\theta^2 q_t(\bar{\theta}_n)$$

and

$$A_n^* = n^{-1}\sum E(\nabla_\theta^2 q_t(\theta_n^*))$$

Given A1-A4, A6, A8, A9(ii)-(iii), by theorem 6.1 in Gallant and White we have

$$\hat{A}_n(\bar{\theta}_n) - A_n^* = (\hat{A}_n(\bar{\theta}_n) - A_n^*(\bar{\theta}_n)) + (A_n^*(\bar{\theta}_n) - A_n^*) = o_{as}(1)$$

It remains to show that,

$$\frac{1}{b_n}\sum \nabla_\theta \bar{q}_t(\theta_n^*) = \frac{1}{b_n}\sum(\nabla_\theta q_t(\theta_n^*) - E(\nabla_\nabla q_t(\theta_n^*))) = O_{as}(1)$$

Basically we shall show that the expression on the left hand side above satisfies Eberlain's (1986) strong invariance principle; the desired outcome will then follow easily. We begin by showing that condition (1.3) in Eberlain is satisfied. Now

$$\|E\left(\sum_{t=m+1}^{m+n} \nabla \bar{q}_t(\theta_{n,m}^*)|F_m\right)\|_2 \leq \|E\left(\nabla \bar{q}_{m+1}(\theta_{n,m}^*)|F_m\right)\|_2$$

$$+\ldots+\|E\left(\nabla\bar{q}_{m+k}(\theta^*_{n,m})|F_m\right)\|_2+\ldots+\|E\left(\nabla\bar{q}_{m+n}(\theta^*_{n,m})|F_m\right)\|_2$$

$$\leq c_{n,m+1}\psi_1+\ldots c_{n,m+k}\psi_k+\ldots+c_{n,m+n}\psi_n\leq\Delta\sum_{j=1}^{\infty}\psi_j\leq\tilde{\Delta},$$

with $\|q_t\|_r=(\sum_{i=1}^d E(q_t^i)^r)^{1/r}$. The first inequality comes from Minkowski inequality, the second and third inequalities, given A6(ii), comes from the fact that $\nabla\bar{q}_t(\theta^*_{n,m})$ is a mixingale of size -1 and constant $c_{n,t}=Max(\|\nabla\bar{q}_t(\theta^*_{n,m})\|_r,1)$ (see lemma 3.14 in Gallant and White (1988)) and $\|\nabla\bar{q}_t(\theta^*_{k,m})\|_r<\Delta$, uniformly in t, because of A4(ii).

Thus condition 1.3 in Eberlain is satisfied and given A1, A4(ii), A6(ii), A11-A13, from his theorem 1 we have that

$$\frac{1}{b_n}\|\sum_{t=1}^{[s]}\nabla\bar{q}_t(\theta^*_n)-B_s\|=o_{as}(1),$$

for $s=nr$, and $r\in[0,1]$, where B is a Brownian motion with variance equal to V^*, as defined in A12. By taking $r=1$, it then follows that

$$\limsup_{n\to\infty}\frac{1}{b_n}\|\sum_{t=1}^{n}\nabla\bar{q}_t(\theta^*_n)\|=\limsup_{s\to\infty}\frac{1}{b_s}\|B_s\|+o_{as}(1),$$

the first term on the RHS above is equal to the square root of the largest eigenvalue of V^*, which is bounded by A12. Thus the left hand side is $O_{as}(1)$. Thus we have

$$\sqrt{\frac{n}{2loglog(n)}}\|\hat{\theta}_n-\theta^*_n\|=\frac{1}{b_n}\|A_n^{*-1}\sum\nabla\bar{q}_t(\theta^*_n)\|+o_{as}(1)=O_{as}(1)$$

□.

PROOF OF THEOREM 2

For notational simplicity we denote the jacobian of $q_t(X^t,\theta)$ with respect to θ, as $\nabla q_t(\theta)$, by dropping the subscript θ. By the mean value expansion of the score around θ^*_n, we have

$$\hat{V}_n=n^{-1}\sum\nabla q_t(\theta^*_n)\nabla q_t(\theta^*_n)' \tag{6}$$

$$+n^{-1}\sum_{\tau=1}^{l_n}w_\tau\sum_{t=1}^{n-\tau}\left(\nabla q_{t+\tau}(\theta^*_n)\nabla q_t(\theta^*_n)'+\nabla q_t(\theta^*_n)\nabla q_{t+\tau}(\theta^*_n)'\right)$$

$$+n^{-1}\sum\nabla^2 q_t(\bar{\theta}_n)(\hat{\theta}_n-\theta^*_n)\nabla q_t(\theta^*_n)'$$

$$+n^{-1}\sum_{\tau=1}^{l_n}w_\tau\sum_{t=1}^{n-\tau}\left(\nabla^2 q_{t+\tau}(\bar{\theta}_n)(\hat{\theta}_n-\theta^*_n)\nabla q_t(\theta^*_n)'+\nabla^2 q_t(\bar{\theta}_n)(\hat{\theta}_n-\theta^*_n)\nabla q_{t+\tau}(\theta^*_n)'\right)$$

$$+n^{-1}\sum \nabla q_t(\theta_n^*)(\hat{\theta}_n-\theta_n^*)'\nabla^2 q_t(\bar{\theta}_n)'$$

$$+n^{-1}\sum_{\tau=1}^{l_n} w_\tau \sum_{t=1}^{n-\tau}\left(\nabla q_{t+\tau}(\theta_n^*)(\hat{\theta}_n-\theta^*)'\nabla^2 q_t(\bar{\theta}_n)' + \nabla q_t(\theta_n^*)(\hat{\theta}_n-\theta_n^*)'\nabla^2 q_{t+\tau}(\bar{\theta}_n)'\right)$$

$$+n^{-1}\sum \nabla^2 q_t(\bar{\theta}_n)(\hat{\theta}_n-\theta_n^*)(\hat{\theta}_n-\theta_n^*)'\nabla^2 q_t(\bar{\theta}_n)'$$

$$+n^{-1}\sum_{\tau=1}^{l_n} w_\tau \sum_{t=1}^{n-\tau}(\nabla^2 q_{t+\tau}(\bar{\theta}_n)(\hat{\theta}_n-\theta_n^*)(\hat{\theta}_n-\theta^*)'\nabla^2 q_t(\bar{\theta}_n)'$$

$$+\nabla^2 q_t(\bar{\theta}_n)(\hat{\theta}_n-\theta_n^*)(\hat{\theta}_n-\theta^*)'\nabla^2 q_{t+\tau}(\bar{\theta}_n))$$

We can write

$$\hat{V}_n = \hat{V}_{1n} + \hat{V}_{2n} + \hat{V}_{3n} + \hat{V}_{4n}$$

We shall organize the proof of this theorem in three different steps. Basically in Step 1 we shall show that $\hat{V}_{1n} = V^* + U^* + o_{as}(1) = B^* + o_{as}(1)$, as defined in A12 and A20, in the following steps we shall show that $\hat{V}_{in} = o_{as}(1)$ for $i = 2, 3, 4$; this gives the desired result.

STEP 1

Let

$$B_n^* = n^{-1}\sum E(\nabla q_t(\theta_n^*)\nabla q_t(\theta_n^*)')$$

$$+n^{-1}\sum_{\tau=1}^{n-1}\sum_{t=1}^{n}\left(E(\nabla q_{t+\tau}(\theta_n^*)\nabla q_t(\theta_n^*)') + E(\nabla q_t(\theta_n^*)\nabla q_{t+\tau}(\theta_n^*)')\right)$$

Now $B_n^* - B^* = o(1)$. Let's also define

$$\ddot{B}_n = n^{-1}\sum E(\nabla q_t(\theta_n^*)\nabla q_t(\theta_n^*)')$$

$$+n^{-1}\sum_{\tau=1}^{l_n} w_\tau \sum_{t=1}^{n-\tau}\left(E(\nabla q_{t+\tau}(\theta_n^*)\nabla q_t(\theta_n^*)') + E(\nabla q_t(\theta_n^*)\nabla q_{t+\tau}(\theta_n^*)')\right)$$

Given $w_\tau = 1 - \frac{\tau}{l_n+1}$, recalling A1-A2, A5(ii), A6(ii),we know, from lemma 6.6 in Gallant White that $\ddot{B}_n - B_n^* = o(1)$. Given A1, A2, A5(ii) and A7(ii), from Corollary 3.4(b) in Gallant and White (1988), we know that $\nabla q_t(\theta_n^*)\nabla q_t(\theta_n^*)'$ is a mixingale of size -1; thus by the same argument used in the proof of theorem 1, we know that it satisfies the (memory) condition 1.3 in Eberlain. Then given A1-A3, A5(ii), A7(ii), A8, A11, A14-A15

$$\hat{V}_{1n} - \ddot{B}_n = o_{as}(1)$$

and so

$$\hat{V}_{1n} - B^* = o_{as}(1)$$

because of Lemma 2 in Corradi (1996). In that lemma it is shown that,

$$\frac{1}{\sqrt{2n \log\log n}} \sum_{t=\tau+1}^{n} \left(\nabla q_t(\theta^*)\nabla q_t(\theta^*)\prime - E\left(\nabla q_t(\theta^*)\nabla q_t(\theta^*)\prime\right)\right) = O_{as}(1),$$

uniformly in τ, as $l_n = o\left(\frac{\sqrt{n}}{\sqrt{2\log\log n}}\right)$, the result follows.

STEP 2

We shall show that $\hat{V}_{2n} = o_{as}(1)$.

It suffices to show that

$$n^{-1}\sum_{\tau=1}^{l_n} w_\tau \sum_{t=1}^{n-\tau} \nabla^2 q_t(\bar{\theta}_n) \otimes \nabla q_{t+\tau}(\theta_n^*)' vec(\hat{\theta}_n - \theta_n^*) = o_{as}(1) \tag{7}$$

For notational convenience, let

$$\nabla^2 q_t(\bar{\theta}_n) = [d_t^{ij}(\bar{\theta}_n)], \forall i, j = 1 \ldots p$$

and

$$\nabla q_t(\theta_n^*) = [s_t^j(\theta_n^*)], \forall j = 1, \ldots p$$

Thus the first element of (7) can be written as

$$n^{-1}\sum_{\tau=1}^{l_n} w_\tau \sum_{t=1}^{n-\tau} (\sum_{j=1}^{p} d_t^{1,j}(\bar{\theta}_n) s_{t+\tau}^1(\theta_n^*)(\hat{\theta}_{jn} - \theta_{jn}^*)))$$

As p is finite is suffices to show that

$$n^{-1}\sum_{\tau=1}^{l_n} w_\tau \sum_{t=1}^{n-\tau} d_t^{11}(\bar{\theta}_n) s_{t+\tau}^1(\theta_n^*)(\hat{\theta}_{1n} - \theta_{1n}^*) = o_{as}(1)$$

Again for notational simplicity we drop out the superscript 1. We can write

$$n^{-1}\sum_{\tau=1}^{l_n} w_\tau \sum_{t=1}^{n-\tau} \left(d_t(\theta_n^*) s_{t+\tau}(\theta_n^*) - E(d_t(\theta_n^*) s_{t+\tau}(\theta_n^*))\right)(\hat{\theta}_{1n} - \theta_{1n}^*) \tag{8}$$

$$+ n^{-1}\sum_{\tau=1}^{l_n} w_\tau \sum_{t=1}^{n-\tau} (E(d_t(\theta_n^*) s_{t+\tau}(\theta_n^*)))(\hat{\theta}_{1n} - \theta_{1n}^*)$$

$$+n^{-1}\sum_{\tau=1}^{l_n} w_\tau \sum_{t=1}^{n-\tau}(d_t(\theta_n^*) - d_t(\bar{\theta}_n))s_{t+\tau}(\theta_n^*)(\hat{\theta}_{1n} - \theta_{1n}^*)$$

Given A1-A2, A5(ii)-(iii) and A7(ii)-(iii) we know, again from Corollary 3.4(b) in Gallant and White, that $d_t(\theta_n^*)s_{t+\tau}(\theta_n^*)$ is a mixingale of size -1.

Then given A1-A3, A5(ii)-(iii), A7(ii)-(iii), A8,A11, A16-A17, the first term in (8) is $o_{as}(1)$, by lemma 2 in Corradi (1995). The second term is $o_{as}(1)$ as $(\hat{\theta}_{1n} - \theta_{1n}^*) = O_{as}(\sqrt{\frac{2loglog(n)}{n}})$ and $\sup_t E(d_t^2(\theta_n^*)) \sup_t E(s_{t+\tau}^2(\theta_n^*)) \leq \Delta$, because of A5(ii)-(iii), and finally $l_n = o(\sqrt{n}/\sqrt{2loglog(n)})$. As to the last term in (8),

$$|n^{-1}\sum_{\tau=1}^{l_n} w_\tau \sum_{t=1}^{n-\tau} d_t(\theta_n^*)s_{t+\tau}(\theta_n^*)(\hat{\theta}_{1n} - \theta_{1n}^*)|$$

$$\leq n^{-1}\sum_{\tau=1}^{l_n}\sum_{t=1}^{n}|d_t(\theta_n^*)s_{t+\tau}(\theta_n^*)||\hat{\theta}_{1n} - \theta_{1n}^*|$$

$$= l_n O_{as}(1) O_{as}\left(\sqrt{\frac{2\log\log(n)}{n}}\right) = o_{as}(1)$$

where the $O_{as}(1)$ term comes from the fact that given A1,A2,A5(ii)-(iii) and A7(ii)-(iii), $d_t(\theta_n^*)s_{t+\tau}(\theta_n^*)$ is a mixingale of size -1 (GW, corollary 3.4(b)) and satisfies a strong law of large numbers. It remains to consider,

$$|n^{-1}\sum_{\tau=1}^{l_n} w_\tau \sum_{t=1}^{n-\tau} d_t(\bar{\theta}_n)s_{t+\tau}(\theta_n^*)(\hat{\theta}_{1n} - \theta_{1n}^*)|$$

$$\leq \sum_{\tau=1}^{l_n}\left(n^{-1}\sum_{t=1}^{n} d_t^2(\bar{\theta}_n)\right)^{1/2}\left(n^{-1}\sum_{t=1}^{n} s_{t+\tau}^2(\theta_n^*)\right)^{1/2}|\hat{\theta}_{1n} - \theta_{1n}^*| \quad .$$

$$= l_n O_{as}(1) O_{as}\left(\sqrt{\frac{2\log\log(n)}{n}}\right)$$

by noting that the first term in brackets satsfies a uniform strong law of large numbers because of A1-A3, A5(iii), A7(iii), A8, A9(iv). The result then follows straightforwardly. Finally $\hat{V}_{3n} = o_{as}(1)$ by the same argument.

STEP 3

It remains to show that $\hat{V}_{4n} = o_{as}(1)$. It suffices to show that

$$n^{-1}\sum_{\tau=1}^{l_n} w_\tau \sum_{t=1}^{n-\tau} \nabla^2 q_t(\bar{\theta}_n) \otimes \nabla^2 q_{t+\tau}(\bar{\theta}_n) vec(\hat{\theta}_n - \theta_n^*)(\hat{\theta}_n - \theta_n^*)\prime = o_{as}(1)$$

As in Step 2, we can write the first element of the term above as

$$\sum_{i=1}^{p}\sum_{j=1}^{p} d_t^{1i}(\bar{\theta}_n) d_t^{1j}(\bar{\theta})(\hat{\theta}_{in}-\theta_{in}^{*})(\hat{\theta}_{jn}-\theta_{jn}^{*})$$

Now by the same argument followed in the proof of Step 2,

$$n^{-1}\sum_{\tau=1}^{l_n} w_\tau \sum_{t=1}^{n-\tau}(d_t^{11})^2(\bar{\theta}_n)(\hat{\theta}_{1n}-\theta_{1n}^{*})^2 = o_{as}(1)$$

The statement in the theorem follows just putting together the 3 steps. □

PROOF OF THEOREM 3: Let $d_n = \sqrt{\frac{n}{2loglog(n)}}$, we have

$$\begin{aligned} d_n\|\hat{V}_n^{-1/2}\hat{A}_n(\hat{\theta}_n-\theta_n^{*})\| &= d_n\|(V^{*}+U^{*})^{-1/2}\hat{A}_n(\hat{\theta}_n-\theta_n^{*}) \\ &\quad +(\hat{V}_n^{-1/2}-(V^{*}+U^{*})^{-1/2})\hat{A}_n(\hat{\theta}_n-\theta_n^{*})\| \\ &= d_n\|(V^{*}+U^{*})^{-1/2}A^{*}(\hat{\theta}_n-\theta_n^{*})+ \\ &\quad +(V^{*}+U^{*})^{-1/2}(\hat{A}_n-A^{*})(\hat{\theta}_n-\theta_n^{*}) \\ &\quad +(\hat{V}_n^{-1/2}-(V^{*}-U^{*})^{-1/2})A^{*}(\hat{\theta}_n-\theta_n^{*}) \\ &\quad +(\hat{V}_n^{-1/2}-(V^{*}+U^{*})^{-1/2})(\hat{A}_n-A^{*})(\hat{\theta}_n-\theta_n^{*})\| \end{aligned}$$

$$\begin{aligned} &\leq d_n\|(V^{*}+U^{*})^{-1/2}A^{*}(\hat{\theta}_n-\theta_n^{*})\| + d_n\|(V^{*}+U^{*})^{-1/2}(\hat{A}_n-A^{*})(\hat{\theta}_n-\theta_n^{*})\| \\ &\quad + d_n\|(\hat{V}_n^{-1/2}-(V^{*}-U^{*})^{-1/2})A^{*}(\hat{\theta}_n-\theta_n^{*})\| \\ &\quad + d_n\|(\hat{V}_n^{-1/2}-(V^{*}+U^{*})^{-1/2})(\hat{A}_n-A^{*})(\hat{\theta}_n-\theta_n^{*})\| \end{aligned}$$

We shall show that all the terms but $d_n\|(V^{*}+U^{*})^{-1/2}A^{*}(\hat{\theta}_n-\theta_n^{*})\|$ are $o_{as}(1)$.
From Theorem 1 we know that $d_n\left(\hat{\theta}_n-\theta_n^{*}\right) = O_{a.s.}(1)$ from Theorem 2, we know that $\hat{V}_n - (V^{*}+U^{*}) = o_{as}(1)$, and given A12 and A20, it also true that $\hat{V}_n^{-1/2}-(V^{*}+U^{*})^{-1/2} = o_{as}(1)$. Thus it follows that

$$d_n\|(\hat{V}_n^{-1/2}-(V^{*}+U^{*})^{-1/2})A^{*}(\hat{\theta}_n-\theta_n^{*})\| = o_{as}(1)$$

Given A1-A4,A5(iii), A7(iii) and A8, we have, from theorem 6.1 in Gallant and White, that $\hat{A}_n - A^{*} = o_{as}(1)$. Thus

$$d_n\|(V^{*}+U^{*})^{-1/2}(\hat{A}_n-A^{*})(\hat{\theta}_n-\theta_n^{*})\| = o_{as}(1)$$

It then follows that

$$d_n\|\hat{V}_n^{-1/2}\hat{A}_n(\hat{\theta}_n - \theta_n^*)\| = d_n\|(V^* + U^*)^{-1/2}A^*(\hat{\theta}_n - \theta_n^*)\| + o_{as}(1)$$

Now

$$d_n\|(V^* + U^*)^{-1/2}A^*(\hat{\theta}_n - \theta_n^*)\| = d_n\|(V^* + U^*)^{-1/2}V^{*1/2}V^{*-1/2}A^*(\hat{\theta}_n - \theta_n^*)\| + o_{as}(1)$$
$$= \frac{1}{b_n}\|(V^* + U^*)^{-1/2}V^{*1/2}V^{*-1/2}\sum_{t=1}^{n}\nabla q_t(\theta_n^*, X^t))\| + o_{as}(1) \quad .$$

By the same argument used in the proof of Theorem 1, we know that given A1-A4, A7(ii), A8, A10-A13 and A20,

$$\limsup_{n\to\infty}\frac{1}{b_n}\|(V^* + U^*)^{-1/2}V^{*1/2}V^{*-1/2}\sum_{t=1}^{n}\nabla q_t(\theta_n^*, X^t)\| = \limsup_{s\to\infty}\frac{1}{b_s}\|B_s\|$$

where $B = (B_s, s \geq 0)$ is Brownian motion with variance equal to $(V^*+U^*)^{-1/2}V^*(V^*+U^*)^{-1/2}$; and so from Mckean (1969, p.96), we know that

$$\limsup_{s\to\infty}\frac{1}{b_s}\|B_s\| = \sqrt{\gamma_d}$$

where γ_d is the largest eigenvalue of $(V^* + U^*)^{-1/2}V^*(V^* + U^*)^{-1/2}$. It remains to show that $\gamma_d \leq 1$.

Now

$$(V^* + U^*)^{-1/2}V^*(V^* + U^*)^{-1/2} \leq I \quad .$$

if

$$V^* \leq (V^* + U^*)$$

but this holds for U^* positive definite, which is guaranteed by A20. Clearly for $U^* = 0$, $\gamma_d = 1$.
□

PROOF OF THEOREM T1:

(i)-(a)

$$d_n\|W_n\| = d_n\|\hat{B}_n^{-1/2}h_n(\theta_n^*) + \hat{B}_n^{-1/2}H_n(\bar{\theta}_n)(\hat{\theta}_n - \theta_n^*)\|$$

$$= d_n\|\hat{B}_n^{-1/2}H_n(\bar{\theta}_n)(-\nabla_\theta^2 Q_n(\bar{\theta}_n))^{-1}n^{-1}\sum\nabla_\theta q_t(\theta_n^*)\| + o_{as}(1)$$

$$= b_n^{-1} \| J_n^* \sum \nabla_\theta q_t(\theta_n^*) \| + o_{as}(1)$$

in fact by Theorem 2,

$$\hat{B}_n^{-1/2} H_n(\bar{\theta}_n)(\nabla_\theta^2 Q_n(\bar{\theta}_n))^{-1} - J_n^* = o_{as}(1)$$

with $J_n^* = B_n^{*-1/2} H_n(\theta_n^*) A_n^{*-1}$. The result then follows from Theorem 3.

(i)-(b)

Given A1-A3,A5, A7, A8, A10-A13, AT1-AT2, under the null, we have that

$$\tilde{\theta}_n - \theta_n^* = o_{as}(1)$$

$$\sqrt{\frac{n}{2 \log\log(n)}} \| \tilde{\theta}_n - \theta_n^* \| = O_{as}(1)$$

From A1-A4, A7 (iii) and A8, A9 (i)-(iii), we have by theorem 6.1 in Gallant and White, that

$$n^{-1} \sum \nabla^2 q_t(\widetilde{\theta}_n) - A_n^0 \to 0 \text{ a.s.} \tag{9}$$

and, as $\bar{\theta}_n \in (\tilde{\theta}_n, \theta_n^0)$,

$$n^{-1} \sum \nabla^2 q_t(\overline{\theta}_n) - A_n^0 \to 0 \text{ a.s.},$$

where $A_n^0 = n^{-1} \sum E(\nabla_\theta^2 q_t(\theta_n^0))$, and by A10,

$$\left(n^{-1} \sum \nabla^2 q_t(\overline{\theta}_n) \right)^{-1} - A_n^{0-1} \to 0 \text{ a.s..}$$

By the Taylor's Theorem, we can write

$$d_n \left(n^{-1} \sum \nabla q_t(\widetilde{\theta}_n) \right) = d_n \left(n^{-1} \sum \nabla q_t(\theta_n^0) \right) + A_n^0 d_n \left(\widetilde{\theta}_n - \theta_n^0 \right) + o_{as}(1)$$

$$d_n h(\widetilde{\theta}_n) = d_n h(\theta_n^0) + H_n(\theta_n^0) d_n \left(\widetilde{\theta}_n - \theta_n^0 \right) + o_{as}(1).$$

Recalling that $h(\widetilde{\theta}_n) \equiv 0$ and $h(\theta_n^0) \equiv 0$ by construction, and that A_n^0 is bounded given A5(iii), we have

$$d_n \left(\widetilde{\theta}_n - \theta_n^0 \right) = \left(A_n^0 \right)^{-1} d_n \left(n^{-1} \sum \nabla q_t(\widetilde{\theta}_n) - n^{-1} \sum \nabla q_t(\theta_n^0) \right) + o_{as}(1) \tag{10}$$

and

$$H_n(\theta_n^0) d_n \left(\widetilde{\theta}_n - \theta_n^0 \right) = o_{as}(1).$$

By premultipling (10) by $H_n(\theta_n^0)$, after some simple manipulations,

$$H_n(\theta_n^0)A_n^{0-1}d_n\left(n^{-1}\sum\nabla q_t(\widetilde{\theta}_n)\right)=H_n(\theta_n^0)A_n^{0-1}d_n\left(n^{-1}\sum\nabla q_t(\theta_n^0)\right)+o_{as}(1). \quad (11)$$

Now, as $d_nh(\theta_n^0)=d_nh(\theta_n^*)+H_n(\theta_n^+)d_n\left(\theta_n^0-\theta_n^*\right)$, with $\theta_n^+\in\left(\theta_n^0,\theta_n^*\right)$, we have that $d_n\left(\theta_n^0-\theta_n^*\right)=o(1)$, so

$$\begin{aligned} d_n\left(n^{-1}\sum\nabla q_t(\theta_n^0)\right) &= d_n\left(n^{-1}\sum\nabla q_t(\theta_n^*)\right)+(A_n^*)\,d_n\left(\theta_n^0-\theta_n^*\right)+o_{as}(1) \quad (12)\\ &= b_n^{-1}\sum\nabla q_t(\theta_n^*)+o_{as}(1)\end{aligned}$$

Substituting (12) into (11),

$$H(\theta_n^0)A_n^{0-1}d_n\left(n^{-1}\sum\nabla q_t(\widetilde{\theta}_n)\right)=H(\theta_n^*)A_n^{*-1}d_n\left(n^{-1}\sum\nabla q_t(\theta_n^*)\right)+o_{as}(1). \quad (13)$$

and given that $\left(\widetilde{\theta}_n-\theta_n^0\right)=o_{as}(1)$, and (13) we have

$$H(\widetilde{\theta}_n)\left(n^{-1}\sum\nabla^2 q_t(\widetilde{\theta}_n)\right)d_n\left(n^{-1}\sum\nabla q_t(\widetilde{\theta}_n)\right)=H(\theta_n^*)A_n^{*-1}d_n\left(n^{-1}\sum\nabla q_t(\theta_n^*)\right)+o_{as}(1).$$

The result follows by part (i)-(a).

(ii)(a) By simple manipulations, we have

$$d_n\|\hat{B}_n^{-1/2}h_n(\hat{\theta}_n)\|\geq d_n\|B_n^{*-1/2}h_n(\theta_n^*)\|$$

$$-d_n\|B_n^{*-1/2}H_n(\theta_n^*)A_n^*n^{-1}\sum\nabla_\theta q_t(\theta_n^*)\|+o_{as}(1)$$

$$=d_n\|B_n^{*-1/2}h_n(\theta_n^*)\|-O_{as}(1)$$

given that under the alternative $h_n(\theta_n^*)\neq 0$. The result follows.

(ii)-(b)

Given A1-A5, A7, A9(ii)(iii) and A11, $n^{-1/2}\sum\nabla_\theta q_t(\theta_n^0)$ satsfy the CLT for mixingales double arrays (Wooldridge 1986) and $\tilde{\theta}_n-\theta_n^0=o_p(1)$ and given AT2, $\tilde{V}_n-(V_n^0+U_n^0)=o_p(1)$, because of theorem 6.8(b) in GW (1988). Thus we have

$$d_n\|LM_n\|=d_n\|B_n^{0-1/2}H_n(\theta_n^0)'A_n^{0-1}n^{-1}\sum\nabla_\theta q_t(\theta_n^0)\|+o_p(1)$$

$$\geq d_n\|B_n^{0-1/2}H_n(\theta_n^0)'n^{-1}\sum\nabla_\theta^2 q_t(\bar{\theta}_n)(\theta_n^0-\theta_n^*)\|$$

$$-d_n\|B_n^{0-1/2}H_n(\theta_n^0)'A_n^{0-1}n^{-1}\sum\nabla_\theta q_t(\theta_n^*)\|+o_p(1)$$

$$= d_n \| B_n^{0-1/2} H_n(\theta_n^0)' A_n^{0-1} n^{-1} \sum \nabla_\theta^2 q_t(\bar{\theta}_n)(\theta_n^0 - \theta_n^*) \| - O_p(1) + o_p(1)$$

where $\bar{\theta}_n \in (\theta_n^*, \theta_n^0)$. Now it will suffice to show that $\forall n > n_0$, and $\eta > 0$,

$$|\theta_n^0 - \theta_n^*| > \eta, \tag{14}$$

but given that

$$h(\theta_n^*) - h(\theta_n^0) = H(\bar{\theta}_n)'(\theta_n^* - \theta_n^0),$$

and $H_n(\theta)$ is uniformly full column rank, $h(\theta_n^0) \equiv 0$ and $|h(\theta_n^*)| > \epsilon$, $\forall n > n_0$ and for some $\epsilon > 0$, then (14) follows.

(iii)-(a)

$$d_n \| W_n \| = d_n \| B_n^{*-1/2} h(\theta_n^*) + B_n^{*-1/2} H_n(\theta_n^*)' A_n^{*-1} n^{-1} \sum \nabla_\theta q_t(\theta_n^*) \| + o_{as}(1)$$

Thus, almost surely

$$\limsup_{n \to \infty} d_n \| W_n \| \leq 1 + \Delta$$

with $0 < \Delta < \infty$.

(iii)-(b)

$$d_n \| LM_n \| = d_n \| B_n^{0-1/2} H_n(\theta_n^0)' A_n^{0-1} n^{-1} \sum \nabla_\theta q_t(\theta_n^0) \| + o_p(1)$$

$$\leq d_n \| B_n^{0-1/2} H_n(\theta_n^0)' n^{-1} \sum \nabla_\theta^2 q_t(\bar{\theta}_n)(\theta_n^0 - \theta_n^*) \|$$

$$+ d_n \| B_n^{0-1/2} H_n(\theta_n^0)' A_n^{0-1} n^{-1} \sum \nabla_\theta q_t(\theta_n^*) \| + o_p(1)$$

as $d_n h(\theta_n^*) = O(1)$, by an argument similar to that one in (ii)-(b), it follows that $d_n(\theta_n^0 - \theta_n^*) = O(1)$ and the desired result follows. □

PROOF OF LEMMA TN1

(i) This proof is similar to the proof of lemma A1 in Andrews (1993), however here we need an almost sure rather than a weak result. Given ATN2 , U-SLLN will imply that

$$\sup_{\pi \in \Pi} \sup_{\theta \in \Theta} |n^{-1} \sum q_t(\theta, \pi) - n^{-1} \sum E(q_t(\theta, \pi))| = o_{a.s.}(1),$$

so it exists a $F_1 \in \Omega$ with $P(F_1) = 1$, where $(\Omega, \Im, P)$ is the underlying probability space, and a $N_1(w, \varepsilon)$ s.t. $\forall w \in F_1$ and $n > N_1(w, \varepsilon)$,

$$\sup_{\pi \in \Pi} \left(n^{-1} \sum q_t(\widehat{\theta}_n(w, \pi), \pi, w) - n^{-1} \sum E(q_t(\widehat{\theta}_n(w, \pi), \pi)) \right) < \delta(\varepsilon)/2.$$

Similarly, it exists a $F_2 \in \Omega$ with $P(F_2) = 1$ and a $N_2(w, \varepsilon)$ s.t. $\forall w \in F_2$ and $n > N_2(w, \varepsilon)$,

$$\sup_{\pi \in \Pi} \left(n^{-1} \sum q_t(\theta^*(\pi), \pi, w) - n^{-1} \sum E(q_t(\theta^*(\pi), \pi)) \right) < \delta(\varepsilon)/2.$$

Now $\forall w \in F = F_1 \cap F_2$ with $P(F) = 1$ and $n > max(N_1(w, \varepsilon), N_2(w, \varepsilon))$, it is true

$$\sup_{\pi \in \Pi} \left(n^{-1} \sum E(q_t(\widehat{\theta}_n(w, \pi), \pi)) - n^{-1} \sum E(q_t(\theta^*(\pi), \pi)) \right) \leq$$

$$\sup_{\pi \in \Pi} \left(n^{-1} \sum q_t(\widehat{\theta}_n(w, \pi), \pi) - n^{-1} \sum E(q_t(\theta^*(\pi), \pi)) \right) +$$

$$\sup_{\pi \in \Pi} \left(n^{-1} \sum q_t(\widehat{\theta}_n(w, \pi), \pi) - n^{-1} \sum E(q_t(\widehat{\theta}_n(\pi), \pi)) \right) \leq$$

$$\sup_{\pi \in \Pi} \left(n^{-1} \sum q_t(\theta_n^*(w, \pi), \pi) - n^{-1} \sum E(q_t(\theta^*(\pi), \pi)) \right) +$$

$$\sup_{\pi \in \Pi} \left(n^{-1} \sum q_t(\widehat{\theta}_n(w, \pi), \pi) - n^{-1} \sum E(q_t(\widehat{\theta}_n(\pi), \pi)) \right)$$

$$\leq 2 \sup_{\pi \in \Pi, \theta \in \Theta} |n^{-1} \sum q_t(\theta, \pi) - \sum E(q_t(\pi, \theta))| \leq \delta(\epsilon)$$

then, because of the identification condition ATN3, the result follows by contradiction. In fact if $\widehat{\theta}_n(\pi) - \theta^*(\pi)$ would not converge, uniformly in π, then the inequality above would give a violation of the identification condition.

(ii) Given ATN6, it is immediate from part (i).□

PROOF OF THEOREM TN1:

(i)-(a) The proof is organized in five steps.

STEP 1

Under the null $\theta^*(\pi) = \theta^*, \forall \pi \in \Pi$; we show that $b_n^{-1} \sum \nabla_\theta q_t(\theta^*, \pi)$ is a.s. asymptotically uniformly stochastic equicontinuous (a.s. AUEC, as in definition 2.3C in Pötscher and Prucha (1994)). For all $\pi \in \Pi$, and for $j = 1, \ldots p$, given ATN4, we have almost surely,

$$b_n^{-1} \| \sum \nabla_\theta q_t^j(\theta^*, \pi) - \sum \nabla_\theta q_t^j(\theta^*, \pi\prime) \| = b_n^{-1} \| \sum (\nabla_\pi (\nabla_\theta q_t^j(\theta^*, \pi)))_{\pi = \bar{\pi}} (\pi - \pi\prime) \|$$

$$\leq b_n^{-1} \| \sum \nabla_\pi (\nabla_\theta q_t^j(\theta^*, \bar{\pi})) \| \| \pi - \pi\prime \| = B_n(\bar{\pi}) \| \pi - \pi\prime \|$$

with $\bar{\pi} \in (\pi, \pi')$. Note that given ATN1, the set of measure zero over each the inequality above does not hold, does not dipend on the specific $\pi^{'}$ we pick. We need to show that (condition 2.10c in Pötscher and Prucha),

$$\sup_{\pi \in \Pi} \limsup_{n \to \infty} B_n(\pi) < \infty$$

is satisfied. The desired result then follows from Lemma A2(a) in Pötscher and Prucha, (1994). Given A1,A2, ATN1, ATN3, ATN5, ATN8, we know that, because of theorem 1 in Eberlain,

$$\limsup_{n \to \infty} B_n(\pi) = \sqrt{\lambda_M(\pi)}$$

where $\lambda_M(\pi)$ is the largest eigenvalue of $V^{**}(\pi)$ and $\sup_{\pi \in \Pi} \lambda_M(\pi) < \Delta$ by ATN8.

STEP 2

In this step it will be shown that

$$\limsup_{n \to \infty} \sup_{\pi} b_n^{-1} \left\| \sum \nabla_\theta q_t(\theta^*, \pi) \right\| = O_{a.s.}(1) \tag{15}$$

which is necessary to extend the a.s. AUEC property to the test statistics W_n. Given that $\left\| b_n^{-1} \sum \nabla_\theta q_t(\theta^*, \pi) \right\|$ is a.s. AUEC, it means that there is a set C', $P(C') = 1$, such that for every $\varpi \in C'$, $\left\| b_n^{-1} \sum \nabla_\theta q_t(\theta^*, \pi, \varpi) \right\|$ is asymptotic uniform equicontinuous. Also by pointwise convergence $\forall \pi_k \in \Pi^0$, where Π^0 is dense in Π, it exists a set C^k, $P\left(C^k\right) = 1$, such that

$$\limsup_{n \to \infty} b_n^{-1} \left\| \sum \nabla_\theta q_t(\theta^*, \pi, \varpi) \right\| = O(1)$$

for $\varpi \in C^k$.

Consider the set $C^* = \bigcap_{k=1}^{\infty} C^k \cap C'$, by construction $P(C^*) = 1$. Given that Π is compact by ATN1, we will define a finite cover of Π by balls of radius δ centered in π_i, $\eta(\pi_i, \delta)$ for $i = 1, ..., m$. For $\varpi \in C^*$ and for $\pi \in \eta(\pi_i, \delta)$,

$$\begin{aligned} b_n^{-1} \left\|\sum \nabla_\theta q_t(\theta^*, \pi, \varpi)\right\| \leq b_n^{-1} \sup_{\pi' \in \eta(\pi_i,\delta)} \left\|\sum \nabla_\theta q_t(\theta^*, \pi', \varpi)\right\| \leq \\ \leq b_n^{-1} \sup_{\pi' \in \eta(\pi_i,\delta)} \left\|\sum \nabla_\theta q_t(\theta^*, \pi', \varpi) - \sum \nabla_\theta q_t(\theta^*, \pi_i, \varpi)\right\| \\ + b_n^{-1} \left\|\sum \nabla_\theta q_t(\theta^*, \pi_i, \varpi)\right\| \end{aligned} \tag{16}$$

Taking the sup over π on both side of (16), given that the cover is finite, we have that

$$\begin{aligned} b_n^{-1} \sup_\pi \left\|\sum \nabla_\theta q_t(\theta^*, \pi, \varpi)\right\| \leq \\ b_n^{-1} \sup_\pi \sup_{\pi' \in \eta(\pi_i,\delta)} \left\|\sum \nabla_\theta q_t(\theta^*, \pi', \varpi) - \sum \nabla_\theta q_t(\theta^*, \pi_i, \varpi)\right\| \\ + b_n^{-1} \max_{1 \leq i \leq m} \left\|\sum \nabla_\theta q_t(\theta^*, \pi_i, \varpi)\right\|. \end{aligned} \tag{17}$$

Because of asymptotic uniform equicontinuity, we have

$$\limsup_{n \to \infty} b_n^{-1} \sup_\pi \sup_{\pi' \in \eta(\pi_i,\delta)} \left\|\sum \nabla_\theta q_t(\theta^*, \pi', \varpi) - \sum \nabla_\theta q_t(\theta^*, \pi_i, \varpi)\right\| = 0$$

and

$$\limsup_{n \to \infty} b_n^{-1} \max_{1 \leq i \leq m} \left\|\sum \nabla_\theta q_t(\theta^*, \pi_i, \varpi)\right\| = O(1)$$

as by ATN7, $b_n^{-1} \|\sum \nabla_\theta q_t(\theta^*, \pi)\|$ satisfies a LIL $\forall \pi \in \Pi$, and m is finite. By taking limsup of both sides of (17), the result follows.

STEP 3

It is possible now to extend the a.s AUEC properties to the test statistic. At this step let us assume that

$$\sup_{\pi \in \Pi} |\hat{V}_n(\pi) - V^*(\pi)| = o_{as}(1) \tag{18}$$

We shall prove (18) in Step 5. Now (18), together with Lemma TN(ii) and step 2, imply that $\sup_\pi d_n \left|\widehat{\theta}_n(\pi) - \theta_n^*(\pi)\right| = O_{as}(1)$. So

$$\begin{aligned} d_n \left|\|W_n(\pi)\| - \|W_n(\pi')\|\right| &\leq d_n \|W_n(\pi) - W_n(\pi')\| \\ &= \Big\| B_n^{*-\frac{1}{2}}(\pi) H(\theta^*) A_n^{*-1}(\pi) b_n^{-1} \sum q_t(\theta^*, \pi) - \\ &\quad - B_n^{*-\frac{1}{2}}(\pi') H(\theta^*) A_n^{*-1}(\pi') b_n^{-1} \sum q_t\left(\theta^*, \pi'\right) \Big\| \\ &\quad + o_{as}(1) O_{as}(1) \end{aligned}$$

where the $o_{as}(1)$ and $O_{as}(1)$ are independent of π. The last equality follows from A1-A2, ATN1-ATN6, because of Lemma TN1(ii) and step 2. It then follows from Step 1 that $d_n\|W_n(\pi)\|$ is a.s.AUEC on Π.

STEP 4

Given that $d_n\|W_n(\pi)\|$ is a.s.AUEC on Π, it is possible to show now that $d_n\|W_n(\pi)\|$ satisfies a LIL by the same type of argument as in STEP 2. In fact given a.s. AUEC, it means that there is a set C', $P(C')=1$, such that for every $\varpi \in C'$, $d_n \|W_n(\varpi,\pi)\|$ is asymptotic uniform equicontinuous. Also by pointwise convergence $\forall \pi_k \in \Pi^0$, where Π^0 is dense in Π, it exists a set C^k, $P\left(C^k\right)=1$, such that

$$\limsup_{n\to\infty} d_n \|W_n(\varpi,\pi)\| = 1$$

for $\varpi \in C^k$.

Consider the set $C^* = \bigcap_{k=1}^{\infty} C^k \cap C'$, by construction $P(C^*)=1$. For $\varpi \in C^*$ and for $\pi \in \eta(\pi_i,\delta)$, defined in STEP 2,

$$\begin{aligned} d_n \|W_n(\varpi,\pi)\| &\leq d_n \sup_{\pi'\in\eta(\pi_i,\delta)} \|W_n(\varpi,\pi')\| \leq \\ &\leq d_n \sup_{\pi'\in\eta(\pi_i,\delta)} \left\|W_n\left(\varpi,\pi'\right) - W_n(\varpi,\pi_i)\right\| + d_n \|W_n(\varpi,\pi_i)\| \end{aligned} \tag{19}$$

Taking the sup over π on both side of (19), given that the cover is finite, we .

$$d_n \sup_{\pi} \|W_n(\varpi,\pi)\| \leq d_n \sup_{\pi} \sup_{\pi'\in\eta(\pi_i,\delta)} \|W_n(\varpi,\pi') - W_n(\varpi,\pi_i)\| + d_n \max_{1\leq i\leq m} \|W_n(\varpi,\pi_i)\| . \tag{20}$$

By asymptotic uniform equicontinuity the limsup of the first term on the LHS is zero and

$$\limsup_{n\to\infty} d_n \max_{1\leq i\leq m} \|W_n(\varpi,\pi_i)\| \leq 1$$

given $d_n\|W_n(\pi)\|$ satisfies a LIL $\forall\pi\in\Pi$, and m is finite. So taking limsup of (20) we prove the result.

STEP 5

By the same argument as in Step 1 of the proof of Theorem 2, we can write

$$\hat{V}_n(\pi) = \hat{V}_{1n}(\pi) + \hat{V}_{2n}(\pi) + \hat{V}_{3n}(\pi) + \hat{V}_{4n}(\pi)$$

We want to show that

$$\sup_{\pi\in\Pi}|\hat{V}_{1n}(\pi) - V^*(\pi)| = o_{as}(1)$$

We begin by showing that $b_n^{-1}Z_{t,\tau}^{i,j}(\theta^*,\pi)$, as defined in ATN7, is a.s.AUEC on Π, $\forall t,\tau$ and $\forall i,j$. Given ATN9,

$$d_n|n^{-1}\sum_{t=1}^{n-\tau}\left(Z_{t,\tau}^{i,j}(\theta^*,\pi) - Z_{t,\tau}^{i,j}(\theta^*,\pi\prime)\right)| \leq d_n^{-1}|n^{-1}\sum_{t=1}^{n-\tau}\nabla_\pi Z_{t,\tau}^{i,j}(\theta^*,\bar{\pi})|\prime|\pi - \pi\prime|$$

$$= d_n|n^{-1}\sum_{t=1}^{n-\tau}DZ_{t,\tau}^{i,j}(\theta^*,\pi)|\prime|\pi - \pi\prime|$$

with $\bar{\pi} \in (\pi,\pi\prime)$. Given ATN11, we know that we can switch expectation and differentiation, that is

$$\nabla_\pi E\left(\nabla_\theta q_t(\theta^*,\pi)\nabla_\theta q_t(\theta^*,\pi)\prime\right)_{ij} = E\left(\nabla_\pi\left(\nabla_\theta q_t(\theta^*,\pi)\nabla_\theta q_t(\theta^*,\pi)\prime\right)\right)_{ij}$$

So that $E(DZ_{t,\tau}^{i,j}(\theta^*,\pi)) = 0, \forall\pi\in\Pi$. Because of ATN10,

$$\sup_{\pi\in\Pi}\limsup_{n\to\infty}||DZ_{t,\tau}^{i,j}(\theta^*,\pi)|| = O_{as}(1)$$

with the $O_{as}(1)$ term independent of τ. Thus, given ATN1, ATN2, ATN7, $d_nZ_{t,\tau}^{i,j}(\theta^*,\pi)$ is a.s.AUEC on Π, by the same argument followed in Step 1. Thus by the same argument used in Step 2 above and in the proof of Step 1 of Theorem 2, it follows that

$$\limsup_{n\to\infty}\sup_{\pi\in\Pi}||n^{-1}\sum Z_{t,\tau}^{ij}(\theta^*,\pi)|| = O_{as}(d_n^{-1})$$

with the right hand side uniform in τ. From lemma 2 in Corradi (1996), it follows that

$$\sup_{\pi\in\Pi}|\hat{V}_{1n}(\pi) - V^*(\pi)| = o_{as}(1)$$

Analogously we can show that $b_n^{-1}\sum_{t=1}^{n-\tau}Y_{t,\tau}^{ij}(\theta^*,\pi)$, $b_n^{-1}\sum_{t=1}^{n-\tau}W_{t,\tau}^{i,j}(\theta^*,\pi)$ are a.s. AUEC on Π, $\forall\tau,\forall i,j$; so

$$\sup_{\pi\in\Pi}\hat{V}_{in}(\pi) = o_{as}(1)$$

for $i = 2,3,4$, by the same arguments used in the proof of Step 2,3,4 in Theorem 2.

(i)(b) Given A1, ATN1-ATN2, ATN3b, ATN4-ATN6, we have that

$$\sup_{\pi\in\Pi}|\tilde{A}_n(\bar{\theta}_n,\pi) - A^0(\pi)| = o_{as}(1)$$

and

$$\sup_{\pi\in\Pi} |\tilde{A}_n^{-1}(\bar{\theta}_n,\pi) - A^{0-1}| = o_{as}(1)$$

where $\tilde{A}_n(\bar{\theta}_n,\pi) = n^{-1}\sum \nabla^2_\theta q_t(\bar{\theta}_n,\pi)$, $\bar{\theta}_n \in (\tilde{\theta}_n,\theta^0)$ and $A^0 = E(n^{-1}\sum \nabla^2_\theta q_t(\theta^0,\pi))$. Also $\sup_{\pi\in\Pi} |A_n^*(\pi) - A_n^0(\pi)| \to 0$. Thus by the same argument followed in the proof of Theorem T1-(ii)(b),

$$\sup_\pi d_n \|LM_n(\pi)\| = \sup_\pi d_n \|W_n(\pi)\| + o_{as}(1)$$

The result then follows from (i)-(a)

(ii)-(a) Given Lemma ATN1, and noting that ATN3, ATN4, ATN7, ATN9-ATN11 hold also under the alternative, by the same argument used in Step 5 of (i)-(a), it follows that

$$\sup_{\pi\in\Pi} |\hat{V}_n(\pi) - V^*(\pi)| = o_{as}(1)$$

In general, under H_A, ATN5 is violated, thus we can no longer show, at least by using the same argument used in the proof of Step 1 in (i)-(a), that $b_n^{-1}\sum \nabla_\theta q_t(\theta^*(\pi),\pi)$ is a.s. AUEC on Π. Nevertheless, given A1-A2, ATN1-ATN3, we know that

$$\sup_{\pi\in\Pi} |n^{-1}\sum \nabla_\theta q_t(\theta^*(\pi),\pi)| = o_{as}(1)$$

because of the uniform SLLN. Recalling Lemma TN1

$$\sup_{\pi\in\Pi} d_n \|\hat{B}_n^{-1/2}(\pi)h(\theta^*(\pi)) + \hat{B}_n^{-1/2}(\pi)H(\bar{\theta}_n(\pi))(\hat{\theta}_n(\pi) - \theta^*(\pi))\|$$

$$= \sup_{\pi\in\Pi} d_n \|B^{*-1/2}(\pi)h(\theta^*(\pi)) - B^{*-1/2}H(\theta^*(\pi))A^{*-1}(\pi)n^{-1}\sum \nabla_\theta q_t(\theta^*(\pi),\pi))\| + o_{as}(d_n)$$

$$= \sup_{\pi\in\Pi} d_n \|B^{*-1/2}(\pi)h(\theta^*(\pi))\| + o_{as}(d_n)$$

with $\bar{\theta}_n(\pi) \in (\hat{\theta}_n(\pi),\theta^*(\pi))$. The result then is given by the fact that under H_A, $\forall \pi \in \Pi'$, where $\mu\left(\Pi'\right) > 0$, $h(\theta^*(\pi)) \neq 0$.

(ii)-(b) Given A1-A2, ATN1-ATN2, ATN3(b), ATN11B, it follows that

$$\sup_{\pi\in\Pi} |\tilde{V}_n(\pi) - V^0(\pi)| = o_{as}(1)$$

$$\sup_{\pi\in\Pi} |\tilde{A}_n(\pi) - A^0(\pi)\| = o_{as}(1)$$

Also, recalling ATN2 and ATN3(b), by the uniform SLLN,

$$\sup_{\pi\in\Pi} |n^{-1}\sum \left(\nabla_\theta q_t(\theta^0,\pi) - E(\nabla_\theta q_t(\theta^0,\pi))\right)| = o_{as}(1)$$

Thus we can write

$$d_n \sup_{\pi\in\Pi} \|LM_n(\pi)\| = \sup_{\pi\in\Pi} d_n\|B^{0-1/2}(\pi)H(\theta^0)\prime A^{0-1}(\pi)n^{-1}\nabla_\theta q_t(\theta^0,\pi)\| + o_{as}(d_n)$$

$$= d_n \sup_{\pi\in\Pi} \|B^{0-1/2}(\pi)H(\theta^0)\prime n^{-1}n^{-1}\sum \nabla_\theta q_t(\theta^*(\pi),\pi)$$
$$+B^{0-1/2}(\pi)H(\theta^0)'A^{0-1}(\pi)H(\bar\theta(\pi))(\theta^0 - \theta^*(\pi))\| + o_{as}(d_n)$$

$$= \sup_{\pi\in\Pi} d_n\|B^{0-1/2}H(\theta^0)'A^{0-1}(\pi)h(\theta^*(\pi))\| + o_{as}(d_n)$$

where $\bar\theta(\pi) \in (\theta^0,\theta^*(\pi))$, and the last equality comes from the fact that $H(\bar\theta(\pi))(\theta^0 - \theta^*(\pi)) = o(d_n)$. The result then follows.

PROOF OF THEOREM TN2.

(i) Let

$$N_l(u) =$$
$$Min(m : \exists\pi_1,\pi_2,\ldots\pi_m;\ s.t.\ \forall\pi\in\Pi,\exists\pi_i\in\Pi,\ s.t.\ \|\nabla q_t(\theta^*(\pi),\pi) - \nabla_\theta q_t(\theta^*(\pi_i),\pi_i)\|_l \le u)$$

$N_l(u)$ is called the L^l bracketing number of $\nabla_\theta q_t(\theta^*(\pi),\pi)$ on Π. By TN1 we know that $\forall l \le p$, $p > 2$, there exists a $\Delta < \infty$, such that for $u > \Delta$, $N_l(u) = 1$. Thus condition (iii) in theorem 3.1 in Arcones (1995) can be replaced by

$$(iii')\int_0^\Delta (\log N_l(u))^{1/2}du < \infty$$

Given ATN17, (iii') is satisfied by the same argument followed in the proof of theorem 1 in Hansen's (1996). Given ATN12-ATN17, the a.s. asymptotic uniform equicontinuity of $b_n^{-1}\nabla_\theta\sum q_t(\theta^*(\pi),\pi)$ follows from theorem (3.1) in Arcones (1995). We need to show that

$$\sup_{\pi\in\Pi} |\hat V_n(\pi) - V^*(\pi)\| = o_{as}(1) \tag{21}$$

The result will then follow by the same argument as in Step 2 and 3 in the proof of Theorem TN1(i). By the same argument as above, given ATN18, the L^l bracketing number of

$Z_{t,\tau}^{i,j}(\theta^*(\pi),\pi)$ satisfies (iii' above). Given ATN12-ATN15, ATN18-ATN19, we know, from theorem 3.1 in Arcones (1995) that $b_n^{-1}\sum_{t=1}^{n-1} Z_{t,\tau}^{i,j}(\theta^*(\pi),\pi)$ is a.s. AUEC on Π. Now from ATN18, we know that $\forall\tau, \forall i, j, \forall\pi \in \Pi$,

$$\left(E\left(Z_{t,\tau}^{i,j}(\pi)\right)^2\right)^{1/2} = v_\tau^{ij}(\pi) \leq v^{i,j}(\pi) \leq v^{i,j}$$

Then, from lemma 2.2 in Arcones (1995), it follows that

$$\limsup_{n\to\infty} d_n^{-1}|n^{-1}\sum_{t=1}^{n-\tau} Z_{t,\tau}^{i,j}(\theta^*(\pi),\pi)| \leq v^{ij}$$

Then (21) follows from lemma 2 in Corradi (1996).

(i)-(b), (ii)(a)-(b) By a similar argument as (i)-(b) and (ii)-(a)(b) in Theorem TN1. □

8 References

Altissimo, F. and G.L.Violante (1996) " Persistence and nonlinearity in U.S. GNP and unemployment: an endogenous delay threshold VAR" , University of Pennsylvania, Economics, manuscript.

Andrews, D.W.K. (1992): "Generic uniform convergence", *Econometric Theory* , 8, 241-257.

Andrews D.W.K. (1993): "Tests for parameter instability and structural change with unknown change point", *Econometrica* , 61, 821-856.

Andrews, D.W.K. and W. Ploberger (1994): "Optimal tests when a nuisance parameter is present only under the alternative", *Econometrica* , 62, 1383-1414.

Arcones, M.A. (1995): "The law of the iterated logarithm for empirical processes under absolute regularity", *Journal of Theoretical Probability* , 8, 433-451.

Bai, J. and P. Perron (1995): "Estimating and testing linear models with multiple structural changes", MIT, Mimeo.

Berger, E. (1990): "An almost sure invariance principle for stationary-ergodic sequences of Banach space valued random variables", *Probability Theory and Related Fields*, 84, 161-201

Bierens, H.B. (1982): "Consistent model specification tests", *Journal of Econometrics*, 20, 105-134.

Bierens, H.B. (1990): "A conditional moment test of functional form", *Econometrica* , 58, 1443-1458

Bierens H.J. and W. Ploberger (1997), Asymptotic theory of integrated conditional moment tests, Mimeo, Pennsylvania State University.

Corradi, V. (1996): "Deciding between $I(1)$ and $I(0)$ via FLIL-based bounds", University of Pennsylvania, Economics, manuscript.

Eberlain, E. (1986): "On strong invariance principles under dependence assumptions", *Annals of Probability* 14, 260-270

Gallant, A.R and H. White (1988): A Unified Theory for Estimation and Inference for Nonlinear Econometric Models, Basil Blackwell, New York

Garcia, R. (1995): "Asymptotic null distribution of the likelihood ratio test in Markov switching models", University of Montreal, Economics, #9510

Hamilton, J.D. (1989): " A new approach to the analysis of nonstationaty time series and business cycle", *Econometrica*, 57, 357-384.

Hansen B.E. (1992) The likelihoodratio test under non-standard conditions: testing the Markov switching model of GNP, *Journal of Applied Econometrics*, 7, S62-S83.

Hansen, B.E. (1996): "Inference when a nuisance parameter is not identified under the null hypothesis", *Econometrica* , 413-430.

Lee, T.H., H. White and C.W.J. Granger (1993): "Testing for neglected nonlinearities in time-series models- a comparison of neural network methods and alternative tests", *Journal of Econometrics* , 56, 269-290.

McKean, H.P. (1969): Stochastic Integrals , Academic Press, New York

Petrucelli, J.D. and S.W. Woolford (1984): "A threshold AR(1) model" *Journal of Applied Probability*, 21, 207-286.

Pham, D.T. and L.T. Tran (1985): "Some mixing properties of time series models", *Stochastic Processes and their Applications*, 19, 297-303.

Phillips, P.C.B. and Ploberger W. (1994): "Posterior odds testing for a unit root with data-based model selection" *Econometric Theory* 10, 774-808.

Phillips, P.C.B. and W. Ploberger (1996): "An asymptotic theory of Bayesian inference for time series", *Econometrica* , 64, 381-412

Pötscher, B.M. and I.R. Prucha (1994): "Generic uniform convergence and equicontinuity concepts for random functions", *Journal of Econometrics* , 60, 23-63

Potter, S.M., (1995) "A nonlinear Approach to US GNP" *Journal of Applied Econometrics*, 10,109-125.

Sin, C.Y. and H. White (1995): "Information criteria for selecting possibly mispecified parametric models", *Journal of Econometrics* 71, 207-225.

Stinchcombe M.B. and H. White (1995): "Consistent specification testing with nuisance parameters present only under the alternative", UCSD Working Paper.

White, H. (1994): Estimation, Inference and Specification Analysis , Cambridge University Press.

Wooldridge, J. (1986): "Asymptotic properties of econometric estimators", University of California, San Diego, Doctoral Dissertation.