Corradi, Valentina; Swanson, Norman R.

Working Paper

Bootstrap Conditional Distribution Tests In the Presence of Dynamic Misspecification

Working Paper, No. 2003-11

Provided in Cooperation with:
Department of Economics, Rutgers University

Suggested Citation: Corradi, Valentina; Swanson, Norman R. (2003) : Bootstrap Conditional Distribution Tests In the Presence of Dynamic Misspecification, Working Paper, No. 2003-11, Rutgers University, Department of Economics, New Brunswick, NJ

This Version is available at:
https://hdl.handle.net/10419/23168

Bootstrap Conditional Distribution Tests In the Presence of Dynamic Misspecification*

Valentina Corradi
Queen Mary, University of London

Norman R. Swanson
Rutgers University

December 2001
this version: September 2003

Abstract

In this paper, we show the first order validity of the block bootstrap in the context of Kolmogorov type conditional distribution tests when there is dynamic misspecification and parameter estimation error. Our approach differs from the literature to date because we construct a bootstrap statistic that allows for dynamic misspecification under both hypotheses. We consider two test statistics; one is the CK test of Andrews (1997), and the other is in the spirit of Diebold, Gunther and Tay (1998). The limiting distribution of both tests is a Gaussian process with a covariance kernel that reflects dynamic misspecification and parameter estimation error. In order to provide valid asymptotic critical values we suggest an extention of the empirical process version of the block bootstrap to the case of non vanishing parameter estimation error. The findings from Monte Carlo experiments show that both statistics have good finite sample properties for samples as small as 500 observations.

JEL classification: C12, C22.
Keywords: block bootstrap, conditional distributions, conditional Kolmogorov tests, dynamic misspecification, parameter estimation error.

*Valentina Corradi, Department of Economics, Queen Mary-University of London, London E14NS, U.K., v.corradi@qmul.ac.uk; and Norman R. Swanson, Department of Economics, Rutgers University, New Brunswick, NJ 08901-1248 , U.S.A., nswanson@econ.rutgers.edu. We are grateful to the organizers (Jean-Marie Dufour and Benoit Perron), as well as the participants of the 2001 C.R.D.E. Conference on Resampling Methods in Econometrics, Université de Montréal for providing many useful comments and suggestions. Additionally, we would like to thank two anomymous referees, Walter Distaso, Marcelo Fernandes, Sílvia Gonçalves, Atsushi Inoue, Lutz Killian, Shinichi Sakata, Paolo Zaffaroni, and seminar participants at Brunel University, Cambridge University, ECARES-ULB, University of Nottingham, and University of Southampton for their helpful comments. Parts of this paper were written while Swanson was visiting the economics department at University of California, San Diego, and he would like to thank the econometrics group there for providing a stimulating research environment within which to work. Corradi gratefully acknowledges financial support from the ESRC (grant code R000230006) and Swanson has benefited from the support of Rutgers University in the form of a Research Council grant.

1 Introduction

In recent years there has been growing interest in providing tests for the correct specification of conditional distributions. One reason for this is that testing for the correct conditional distribution is equivalent to jointly evaluating many conditional features of a process, including the conditional mean, variance, and symmetry. Along these lines, Bai and Ng (2001) construct tests for conditional asymmetry. Just as importantly, these sorts of tests allow for the evaluation of predictive densities, thus generalizing the evaluation of point and interval forecasts.[1]

In this paper, we show the first order validity of the block bootstrap in the context of Kolmogorov type conditional distribution tests when there is dynamic misspecification and parameter estimation error. Our approach differs from the literature to date because we construct a bootstrap statistic that allows for dynamic misspecification under both hypotheses, rather than assuming correct dynamic specification under the null hypothesis. This difference between our approach and that taken elsewhere can be most easily motivated within the framework used by Diebold, Gunther and Tay (DGT: 1998), Bai (2003), and Hong (2002). In their paper, DGT use the probability integral transform (see e.g. Rosenblatt (1952)) to show that $F_t(y_t|\Im_{t-1}, \theta_0)$, is identically and independently distributed as a uniform random variable on $[0, 1]$, where $F_t(\cdot|\Im_{t-1}, \theta_0)$ is a parametric distribution with underlying parameter θ_0, y_t is the random variable of interest, and $\Im_{t-1}$ is the information set containing all "relevant" past information (see below for further discussion). They thus suggest using the difference between the empirical distribution of $F_t(y_t|\Im_{t-1}, \widehat{\theta}_T)$ and the $45°-$degree line as a measure of "goodness of fit", where $\widehat{\theta}_T$ is some estimator of θ_0. This approach has been shown to be very useful for financial risk management (see e.g. Diebold, Hahn and Tay (1998)), as well as for macroeconomic forecasting (see e.g. Diebold, Tay and Wallis (1998) and Clements and Smith (2000,2002)). Likewise, Bai (2003) proposes a Kolmogorov type test of $F_t(u|\Im_{t-1}, \theta_0)$ based on the comparison of $F_t(y_t|\Im_{t-1}, \widehat{\theta}_T)$ with the CDF of a uniform on $[0, 1]$. As a consequence of using estimated parameters, the limiting distribution of his test reflects the contribution of parameter estimation error and is not nuisance parameter free. To overcome this problem, Bai (2003) uses a novel approach based on a martingalization argument to construct a modified Kolmogorov test which has a nuisance parameter free limiting distribution. This test has power against violations of uniformity but not against violations of independence. Two features differentiate our approach from that taken in the above papers. First, we assume strict stationarity, while they do not. Second,

[1] A few recent contributions in the area of predictive evaluation include: Diebold and Mariano (1995), West (1996), Christoffersen (1998), McCracken (2000), White (2000), Chao, Corradi and Swanson (2001), Corradi, Swanson, and Olivetti (2001), and Clark and McCracken (2001).

we allow for dynamic misspecification under the null hypothesis, while they do not. While our approach is clearly less general because of the first feature, the second feature allows us to obtain asymptotically valid critical values even when the conditioning information set does not contain all of the relevant past history. More precisely, we are interested in testing for correct specification, given a particular information set which may or may not contain all of the relevant past information. This is relevant when a Kolmogorov test is constructed, as one is generally faced with the problem of defining $\Im_{t-1}$. If enough history is not included, then there may be dynamic misspecification. Additionally, finding out how much information (e.g. how many lags) to include may involve pre-testing, hence leading to a form of sequential test bias. By allowing for dynamic misspecification, we do not require such pre-testing. Another key feature of our approach concerns the fact that the limiting distribution of Kolmogorov type tests is affected by dynamic misspecification. Critical values derived under correct specification given $\Im_{t-1}$ are not in general valid in the case of correct specification given a subset of $\Im_{t-1}$. Consider the following example. Assume that we are interested in testing whether the conditional distribution of $y_t|y_{t-1}$ is $N(\alpha_1^\dagger y_{t-1}, \sigma_1)$. Suppose also that in actual fact the "relevant" information set has $\Im_{t-1}$ including both y_{t-1} and y_{t-2}, so that the true conditional model is $y_t|\Im_{t-1} = y_t|y_{t-1}, y_{t-2} = N(\alpha_1 y_{t-1} + \alpha_2 y_{t-2}, \sigma_2)$, where $\alpha_1^\dagger$ differs from α_1. In this case, we have correct specification with respect to the information contained in y_{t-1}; but we have dynamic misspecification with respect to y_{t-1}, y_{t-2}. Even without taking account of parameter estimation error, the critical values obtained assuming correct dynamic specification are invalid, thus leading to invalid inference. Stated differently, tests that are designed to have power against both uniformity and independence violations (i.e. tests that assume correct dynamic specification under H_0) will reject; an inference which is incorrect, at least in the sense that the "normality" assumption is *not* false. In summary, if one is interested in the particular problem of testing for correct specification for a given information set, then our approach is appropriate.

We consider two Kolmogorov type test statistics; one is the CK test of Andrews (1997), and the other is based on the arguments presented in DGT (1998), and is similar to the statistic proposed by Bai (2003). The limiting distribution of both tests is a Gaussian process with a covariance kernel that reflects dynamic misspecification and parameter estimation error.[2] Therefore, critical values are data dependent and cannot be tabulated. In addition to the generalized spectrum test mentioned above, Hong (2002) also proposes a test for uniformity that is robust to non independence, and that is based on the comparison of a kernel density estimator and the uniform density. His a test has a normal limiting distribution, but converges at a nonparametric rate. The tests suggested here converge instead at a parametric rate and do not require

[2] As we allow for possible dynamic mispecification, Bai's (2003) martingalization argument does not apply in our context.

the choice of the bandwidth, although nuisance parameters free limiting distributions do not obtain. With regard to the CK test, for the case of non vanishing parameter estimation error and independent observations, Andrews (1997) suggests a parametric bootstrap based on drawing observations from the distribution implied under the null, which is in turn evaluated at some given estimated parameters, conditional on observed covariates, say X_t. If our null is correct dynamic specification (i.e. if $X_t = \Im_{t-1}$), then we can still use Andrews' parametric bootstrap and draw observations from $F(y_t|X_t, \widehat{\theta}_T)$. However, if instead $X_t \subset \Im_{t-1}$, then the long run variance of the resampled statistic does not properly mimic the long run variance of the original statistic, thus leading to invalid asymptotic critical values. In the case of dependent observations and dynamic misspecification, but no parameter estimation error, we could almost straightforwardly apply an empirical process version of either the block bootstrap (see e.g. Buhlmann (1994), Naik-Nimbalkar and Rajarshi (1994) or Peligrad (1998)) or the stationary bootstrap of Politis and Romano (1994a,b), as the only difference is that we are evaluating conditional rather than marginal distributions. In the present context, though, we require a bootstrap that is valid for dependent observations, possible dynamic misspecification under both hypotheses, and non vanishing parameter estimation error. One possibility in this case is to use the conditional p-value approach of Corradi and Swanson (2002), which extends Inoue's (2001) approach to the case of parameter estimation error.[3] A drawback of this approach is that the simulated critical values under the alternative are of order $O(l)$, (where l plays the same role as the block length in the block bootstrap) and so the finite sample power can be somewhat low with small and medium size samples. Another possibility, which we examine in this paper, is the use of an extension of the empirical process version of the block bootstrap to the case of non vanishing parameter estimation error.

The rest of this paper is organized as follows. In Section 2 we describe our setup, and examine the asymptotic behavior of the two statistics. In Section 3 we show the first order validity of the block bootstrap in our context. The fourth section contains the results of a small Monte Carlo study, and Section 5 concludes. All proofs are contained in the appendix.

2 Setup and Asymptotic Behavior of the Tests

Before stating the hypotheses and defining the test statistics, it is worthwhile to sketch some examples of conditional distributions which are correctly specified for a given information set, but misspecified for a larger information set.

[3]Inoue (2001) in turn extends Hansen's (1996) conditional p-value approach to the case of non martingale difference scores.

Assume that $Y^t = (y_t, y_{t-1}, ..., y_{t-p+1})'$ is jointly elliptically distributed; then the density can be expressed as (see e.g. Ingersoll (1987), Ch.4, Appendix B)

$$f_p(y) = |\Sigma|^{-1/2} g\left((y-\mu)' \Sigma^{-1}(y-\mu); p\right), \tag{1}$$

where μ is the mean vector and Σ is a positive definite matrix, proportional to the covariance matrix if the latter exists. Then, by Cambanis, Huang and Simons (1981), for any $m < p$, the distribution of $y_t, y_{t-1}, ..., y_{t-m}$ conditional on $y_{t-m-1}, ..., y_{t-m-p+1}$ belongs to the same elliptical family as in (1). Furthermore, for any $m < p$, the distribution of $y_t, y_{t-1}, ..., y_{t-m}$ is given by

$$f_m(y_1) = |\Sigma_1|^{-1/2} g\left((y_1-\mu_1)' \Sigma_1^{-1}(y_1-\mu_1); m\right),$$

and also $y_t | y_{t-1}, ..., y_{t-m}$ is elliptically distributed, with a density belonging to the same family as in (1).

We can now specialize the argument above to two well known elliptic distributions, the normal and the Student's-t. Suppose,

$$\text{(a)} \begin{pmatrix} y_t \\ y_{t-1} \\ y_{t-2} \end{pmatrix} \sim N(\mu, \Sigma) \text{ and (b)} \begin{pmatrix} y_t \\ y_{t-1} \\ y_{t-2} \end{pmatrix} \sim St(\mu, \Sigma, v),$$

where $v > 2$ denotes the degree of freedom, and

$$\mu = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \ \Sigma = \begin{pmatrix} \sigma^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma^2 \end{pmatrix}.$$

Also, let $\Sigma_{22} = \begin{pmatrix} \sigma^2 & \sigma_{23} \\ \sigma_{23} & \sigma^2 \end{pmatrix}$, $\Sigma_{12} = (\sigma_{12}, \sigma_{13})'$, and $\Sigma_{21} = \Sigma_{12}'$. Now, in the normal and Student's-t cases, respectively,

$$y_t | y_{t-1}, y_{t-2} \sim N\left(\Sigma_{12}\Sigma_{22}^{-1} \begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix}, \left(\sigma^2 - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)\right), \text{ and}$$

$$y_t | y_{t-1}, y_{t-2} \sim St\left(\Sigma_{12}\Sigma_{22}^{-1} \begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix}, \left(\frac{v}{v-2+2}\left(1 + \begin{pmatrix} y_{t-1} & y_{t-2} \end{pmatrix} \Sigma_{22}^{-1} \begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix}\right)\left(\sigma^2 - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)\right); v\right).$$

Because of ellipticity, it also follows that

$$\begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma_{12} \\ \sigma_{12} & \sigma^2 \end{pmatrix}\right),$$

$$\begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix} \sim St\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma_{12} \\ \sigma_{12} & \sigma^2 \end{pmatrix}; v\right),$$

and (see e.g. Spanos (1999)), for the normal and Student's t cases, respectively,

$$y_t | y_{t-1} \sim N\left(\frac{\sigma_{12}}{\sigma^2} y_{t-1}, \left(\sigma^2 - \frac{\sigma_{12}^2}{\sigma^2}\right)\right), \text{ and} \tag{2}$$

$$y_t|y_{t-1} \sim St\left(\frac{\sigma_{12}}{\sigma^2}y_{t-1}, \left(\frac{v}{v-1}\left(1+\frac{y_{t-1}^2}{\sigma^2}\right)\left(\sigma^2-\frac{\sigma_{12}^2}{\sigma^2}\right)\right); v\right). \tag{3}$$

Therefore, (2) and (3) are correctly specified for the distribution $y_t|y_{t-1}$, but are misspecified for the distribution of $y_t|y_{t-1}, y_{t-2}$. Given (3), a joint t-distribution seems appropriate when the conditional mean is a linear function and the conditional variance is a quadratic function of the conditioning variables. For example, financial series characterized by conditional heteroskedasticity and heavy tails might be suitably modelled using a multivariate t distribution with "few" degrees of freedom.

Turning now to the main topic of this section, assume that we have a sample $\{y_t, X_t\}_{t=1}^T$ from an underlying (stationary) process $\{y_t, X_t\}_{t=-\infty}^{\infty}$. We wish to test whether the distribution of $y_t|X_t$ has a given parametric form, where X_t may contains lags of y_t and/or lags of some other variables. Hereafter, let $\Im_{t-1}$ denote the information set containing all relevant past information, such that for any set $G_{t-1} \supset \Im_{t-1}$, $Y|\Im_{t-1} \equiv Y|G_{t-1}$, where $\equiv$ denotes equality in distribution. We define dynamic misspecification as the case in which $X_t \subset \Im_{t-1}$ and $y_t|X_t$ not$\equiv$ $y_t|\Im_{t-1}$, where "not$\equiv$ " denotes non equality in distribution.

The null and alternative hypotheses of interest are:

$$H_0 : \Pr(y_t \leq y|X_t, \theta_0) = F(y|X_t, \theta_0), \text{ a.s. for some } \theta_0 \in \Theta \tag{4}$$

$$H_A : \text{ the negation of } H_0 \tag{5}$$

The DGT type test is:

$$V_{1T} = \sup_{r\in[0,1]} |V_{1T}(r)|, \tag{6}$$

where,

$$V_{1T}(r) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(1\{F(y_t|X_t, \widehat{\theta}_T) \leq r\} - r\right),$$

and

$$\widehat{\theta}_T = \arg\max_{\theta\in\Theta}\frac{1}{T}\sum_{t=1}^{T}\ln f(y_t|X_t, \theta).$$

The standard rationale underlying the above test, which is known to hold when $X_t = \Im_{t-1}$, is that under H_0, $F(y_t|X_t, \theta_0)$ is distributed independently and uniformly on $[0, 1]$. The uniformity result also holds under dynamic misspecification. To see this, let $c_f^r(X_t)$ be the $r-th$ critical value of $f(\cdot|X_t, \theta_0)$, where f is the density associated with $F(\cdot|X_t, \theta_0)$ (i.e. the conditional distribution under the null)[4]. We have,

$$\begin{aligned}\Pr(F(y_t|X_t, \theta_0) &\leq r) = \Pr\left(\int_{-\infty}^{y_t} f(y|X_t, \theta_0)dy \leq r\right) \\ &= \Pr\left(1\{y_t \leq c_f^r(X_t)\} = 1|X_t\right) = r, \text{ for all } r \in [0, 1],\end{aligned}$$

[4]For example, if $f(Y|X_t, \theta_0) \sim N(\alpha X_t, \sigma^2)$, then $c_f^{0.95}(X_t) = 1.645 + \sigma\alpha X_t$.

if $y_t|X_t$ has density $f(\cdot|X_t,\theta_0)$. Now, if the density of $y_t|\ X_t$ is different from $f(\cdot|X_t,\theta_0)$, then,

$$\Pr\left(1\{y_t \leq c_f^r(X_t)\} = 1|X_t\right) \neq r$$

for some r with nonzero Lebesgue measure on $[0,1]$. However, under dynamic misspecification, $F(y_t|X_t,\theta_0)$ is no longer independent (or even martingale difference), in general, and this will clearly affect the covariance structure of the limiting distribution of the statistic.[5]

Notice that, if the statistic V_{1T} in (6) is constructed using the residuals from a regression, then we cannot rule out the possibility that uniformity is not violated even if $F(y|X_t,\theta_0)$ is not correctly specified for the conditional distribution of $y_t|X_t$. The intuitive reason for this is that we might misspecify the conditional mean and/or conditional variance and/or the marginal distribution of the "true" error, but still get the right marginal for the "wrong" error. Broadly speaking, if we correctly specify the marginal of $(y_t - \mu(y_{t-1};\theta^\dagger))/\sigma^2(y_{t-1};\theta^\dagger)$, say, then uniformity is not violated even if $\mu(y_{t-1};\theta^\dagger)$ and/or $\sigma^2(y_{t-1};\theta^\dagger)$ are not correctly specified for the conditional mean and variance respectively.

The other test that we examine is the CK test of Andrews (1997).[6] This test is constructed by comparing the empirical joint distribution of y_t and X_t with the product of the distribution of $y_t|X_t$ and the empirical CDF of X_t. In practice, the empirical joint distribution, say $\widehat{H}_T(u,v) = \frac{1}{T}\sum_{t=1}^T 1\{y_t \leq u\}1\{X_t < v\}$, and the semi-empirical/semi-parametric analog of $F(u,v,\theta_0)$, say $\widehat{F}_T(u,v,\widehat{\theta}_T) = \frac{1}{T}\sum_{t=1}^T F(u|X_t,\widehat{\theta}_T)1\{X_t < v\}$ are used, and the test statistic is:

$$V_{2T} = \sup_{u\times v \in U\times V} |V_{2T}(u,v)|, \tag{7}$$

where U and V are compact subsets of $\Re$ and $\Re^d$, respectively, and

$$V_{2T}(u,v) = \frac{1}{\sqrt{T}}\sum_{t=1}^T \left((1\{y_t \leq u\} - F(u|X_t,\widehat{\theta}_T))1\{X_t \leq v\}\right).$$

Note that V_{2T} is given in equation (3.9) of Andrews (1997).[7] Note also that when computing this statistic, a grid search over $U \times V$ may be computationally demanding when V is high-dimensional. To avoid this

[5]The above explanation of the uniformity result is rather long-winded, although we still include it for the sake of clarity. As pointed out by an anonymous referee, a simpler way of stating the result is to simply note that if Z has a continuous CDF, say $F(z)$, then $F(Z)$ is uniform. This is universally true for any distribution, including conditional distributions, for example.

[6]In a related paper, Li and Tkacz (2004) discuss an interesting approach to testing for correct specification of the conditional density which involves comparing a nonparametric kernel estimate of the conditional density with the density implied under the null hypothesis. Also, Whang (2000,2001) proposed a version of Andrews CK test for the correct specification of the conditional mean.

[7]Andrews (1997), for the case of *iid* observations, actually addresses the more complex situation where U and V are unbounded sets in R and R^d, respectively. We believe that an analogous result for the case of dependent observations holds, but showing this involves proofs for stochastic equicontinuity in Theorems 2 and 4 which are much more demanding than those included in this paper. This subject is left to future research.

problem, Andrews shows that when all (u, v) combinations are replaced with (y_t, X_t) combinations, the resulting test is asymptotically equivalent to $V_{2T}(u, v)$.

Theorems 1 and 2 below rely on the following assumptions.

Assumption A1: (y_t, X_t), are jointly strictly stationary and strong mixing with size $-4(4+\psi)/\psi$, $0 < \psi < 1/2$.

Assumption A2: (i) $F(y_t|X_t, \theta)$ is twice continuously differentiable on the interior of $\Theta \subset R^p$, Θ compact; (ii) $E(\sup_{\theta\in\Theta} |\nabla_\theta F(y_t|X_t, \theta)_i|^{5+\psi}) \leq C < \infty$, $i = 1, ..., p$, where ψ is the same positive constant defined in A1, and $\nabla_\theta F(y_t|X_t, \theta)_i$ is the $i-$th element of $\nabla_\theta F(y_t|X_t, \theta)$; (iii) $F(u|X_t, \theta)$ is twice differentiable on the interior of $U \times \Theta$, where U and Θ are compact subsets of $\Re$ and $\Re^p$ respectively; and (iv) $\nabla_\theta F(u|X_t, \theta)$ and $\nabla_{u,\theta} F(u|X_t, \theta)$ are jointly continuous on $U \times \Theta$ and $4s-$dominated on $U \times \Theta$ for $s > 3/2$.

Assumption A3: (i) $\theta^\dagger = \arg\max_{\theta\in\Theta} E(\ln f(y_1|X_1, \theta))$ is uniquely identified, (ii) $f(y_t|X_t, \theta)$ is twice continuously differentiable in θ in the interior of Θ, (ii) the elements of $\nabla_\theta \ln f(y_t|X_t, \theta)$ and of $\nabla^2_\theta \ln f(y_t|X_t, \theta)$ are $4s-$dominated on Θ, with $s > 3/2$, $E\left(-\nabla^2_\theta \ln f(y_t|X_t, \theta)\right)$ is positive definite uniformly in Θ.[8]

A2 imposes mild smoothness and moment restrictions on the cumulative distribution function under the null, and is thus easily verifiable. We use A2(i)-(ii) in the study of the limiting behavior of V_{1T} and A2(iii)-(iv) in the study of V_{2T}.[9]

Theorem 1: Let A1, A2(i)–(ii) and A3 hold. Then: (i) Under H_0, $V_{1T} \Rightarrow \sup_{r\in[0,1]} |Z(r)|$, where Z is a zero mean Gaussian process with covariance kernel $K_1(r, r')$ given by:

$$E(Z(r)Z(r')) = K_1(r, r') = E(\sum_{s=-\infty}^{\infty} (1\{F(y_0|X_0, \theta_0) \leq r\} - r)\,(1\{F(y_s|X_s, \theta_0) \leq r'\} - r'))$$

$$+E(\nabla_\theta F(y_t|X_t, \theta_0))' A(\theta_0) \sum_{s=-\infty}^{\infty} E(q_0(\theta_0) q_s(\theta_0)') A(\theta_0) E(\nabla_\theta F(y_t|X_t, \theta_0))$$

$$-2E(\nabla_\theta F(y_t|X_t, \theta_0))' A(\theta_0) \sum_{s=-\infty}^{\infty} E((1\{F(y_0|X_0, \theta_0) \leq r\} - r)\, q_s(\theta_0)')$$

[8] Let $\nabla_\theta \ln f(y_t|X_t, \theta)_i$ be the $i-th$ element of $\nabla_\theta \ln f(y_t|X_t, \theta)$. For $4s-$domination on Θ, we require $|\nabla_\theta \ln f(y_t|X_t, \theta)_i| \leq m(X_t)$, for all i, with $E((m(X_t))^{4s}) < \infty$, for some function m.

[9] For the case of the CK test, a "natural" estimator can be defined as

$$\widetilde{\theta}_T = \arg\min_{\theta\in\Theta} \int_U \int_V \frac{1}{T} \sum_{t=1}^{T} ((1\{y_t \leq u\} - F(u|X_t, \theta))\, 1\{X_t \leq v\})^2 \, dudv$$

Under mild regularity conditions, $\widetilde{\theta}_T \xrightarrow{pr} \theta^\ddagger$, where $\theta^\ddagger$ satisfies the moment conditions

$$\int_U \int_V E\left(\left(1\{y_t \leq u\} - F(u|X_t, \theta^\ddagger)\right) 1\{X_t \leq v\} \nabla_\theta F(u|X_t, \theta^\ddagger)\right) = 0.$$

with $q_s(\theta_0) = \nabla_\theta \ln f_s(y_s|X_s,\theta_0)$, and $A(\theta_0) = \left(E\left(\nabla_\theta q_s(\theta_0)\nabla_\theta q_s(\theta_0)'\right)\right)^{-1}$.

(ii) Under H_A, there exists an $\varepsilon > 0$ such that $\lim_{T\to\infty} \Pr(\frac{1}{T^{1/2}} V_{1T} > \varepsilon) = 1$.

Theorem 2: Let A1, A2(iii)-(iv) and A3 hold. Then: (i) Under H_0, $V_{2T} \Rightarrow \sup_{u\times v \in U\times V} |Z(u,v)|$, where V_{2T} is defined in (7) and Z is a zero mean Gaussian process with covariance kernel $K_2(u,v,u',v')$ given by:

$$E(\sum_{s=-\infty}^{\infty}((1\{y_0 \le u\} - F(u|X_0,\theta_0))1\{X_0 \le v\})((1\{y_s \le u'\} - F(u|X_s,\theta_0))1\{X_s \le v'\}))$$

$$+E(\nabla_\theta F(u|X_0,\theta_0)'1\{X_0 \le v\})A(\theta_0)\sum_{s=-\infty}^{\infty} q_0(\theta_0)q_s(\theta_0)'A(\theta_0)E(\nabla_\theta F(u'|X_0,\theta_0)1\{X_0 \le v'\})$$

$$-2\sum_{s=-\infty}^{\infty}((1\{y_0 \le u\} - F(u|X_0,\theta_0))1\{X_0 \le v\})E(\nabla_\theta F(u'|X_0,\theta_0)'1\{X_0 \le v'\})A(\theta_0)q_s(\theta_0)).$$

(ii) Under H_A, there exists an $\varepsilon > 0$ such that $\lim_{T\to\infty} \Pr(\frac{1}{T^{1/2}} V_{2T} > \varepsilon) = 1$.

Notice that in both cases, the limiting distribution is a zero mean Gaussian process, with a covariance kernel that reflects both dynamic misspecification as well as the contribution of parameter estimation error. Thus, the limiting distribution is not nuisance parameter free and so critical values cannot be tabulated. In the next section we outline a bootstrap procedure that takes into account the joint presence of parameter estimation error and possible dynamic misspecification.

3 Validity of the Block Bootstrap

Given that the limiting distributions of V_{1T} and V_{2T} are not nuisance parameters free, our approach is to construct bootstrap critical values for the tests. In order to show the first order validity of the bootstrap, we shall obtain the limiting distribution of the bootstrapped statistic and show that it coincides with the limiting distribution of the actual statistic under H_0. Then, a test with correct asymptotic size and unit asymptotic power can be obtained by comparing the value of the original statistic with bootstrapped critical values.

As discussed above, if the data consists of *iid* observations, we should consider proceeding along the lines of Andrews (1997), by drawing B samples of T *iid* observations from the distribution under H_0, conditional on the observed values for the covariates, X_t. The same approach could also be used in the case of dependence, if H_0 were correct dynamic specification, (i.e. if $X_t = \Im_{t-1}$); in fact, in that case we could use a parametric bootstrap and draw observations from $F(y_t|X_t,\widehat{\theta}_T)$. However, if instead $X_t \subset \Im_{t-1}$, using the parametric bootstrap procedure based on drawing observations from $F(y_t|X_t,\widehat{\theta}_T)$ does not ensure that the

long run variance of the resampled statistic properly mimics the long run variance of the original statistic; thus leading in general to the construction of invalid asymptotic critical values.

Our approach is to compare the empirical CDF of the resampled series, evaluated at the bootstrap estimator, with the empirical CDF of the actual series, evaluated at the estimator based on the actual data. For this, we use the overlapping block resampling scheme of Künsch (1989), as follows:[10] At each replication, draw b blocks (with replacement) of length l from the sample $W_t = (y_t, X_t)$, where $T = lb$. Thus, the first block is equal to $W_{i+1}, ..., W_{i+l}$, for some i, with probability $1/(T - l + 1)$, the second block is equal to $W_{i+1}, ..., W_{i+l}$, for some i, with probability $1/(T - l + 1)$, and so on for all blocks. More formally, let I_k, $k = 1, ..., b$ be *iid* discrete uniform random variables on $[0, 1, ..., T - l]$, and let $T = bl$. Then, the resampled series, $W_t^* = (y_t^*, X_t^*)$, is such that $W_1^*, W_2^*, ..., W_l^*, W_{l+1}^*, ..., W_T^* = W_{I_1+1}, W_{I_1+2}, ..., W_{I_1+l}, W_{I_2}, ..., W_{I_b+l}$, and so a resampled series consists of b blocks that are discrete *iid* uniform random variables, conditional on the sample. Also, let $\widehat{\theta}_T^*$ be the estimator constructed using the resampled series. For V_{1T}, the bootstrap statistic is:

$$V_{1T}^* = \sup_{r \in [0,1]} |V_{1T}^*(r)|,$$

where

$$V_{1T}^*(r) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(1\{F(y_t^*|X_t^*, \widehat{\theta}_T^*) \leq r\} - 1\{F(y_t|X_t, \widehat{\theta}_T) \leq r\} \right), \tag{8}$$

and

$$\widehat{\theta}_T^* = \arg\max_{\theta \in \Theta} \frac{1}{T} \sum_{t=1}^{T} \ln f(y_t^*|X_t^*, \theta).$$

The rationale behind the choice of (8) is the following. By a mean value expansion we can show (see the appendix) that,

$$\begin{aligned} V_{1T}^*(r) &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(1\{F(y_t^*|X_t^*, \theta^\dagger) \leq r\} - 1\{F(y_t|X_t, \theta^\dagger) \leq r\} \right) \\ &\quad - \frac{1}{T} \sum_{t=1}^{T} \nabla_\theta F(y_t|X_t, \theta^\dagger) \sqrt{T}(\widehat{\theta}_T^* - \widehat{\theta}_T) + o_{P^*}(1), \ \Pr - P, \end{aligned} \tag{9}$$

[10] The main difference between the block bootstrap and the stationary bootstrap of Politis and Romano (PR: 1994a) is that the former uses a deterministic block length, which may be either overlapping as in Künsch (1989) or non-overlapping as in Carlstein (1986), while the latter resamples using blocks of random length. One important feature of the PR bootstrap is that the resampled series, conditional on the sample, is stationary, while a series resampled from the (overlapping or non overlapping) block bootstrap is nonstationary, even if the original sample is strictly stationary. However, Lahiri (1999) shows that all block boostrap methods, regardless of whether the block length is deterministic or random, have a first order bias of the same magnitude, but the bootstrap with deterministic block length has a smaller first order variance. In addition, the overlapping block boostrap is more efficient than the non overlapping block bootstrap.

where P^* denotes the probability law of the resampled series, conditional on the sample; P denotes the probability law of the sample; and where "$o_{P^*}(1)$, $\Pr - P$ ", means a term approaching zero according to P^*, conditional on the sample and for all samples except a set of measure approaching zero. Now, the first term on the RHS of (9) can be treated via the empirical process version of the block bootstrap, suggesting that the term has the same limiting distribution as $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(1\{F(y_t|X_t,\theta^\dagger)\leq r\} - E\left(1\{F(y_t|X_t,\theta^\dagger)\leq r\}\right)\right)$, where $E\left(1\{F(y_t|X_t,\theta^\dagger)\leq r\}\right) = r$ under H_0, and is different from r under H_A, conditional of the sample. If $\sqrt{T}(\widehat{\theta}_T^* - \widehat{\theta}_T)$ has the same limiting distribution as $\sqrt{T}(\widehat{\theta}_T - \theta^\dagger)$, conditionally on the sample and for all samples except a set of measure approaching zero, then the second term on the RHS of (9) will properly capture the contribution of parameter estimation error to the covariance kernel. For the case of dependent observations, the limiting distribution of $\sqrt{T}(\widehat{\theta}_T^* - \widehat{\theta}_T)$ for a variety of quasi maximum likelihood (QMLE) and GMM estimators has been examined in numerous papers in recent years.

For example, Hall and Horowitz (1996) and Andrews (2002) show that the block bootstrap provides improved critical values, in the sense of asymptotic refinement, for "studentized" GMM estimators and for tests of overidentifying restrictions, in the case where the covariance across moment conditions is zero after a given number of lags.[11] In addition, Inoue and Shintani (2004) show that the block bootstrap provides asymptotic refinements for linear overidentified GMM estimators for general mixing processes. In the present context, however, we cannot "studentize" the statistic, and we are thus unable to show second order refinement, as mentioned above. Instead, and again as mentioned above, our approach is to show first order validity of $\sqrt{T}(\widehat{\theta}_T^* - \widehat{\theta}_T)$. An important recent contribution which is useful in our context is that of Goncalves and White (2002a,b), who show that for QMLE estimators, the limiting distribution of $\sqrt{T}(\widehat{\theta}_T^* - \widehat{\theta}_T)$ provides a valid first order approximation to that of $\sqrt{T}(\widehat{\theta}_T - \theta^\dagger)$ for heterogeneous and near epoch dependent series.

Theorem 3: Let A1, A2(ii)–(iii) and A3 hold, and let $T = bl$, with $l = l_T$, such that as $T \to \infty$, $l_T^2/T \to 0$. Then,

$$P\left(\omega : \sup_{x\in\Re}\left|P^*\left[V_{1T}^*(\omega)\leq u\right] - P\left[\sup_{r\in[0,1]}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(1\{F(y_t|X_t,\widehat{\theta}_T)\leq r\} - E\left(1\{F(y_t|X_t,\theta^\dagger)\leq r\}\right)\right)\leq x\right]\right| > \varepsilon\right) \to 0.$$

Thus, V_{1T}^* has a well defined limiting distribution under both hypotheses, which under the null coincides with the same limiting distribution of V_{1T} , Pr - P, as $E(1\{F(y_t|X_t,\theta^\dagger)\leq r\}) = r$. Now, define $V_{2T}^* =$

[11] Andrews (2002) shows first order validity and asymptotic refinements of the equivalent $k-$step estimator of Davidson and MacKinnon (1999), which only requires the construction of a closed form expression at each bootstrap replication, thus avoiding nonlinear optimization at each replication.

$\sup_{u\times v\in U\times V}|V_{2T}^*(u,v)|$, where

$$V_{2T}^*(u,v)=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big((1\{y_t^*\leq u\}-F(u|X_t^*,\widehat{\theta}_T^*))1\{X_t^*\leq v\}-(1\{y_t\leq u\}-F(u|X_t,\widehat{\theta}_T))1\{X_t\leq v\}\Big).$$

Theorem 4: Let A, A2(iii)–(iv) and A3 hold, and let $T=bl$, with $l=l_T$, such that as $T\to\infty$, $l_T^2/T\to 0$. Then,

$$\begin{aligned}&P\Bigg(\omega:\sup_{x\in\Re}|P^*[V_{2T}^*(\omega)\leq x]\\&-\ P\left[\sup_{u\times v\in U\times V}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}((1\{y_t\leq u\}-F(u|X_t,\widehat{\theta}_T))1\{X_t\leq v\}-E((1\{y_t\leq u\}-F(u|X_t,\theta^\dagger))1\{X_t\leq v\}))\leq x\right]\Bigg|>\varepsilon\Bigg)\\\to\ &0.\end{aligned}$$

In summary, from Theorems 3 and 4, we know that $V_{1T}^*(\omega)$ (resp. $V_{2T}^*(\omega)$) has a well defined limiting distribution, conditional on the sample and for all samples except a set of probability measure approaching zero. Furthermore, the limiting distribution coincides with that of V_{1T} (resp. V_{2T}), under H_0. The above results suggest proceeding in the following manner. For any bootstrap replication, compute the bootstrapped statistic, V_{1T}^* (resp. V_{2T}^*). Perform B bootstrap replications (B large) and compute the percentiles of the empirical distribution of the B bootstrapped statistics. Reject H_0 if V_{1T} (V_{2T}) is greater than the $(1-\alpha)th$-percentile. Otherwise, do not reject H_0. Now, for all samples except a set with probability measure approaching zero, V_{1T} (V_{2T}) has the same limiting distribution as the corresponding bootstrapped statistic, under H_0. Thus, the above approach ensures that the test has asymptotic size equal to α. Under the alternative, V_{1T} (V_{2T}) diverges to infinity, while the corresponding bootstrap statistic has a well defined limiting distribution. This ensures unit asymptotic power. Note that the validity of the bootstrap critical values is based on an infinite number of bootstrap replications, although in practice we need to choose B. Andrews and Buchinsky (2000) suggest an adaptive rule for choosing B, Davidson and McKinnon (2000) suggest a pretesting procedure ensuring that there is a "small probability" of drawing different conclusions from the ideal bootstrap and from the bootstrap with B replications, for a test with a given level. However, in our case, the limiting distribution is a functional of a Gaussian process, so that we do not know the explicit density function; and thus we cannot directly apply the approaches suggested in the papers above. In the Monte Carlo section below, we analyze the robustness of our findings to the choice of B, and find that even for values of B as small as 100, the bootstrap has good finite sample properties.

Needless to say, if the parameters are estimated using T observations, and the statistic is constructed using only R observations, with $R=o(T)$, then the contribution of parameter estimation error to the

covariance kernel is asymptotically negligible. In this case, we do not need to compute $\widehat{\theta}_T^*$. For example, when bootstrapping critical values for a statistic analogous to V_{1T}, but constructed using R observations, say V_{1R}, we could instead construct V_{1R}^* as follows:

$$V_{1R}^* = \sup_{r\in[0,1]} \frac{1}{\sqrt{R}} \sum_{t=1}^{R} \left(1\{F(y_t^*|X_t^*, \widehat{\theta}_T) \leq r\} - 1\{F(y_t|X_t, \widehat{\theta}_T) \leq r\} \right). \tag{10}$$

The intuition for this statistic is that $\sqrt{R}(\widehat{\theta}_T - \theta^\dagger) = o_p(1)$, and so we do not need to use the bootstrap estimator of θ to mimic the distribution of $\sqrt{T}(\widehat{\theta}_T - \theta^\dagger)$. Analogs of V_{1R} and V_{1R}^* can similarly be constructed for V_{2T}. We do not suggest using this approach because of the cost to finite sample power, and also because we do not have an adaptive, data-driven rule for choosing R.

4 Monte Carlo Results

In this section we report the results of a small Monte Carlo study of V_{1T} and V_{2T}. Data are generated according to the following processes:

Size Experiments:

*Size*1: Generate $y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \epsilon_t = 0.2y_{t-1} + \epsilon_t$, $\epsilon_t \sim iidN(0, \sigma_\epsilon^2)$ with $\sigma_\epsilon^2 = 1$. Estimate an AR(1)model..

*Size*2: Generate $y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \epsilon_t = 0.2y_{t-1} + 0.2y_{t-2} + \epsilon_t$, $\epsilon_t \sim iidN(0, \sigma_\epsilon^2)$ with $\sigma_\epsilon^2 = 1$. Estimate an AR(1)model.

The null hypothesis is that $y_t|y_{t-1} \sim N(\alpha_1^\dagger y_{t-1}, \sigma_\epsilon^{\dagger,2})$; that is $H_0 : \Pr(y_t \leq y|y_{t-1}, \theta_0) = \frac{1}{\sqrt{2\pi}\sigma_\epsilon^\dagger} \int_{-\infty}^{y} \exp((u - \alpha_1^\dagger y_{t-1})^2/2\sigma_\epsilon^{\dagger,2})du$, while the alternative is the negation of H_0. Notice that we allow for dynamic misspecification in *Size*2, as $\Pr(y_t \leq y|y_{t-1}, \theta_0) \neq \Pr(y_t \leq y|y_{t-1}, y_{t-2}, \theta^0)$, where $\theta_0 \neq \theta^0$. On the other hand, in *Size*1, the AR(1) model is dynamically correctly specified. As $y_t|y_{t-1}$ is conditionally normal in the size experiments, the parameter in the conditional mean expression can be consistently estimated by least squares (i.e. $\widehat{\alpha}_T = \frac{\sum_{t=1}^{T} y_t y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2}$), and a consistent estimator for $\sigma_\epsilon^{\dagger,2}$ is $\widehat{\sigma}_{\epsilon,T}^2 = \frac{1}{T}\sum_{t=1}^{T}(y_t - \widehat{\alpha}_T y_{t-1})^2$. Also, $\widehat{\alpha}_T^*$ and $\widehat{\sigma}_{\epsilon,T}^{2*}$ are defined in an analogous way, using appropriately resampled observations (as discussed above).

Power Experiments:

*Power*1: Generate $y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \epsilon_t = 0.2y_{t-1} + 0.2y_{t-2} + \epsilon_t$, $\epsilon_t \sim iid\ t_3$, and estimate an AR(1) model.

*Power*2: Generate $y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \epsilon_t = 0.2y_{t-1} + 0.2y_{t-2} + \epsilon_t$, $\epsilon_t \sim iid\ t_5$, and estimate an AR(1) model.

*Power*3: Generate $y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-1} 1\{y_{t-1} \geq 0\} + \epsilon_t = 0.2 y_{t-1} + 0.2 y_{t-1} 1\{y_{t-1} \geq 0\} + \epsilon_t$, $\epsilon_t \sim iidN(0,1)$, and estimate an AR(1) model.

*Power*4: Generate $y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-1} 1\{y_{t-1} \geq 0\} + \epsilon_t = 0.2 y_{t-1} + 0.4 y_{t-1} 1\{y_{t-1} \geq 0\} + \epsilon_t$, $\epsilon_t \sim iidN(0,1)$, and estimate an AR(1) model.

*Power*5: Generate $y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-1} 1\{y_{t-1} \geq 0\} + \epsilon_t = 0.2 y_{t-1} + 0.4 y_{t-1} 1\{y_{t-1} \geq 0\} + \epsilon_t$, $\epsilon_t \sim iid\ t_3$, and estimate an AR(1) model.

*Power*6: Generate $y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-1} 1\{y_{t-1} \geq 0\} + \epsilon_t = 0.2 y_{t-1} + 0.4 y_{t-1} 1\{y_{t-1} \geq 0\} + \epsilon_t$, $\epsilon_t \sim iid\ t_5$, and estimate an AR(1) model.

Notice that in these experiments, rejection arises because of misspecification of the marginal distribution of ϵ_t (*Power*1 and *Power*2), misspecification of the conditional mean (*Power*3 and *Power*4), and misspecification of the marginal distribution of ϵ_t and of the conditional mean (*Power*5 and *Power*6).

For both size and power experiments, we construct the statistics

$$V_{1T} = \sup_{r \in [0,1]} \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(1\left\{ \int_{-\infty}^{y_t} \frac{1}{\sqrt{2\pi \widehat{\sigma}^2_{\epsilon,T}}} \exp\left(\frac{-(y - \widehat{\alpha}_T y_{t-1})^2}{2\widehat{\sigma}^2_{\epsilon,T}} \right) dy \leq r \right\} - r \right) \right|$$

and

$$V_{2T} = \sup_{u \times v \in U \times V} \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(\left(1\{y_t \leq u\} - \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi \widehat{\sigma}^2_{\epsilon,T}}} \exp\left(\frac{-(y - \widehat{\alpha}_T y_{t-1})^2}{2\widehat{\sigma}^2_{\epsilon,T}} \right) dy \right) 1\{X_t \leq v\} \right) \right|,$$

The bootstrap statistics are constructed in an analogous manner using

$$F(y_t^* | X_t^*, \widehat{\theta}_T^*) = \int_{-\infty}^{y_t} \frac{1}{\sqrt{2\pi \widehat{\sigma}^{2*}_{\epsilon,T}}} \exp\left(\frac{-(y - \widehat{\alpha}_T^* y_{t-1}^*)^2}{2\widehat{\sigma}^{2*}_{\epsilon,T}} \right).$$

Rejection frequencies for experiments with $B = 100$, $T = 500, 1000$, and $l = 25, 40, 50$ are reported in Tables 1 (V_{1T}) and 2 (V_{2T}).[12] Notice first that empirical level figures are rather close to the respective nominal levels of the tests, regardless of the value of l, and regardless of whether there is dynamic misspecification (*Size*2) or not (*Size*1). Additionally, empirical level is neither consistently above or consistently below nominal level. This may be due to the fact that only 500 Monte Carlo replications are carried out and B is set equal to 100. As expected, empirical level is in general closer to nominal level when 1000 instead of 500

[12]Further results for different values of B, T, and l are available upon request from the authors. Although the choice of block length is important, and there exist data-dependent methods for block length selection, they may not work in our context, for the reasons discussed above and in Inoue (2001). For this reason, we simply use $l = 25$, 40 and 50. Due to computational costs, more detailed simulations are left to future research. Additionally, and also due to computational costs, all rejection frequencies reported in the tables are based on 500 Monte Carlo replications. A GAUSS program implementing the tests and the bootstrap is available upon request.

observations are used. Turning to the empirical power figures reported in the tables, a number of clear-cut conclusions emerge. First, power increases with sample, as expected, and is in many cases 1.0 (or very close to it) when $T = 1000$. Second, there are marked improvements in finite sample power when: (i) the errors are distributed as t_3 (see $Power1$) as opposed to when they are distributed as t_5 (see $Power2$); and when (ii) the magnitude of the coefficient associated with the nonlinear component is 0.4 (see $Power4$) rather than 0.2 (see $Power3$). Third, when the departure from normality enters both through the error being non-normal and the conditional mean being nonlinear (see $Power5$ and $Power6$), rejection frequencies are the highest, as expected; and in some cases are 1.0, even for $T = 500$. Finally, there again appears to be little to choose between the two tests, although V_{1T} appears to have slightly higher power overall when comparing results for the cases where the tests do worst (i.e. $Power2$ and $Power3$). In summary, both tests appear to perform reasonably well in finite samples, although it should be stressed that the results here are only preliminary, as they are based on a limited number of Monte Carlo experiments.

5 Concluding Remarks

We propose an extension of two conditional Kolmogorov type tests to the case of dynamic misspecification under both the null and alternative hypotheses. We additionally outline conditions under which a version of the block bootstrap can be used to construct valid critical values for the tests, in the context of parameter estimation error and dynamic misspecification. Our approach is useful because critical values derived under correct (dynamic) specification given $\Im_{t-1}$ are not in general valid in the case of correct specification given a subset of $\Im_{t-1}$, where $\Im_{t-1}$ is taken to mean the "relevant" information set. Allowing for dynamic misspecification enables us to test for correct specification given a particular information set, rather than given the "correct" information or given the "entire" history. The findings from a Monte Carlo study illustrate that the finite sample properties of both suggested statistics are satisfactory, for samples as small as 500 observations.

6 Appendix

In the sequel, the notation "$o_{P^*}(1)$, $\Pr - P$" and "$\overset{p*}{\to} 0$, $\Pr - P$" is used to denote convergence in probability to zero, according to P^*, the probability law governing the resampled series, conditional on the sample, and for all samples except a subset with probability measure approaching zero. Also, E^* and Var^* denote expectation and variance operators, with respect to the probability measure P^*.

Proof of Theorem 1: (i) Given A1 and A3, $\widehat{\theta}_T \overset{a.s.}{\to} \theta^\dagger$, with $\theta^\dagger = \theta_0$, under H_0. Give A2(i), by a mean value expansion of $F(y_t|X_t, \widehat{\theta}_T)$ around θ_0, we have,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(1\{F(y_t|X_t,\widehat{\theta}_T)\le r\}-r) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(1\{F(y_t|X_t,\theta_0)\le r-\nabla_\theta F(y_t|X_t,\overline{\theta}_T)(\widehat{\theta}_T-\theta_0)\}-r)$$

$$\begin{aligned} = \quad & \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(1\{F(y_t|X_t,\theta_0)\le r-\nabla_\theta F(y_t|X_t,\overline{\theta}_T)(\widehat{\theta}_T-\theta_0)\}-(r-\nabla_\theta F(y_t|X_t,\overline{\theta}_T)(\widehat{\theta}_T-\theta_0))) \\ & -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\nabla_\theta F(y_t|X_t,\overline{\theta}_T)(\widehat{\theta}_T-\theta_0), \end{aligned}$$

with $\overline{\theta}_T \in (\widehat{\theta}_T, \theta_0)$. We begin by showing that,

$$\begin{aligned} \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(1\{F(y_t|X_t,\theta_0) \quad & \le \quad r-\nabla_\theta F(y_t|X_t,\overline{\theta}_T)(\widehat{\theta}_T-\theta_0)\}-(r-\nabla_\theta F(y_t|X_t,\overline{\theta}_T)(\widehat{\theta}_T-\theta_0)) \\ & = \quad \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(1\{F(y_t|X_t,\theta_0)\le r\}-r)+o_P(1), \end{aligned} \tag{11}$$

where the $o_P(1)$ term holds uniformly in r. Now, (11) follows, once we have shown (12) and (13) below,

$$\Pr(\sup_{r\in[0,1]}\sup_{r':|r-r'|<\delta}|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}((1\{F(y_t|X_t,\theta_0)\le r\}-r)-(1\{F(y_t|X_t,\theta_0)\le r'\}-r'))|>\varepsilon)\to 0, \text{ as } \delta\to 0 \tag{12}$$

$$\sup_t \nabla_\theta F(y_t|X_t,\overline{\theta}_T)(\widehat{\theta}_T-\theta_0)\overset{a.s.}{\to}0. \tag{13}$$

Given that $r\in[0,1]$, the indicator function is bounded, and $F(y_t|X_t)$ is a strictly stationary strong mixing process, with size given in A1, the stochastic equicontinuity of $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(1\{F(y_t|X_t,\theta_0)\le r\}-r)$ (and hence equation (12)) follows from Andrews ((1993), Example 2b, p. 199), once we notice that his L^p- continuity condition (p. 201) is immediately satisfied. Notice that via a mean value expansion around θ_0,

$$\sqrt{T}\left(\widehat{\theta}_T-\theta_0\right) = \left(-\frac{1}{T}\sum_{t=1}^{T}\nabla_\theta q_s(\overline{\theta}_T)\nabla_\theta q_s(\overline{\theta}_T)'\right)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\nabla_\theta q_s(\theta_0),$$

with $\overline{\theta}_T \in \left(\widehat{\theta}_T, \theta_0\right)$. Given A1 and A3(ii), $T^{\vartheta}(\widehat{\theta}_T - \theta_0) \overset{a.s.}{\to} 0$, for all $\vartheta < 1/2$, as direct consequence of the strong law of large numbers and of the law of iterated logarithms for stationary strong mixing process (e.g. Eberlain (1986)). Also,

$$\Pr\left(T^{-7/15}\sup_t |\nabla_\theta F(y_t|X_t,\overline{\theta}_T)| > \varepsilon\right) \leq \sum_{t=1}^{T} \Pr\left(T^{-7/15}|\nabla_\theta F(y_t|X_t,\overline{\theta}_T)| > \varepsilon\right)$$

$$\leq \sum_{t=1}^{T} \frac{1}{\varepsilon^5 T^{7/3}} E\left(\sup_{\theta\in\Theta} |\nabla_\theta F(y_t|X_t,\theta|)\right)^5 \leq CT^{-4/3},$$

given A2(ii). Thus, (13) follows from the Borel Cantelli Lemma. Hence, recalling that $\overline{\theta}_T \in (\widehat{\theta}_T, \theta_0)$, it is immediate to see that,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(1\{F(y_t|X_t,\widehat{\theta}_T) \quad \leq \quad r\} - r) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(1\{F(y_t|X_t,\theta_0) \leq r\} - r)$$

$$-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\nabla_\theta F(y_t|X_t,\theta_0)(\widehat{\theta}_T - \theta_0) + o_P(1), \tag{14}$$

with the $o_P(1)$ term holding uniformly in r. Given A1 and A3, pointwise convergence follows from the central limit theorem for strong mixing processes and from application of the Cramer Wold device. The first term on the RHS of (14) is stochastic equicontinuous, given Example 2b in Andrews (1993). The desired result then follows by the continuous mapping theorem.

(ii) Under H_0, $\theta^\dagger \neq \theta_0$. Let $c_f^r(X_t,\theta^\dagger)$ denote the $r-th$ critical value of the density associated with F, conditional on X_t. Note that $E(1\{F(y_t|X_t,\theta^\dagger)\} \leq r) = \Pr(y_t \leq c_f^r(X_t,\theta^\dagger)) \neq r$, for some subset of r, with nonzero Lebesgue measure, if the true conditional distribution differs from $F(\cdot|X_t,\theta^\dagger)$. The results then follows by noting that,

$$V_{1T}(r) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(1\{F(y_t|X_t,\widehat{\theta}_T) \leq r\} - E(1\{F(y_t|X_t,\widehat{\theta}_T) \leq r\})) + \sqrt{T}(E(1\{F(y_t|X_t,\widehat{\theta}_T) \leq r\}) - r),$$

as the first term is bounded in probability uniformly in r, and $E(1\{F(y_t|X_t,\widehat{\theta}_T) \leq r\}) - r \neq 0$ for some r with nonzero Lebesgue measure.

Proof of Theorem 2: Given A2(iii),

$$\frac{1}{T^{1/2}}\sum_{t=1}^{T}(1\{y_t \leq u\} - F(u|X_t,\widehat{\theta}_T))1\{X_t \leq v\}$$

$$= \quad \frac{1}{T^{1/2}}\sum_{t=1}^{T}(1\{y_t \leq u\} - F(u|X_t,\theta_0))1\{X_t \leq v\} -$$

$$\frac{1}{T}\sum_{t=1}^{T}\left(\nabla_\theta F(u|X_t,\overline{\theta}_T)' 1\{X_t \leq v\}\right) T^{1/2}(\widehat{\theta}_T - \theta_0),\ \overline{\theta}_T \in (\widehat{\theta}_T, \theta_0).$$

We first show pointwise convergence in distribution for any (u, v). Note that under the null,

$$E\left(\left(1\{y_t \leq u\}1\{X_t \leq v\}\right)|X_t\right) = F(u|X_t, \theta_0)1\{X_t \leq v\}.$$

Given Assumption A1, A2(iii)-(iv) and A3, and recalling that under H_0, $\theta^\dagger = \theta_0$, pointwise convergence in distribution follows from the central limit theorem for stationary strong mixing processes. Convergence of the finite dimensional distribution follows straightforwardly from the multivariate central limit theorem and application of the Cramer Wold device. We now need to show stochastic equicontinuity, the desired result will then follow from the continuous mapping theorem. In order to show that,

$$\lim_{T\to\infty} \sup \Pr\left(\sup_{u\times v\in U\times V} \sup_{\substack{u':|u-u'|<\delta \\ v':|v-v'|<\delta}} |V_{2T}(u,v) - V_{2T}(u',v')| > \varepsilon\right) = 0,$$

it suffices to show that,

$$\lim_{T\to\infty} \sup \Pr\left(\sup_{u\times v\in U\times V} \sup_{\substack{u':|u-u'|<\delta \\ v':|v-v'|<\delta}} \left|\frac{1}{T^{1/2}}\sum_{t=1}^{T}\left(m_t(u,v) - m_t(u',v')\right)\right| > \varepsilon/2\right) = 0, \tag{15}$$

where $m_t(u,v) = (1\{y_t \leq u\} - F(u|X_t,\theta_0))1\{X_t \leq v\}$, and that

$$\lim_{T\to\infty} \sup \Pr\left(\sup_{u\times v\in U\times V} \sup_{\substack{u':|u-u'|<\delta \\ v':|v-v'|<\delta}} \left|\frac{1}{T}\sum_{t=1}^{T}\left(\nabla_\theta F(u|X_t,\overline{\theta})'1\{X_t \leq v\} - \nabla_\theta F(u'|X_t,\overline{\theta})'1\{X_t \leq v'\}\right)\right.\right.$$

$$\left.\left. \times\sqrt{T}(\widehat{\theta}_T - \theta_0)\right| \geq \varepsilon/2\right) = 0. \tag{16}$$

We begin by showing (15). First, almost surely,

$$\sup_{\substack{u':|u-u'|<\delta \\ v':|v-v'|<\delta}} |m_t(u,v)| = 1. \tag{17}$$

Now, as U and V are assumed to be compact, it suffices to show that the L^p-continuity condition in Andrews (1993, p. 201) holds, so for $p \geq 2$,

$$\sup_{t\leq T, T>1}\left(E\left(\sup_{\substack{u':|u-u'|<\delta \\ v':|v-v'|<\delta}} |m_t(u,v) - m_t(u',v')|^p\right)\right)^{1/p} \tag{18}$$

$$\leq \sup_{t\leq T, T>1}\left(E\left(\sup_{\substack{u':|u-u'|<\delta \\ v'\in V}} |(1\{y_t \leq u\} - 1\{y_t \leq u'\})1\{X_t \leq v'\}|^p\right)\right)^{1/p}$$

$$+\sup_{t\leq T,T>1}\left(E\left(\sup_{\substack{u':|u-u'|<\delta\\ v'\in V}}|(F(u|X_t,\theta_0)-F(u'|X_t,\theta_0))1\{X_t\leq v'\}|^p\right)\right)^{1/p}$$

$$+\sup_{t\leq T,T>1}\left(E\left(\sup_{\substack{u\in U\\ v':|v-v'|<\delta}}|(1\{X_t\leq v\}-1\{X_t\leq v'\})(1\{y_t\leq u\}-F(u|X_t,\theta_0))|^p\right)\right)^{1/p}$$

$$\leq (\sup_{u\in U}\left|\int_u^{u+\delta}f_y(s)ds\right|)^{1/p}+\left|E\left(\sup_{u\in U}\nabla_u F(u,\theta_0)\right)^p\right|^{1/p}\delta+(\sup_{v\in V}\left|\int_v^{v+\delta}f_x(s)ds\right|)^{1/p}\leq C\delta \tag{19}$$

where f_x, f_y denote the marginal densities of X and y respectively. Equation (15) then follows by Example 2a in Andrews (1993). In fact, condition (i) in that example is satisfied given A1; condition (ii) is ensured by (17) and condition (iii) is implied by the fact that (18) is majorized by (19). As for (16),

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\left(\nabla_\theta F(u|X_t,\overline{\theta})'1\{X_t\leq v\}-\nabla_\theta F(u'|X_t,\overline{\theta})'1\{X_t\leq v'\}\right)\\
=\;&\frac{1}{T}\sum_{t=1}^{T}(\nabla_\theta F(u|X_t,\overline{\theta})'-\nabla_\theta F(u'|X_t,\overline{\theta})')1\{X_t\leq v\} &(20)\\
&+\frac{1}{T}\sum_{t=1}^{T}\left(\nabla_\theta F(u'|X_t,\overline{\theta}))'(1\{X_t\leq v\}-1\{X_t\leq v'\}\right). &(21)
\end{aligned}$$

it suffices to show that the $\sup_{u':|u-u'|<\delta}$ of equation (20) converges to zero in probability and that the $\sup_{v':|v-v'|<\delta}$ of (21) also converges to zero in probability. The summands in (20) satisfy condition WLIP in Andrews (1992), as $\left|(\nabla_\theta F(u|X_t,\overline{\theta})'-\nabla_\theta F(u'|X_t,\overline{\theta})')1\{X_t\leq v\}\right|\leq|\nabla_{u,\theta}F(\overline{u}|X_t,\overline{\theta})||u-u'|$, $\overline{u}\in(u,u')$, and $E|\nabla_{u,\theta}F(u|X_t,\theta)|<\infty$ uniformly in $\Theta\times U$ because of A2(iv). This ensures that $\sup_{u':|u-u'|<\delta}$ of (20) converges to zero in probability. Finally the summands in (21) satisfy conditions TSE and DM in Andrews (1992). In particular,

$$\lim_{\delta\to 0}\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\Pr\left(\sup_{v':|v-v'|<\delta}\left|\nabla_\theta F(u|X_t,\overline{\theta})'\right|1\{v\leq X_t\leq v'\}>\epsilon\right)=0,\ v<v'$$

as $\frac{1}{T}\sum_{t=1}^{T}(\nabla_\theta F(u|X_t,\overline{\theta})'=O_p(1)$ uniformly in $U\times\Theta$, given A1 and A2(iii), thus satisfying TSE. Furthermore, DM is trivially satisfied because of A2(iv). This ensures that $\sup_{v':|v-v'|<\delta}$ of (21) also converges to zero in probability.

Proof of Theorem 3: By a mean value expansion around $\theta^\dagger$,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(1\{F(y_t^*|X_t^*,\widehat{\theta}_T^*)\leq r\}-1\{F(y_t|X_t,\widehat{\theta}_T)\leq r\})=$$

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(1\{F(y_t^*|X_t^*,\theta^\dagger) \leq r-\nabla_\theta F(y_t^*|X_t^*,\overline{\theta}_T^*)(\widehat{\theta}_T^*-\theta^\dagger)\}-(r-\nabla_\theta F(y_t^*|X_t^*,\overline{\theta}_T^*)(\widehat{\theta}_T^*-\theta^\dagger))$$

$$-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(1\{F(y_t|X_t,\theta^\dagger) \leq r-\nabla_\theta F(y_t|X_t,\overline{\theta}_T)(\widehat{\theta}_T-\theta^\dagger)\}+(r-\nabla_\theta F(y_t|X_t,\overline{\theta}_T)(\widehat{\theta}_T-\theta^\dagger))$$

$$-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\nabla_\theta F(y_t^*|X_t^*,\overline{\theta}_T^*)(\widehat{\theta}_T^*-\theta^\dagger)-\nabla_\theta F(y_t|X_t,\overline{\theta}_T)(\widehat{\theta}_T-\theta^\dagger)\right), \tag{22}$$

where $\overline{\theta}_T^* \in (\widehat{\theta}_T^*,\theta^\dagger)$ and $\overline{\theta}_T \in (\widehat{\theta}_T,\theta^\dagger)$. We begin by showing that the last term on the right hand side of (22) has the same limiting distribution as $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\nabla_\theta F(y_t|X_t,\theta_0)(\widehat{\theta}_T-\theta^\dagger)$ conditionally on the sample, and for all samples except a subset with probability measure approaching zero. We can write the last term on the RHS of (22) as,

$$\frac{1}{T}\sum_{t=1}^{T}\left(\nabla_\theta F(y_t|X_t,\overline{\theta}_T)-\nabla_\theta F(y_t^*|X_t^*,\overline{\theta}_T^*)\right)\sqrt{T}(\widehat{\theta}_T^*-\theta^\dagger)$$
$$-\frac{1}{T}\sum_{t=1}^{T}\nabla_\theta F(y_t|X_t,\overline{\theta}_T)\sqrt{T}(\widehat{\theta}_T^*-\widehat{\theta}_T). \tag{23}$$

Given A1 and A3, by Theorem 2.2 in Goncalves and White (2002(b))there exists an $\varepsilon>0$ such that,

$$\Pr\left(\sup_{x\in\Re^p}\left|P^*\left(\sqrt{T}\left(\widehat{\theta}_T^*-\widehat{\theta}_T\right)\leq x\right)-P\left(\sqrt{T}\left(\widehat{\theta}_T-\theta^\dagger\right)\leq x\right)\right|>\varepsilon\right)\to 0, \tag{24}$$

thus the second term on the RHS of (23) has the same limiting distribution of $\frac{1}{T}\sum_{t=1}^{T}\nabla_\theta F(y_t|X_t,\overline{\theta}_T)\sqrt{T}(\widehat{\theta}_T^*-\widehat{\theta}_T)$, conditionally on the sample and for all samples except a subset with probability measure approaching zero. We also need to show that the first term on the RHS of (23) is $o_{P^*}(1)$ $\Pr-P$. First note that, $\sqrt{T}(\widehat{\theta}_T^*-\theta^\dagger)=\sqrt{T}(\widehat{\theta}_T^*-\widehat{\theta}_T)+\sqrt{T}(\widehat{\theta}_T-\theta^\dagger)=O_{P^*}(1)+O(1)=O_{P^*}(1)$, $\Pr-P$. As $\overline{\theta}_T^*\in(\widehat{\theta}_T^*,\theta^\dagger)$ and $\overline{\theta}_T\in(\widehat{\theta}_T,\theta^\dagger)$, it follows that $\overline{\theta}_T^*-\overline{\theta}_T=o_{P^*}(1)$, $\Pr-P$. Thus, by the same argument used in Lemma A4 of Goncalves and White (2000(b)), it suffices to show that,

$$\sup_{\theta\in\Theta}\left|\frac{1}{T}\sum_{t=1}^{T}\left(\nabla_\theta F(y_t|X_t,\theta)-\nabla_\theta F(y_t^*|X_t^*,\theta)\right)\right|\xrightarrow{P^*}0,\ \ \Pr-P.$$

For any given θ,

$$\frac{1}{T}\sum_{t=1}^{T}\left(\nabla_\theta F(y_t^*|X_t^*,\theta)-\nabla_\theta F(y_t|X_t,\theta)\right)=\frac{1}{T}\sum_{t=1}^{T}\left(\nabla_\theta F(y_t^*|X_t^*,\theta)-E^*(\nabla_\theta F(y_t^*|X_t^*,\theta))\right) \tag{25}$$

$$-\frac{1}{T}\sum_{t=1}^{T}\left(\nabla_\theta F(y_t|X_t,\theta)-E^*(\nabla_\theta F(y_t^*|X_t^*,\theta))\right) \tag{26}$$

Recall that I_k is an *iid* uniform random variable on $[0, 1, ..., T - l_T + 1]$, and $T = l_T b$, where b is the number of blocks. Thus, $\frac{1}{T}\sum_{t=1}^{T} \nabla_\theta F(y_t^*|X_t^*, \theta) = \frac{1}{T}\sum_{k=1}^{b}\sum_{i=1}^{l_T} \nabla_\theta F(y_{I_k+i}|X_{I_k+i}, \theta)$, and so, pointwise in θ,

$$\begin{aligned} E^*\left(\frac{1}{T}\sum_{t=1}^{T} \nabla_\theta F(y_t^*|X_t^*, \theta)\right) &= \frac{1}{l_T(T - l_T + 1)}\sum_{t=0}^{T-l}\sum_{i=1}^{l_T} \nabla_\theta F(y_{t+i}|X_{t+i}, \theta) + O(l_T/T), \ \Pr - P \\ &= \frac{1}{T}\sum_{t=0}^{T} \nabla_\theta F(y_t|X_t, \theta) + O(l_T/T), \ \Pr - P, \end{aligned}$$

where the last equality follows from Lemma A1 in Fitzenberger (1997). Thus, the term in (26) is $O(l_T/T)$, $\Pr - P$, pointwise in θ. With respect to the RHS in (25), by the Chebyshev inequality,

$$P^*\left(\left|\frac{1}{T}\sum_{t=1}^{T}(\nabla_\theta F(y_t^*|X_t^*, \theta) - E^*(\nabla_\theta F(y_t^*|X_t^*, \theta)))\right| > \varepsilon\right) \leq \frac{1}{T\varepsilon^2} Var^*\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \nabla_\theta F(y_t^*|X_t^*, \theta)\right). \quad (27)$$

Now,

$$Var^*\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \nabla_\theta F(y_t^*|X_t^*, \theta)\right) = Var^*\left(\frac{1}{\sqrt{T}}\sum_{k=1}^{b}\sum_{i=1}^{l_T} \nabla_\theta F(y_{I_k+i}|X_{I_k+i}, \theta)\right)$$

$$\begin{aligned} &= E^*\left(\frac{1}{T}\sum_{k=1}^{b}\sum_{i=1}^{l_T}\sum_{j=1}^{l_T} \nabla_\theta F(y_{I_k+i}|X_{I_k+i}, \theta)\nabla_\theta F(y_{I_k+j}|X_{I_k+j}, \theta)\right) \\ &= \frac{1}{T - l_T + 1}\frac{1}{l}\sum_{t=0}^{T-l_T}\sum_{i=1}^{l_T}\sum_{j=1}^{l_T} \nabla_\theta F(y_{t+j}|X_{t+j}, \theta)\nabla_\theta F(y_{t+i}|X_{t+i}, \theta)' = O(1), \ \Pr - P, \end{aligned}$$

pointwise in θ. Thus, the LHS of (25) is $o_{P^*}(1)$, $\Pr - P$, pointwise in θ. By noting that given A2(i)-(ii), $\nabla_\theta F(y_t|X_t, \theta)$ is almost surely Lipschitz (as in Andrews (1992), Definition S-Lip, p. 248), the LHS of (25) is $o_P^*(1)$, $\Pr - P$ uniformly in θ, given Lemma A.2 in Goncalves and White (2000(b)). We now need to show that,

$$\begin{aligned} \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(1\{F(y_t^*|X_t^*, \theta^\dagger) &\leq r - \nabla_\theta F(y_t^*|X_t^*, \overline{\theta}_T^*)(\widehat{\theta}_T^* - \theta^\dagger)\} - (r - \nabla_\theta F(y_t^*|X_t^*, \overline{\theta}_T^*)(\widehat{\theta}_T^* - \theta^\dagger)) \\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(1\{F(y_t^*|X_t^*, \theta^\dagger) \leq r\} - r) + o_{P^*}(1), \ \Pr - P. \end{aligned}$$

Along the same lines as in the proof of Theorem 1, we first show that $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(1\{F(y_t^*|X_t^*, \theta^\dagger) \leq r\} - r)$ is P^*-stochastic equicontinuous in r, and then that $\nabla_\theta F(y_t^*|X_t^*, \overline{\theta}_T^*)(\widehat{\theta}_T^* - \theta^\dagger)$ is $o_{a.s.^*}(1)$, $a.s. - P$. Without loss of generality suppose that $r' < r$. Then,

$$\begin{aligned} &P^*\left(\sup_{r\in[0,1]}\sup_{r':|r-r'|<\delta}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(1\{r' \leq F(y_t^*|X_t^*, \theta^\dagger) \leq r\} > \varepsilon\right) \\ \leq \ &P^*\left(\sup_{r\in[0,1]}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(1\{r - \delta \leq F(y_t^*|X_t^*, \theta^\dagger) \leq r\} > \varepsilon)\right) \end{aligned}$$

$$
\begin{aligned}
&\leq \frac{1}{\varepsilon^2} E^* \left(\sup_{r\in[0,1]} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} 1\{r-\delta \leq F(y_t^*|X_t^*,\theta^\dagger) \leq r\} \right)^2 \\
&= \sup_{r\in[0,1]} \frac{1}{\varepsilon^2} \frac{1}{T-l_T+1} \frac{1}{l_T} \sum_{t=0}^{T-l_T} \sum_{i=1}^{l_T} \sum_{j=1}^{l_T} 1\{r-\delta \leq F(y_{t+j}|X_{t+j},\theta^\dagger) \leq r\} 1\{r-\delta \leq F(y_{t+i}|X_{t+i},\theta^\dagger) \leq r\} \\
&\leq C\delta, \ \Pr - P.
\end{aligned}
$$

The fact that $\nabla_\theta F(y_t^*|X_t^*,\overline{\theta}_T^*)(\widehat{\theta}_T^* - \theta^\dagger)$ is $a.s.*(1)$, $a.s.-P$ follows by a similar argument to that used in the proof of Theorem 1. Thus,

$$
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (1\{F(y_t^*|X_t^*,\widehat{\theta}_T^*) \leq r\} - 1\{F(y_t|X_t,\widehat{\theta}_T) \leq r\}) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (1\{F(y_t^*|X_t^*,\theta^\dagger) \leq r\} - 1\{F(y_t|X_t,\theta^\dagger) \leq r\}) \tag{28}
$$

$$
-\frac{1}{T} \sum_{t=1}^{T} \nabla_\theta F(y_t|X_t,\theta^\dagger) \sqrt{T}(\widehat{\theta}_T^* - \widehat{\theta}_T) + o_{P^*}(1), \ \Pr - P. \tag{29}
$$

Given (24), it is immediate to see that the first term in (29) has the same limiting distribution as $E(\nabla_\theta F(y_t|X_t,\theta^\dagger))\sqrt{T}(\widehat{\theta}_T - \widehat{\theta}^\dagger)$, $\Pr - P$, given the law of large numbers. Now, consider the term on the RHS of (28), which can be written as,

$$
\begin{aligned}
&\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (1\{F(y_t^*|X_t^*,\theta^\dagger) \ \leq \ r\} - E^* \left(1\{F(y_t^*|X_t^*,\theta^\dagger) \leq r\}\right) \\
&-\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (1\{F(y_t|X_t,\theta^\dagger) \ \leq \ r\} - E^* \left(1\{F(y_t^*|X_t^*,\theta^\dagger) \leq r\}\right),
\end{aligned} \tag{30}
$$

where

$$
\begin{aligned}
E^*(1\{F(y_t^*|X_t^*,\theta^\dagger) \ \leq \ r\}) &= \frac{1}{l_T(T-l_T+1)} \sum_{t=0}^{T-l} \sum_{i=1}^{l_T} 1\{F(y_{t+i}|X_{t+i},\theta^\dagger) \leq r\} \\
&= \frac{1}{T} \sum_{t=0}^{T} 1\{F(y_{t+i}|X_{t+i},\theta^\dagger) \leq r\} + O(l_T/T), \ \Pr - P,
\end{aligned}
$$

uniformly in r. Thus, the second term in (30) is $o(1)$, $\Pr - P$. We now need to show that the first term in (30) has the same limiting distribution as $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (1\{F(y_t|X_t,\theta^\dagger) \leq r\} - E\left(1\{F(y_t|X_t,\theta^\dagger) \leq r\}\right)$, uniformly in r, $\Pr - P$. This follows from the empirical process version of the block bootstrap of Naik-Nimbalkar and Rajarshi (1994, Theorem 2.1), once we note that given $\psi < 1/2$, A1 and A2(i)-(ii) imply their conditions (a) and (b).[13]

Proof of Theorem 4: Given A2(iii),

$$
V_{2T}^*(u,v) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left((1\{y_t^* \leq u\} - F(u|X_t^*,\theta^\dagger))1\{X_t^* \leq v\} - (1\{y_t \leq u\} - F(u|X_t,\theta^\dagger))1\{X_t \leq v\}\right)
$$

[13]The almost sure validity of the block bootstrap for empirical processes is obtained by Naik-Nimbalkar and Rajarshi (1994) assuming that $l_T = o(T^{1/4})$, and by Peligrad (1998) under milder mixing conditions and assuming $l_T = o(T^{1/3})$.

$$-\frac{1}{T}\sum_{t=1}^{T}\left(\left(\nabla_\theta F(u|X_t^*,\overline{\theta}_T^*)'1\{X_t^*\leq v\}\right)T^{1/2}(\widehat{\theta}_T^*-\theta^\dagger)-\left(\nabla_\theta F(u|X_t,\overline{\theta}_T)'1\{X_t\leq v\}\right)T^{1/2}(\widehat{\theta}_T-\theta^\dagger)\right). \tag{31}$$

Given A2(iii)-(iv) and (24), by the same argument as that used in the proof of Theorem 3, the term in (31) has the same limiting distribution as $\frac{1}{T}\sum_{t=1}^{T}\left(\nabla_\theta F(u|X_t,\theta^\dagger)'1\{X_t\leq v\}\right)T^{1/2}(\widehat{\theta}_T-\theta^\dagger)$, conditional on the samples and for all samples except a subset with probability measure approaching zero. Again by the same argument as that used in the proof of Theorem 3,

$$\begin{aligned}&\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(1\{y_t\leq u\}-F(u|X_t,\theta^\dagger)\right)1\{X_t\leq v\}-E^*\left(1\{y_t^*\leq u\}-F(u|X_t^*,\theta^\dagger)\right)1\{X_t^*\leq v\}\right)\\ =\;& o(1),\ \Pr-P,\ \text{for } l_T=o(T^{1/2}).\end{aligned}$$

Now, P^*-convergence in distribution, $\Pr-P$, pointwise in u and v of

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(1\{y_t^*\leq u\}-F(u|X_t^*,\theta^\dagger)\right)1\{X_t^*\leq v\}-E^*\left(1\{y_t^*\leq u\}-F(u|X_t^*,\theta^\dagger)\right)1\{X_t^*\leq v\}\right) \tag{32}$$

follows from Theorem 3.5 in Künsch (1989). Finally, along the same lines as used in the proof of Theorem 2, it can be shown that (32) is P^*-stochastic equicontinuous in $U\times V$, $\Pr-P$. As,

$$\begin{aligned}&Var^*\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}1\{y_t^*\leq u\}-F(u|X_t^*,\theta^\dagger)\right)1\{X_t^*\leq v\}\\ =\;&\frac{1}{T}\sum_{t=l_T}^{T-l_T}\sum_{j=-l_T}^{l_T}\left(1\{y_t\leq u\}-F(u|X_t,\theta^\dagger)\right)\left(1\{y_{t-j}\leq u\}-F(u|X_{t-j},\theta^\dagger)\right)+O\left(\frac{l_T}{\sqrt{T}}\right),\end{aligned}$$

the desired results follows from application of the continuous mapping theorem.

7 References

Andrews, D.W.K., (1991), Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation, *Econometrica*, 59, 817-858.

Andrews, D.W.K., (1992), Generic Uniform Convergence, *Econometric Theory,* 8, 241-257.

Andrews, D.W.K., (1993), An Introduction to Econometric Applications of Empirical Process Theory for Dependent Random Variables, *Econometric Reviews,* 12, 183-216.

Andrews, D.W.K., (1997), A Conditional Kolmogorov Test, *Econometrica,* 65, 1097-1128.

Andrews, D.W.K. and M. Buchinsky, (2000), A Three Step Method for Choosing the Number of Bootstrap Replications, *Econometrica,* 68, 23-52.

Andrews, D.W.K., (2002), Higher-Order Improvements of a Computationally Attractive $k-$step Bootstrap for Extremum Estimators, *Econometrica,* 70, 119-162.

Bai, J., (2003), Testing Parametric Conditional Distributions of Dynamic Models, *Review of Economics and Statistics,* 85, 531-549.

Bai, J. and S. Ng, (2001), A Consistent test for Conditional Symmetry in Time Series Models, *Journal of Econometrics,* 103, 225-258.

Bühlmann, P., (1995), The Blockwise Bootstrap for General Empirical Processes of Stationary Sequences, *Stochastic Processes and their Applications,* 58, 2, 247-266.

Cambanis, S., S. Huang and G. Simons, (1981), On the Theory of Elliptically Contoured Distributions, *Journal of Multivariate Analysis, 11, 368-385.*

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

Chao, J., V. Corradi and N.R. Swanson, 2001, An Out of Sample Test for Granger Causality, *Macroeconomic Dynamics*, 5, 598-620.

Christoffersen, P.F., (1998), Evaluating Interval Forecasts, *International Economic Review*, 39, 841-862.

Clark, T.E. and M.W. McCracken, (2001), Tests of Equal Forecast Accuracy and Encompassing for Nested Models, *Journal of Econometrics,* 105, 85-110.

Clements, M.P. and J. Smith, (2000), Evaluating the Forecast Densities of Linear and Nonlinear Models: Applications to Output Growth and Unemployment, *Journal of Forecasting,* 19, 255-276.

Clements, M.P. and J. Smith, (2002), Evaluating Multivariate Forecast Densities: A Comparison of Two Approaches, *International Journal of Forecasting,* 18, 397-407.

Corradi, V., N.R. Swanson and C. Olivetti, (2001), Predictive Ability with Cointegrated Variables, *Journal of Econometrics,* 104, 315-358.

Corradi, V. and N.R. Swanson, (2002), A Consistent Test for Out of Sample Nonlinear Predictive Ability, *Journal of Econometrics,* 110, 353-381, 2002.

Davidson, R., and J.G. MacKinnon, (1999), Bootstrap Testing in Nonlinear Models, *International Economic Review,* 40, 487-508.

Davidson, R., and J.G. MacKinnon, (2000), Bootstrap Tests: How Many Bootstraps, *Econometric Reviews, 19, 55-68.*

Diebold, F.X., T. Gunther and A.S. Tay, (1998), Evaluating Density Forecasts with Applications to Finance and Management, *International Economic Review,* 39, 863-883.

Diebold, F.X., J. Hahn and A.S. Tay, (1999), Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High Frequency Returns on Foreign Exchange, *Review of Economics and Statistics*, 81, 661-673.

Diebold, F.X., and R.S. Mariano, (1995), Comparing Predictive Accuracy, *Journal of Business and Economic Statistics*, 13, 253-263.

Diebold, F.X., A.S. Tay and K.D. Wallis, (1998), Evaluating Density Forecasts of Inflation: The Survey of Professional Forecasters, *in Festschrift in Honor of C.W.J. Granger*, eds. R.F. Engle and H. White, Oxford University Press, Oxford. Eberlain, E., (1986), On Strong Invariance Principles under Dependence

Assumptions, *Annals of Probability,* 14, 260-270.

Fitzenberger, B. (1997), The Moving Block Bootstrap and Robust Inference for Linear Least Square and Quantile Regressions, *Journal of Econometrics*, 82, 235-287.

Goncalves, S., and H. White, (2002(a)), The Bootstrap of the Mean for Dependent and Heterogeneous Arrays, *Econometric Theory, 18, 1367-1384.*

Goncalves, S., and H. White, (2002(b)), Maximum Likelihood and the Bootstrap for Nonlinear Dynamic Models, Mimeo, Université de Montréal.

Hall, P., (1992), *The Bootstrap and Edgeworth Expansion,* Springer and Verlag, New York.

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

Hansen, B.E., (1996), Inference When a Nuisance Parameter is Not Identified Under the Null Hypothesis, *Econometrica,* 64, 413-430.

Hong, Y., (2002), Evaluation of Out of Sample Probability Density Forecasts with Applications to S&P 500 Stock Prices, Mimeo, Cornell University.

Ingersoll, J.E., (1987), *Theory of Financial Decision Making,* Rowan and Littlefield, New York.

Inoue, A., (2001), Testing for Distributional Change in Time Series, *Econometric Theory*, 17, 156-187.

Inoue, A., and M. Shintani, (2004), Bootstrapping GMM Estimators for Time Series, *Journal of Econometrics*, forthcoming.

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

Lahiri, S.N., (1999), Theoretical Comparisons of Block Bootstrap Methods, *Annals of Statistics,* 27, 386-404.

Li, F. and G. Tkacz, (2004), A Consistent Test for Conditional Density Functions with Time Dependent Data, *Journal of Econometrics,* forthcoming.

McCracken, M.W., (2000), Robust Out of Sample Inference, *Journal of Econometrics,* 99, 195-223.

Naik-Nimbalkar U.V. and M.B. Rajarshi, (1994), Validity of Blockwise Bootstrap for Empirical Processes with Stationary Observations, *Annals of Statistics,* 22, 980-994.

Peligrad M., (1998), On the Blockwise Bootstrap for Empirical Processes for Stationary Sequences, *Annals of Probability,* 26, 877-901.

Politis, D.N. and J.P. Romano, (1994a), The Stationary Bootstrap, *Journal of the American Statistical Association,* 89, 1303-1313.

Politis, D.N. and J.P. Romano, (1994b), Limit Theorems for Weakly Dependent Hilbert Space Valued Random Variables with Application to the Stationary Bootstrap, *Statistica Sinica*, 4, 461-476.

Spanos, A., (1999), *Probability Theory and Statistical Inference: Econometric Modelling with Observational Data,* Cambridge University Press.

Rosenblatt, M., (1952), Remarks on a Multivariate Transformation, *Annals of Mathematical Statistics*, 23, 470-472.

West, K., (1996), Asymptotic Inference About Predictive Ability, *Econometrica*, 64, 1067-1084.

Whang, Y.J., (2000), Consistent Bootstrap Tests of Parametric Regression Functions, *Journal of Econometrics,* 27-46.Whang, Y.J., (2001), Consistent Specification Testing for Conditional Moment Restrictions,

Economics Letters, 71, 299-306.

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

White, H., (2000), A Reality Check for Data Snooping, *Econometrica,* 68, 1097-1126.

Table 1: Monte Carlo Rejection Frequencies - V_{1T}

Model	l=25		l=40		l=50	
	10%	5%	10%	5%	10%	5%
Panel A: Sample Size = 500 Observations						
Size1	0.066	0.034	0.098	0.060	0.088	0.046
Size2	0.092	0.058	0.104	0.064	0.098	0.060
Power1	0.934	0.854	0.920	0.806	0.926	0.810
Power2	0.420	0.250	0.402	0.280	0.414	0.276
Power3	0.434	0.324	0.428	0.320	0.442	0.316
Power4	0.944	0.880	0.928	0.902	0.966	0.928
Power5	1.000	0.998	1.000	1.000	1.000	1.000
Power6	0.998	0.990	0.996	0.992	1.000	0.994
Panel B: Sample Size = 1000 Observations						
Size1	0.076	0.040	0.106	0.052	0.104	0.062
Size2	0.102	0.042	0.108	0.058	0.122	0.070
Power1	1.000	0.996	1.000	0.998	1.000	0.994
Power2	0.738	0.546	0.752	0.576	0.760	0.578
Power3	0.676	0.590	0.700	0.594	0.682	0.580
Power4	1.000	0.996	1.000	0.998	1.000	0.998
Power5	1.000	1.000	1.000	1.000	1.000	1.000
Power6	1.000	1.000	1.000	1.000	1.000	1.000

Notes: All entries are rejection frequencies of the null hypothesis correct conditional specification, given a particular information set, based on 5% and 10% critical values constructed from the empirical bootstrap distribution, and based on the block bootstrap with blocks of length 25, 40, and 50 observations. Model descriptions are outlined above. All rejection frequencies are based on 500 Monte Carlo replications, and B=100 bootstrap replications. See above for further details.

Table 2: Monte Carlo Rejection Frequencies - V_{2T}

Model	l=25		l=40		l=50	
	10%	5%	10%	5%	10%	5%
Panel A: Sample Size = 500 Observations						
Size1	0.114	0.072	0.112	0.068	0.144	0.068
Size2	0.116	0.072	0.114	0.068	0.132	0.082
Power1	0.916	0.846	0.910	0.766	0.912	0.792
Power2	0.424	0.270	0.414	0.262	0.422	0.268
Power3	0.392	0.274	0.410	0.290	0.418	0.310
Power4	0.900	0.808	0.906	0.826	0.910	0.852
Power5	0.998	0.998	0.998	0.992	0.996	0.986
Power6	0.984	0.970	0.982	0.952	0.982	0.942
Panel B: Sample Size = 1000 Observations						
Size1	0.108	0.058	0.140	0.088	0.128	0.074
Size2	0.136	0.068	0.126	0.076	0.144	0.090
Power1	0.998	0.994	1.000	0.998	1.000	0.992
Power2	0.716	0.526	0.728	0.528	0.706	0.524
Power3	0.644	0.508	0.660	0.534	0.620	0.522
Power4	0.998	0.988	0.996	0.992	0.990	0.974
Power5	1.000	1.000	1.000	1.000	1.000	1.000
Power6	1.000	1.000	1.000	1.000	1.000	1.000

Notes: See notes to Table 1.