Economics Department Discussion Papers Series

ISSN 1473 – 3307

Bootstrap Specification Tests with Dependent Observations and Parameter Estimation Error

Valentina Corradi and Norman R. Swanson

Paper number 01/01

URL: http://business-school.exeter.ac.uk/economics/papers/

Bootstrap Specification Tests with Dependent Observations and Parameter Estimation Error*

Valentina Corradi
University of Exeter, U.K.

Norman R. Swanson
Texas A&M University, U.S.A.

February 2001

Abstract

This paper introduces a parametric specification test for diffusion processes which is based on a bootstrap procedure that accounts for data dependence and parameter estimation error. The proposed bootstrap procedure additionally leads to straightforward generalizations of the conditional Kolmogorov test of Andrews (1997) and the conditional mean test of Whang (2000) to the case of dependent observations. The bootstrap hinges on a twofold extension of the Politis and Romano (1994) stationary bootstrap. First we provide an empirical process version of this bootstrap, and second, we account for parameter estimation error. One important feature of this new bootstrap is that one need not specify the conditional distribution given the entire history of the process when forming conditional Kolmogorov tests. Hence, the bootstrap, when used to extend Andrews (1997) conditional Kolmogorov test to the case of data dependence, allows for dynamic misspecification under both hypotheses. An example based on a version of the Cox, Ingersol and Ross square root process is outlined and related Monte Carlo experiments are carried out. These experiments suggest that the boostrap has excellent finite sample properties, even for samples as small as 500 observations when tests are formed using critical values constructed with as few as 100 bootstrap replications.

JEL classification: C12, C22.
Keywords: diffusion process, parameter estimation error, specification test, stationary bootstrap.

*Valentina Corradi, Dept. of Economics, University of Exeter, Exeter EX4 4PU, U.K., v.corradi©exeter.ac.uk. Norman R. Swanson, Dept. of Economics, Texas A&M University, College Station, Texas 77843-4228, U.S.A., nswanson©econ.tamu.edu. Parts of this paper were written while the second author was visiting the University of California, San Diego. We owe special thanks to Bent Sorensen for very helpful suggestions. We also wish to thank Filippo Altissimo, Oliver Linton, Allan Timmerman, Paolo Zaffaroni and seminar participants at Ente Einaudi-Roma, ESRC - Econometrics Group Bristol, Queen Mary and Westfield College, University of London, Université Libre Bruxelles, and the 2000 Winter Meetings of the Econometric Society for their useful comments. Swanson thanks the Private Enterprise Research Center at Texas A&M University for research support.

1 Introduction

This paper introduces a parametric specification test for diffusion processes which is based on a bootstrap procedure that accounts for data dependence and parameter estimation error. The proposed bootstrap procedure additionally leads to straightforward generalizations of the conditional Kolmogorov test of Andrews (1997) and the conditional mean test of Whang (2000) to the case of dependent observations.

The diffusion specification test discussed below is closest to the nonparameteric test introduced by Aït-Sahalia (1996), in the sense that both procedures determine whether the drift and variance components of a particular continuous time model are correctly specified. The main feature which differentiates these tests is that we compare cumulative distribution functions (CDFs), while Aït-Sahalia compares densities. Thus, his approach requires the use of a nonparametric density estimator, and so the choice of the bandwith parameter and is characterized by a nonparametric rate. On the other hand our test has power against $1/\sqrt{T}$ local alternatives and does not require the adaptive choice of bandwith parameters. This said, it should be noted that neither test has power against *i.i.d.* alternatives that are generated by the same marginal density as that implied under H_0. In fact, while the knowledge of the drift and variance implies the knowledge of the invariance density, it does not in general imply the knowledge of the transition density. Thompson (2000) discusses diffusion specification testing based on the use of conditional distribution functions. For the cases in which the conditional distribution is known, he provides valid asymptotic upper bounds for the critical values. He also suggests an ingenious device (based on the use of an Euler scheme) for the approximation of the transition function, when the latter is unknown. However, the approximated conditional distribution function is not differentiable over the parameter space, so that contribution of parameter estimation error to the limiting distribution cannot be accounted for using standard techniques.

As is commonly the case with tests based on (conditional) distributions, the diffusion specification test which we propose has a Gaussian limiting distribution; but critical values are data dependent, and hence cannot be tabulated.[1] It is for this reason that we propose a valid bootstrap

[1]The reason why data dependence (and parameter estimation error) must be accounted for when forming critical values for our diffusion specification test is that the empirical CDF is constructed using a discrete sample from the underlying diffusion, that is assumed to be geometric ergodic under both hypotheses, and is evaluated at the estimated parameters. An alternative Kolmogorov type test for diffusion processes has been previously suggested by

procedure (for which a well defined limiting distribution obtains under both the null and alternative hypothesis) which allows for both data dependence and parameter estimation error. It should be noted, though, that various tests have been proposed in recent years which at least partially address data dependence and parameter estimation error, although they do not consider continuous time models. An incomplete list of papers which discuss specification testing via the use of conditional distributions (except where noted) includes: Beran and Millar (1989), who construct a test for parametric families of multivariate distributions, assuming *i.i.d.* observations; Andrews (1997) who generalizes the Kolmogorov test of goodness of fit (see e.g. Kolmogorov (1933) and Smirnov (1939)) to the case of conditional parameteric models with covariates, assuming *i.i.d.* observations; Whang (2000) who considers tests for the correct specification of the conditional mean, assuming *i.i.d.* observations; Bai (1998) who uses a novel approach to ensure that the effect of parameter estimation error does not enter the limiting distribution of the test statistic; and Inoue (1999) who studies different conditional features, such as conditional symmetry and conditional quantiles. Although all of these papers take into account the effect of parameter estimation error, only the last two allow for data dependence. In particular, Inoue (1999) accounts for data dependence by constructing critical values along the lines of the 'upper bounds' approach suggested by Bierens and Ploberger (1997). An alternative to this approach is the conditional p-value approach of Corradi and Swanson (2000) (which extends Inoue (1998) to the case of parameter estimation error), although their approach has the drawback that simulated critical values diverge at rate l (where l plays the role of the blocksize length) under the alternative. On the other hand, Bai (1998) can be applied directly to the problem of jointly accounting for parameter estimation error and data dependence. One feature of Bai's setup, though, is that his null hypothesis is the correct specification of the conditional distribution given the entire history. We instead consider the null of correct specification of the conditional distribution given a (sub)set of covariates. In particular, the bootstrap proposed below, which is an empirical process version of the Politis and Romano (1994) stationary bootstrap that allows for data dependence and non-vanishing parameter estimation error, can be used to test the correct specification of the conditional distribution without specifying the entire history. Hence, our test and bootstrap procedure, when used to extend Andrews (1997) conditional Kolmogorov test to the case of data dependence, allow for dynamic misspecification under both

Fournie (1993), for the case in which we observe the continuous trajectories.

hypotheses. In addition to generalizing Andrews (1997), we also generalize Whang (2000).

The potential usefulness of our proposed bootstrap procedure is examined via a series of Monte Carlo experiments in which our diffusion specification test is applied to the problem of testing the goodness of fit of a square root diffusion process. In particular, data are generated according to a square root diffusion (e.g. see Cox, Ingersoll and Ross (1985)) under the null hypothesis. Under the alternative, logged data are generated according to an Ornstein-Uhlenbeck process, so that data are lognormal. For samples of 500 and 1000 observations, and based on the use of bootstrap critical values constructed using as few as 100 replications, rejection rates under the null are quite close to the nominal, and rejection rates under the alternative are generally above 0.90. It is worth noting that the joint problem of simulating paths and simulating bootstrap replications make this Monte Carlo study rather computationally intensive. We are not aware of other simulation studies which analyze the performance of bootstrap tests for diffusion processes.

The rest of the paper is organized as follows. In Section 2, we outline our parametric diffusion specification test and analyze its asymptotic behavior. Section 3 outlines the bootstrap procedure which we propose, and establishes its asymptotic validity. In Section 4, we show that the simulated generalized methods of moments estimator of Duffie and Singleton (1993) satisfies the assumptions required in order to ensure that parameter estimation error is properly accounted for in the bootstrap procedure. Section 5 contains theorems which generalize Andrews (1997) and Whang (2000) to the case of dependent data. Section 6 contains a detailed example which is used to illustrate how to construct the diffusion specification test in practice, and reports the results from a series of Monte Carlo experiments based on the example. Concluding remarks are contained in Section 7. All proofs are collected in an appendix.

2 Parametric Diffusion Specification Test

In this section we outline a test for the joint correct specification of the drift and variance in stationary ergodic diffusion processes, $\{X(t),\, t \geq 0\}$. Consider the following hypotheses:
$H_0 : X(t)$ is a diffusion process solution to the following stochastic differential equation:

$$dX(t) = b(X(t), \theta)dt + \sigma(X(t), \theta)dW(t),$$

for some $\theta = \theta_0 \in \Theta$, where $\Theta \in \Re^k$ is compact.

H_A : The negation of H_0.

It is known (e.g. see Karlin and Taylor (1981) pp. 241) that for a given initial condition, the drift and variance terms ($b(\cdot)$ and $\sigma^2(\cdot)$, respectively) uniquely determine the stationary density associated with the invariant probability measure of the above diffusion process, say $f(x,\theta)$. In particular,

$$f(x,\theta) = \frac{\exp\left(\int_l^x [-2b(\xi,\theta)/\sigma^2(\xi,\theta)]d\xi\right)}{\sigma^2(x,\theta)\int_l^r (\sigma^2(\xi,\theta)m(\xi,\theta))^{-1}d\xi}, \tag{1}$$

where l and r are the lower and upper bound of the support of the diffusion, respectively, $m(x,\theta) = 1/(\sigma^2(x,\theta)s(x,\theta))$, and $s(x,\theta) = \exp\left(-2\int_l^x [b(\xi,\theta)/\sigma^2(\xi,\theta)]d\xi\right)$. Now, suppose that we observe a discrete sample (skeleton) of size T, say $(X_1, X_2, \ldots, X_T)'$, of the underlying diffusion $X(t)$, and construct an estimator of θ_0, say $\widehat{\theta}_T$, which is based on the skeleton. Hereafter we use the notation $X(t)$ for the continuous time process and the notation X_t for the skeleton. In addition, let $F_0(u)$ be the cumulative distribution function of the underlying diffusion. Given (1), H_0 is equivalent to $\widetilde{H}_0 : F(u,\theta_0) = F_0(u)\ \forall u \in U$,[2] and H_A is equivalent to $\widetilde{H}_A : F(u,\theta) \neq F_0(u)\ \forall \theta \in \Theta$, for some $u \in U$. For this reason we can test H_0 versus H_A by properly comparing the CDF associated with the density implied by the parametric specification under the null with the empirical CDF. Of course, as we do not know θ_0, we need to replace it with an estimator. This suggests using the following test statistic:

$$S_T^2 = \int_U S_T^2(u)\pi(u)du, \tag{2}$$

where $S_T(u) = \frac{1}{\sqrt{T}}\sum_{t=1}^T \left(1\{X_t \leq u\} - F(u,\widehat{\theta}_T)\right)$, with U the compact interval defined below (see Assumption 2), and $\int_U \pi(u)du = 1$. In the simulation exercise reported below, we shall also consider $|S_T| = \int_U |S_T(u)|\pi(u)du$, and $S_T = \sup_{u\in U}|S_T(u)|$. Note that S_T^2 does not have power against an *i.i.d.* alternative having a CDF equal to $F(u,\theta_0)$. This is also the case, for example, in Aït-Sahalia (1996). Furthermore, this case is ruled out by Assumption 1 (below), which requires that $X_1, \ldots X_T$ is a sample drawn from a geometric ergodic diffusion under both hypotheses. Finally, it is worth noting that a "conditional" version of our test does not follow straightforwardly in our framework, given that knowledge of the variance and drift terms does not imply knowledge of the (conditional) transition density. The following assumptions are used in the sequel.

[2]Note that $F(u,\theta_0)$ is the CDF implied by the model (i.e. $\int_l^u f(x,\theta_0)dx = F(u,\theta_0)$).

Assumption A1 (A1): $X(t), t \in \Re^+$ is a strictly stationary, geometric ergodic diffusion, under both the null and the alternative hypotheses. Under the null, the invariant density is $f(\cdot, \theta_0)$, with cumulative distribution function $F(\cdot, \theta_0)$.

Assumption A2 (A2): $F(u, \theta)$ is twice continuously differentiable in the interior of $\Theta \times U$, where Θ and U are compact subsets of $\Re^k$ and of $\Re$, respectively. Also, $\nabla_\theta F(u, \theta)$, $\nabla^2_\theta F(u, \theta)$ and $\nabla_{\theta,u} F(u, \theta)$ are jointly continuous on the interior of $\Theta \times U$.

Assumption A3 (A3): $\exists\ \theta^\dagger \in \Theta$ such that under both alternatives: (i) $\widehat{\theta}_T - \theta^\dagger = o_{a.s.}(1)$; hereafter let $q_t(\theta^\dagger) = q(X_t, \ldots, X_{t-m}, \theta^\dagger)$, m finite (ii) $\sqrt{T}(\widehat{\theta}_T - \theta^\dagger) = A_T(\overline{\theta}_T)\frac{1}{\sqrt{T}}\sum_{t=1}^T q_t(\theta^\dagger)$, where $\overline{\theta}_T \in (\widehat{\theta}_T, \theta^\dagger)$, $\frac{1}{\sqrt{T}}\sum_{t=1}^T q(X_t, \theta^\dagger) \xrightarrow{d} N(0, V_{q,\dagger})$, and $\sup_{\theta \in \Theta} |A_T(\theta) - A(\theta)| = o_{a.s.}(1)$; and (iii) $\sqrt{\frac{T}{2 \log \log T}}(\widehat{\theta}_T - \theta^\dagger) = O_{a.s.}(1)$, where $\theta^\dagger = \theta_0$ under H_0, and $\theta^\dagger \neq \theta_0$ under H_A.

Assumption A4 (A4): $q(X_t, \ldots, X_{t-m,}\theta)$ is continuosly differentiable on the interior of Θ, and the elements of $\nabla_\theta q(X_t, \ldots X_{t-m}, \theta)$ and of $q(X_t, \ldots, X_m, \theta)$ are $4r-$dominated, uniformly on Θ, with $r > 3/2$.[3]

Assumption A1 requires the diffusion to be geometric ergodic, under both hypotheses. Note also that A1 ensures that the skeleton is strong mixing with mixing coefficients decaying at a geometric rate. A2 imposes very mild smoothness requirements on the cumulative distribution function under the null, and is thus easily verified. A3 requires strong consistency of the estimator, asymptotic normality of $\frac{1}{\sqrt{T}}\sum_{t=1}^T q(X_t, \ldots X_{t-m}, \theta^\dagger)$, and an almost sure log log rate, under both hypotheses. In Section 4, we shall show that under quite primitive assumptions, (simulated) method of moments estimators satisfy A3**.** A4 is a standard moment condition. The asymptotic behavior of S_T^2 is summarized in the following theorem. Hereafter, let $q_t(\theta) = q(X_t, \ldots X_{t-m,}\theta)$. The following results then obtain.

Theorem 1: Let A1,A2,A3 and A4 hold.

(i) Under H_0,

$$S_T^2 \Rightarrow \int_U Z^2(u)\pi(u)du,$$

[3] Let $\nabla_\theta q(X_t, \ldots X_{t-m}, \theta)_i$ be the $i-th$ element of $\nabla_\theta q(X_t, \ldots X_{t-m}, \theta)$, $i = 1, \ldots m$. We require $\sup_{\theta \in \Theta} |\nabla_\theta q(X_t, \ldots X_{t-m}, \theta)_i| \leq g(X_t)$, with $E((g(X_t))^{4r}) < \infty$.

where Z is a Gaussian process with covariance kernel given by:

$$
\begin{aligned}
K(u,u') &= E(\sum_{s=-\infty}^{\infty}(1\{X_0 \leq u\} - F(u,\theta_0))(1\{X_s \leq u'\} - F(u',\theta_0)) \\
&\quad + \sum_{s=-\infty}^{\infty}(\nabla_\theta F(u,\theta_0)' A(\theta_0) q_0(\theta_0))(q_s(\theta_0)' A(\theta_0) \nabla_\theta F(u',\theta_0)) \\
-2\sum_{s=-\infty}^{\infty}(1\{X_0 &\leq u\} - F(u,\theta_0))(\nabla_\theta F(u',\theta_0)' A(\theta_0) q_s(\theta_0))) \qquad (3)
\end{aligned}
$$

(ii) Under H_A, there exists an $\varepsilon > 0$ such that $\forall \gamma < 1$,

$$\lim_{T\to\infty} \Pr(\frac{1}{T^\gamma} S_T^2 > \varepsilon) = 1.$$

As the estimated parameters are $\sqrt{T}$ consistent, parameter estimation error does not vanish asymptotically, but instead enters into the asymptotic covariance kernel (i.e. the last two lines in (3) summarize the contribution of parameter estimation error to the kernel). By noting that the parameter estimation error terms vanish asymptotically if the statistic is constructed using only R observations, with $R = o(T)$, we might consider forming

$$S_{R,T}^2 = \int_U S_{R,T}^2(u)\pi(u)du \qquad (4)$$

where $S_{R,T}(u) = \frac{1}{\sqrt{R}}\sum_{t=1}^{R}(1\{X_t \leq u\} - F(u,\widehat{\theta}_T))$.

Corollary 2: Under the same assumptions used in Theorem 1, if $R/T \to 0$ as $T \to \infty$, $R \to \infty$, then: (i) under H_0,

$$S_{R,T}^2 \Rightarrow \int_U \widetilde{Z_R}^2(u)\pi(u)du,$$

where $\widetilde{Z}$ is a Gaussian process with covariance kernel given by

$$\widetilde{K}(u,u') = E(\sum_{s=-\infty}^{\infty}(1\{X_0 \leq u\} - F(u,\theta_0))(1\{X_s \leq u'\} - F(u',\theta_0))), \qquad (5)$$

and (ii) under H_A, there exists an $\varepsilon > 0$, such that $\forall \gamma < 1$,

$$\lim_{R\to\infty} \Pr(\frac{1}{R^\gamma} S_{R,T}^2 > \varepsilon) = 1.$$

Note that we do not address the issue of power against $1/\sqrt{T}$ local alternatives. The reason is that the sequence of alternatives, say $\widetilde{b}(x) = b(x,\theta_0) + d(x)/\sqrt{T}$ and/or $\widetilde{\sigma}^2(x) = \sigma^2(x,\theta_0) + e(x)/\sqrt{T}$ does not necessarily imply, given (1) that the true density is say $\widetilde{f}(x) = f(x,\theta_0) + \eta(x)/\sqrt{T}$.

3 Bootstrap Critical Values

As the limiting distributions of S_T^2 and $S_{R,T}^2$ under H_0 are non standard, the usual approach is to bootstrap the critical values of the test.[4] In order to show the validity of the bootstrap, we shall obtain the limiting distribution of the bootstrapped statistic under both hypotheses and show that it coincides with the limiting distribution of the actual statistic under the null. Then, a test with correct asymptotic size and unit asymptotic power can be obtained by comparing the value of the original statistic with the bootstrapped critical values.

In our framework, we need to take into account both the dependence structure of the data and parameter estimation error when constructing bootstrap critical values. If the data consisted of *i.i.d.* observations, we could have proceeded along the lines of Andrews (1997), by drawing B samples of T *i.i.d.* obsevations from $F(u, \widehat{\theta}_T)$. In this case, the bootstrapped sample has limiting distribution $F(u, \theta^{\dagger})$, with $\theta^{\dagger} = \theta_0$ under the null and $\theta^{\dagger} \neq \theta_0$ under the alternative. However, our data are discrete realizations from a geometric ergodic diffusion, and so are strong mixing. In principle, we could instead follow the approach of Andrews (1997), and draw observations from the transition density. However, as mentioned above, we do not have general knowledge of the functional form of the transition density under the null hypothesis.

For the case of dependent observations, two approaches are available. One is the blockwise bootstrap of Künsch (1989). The validity of this particular bootstrap for empirical processes has recently been shown by Bühlmann (1995), and Radulovic (1996), Naik-Nimbalkar and Rajarshi (1994), and Peligrad (1998). A second is the conditional p-value approach proposed by Corradi and Swanson (2000) which extends Inoue (1998) to the case of non vanishing parameter estimation error, which in turn extends Hansen (1996) to the case of dependent observations. However, a drawback of this approach is that the simulated critical values diverge at rate l (where l plays the role of the blocksize lenght) under the alternative.

Our approach is to propose a stationary bootstrap procedure which is similar in the spirit to the stationary bootstrap of Politis and Romano (PR: 1994). The asymptotic validity of the PR bootstrap has been established under the assumption that the original statistic has a normal limiting distribution and does not contain estimated parameters. White (2000) generalizes the PR bootstrap

[4]As the limiting distributions in Theorem 1 and Corollary 2 are not pivotal, the bootstrap critical values do not provide any refinement of first order asymptotics (see e.g. Hall (1992) ch.3).

to statistics which contain estimated parameters, although he assumes that parameter estimation error vanishes asymptotically, while in our framework it does not. We begin by briefly outlining the PR bootstrap. Thereafter, we extend the PR bootstrap to the case where the distribution of the original statistic under the null is a functional of a Gaussian process whose covariance kernel contains terms which are affected by non vanishing parameter estimation error.

Let I_i be a discrete uniform random variable which can take values $1, 2, \ldots T, \forall i$, and let L_i be a geometrically distributed random variable (i.e. $\Pr(L_i = m) = (1-p)^{m-1}p$, $m = 1, 2, \ldots$). Also, I_i and L_i are independent of each other and independent of the sample $X_1, X_2, \ldots, X_T$, $\forall i$. Pseudo time series, X_t^*, are constructed as follows. First, draw a realization of I_i, say I_1, and a corresponding realization of L_i, say L_1. Then, form the sequence $X_1^* = X_{I_1}, \ldots, X_{L_1}^* = X_{I_1+L_1-1}$, which corresponds to the first L_1 observations of a single pseudo time series. Second, draw I_2 and L_2 and so on until there are T observations in the sequence X_t^*. This procedure is repeated B times, yielding B pseudo time series (for use in the construction of B pseudo statistics and the corresponding empirical distribution of these statistics). An important feature of the PR bootstrap is the treatment of end effects. In order to ensure the stationarity of each pseudo time series, whenever $I_m = T - m$, say, and $L_m = \widetilde{m} > m$, so that $I_m + L_m > T$, one has to wrap around to the beginning of the actual series when forming the block of pseudo observations based on I_m and L_m.[5]

We extend the PR bootstrap in two ways. First, note that PR show that conditional on the sample and for a generic measurable function, f, the limiting distribution of $\frac{1}{\sqrt{T}}\sum_{t=1}^T (f(X_t^*) - f(X_t))$ is the same as the limiting distribution of $\frac{1}{\sqrt{T}}\sum_{t=1}^T (f(X_t) - E(f(X_1)))$, provided that the latter is asymptotically normal. We, on the other hand, require that conditional on the sample and for a generic measurable function, f, $\frac{1}{\sqrt{T}}\sum_{t=1}^T (f(X_t^*(\cdot)) - f(X_t(\cdot)))$ and $\frac{1}{\sqrt{T}}\sum_{t=1}^T (f(X_t(\cdot)) - E(f(X_1(\cdot))))$ weakly converge to Gaussian processes with the same covariance structure. Thus, we require an empirical process version of the PR bootstrap. This is accomplished by showing that $\frac{1}{\sqrt{T}}\sum_{t=1}^T (f(X_t^*(\cdot)) - f(X_t(\cdot)))$ is stochastically equicontinuous on some compact set, conditional on the sample. Second, we require a bootstrap which takes parameter estimation error into account. An "apparently" natural extension of the PR bootstrap to the case of non vanishing parameter

[5]In the case of stationary mixing observations, the advantage of the PR bootstrap over Künsch (1989) blockwise bootstrap is that it ensures the stationarity of the (resampled) pseudo time series.

estimation error is based on the following statistic,

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

where $\widehat{\theta}_T^*$ is the (simulated) GMM estimator constructed using the pseudo time series. A mean value expansion around $\theta^\dagger$ yields

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(1\{X_t^*\leq u\}-1\{X_t\leq u\})-\left(\nabla_\theta F(u,\overline{\theta}_T\right)\sqrt{T}(\widehat{\theta}_T^*-\theta^\dagger)+\left(\nabla_\theta F(u,\widetilde{\theta}_T\right)\sqrt{T}(\widehat{\theta}_T-\theta^\dagger),$$

where $\overline{\theta}_T\in(\widehat{\theta}_T^*,\theta^\dagger)$ and $\widetilde{\theta}_T\in(\widehat{\theta}_T,\theta^\dagger)$. If, conditional on the sample, and for all samples except a subset of probability measure approaching zero, $\sqrt{T}(\widehat{\theta}_T^*-\widetilde{\theta}_T)\overset{P_*}{\to}0$, where P_* denotes the probability law governing the pseudo time series, then the contribution of parameter estimation error vanishes asymptotically. On the other hand, if $\sqrt{T}(\widehat{\theta}_T^*-\widetilde{\theta}_T)\overset{P_*}{\nrightarrow}0$, as is likely to be the case when we perform simulated GMM (this has been shown for the (non simulated) GMM case by Hall and Horowitz (1996) pp. 897-898), then the sum of the last two terms does not vanish, and in general does not mimic the contribution of parameter estimation error to the limiting distribution of the actual statistic. Thus, we need to follow a different route. We shall establish that $A_T(\theta^\dagger)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(q(X_t^*,\theta^\dagger)-q(X_t,\theta^\dagger))$ has the same limiting distribution as $A_T(\theta^\dagger)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}q(X_t,\theta^\dagger)$, provided that $E(q(X_1,\theta^\dagger))=0$. Thus, by taking this extra term into consideration we can account for the contribution of parameter estimation error to the covariance kernel of the Gaussian process to which S_T^2 weakly converges. Of course, we do not know $\theta^\dagger$, and if we replace it with an estimator we generate another source of estimation error. However, if we estimate θ using T observations and compute the bootstrapped statistic using R observations (and so we resample from $X_1,\ldots X_R$), with $R=o(T)$, we can account for the original parameter estimation error by adding the term $A_T(\widehat{\theta}_T)\frac{1}{\sqrt{R}}\sum_{t=1}^{R}(q(X_t^*,\widehat{\theta}_T)-q(X_t,\widehat{\theta}_T))$ to the "basic" bootstrap statistic, $\frac{1}{\sqrt{R}}\sum_{t=1}^{R}(1\{X_t^*\leq u\}-1\{X_t\leq u\})$. In fact, given **A3** and **A4,**

$$\frac{1}{\sqrt{R}}\sum_{t=1}^{R}(q(X_t^*,\widehat{\theta}_T)-q(X_t,\widehat{\theta}_T))=\frac{1}{\sqrt{R}}\sum_{t=1}^{R}(q(X_t^*,\theta^\dagger)-q(X_t,\theta^\dagger))$$

$$+\frac{1}{\sqrt{T}\sqrt{R}}\sum_{t=1}^{R}\nabla_\theta(q(X_t^*,\overline{\theta}_T)-q(X_t,\overline{\theta}_T))\sqrt{T}(\widehat{\theta}_T-\theta^\dagger) \qquad (6)$$

with the term in (6) approaching zero, conditional on the sample, and for all samples except a subset of probability measure approaching zero. For this reason, we shall compute the bootstrapped statistic using R observations, where $R = o(T)$. Broadly speaking, we have two alternative strategies. (I) Compute the actual statistic using T observations, and then compute the bootstrap statistic using R observations, adding extra terms to account for non vanishing parameter estimation error. (II) Compute the actual statistic using R observations, so that there is no parameter estimation error, and then compute the bootstrap statistic using T observations, without adding extra terms.[6] The advantage of (I) over (II) is that under the alternative hypothesis, the actual statistic diverges at rate $\sqrt{T}$, instead of $\sqrt{R}$, hence ensuring better finite sample power. The disadvantage is that the bootstraped critical values are less accurate "approximations" of the asymptotic critical values, as they have been computed from a sample of R instead of T observations. The bootstrap statistics associated with (I) and (II) can be written as follows.[7] For (I),

$$Z^2_{R,T}(\omega) = \int_U Z^2_R(u,\omega)\pi(u)du, \tag{7}$$

with $\int_U \pi(u)du = 1$, and

$$\begin{aligned} Z_{R,T}(u,\omega) &= \frac{1}{\sqrt{R}}\sum_{t=1}^{R}\left(1\{X^*_t \le u\} - 1\{X_t \le u\}\right) \\ &\quad - \left(\nabla_\theta F(u,\widehat{\theta}_T)' A_T(\widehat{\theta}_T)\right)\frac{1}{\sqrt{R}}\sum_{t=1}^{R}\left(q(X^*_t,\widehat{\theta}_T) - q(X_t,\widehat{\theta}_T)\right), \end{aligned} \tag{8}$$

where $q(\cdot,\cdot)$ is a score vector and $A_T(\widehat{\theta}_T)$ is a matrix of second derivatives, both of which depend on the null model and on the estimator used (e.g. which moment conditions are used). The reader is referred to Section 6 for a detailed example. Additionally, note that ω is used to stress sample dependence. So far we have dealt with the case in which $q_t(\theta)$ depends only on the current value, X_t. In practice, it may also depend on lagged values of X_t (see Assumption 3). In such

[6] One might think of avoiding resampling by instead simulating B trajectories via an Euler scheme, for example, using $\widehat{\theta}_T$ (e.g. call $X^{\widehat{\theta}_T}_{t,i}$ $i = 1,\ldots,B$ the simulated trajectory sampled at the same frequency of the data), constructing B statisticics based on

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(1\{X_t \le u\} - 1\{X^{\widehat{\theta}_T}_{t,i} \le u\}),\ i = 1,\ldots,B,$$

and obtaining the empirical distribution. However, critical values constructed in this manner diverge under the alternative, and so are not valid.

[7] In practice, it might be useful to compute both statistics.

a case, we should construct the pseudo time series in the following way. Suppose that $q_t(\widehat{\theta}_T) = q(X_t \ldots X_{t-m}, \widehat{\theta}_T)$, let I_i be *i.i.d.* uniform on $m, m+1, \ldots, R$, let L_i be as defined above, and let $z_t = 1\{X_t \leq u\} - \nabla_\theta F(u, \widehat{\theta}_T)' A_T(\widehat{\theta}_T) q(X_t \ldots X_{t-m}, \widehat{\theta}_T)$. Thus, the pseudo time series z_t^* is defined as follows: $z_m^* = 1\{X_{I_1} \leq u\} - \nabla_\theta F(u, \widehat{\theta}_T)' A_T(\widehat{\theta}_T) q(X_{I_1} \ldots X_{I_{1-m}}, \widehat{\theta}_T)$, $z_{m+L_1}^* = 1\{X_{I_1+L_1-1} \leq u\} - \nabla_\theta F(u, \widehat{\theta}_T)' A_T(\widehat{\theta}_T) q(X_{I_1+L_1} \ldots X_{I_{1+L_1-m-1}}, \widehat{\theta}_T)$ and so on. In this case we have

$$Z_{R,T}(u, \omega) = \frac{1}{\sqrt{R}} \sum_{t=m}^{R} (z_t^* - z_t)\,. \tag{9}$$

The statistic associated with (II) (vanishing parameter estimation error) is,

$$\widetilde{Z}_T^2(\omega) = \int_U \widetilde{Z}_T^2(u, \omega) \pi(u) du,$$

where

$$\widetilde{Z}_T(u, \omega) = \frac{1}{\sqrt{R}} \sum_{t=1}^{R} \left(1\{X_t^* \leq u\} - 1\{X_t \leq u\}\right).$$

Theorem 3: Let A1, A2, A3 and A4 hold. If (i) $R/T \to 0$ as $T \to \infty$, $R \to \infty$, and (ii) $p = p_R = R^{-\delta}$, $0 < \delta < 1$, then:[8] under H_0, for $\varepsilon > 0$ and arbitrarily small,

$$\lim_{R \to \infty} P\left\{\omega : \rho\left(L(Z_{R,T}^2(\omega)), L(S_T^2)\right) > \varepsilon\right\} = 0,$$

where $L(Z_{R,T}^2(\omega))$ and $L(S_T^2)$ denote the probability laws of $Z_{R,T}^2(\omega)$ and $L(S_T^2)$ respectively, with $Z_{R,T}(u, \omega)$ defined either as in (8) or as in (9), and where ρ is any metric metrizing convergence in distribution. Under H_A,

$$\lim_{R \to \infty} P\left\{\omega : \rho\left(\begin{array}{c} L(Z_{R,T}^2(\omega)), L(\int_U (\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (1\{X_t \leq u\} - E(1\{X_t \leq u\})) \\ -\nabla_\theta F(u, \theta^\dagger)' A(\theta^\dagger) q_t(X_t, \theta^\dagger))^2 \pi(u) du) \end{array}\right) > \varepsilon\right\} = 0.$$

Thus, under the null the bootstrapped statistic has the same limiting distribution as the actual statistic, on the other hand under the alternative the actual statistic diverges to infinity at rate T while the bootstrapped statistic converges in distribution.

Corollary 4: Let A1, A2, A3 and A4 hold. Then, as $T \to \infty$, $p = p_T = T^{-\delta}$, $0 < \delta < 1$, under H_0,

$$\lim_{T \to \infty} P\left\{\omega : \rho\left(L(\widetilde{Z}_T^2(\omega)), L(S_{R,T}^2)\right) > \varepsilon\right\} = 0$$

[8] Recall that p is the parameter governing the geometric distribution from which the random block lengths are drawn. Note also that we require $p = p_R$ to approach zero as $R \to \infty$, but not at *too* fast rate.

and under H_A,

$$\lim_{R\to\infty} P\left\{\omega : \rho\left(L(\widetilde{Z}_T^2(\omega)), L(\int_U (\frac{1}{\sqrt{R}}\sum_{t=1}^{R}(1\{X_t \leq u\} - E(1\{X_t \leq u\})))^2 \pi(u)du)\right) > \varepsilon\right\} = 0.$$

Thus, we have that under the null the actual and the bootstrap statistics have the same limiting distribution, while under the alternative the actual statistic diverges at rate R, while the boostrap statistic converges in distribution.

The above results suggest proceeding in the following manner. For any bootstrap replication, compute the bootstrapped statistic, $Z_{R,T}^2(\omega)$ ($\widetilde{Z}_T^2(\omega)$). Perform B bootstrap replications (B large) and compute the percentiles of the empirical distribution of the B bootstrapped statistics. Reject H_0 if S_T^2 ($S_{R,T}^2$) is greater than the $(1-\alpha)th$-percentile of $Z_{R,T}^2(\omega)$ ($\widetilde{Z}_T^2(\omega)$). Otherwise, do not reject H_0. Now, for all samples, except a set with probability measure approaching zero, S_T^2 ($S_{R,T}^2$) has the same distribution as the bootstrapped statistic $Z_{R,T}^2(\omega)$ ($\widetilde{Z}_T^2(\omega)$), under the null. Thus, the above approach ensures that the test has asymptotic size equal to α. Under the alternative, S_T^2 ($S_{R,T}^2$) diverge to plus infinity, while $Z_{R,T}^2(\omega)$ ($\widetilde{Z}_T^2(\omega)$) 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. In our case the limiting distribution is a functional of a Gaussian process, so that we do not know the explicit density function. Thus, we cannot directly apply Andrews and Buchinsky approach; nevertheless in the Monte Carlo section below, we carefully 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 surprisingly good finite sample properties.

4 Simulation Based Estimators

In this section, we provide sufficient conditions under which simulated method of moments (SGMM) estimators satisfy A3.[9]

In order to construct simulated estimators, we require simulated paths. Assume, for example

[9]For the sake of simplicity, we focus on SGMM estimators. However, under analogous primitive conditions, Indirect Inference (II: Gourieroux, Monfort and Renault (1993)) and Efficient Method of Moment (EMM: Gallant and Tauchen (1996)) estimators satisfy A3 too. A unified framework for simulation based estimators, which nests SGMM, II and EMM is provided in Dridi (1999).

that diffusion paths are generated under H_0 using the Euler scheme.[10] Let X^{θ}_{kh} be the simulated path, with

$$X^{\theta}_{kh} - X^{\theta}_{(k-1)h} = b(X^{\theta}_{(k-1)h}, \theta)h + \sigma(X^{\theta}_{(k-1)h}, \theta)\epsilon_{kh},$$

where $\epsilon_{kh} \overset{iid}{\sim} N(0, h)$, $k = 1, \ldots, N$, and $Nh = T$. Thus, we are imposing (without loss of generality) that the simulated time span and the time span of the data are the same. In addition, let $N = T^2$, so that $h = T^{-1}$. This ensures that $h^2 T \to 0$ as $T \to \infty$, $Nh \to \infty$, and $h \to 0$. For any k such that $kh = t = 1, 2, \ldots, T$, it follows from Pardoux and Talay (1985 pp. 33 and pp. 37) that,

$$E|X^{\theta}_{kh} - X_t|^2 = O(h), \ \forall \theta \in \Theta,$$

and that for any g Lipschitz continuous,

$$E(g(X^{\theta}_{kh})) - E(g(X_t)) = O(h)$$

Hereafter, let X^{θ}_t, $t = 1, \ldots, T$ be a simulated path sampled at the same frequency as the data. (For notational simplicity we do not explicitly note that X^{θ}_t is dependent on h.) In addition, define the SGMM estimator,

$$\widehat{\theta}_T = \arg\min_{\theta \in \Theta} G_T(\theta)' W_T G_T(\theta),$$

where,

$$G_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} \left(g_t - g^{\theta}_t \right),$$

with $g_t = g(X_t)$ and $g^{\theta}_t = g(X^{\theta}_t)$. In order to obtain the desired result we shall need the following assumptions.

Assumption B1 (B1): For any fixed h and $\forall \theta \in \Theta$, X^{θ}_{kh} is geometrically ergodic and strictly stationary.[11]

Assumption B2 (B2): $W_T \overset{a.s.}{\to} W_0 = \sum_0^{-1}$, where, $\sum_0 = \sum_{j=-\infty}^{\infty} E\left((g_1 - E(g_1))(g_{1+j} - E(g_{1+j}))'\right).$

[10] As an alternative to the Euler scheme, the Milshtein scheme could be used, for example, possibly leading to a more accurate approximation of the paths (see e.g. Pardoux and Talay (1985) or Gard (1988)). In our Monte Carlo experiments (see Section 5), the Milshtein scheme yields paths that are the same as the Euler scheme, up to three decimal places.

[11] Stramer and Tweedie (1997) propose a new algorithm for simulating the path of a diffusion which ensures that the geometric ergodicity of the underlying diffusion is inheredited by the simulated paths. This is in general the case for the Euler scheme, when the drift grows at most at a linear rate and drift and variance are not "too big".

Assumption B3 (B3): $\forall\theta \in \Theta$, $||g_t^\theta||_{2+\delta} < C < \infty$, g_t^θ is almost surely Lipschitz, uniformly on Θ, and $\theta \to E(g_1^\theta)$ is continuous, g_t, g_t^θ are $2r-$dominated (the latter uniformly on Θ) for $r > 3/2$.
Assumption B4 (B4): Let $C_\infty(\theta) = G_\infty(\theta)'W_0G_\infty(\theta)$, with $G_T(\theta) \overset{a.s.}{\to} G_\infty(\theta)$. Assume that $C_\infty(\theta^\dagger) < C_\infty(\theta)$, $\forall\theta \in \Theta$ and $\theta \neq \theta^\dagger$.
Assumption B5 (B5): (i) $\widehat{\theta}_T$ and $\theta^\dagger$ are in the interior of Θ. (ii) g_t^θ is twice continuously differentiable in the interior of Θ. (iii) $D_0 = E(\partial g_1^\theta/\partial\theta|_{\theta=\theta^\dagger})$ exists and is of full rank.
Assumption B6 (B6): $\frac{1}{\sqrt{2T\log\log T}}\sum_{t=1}^{T}((g_t - E(g_1)) - (g_t^{\theta^\dagger} - E(g_1^{\theta^\dagger}))) = O_{a.s.}(1)$.[12]
Assumption B7 (B7): $E(g_1) - E(g_1^{\theta^\dagger}) = O(h)$ under both H_0 and H_A.
Proposition 5: Let A1 and B1-B7 hold. If $h^2T \to 0$, $h \to 0$, and $Nh \to 0$ as $T \to \infty$, then A3 is satisfied.

5 Generalization of Andrews (1997) and Whang (2000)

The purpose of this section is to illustrate that the proposed bootstrap procedure is generally applicable to a wide variety of tests. In particular, it can be used to obtain critical values whenever the limiting distribution of a test is a Gaussian process with covariance kernel reflecting both the time series nature of the data and non vanishing parameter estimation error. Along these lines, we generalize Andrews' (1997) conditional Kolmogorov test and Whang's (2000) conditional expectation test to the case of dependent observations. We shall consider only the case in which the statistic is computed using T observations and the bootstrap statistic is computed using R observations and contains a term which captures the contribution of parameter estimation error to the limiting distribution. The case of vanishing parameter estimation error would follow as an immediate corollary

5.1 Andrews' (1997) Conditional Kolmogorov Test

Andrews (1997) proposes a test for the null of correct specification of the conditional distribution in the case of $i.i.d.$ observations. Let $H(y_t|X_t)$ be the distribution of y_t conditional on X_t (in the sequel we shall assume that y_t scalar and that X_t is a qx1 vector). The null hypothesis is

$$H_0 : H(\cdot|\cdot) = F(\cdot|\cdot, \theta_0),\ \text{for some}\ \theta_0 \in \Theta, \tag{10}$$

[12]Sufficient conditions for a strong mixing process to satisfy B6 are given in Sin and White (1996) and in Altissimo and Corradi (2000), for example.

and the alternative is the negation of H_0. Here we generalize Andrews' test to the case of stationary dependent observations (see Assumptiom C1(i) below). It is important to point out that our purpose is to test for the correct specification of the conditional distribution of y_t given X_t, and not to test for the correct specification of the conditional distribution of y_t given all the past history, say $\Im_t$. Thus, we allow for possible dynamic misspecification, under both hypotheses. As discussed above, Bai (1998) develops a test for the null hypothesis of the correct specification of the distribution of y_t given $\Im_t$ which has a nuisance parameter free limiting distribution, so that critical value can be tabulated.[13]

In order to generalize the conditional Kolmogorov test, we rely on the empirical analog of $H(z)$, $z = (u, v)$, 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 $F(z, \theta_0)$, say $\widehat{F}_T(u, v, \widehat{\theta}_T) = \frac{1}{T}\sum_{t=1}^{T} F(u|X_t, \widehat{\theta}_T)1\{X_t < v\}$. We can then specify tests of the following sort:

$$KS_T = \sup_{u \times v \in U \times V} |KS_T(u, v)|, \tag{11}$$

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

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

Note that KS_T is the same statistic as given in equation (3.9) of Andrews (1997). A grid serach over $U \times V$ may be computationally demanding when V is high-dimensional. Andrews has shown that by taking the maximum over $Z_t = (y_t, X_t)$ $t = 1, \ldots T$ we get an asymptotically equivalent test. We cannot show that this is true also in the dependent case, as Andrews' Lemma A6 does no longer hold. In the sequel we shall rely on the following assumption.

Assumption C1 (C1): (i) (y_t, X_t), with y_t scalar and X_t $\Re^q-$valued, are stationary and strong mixing with size $-\frac{3(6+\psi)}{\psi}$, $\psi > 0$. (ii) $F(u|X_t, \theta)$ is differentiable on $U \times \Theta$, where U and Θ are compact subsets of $\Re$ and $\Re^k$ respectively, $\nabla_\theta F(u|X_t, \theta)$ and $\nabla_{u,\theta} F(u|X_t, \theta)$ are jointly continuous on $U \times \Theta$ and the elements of $\nabla_\theta F(u|X_t, \theta)$, $\nabla_u F(u|X_t, \theta)$, and $\nabla_{u,\theta} F(u|X_t, \theta)$ are $4r-$dominated uniformly on $U \times \Theta$ for $r > 3/2$. (iii) $\exists\ \theta^\dagger \in \Theta$ such that under both alternatives: (iiia) $\widehat{\theta}_T - \theta^\dagger = o_{a.s.}(1)$; (iiib) $\sqrt{T}(\widehat{\theta}_T - \theta^\dagger) = A_T(\overline{\theta}_T)\frac{1}{\sqrt{T}}\sum_{t=1}^{T} q(y_t, X_t, \theta^\dagger)$, where $\overline{\theta}_T \in (\widehat{\theta}_T, \theta^\dagger)$,

[13] Note that Diebold, Gunther and Tay (1998) show that, under H_0, $F(y_t|\Im_t, \theta_0)$ is distributed as an *i.i.d.* uniform random variable on $[0, 1]$. Bai (1998) allows for the case in which θ_0 is replaced by $\widehat{\theta}_T$, and, using a martingalization argument, provides a test with a nuisance parameter free limiting distribution.

$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} q(y_t, X_t, \theta^\dagger) \xrightarrow{d} N(0, V_{q,\dagger})$, and $\sup_{\theta\in\Theta} |A_T(\theta)-A(\theta)| = o_{a.s.}(1)$; (iiic) $\sqrt{\frac{T}{2\log\log T}}(\widehat{\theta}_T-\theta^\dagger) = O_{a.s.}(1)$, where $\theta^\dagger = \theta_0$ under H_0, and $\theta^\dagger \neq \theta_0$ under H_A; and (iiid) $q(y_t, X_t, \theta)$ is $4r-$dominated uniformly on Θ for $r > 3/2$.

We now have,

Theorem 6: Let C1 hold, then: (i) Under H_0, $KS_T \Rightarrow \sup_{u\times v\in U\times V} |Z(u, v)|$, where KS_T is defined in (11) and Z is a Gaussian process with covariance kernel $K(u, v, u', v')$ given by:

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

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

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

where $q_s(\theta_0) = q(y_s, X_s, \theta_0)$. (ii) Under H_A, there exists an $\varepsilon > 0$ such that $\forall\gamma < 1/2$, $\lim_{T\to\infty} \Pr(\frac{1}{T^\gamma}KS_T > \varepsilon) = 1$.

We can now use the procedure outlined in Section 3 to provide asymptotically valid critical values for the limiting distribution of KS_T. In particular, let

$$z_t(u, v) = (1\{y_t \leq u\} - F(u|X_t, \widehat{\theta}_T))1\{X_t \leq v\}$$
$$-\left(\left(\frac{1}{T}\sum_{t=1}^{T}\nabla_\theta F(u|X_t, \widehat{\theta}_T)'1\{X_t \leq v\}\right)A_T(\widehat{\theta}_T)q_t(y_t, X_t, \widehat{\theta}_T)\right),\ t = 1, \ldots R$$

and resample z_t according to the PR stationary bootstrap, so that $z_1^*(u, v) = z_{I_1}(u, v), \ldots, z_{L_1}^*(u, v) = z_{I_1+L_1-1}(u, v), \ldots, z_{L_1+1}^*(u, v) = z_{I_2}(u, v)$ and so on, where I_i and L_i are defined as above. Additionally, define

$$KS_{R,T}^*(\omega) = \sup_{u\times v\in U\times V}|KS_{R,T}^*(\omega, u, v)|$$

with

$$KS_{R,T}^*(\omega, u, v) = \frac{1}{R^{1/2}}\sum_{t=1}^{R}(z_t^*(u, v) - z_t(u, v)).$$

Theorem 7: Let C1 hold. If (i) $R/T \to 0$ as $T \to \infty$, $R \to \infty$, and (ii) $p = R^{-\delta}$, $0 < \delta < 1$, then:[14] under H_0, for $\varepsilon > 0$ and arbitrarily small, $\lim_{R\to\infty} P\left\{\omega : \rho\left(L(KS_{R,T}^*(\omega)), L(KS_T)\right) > \varepsilon\right\} = 0$,

[14] Recall that p is the parameter governing the geometric distribution from which the random block lengths are drawn.

where $L(KS^*_{R,T}(\omega))$ and $L(KS_T)$ denote the probability laws of $KS^*_{R,T}(\omega)$ and KS_T, respectively, and ρ is any metric metrizing convergence in distribution. Under H_A,

$$\lim_{R\to\infty} P\left\{\omega : \rho\left(L(KS^*_{R,T}(\omega)), L(\sup_{u\times v\in U\times V}\left|\frac{1}{R^{1/2}}\sum_{t=1}^{R}(z_t(u,v)-E(z_t(u,v)))\right|)\right) > \varepsilon\right\} = 0.$$

Thus, the bootstrapped statistic has the same limiting distribution as the actual statistic under H_0, and the actual statistic diverges at rate T, while the bootstrapped statistic converges in distribution, under H_A.

5.2 Whang's (2000) Test for Parametric Regression Functions

Whang's (2000) test for the correct specification of a parametric regression function is a generalization of Kolmogorov test to the regression framework, for the case of *i.i.d.* observations. The null hypothesis is formulated as:

$$\begin{aligned} H_0 &: \Pr(E(y_t|X_t) = g(X_t,\theta_0)) = 1, \text{ for some } \theta_0 \in \Theta, \text{ and} \\ H_A &: \Pr(E(y_t|X_t) = g(X_t,\theta)) < 1, \ \forall \theta \in \Theta. \end{aligned}$$

Our objective is to generalize Whang's test to the case of dependent observations. Note that in this example, H_0 implies correct specification of the mean of y_t conditional on X_t, and does not rule out the possibility of dynamic mispecification. Thus, $y_t - g(X_t,\theta^\dagger)$ (with $\theta^\dagger = \theta_0$ under H_0) is not assumed to be a martingale difference sequence under either H_0 or H_A. Define,

$$K_T = \sup_{v\times V} |K_T(v)|, \tag{12}$$

where V is a compact subset of $\Re^q$, and

$$K_T(v) = \frac{1}{T^{1/2}}\sum_{t=1}^{T}\left((y_t - g(X_t,\widehat{\theta}_T))1\{X_t \leq v\}\right)$$

Note that K_T is the same statistic as given in equation (5) of Whang (2000). Further, assume that the following holds:

Assumption C1′: (i) (y_t, X_t), with y_t scalar and X_t $\Re^q-$valued, are stationary and absolutely regular mixing processes (i.e. $\beta-$mixing) with size $-\frac{3(6+\psi)}{\psi}$, $\psi > 0$.[15] (ii) $E(y_t^6) < \infty$, $g(X_t,\theta)$ is

[15] β-mixing is a stronger requirement than $\alpha-$mixing, but weaker than $\phi-$mixing. In fact for any given two sigma fields $\mathcal{A}$ and $\mathcal{B}$, $2\alpha(\mathcal{A},\mathcal{B}) \leq \alpha(\mathcal{A},\mathcal{B}) \leq \phi(\mathcal{A},\mathcal{B})$ (e.g. see Doukhan, Massart and Rio (1995 pp.397)).

continuously differentiable on the interior of Θ, where Θ is a compact subset of $\Re^k$, $g(X_t, \theta)$ and $\nabla_\theta g(X_t, \theta)$ are $4r-$dominated uniformly on Θ for $r > 3/2$. (iii) as in C1(iii)(a)-(iii)(d) above. Note that in C$'$(i) we require y_t, X_t to be absolutely regular mixing (or $\beta-$mixing), while in C(i) above we just require strong mixing. The reason is that neither y_t nor $g(X_t, \theta)$ are bounded sequences, so we require absolute regularity in proving the stochastic equicontinuity of the (bootstrap) statistic.

Theorem 8: Let C1' hold. Then: (i) Under H_0, $K_T \Rightarrow \sup_{v \in V} |Z(v)|$, where K_T is defined in (12) and Z is a Gaussian process with covariance kernel $K(v, v')$ given by:

$$E(\sum_{s=-\infty}^{\infty} ((y_0 - g(X_0, \theta_0))1\{X_0 \leq v\})((y_s - g(X_s, \theta_0))1\{X_s \leq v'\}))$$

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

$$-2 \sum_{s=-\infty}^{\infty} ((y_0 - g(X_0, \theta_0))1\{X_0 \leq v\})E(\nabla_\theta g(X_s, \theta_0)'1\{X_s \leq v'\})A(\theta_0)q_s(\theta_0)),$$

where $q_s(\theta_0) = q(y_s, X_s, \theta_0)$. (ii) Under H_A, there exists an $\varepsilon > 0$, such that $\forall \gamma < 1/2$, $\lim_{T \to \infty} \Pr(\frac{1}{T^\gamma} K_T > \varepsilon) = 1$.

As above, we can now use the procedure outlined in Section 3 to provide asymptotically valid critical values for the limiting distribution of K_T. In particular, let

$$z_t(v) = (y_t - g(X_t, \widehat{\theta}_T))1\{X_t \leq v\}$$
$$-\left(\left(\frac{1}{T}\sum_{t=1}^{T} \nabla_\theta g(X_t, \widehat{\theta}_T)'1\{X_t \leq v\}\right) A_T(\widehat{\theta}_T)q_t(y_t, X_t, \widehat{\theta}_T)\right)$$

and resample z_t according to the PR stationary bootstrap, so that $z_1^*(v) = z_{I_1}(v), \ldots, z_{L_1}^*(v) = z_{I_1+L_1-1}(v), \ldots, z_{L_1+1}^*(v) = z_{I_2}(v)$ and so on, where I_i and L_i are defined as above. Additionally, define

$$K_{R,T}^*(\omega) = \sup_{v \in V} |K_{R,T}^*(v)|$$

with

$$K_{R,T}^*(\omega, v) = \frac{1}{R^{1/2}} \sum_{t=1}^{R} (z_t^*(v) - z_t(v))$$

Theorem 9: Let C1$'$ hold. If (i) $R/T \to 0$ as $T \to \infty$, $R \to \infty$, and (ii) $p = R^{-\delta}$, $0 < \delta < 1$, then under H_0, for $\varepsilon > 0$ and arbitrarily small, $\lim_{R \to \infty} P\left\{\omega : \rho\left(L(K_{R,T}^*(\omega)), L(K_T)\right) > \varepsilon\right\} = 0$, where $L(K^*_{R,T}(\omega))$ and $L(K_T)$ denote the probability laws of $K^*_{R,T}(\omega)$ and KS_T respectively, and ρ is any metric metrizing convergence in distribution. Under H_A,

$$\lim_{R\to\infty} P\left\{\omega : \rho\left(L(K^*_{R,T}(\omega)), L(\sup_{v\in V}\left|\frac{1}{R^{1/2}}\sum_{t=1}^{R}(z_t(v) - E(z_t(v)))\right|)\right) > \varepsilon\right\} = 1.$$

It is apparent that Theorem 9 is a conditional mean version of Theorem 7. In the next section, an application of the diffusion specification test outlined in Section 2 is discussed, and some Monte Carlo findings are presented.

6 An Example

6.1 Testing the Null Hypothesis of a Square Root Process

In this section we discuss implementation of the diffusion specification test when a square root process such as that used by Cox, Ingersoll and Ross (1985) is assumed to hold under H_0. Under H_A, it is assumed that the logarithm of the diffusion follows an Ornstein-Uhlenbeck process, so that X_t is lognormal $\forall t$ provided that X_0 is drawn from a lognormal density. We shall show that the assumptions used in Proposition 5 hold, so that S^2_T $(S^2_{R,T})$ and $Z^2_{R,T}(\omega)$ $(\widetilde{Z}^2_T(\omega))$ can be constructed and used as described above. The diffusion under the null is obtained by defining the drift and the variance as in equations (10) and (11) in Wong (1964) and setting $\beta = 1$, $c = 0$, $d = c_1$, $e = 0$ in equation 10, and $a = -1$, $b = -a$ in equation (11). The resulting diffusion is indeed a square root process. In particular, define

H_0 : $X(t)$ is a diffusion process solution to the following stochastic differential equation:

$$dX(t) = ((c_1 - a) - X(t))dt + \sqrt{c_1 X(t)}dW(t),\ \ c_1 > 0 \text{ and } c_1 - a > 0, \tag{13}$$

for some $\theta = \theta_0 = (a_0, c_{10}) \in \Theta$, where $\Theta \in R^2$ is compact.[16]

H_A : The negation of H_0.

From Wong (1964, pp. 264-265), we know that the stationary density of $X(t)$ under H_0 belongs to the linear exponential (or Pearson) family. The process has a non-central chi-squared transition

[16]As discussed below, the conditions $c_1 > 0$ and $c_1 - a > 0$ are imposed in order to ensure that X_t is geometric ergodic and has an invariant gamma density under H_0. When defining the Cox, Ingersoll and Ross model in his Table 3, Aït-Sahalia (1996 pp. 403) additionally requires that $c_1 > 2a$. However, as we set $a < 0$ in all of our Monte Carlo experiments, this additional restriction is not needed in order to ensure that Properties 1-5 of Aït-Sahalia (1996 Section 2) hold for our null model.

density, and the invariant density is a gamma. Thus, in this case the functional form of the transition function is known and we could have potentially implement the approach developed by Bai (1998). Nevertheless, the main purpose of our Monte Carlo studies is to analyze the finite sample properties of the bootstrap procedure we have suggested. The invariant density can be written as[17]:

$$f(x;a,c_1)=\frac{(\frac{c_1}{2})^{-2(1-a/c_1)}x^{2(1-a/c_1)-1}\exp(-x/(\frac{c_1}{2}))}{\Gamma(2(1-a/c_1))},\quad \Gamma(2(1-a/c_1))=\int_0^{\infty}t^{2(1-a/c_1)-1}\exp(-t)dt.$$

It follows that,

$$F(u;a,c_1)=\frac{\int_0^u(\frac{c_1}{2})^{-2(1-a/c_1)}x^{2(1-a/c_1)-1}\exp(-x/(\frac{c_1}{2}))dx}{\Gamma(2(1-a/c_1))}.$$

The simplest moment conditions which ensure (exact) identification when applying SGMM in this example are the sample mean and variance, minus their respective probability limits. Note that in this example we have analytical expressions for these moments. In particular, the mean is (c_1-a) and the variance is $\frac{c_1}{2}(c_1-a)$, (see e.g. Johnson, Kotz and Balakrishnan ((1994), ch. 17). Thus, we can apply GMM rather than SGMM, with

$$(\widehat{a},\ \widehat{c}_1)'=\arg\min_{a,c_1\in\Theta}G'(\theta)W_TG(\theta),\tag{14}$$

where

$$G'(\theta)=\left(\ \left(\frac{1}{T}\sum_{t=1}^T X_t-(c_1-a)\right),\quad \left(\frac{1}{T}\sum_{t=1}^T(X_t-\overline{X})^2-\frac{c_1}{2}(c_1-a)\right)\ \right).$$

and

$$\begin{aligned}W_T^{-1} &= \frac{1}{T}\sum_{t=1}^T(f_t-\overline{f})(f_t-\overline{f})'\\ &+\frac{1}{T}\sum_{\tau=1}^{l_T}w_\tau\sum_{t=\tau+t}^T((f_t-\overline{f})(f_{t-\tau}-\overline{f})'+(f_{t-\tau}-\overline{f})(f_t-\overline{f})'),\end{aligned}$$

where $f_t=(f_{1t},f_{2t})'$, $f_{1t}=X_t$, $f_{2t}=(X_t-\overline{X})^2$, $\overline{f}=\frac{1}{T}\sum_{t=1}^T f_t$, and $w_\tau=1-\frac{\tau}{l_T+1}$. Thus, $(\widehat{a},\widehat{c}_1)'$ is the solution to:

$$\begin{pmatrix}1 & \frac{\widehat{c}_1}{2}\\ -1 & -\widehat{c}_1+\frac{\widehat{a}}{2}\end{pmatrix}W_T\begin{pmatrix}\frac{1}{T}\sum_{t=1}^T X_t-(\widehat{c}_1-\widehat{a})\\ \frac{1}{T}\sum_{t=1}^T(X_t-\overline{X})^2-\frac{\widehat{c}_1}{2}(\widehat{c}_1-\widehat{a})\end{pmatrix}=0$$

[17]Note that the invariant density under the null hypothesis can be written as $f(x;\alpha_1,\alpha_2)=\frac{\alpha_2^{-\alpha_1}x^{\alpha_1-1}\exp(-x/\alpha_2)}{\Gamma(\alpha_1)}$,which is the standard form of the density (see e.g. Johnson, Kotz and Balakrishnan (1994)). In our case, $\alpha_1=2(c_1-a)/c_1$ and $\alpha_2=c_1/2$.

Now, take a first order expansion around the "true" parameter values, say a_0 and $c_{1,0}$. Then, the above solution can be written as:

$$\begin{pmatrix} 1 & \frac{\widehat{c}_1}{2} \\ -1 & -\widehat{c}_1 + \frac{\widehat{a}}{2} \end{pmatrix} W_T \begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T} X_t - (c_0 - a_0) \\ \frac{1}{T}\sum_{t=1}^{T} (X_t - \overline{X})^2 - \frac{c_{10}}{2}(c_{10} - a_0) \end{pmatrix}$$

$$+ \begin{pmatrix} 1 & \frac{\widehat{c}_1}{2} \\ -1 & -\widehat{c}_1 + \frac{\widehat{a}}{2} \end{pmatrix} W_T \begin{pmatrix} 1 & -1 \\ \frac{\widetilde{c}_1}{2} & -\widetilde{c}_1 + \frac{\widetilde{a}}{2} \end{pmatrix}' \sqrt{T} \begin{pmatrix} \widehat{a} - a_0 \\ \widehat{c}_1 - c_{10} \end{pmatrix} = 0,$$

where $\widetilde{c}_1 \in (c_{10}, \widehat{c}_1), \widetilde{a} \in (a_0, \widehat{a})$. Let,

$$J_T = \begin{pmatrix} 1 & \frac{\widehat{c}_1}{2} \\ -1 & -\widehat{c}_1 + \frac{\widehat{a}}{2} \end{pmatrix} W_T \begin{pmatrix} 1 & -1 \\ \frac{\widetilde{c}_1}{2} & -\widetilde{c}_1 + \frac{\widetilde{a}}{2} \end{pmatrix}'$$

Then,

$$\begin{aligned} \sqrt{T} \begin{pmatrix} \widehat{a} - a \\ \widehat{c}_1 - c_1 \end{pmatrix} &= -J_T^{-1} \begin{pmatrix} 1 & \frac{\widehat{c}_1}{2} \\ -1 & -\widehat{c}_1 + \frac{\widehat{a}}{2} \end{pmatrix} W_T \begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^{T} (X_t - (c_{10} - a_0)) \\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T} ((X_t - \overline{X})^2 - \frac{c_{10}}{2}(c_{10} - a_0)) \end{pmatrix} \\ &= A_T \frac{1}{\sqrt{T}} \sum_{t=1}^{T} q_t(\theta_0) \end{aligned}$$

where

$$A_T = -J_T^{-1} \begin{pmatrix} 1 & \frac{\widehat{c}_1}{2} \\ -1 & -\widehat{c}_1 + \frac{\widehat{a}}{2} \end{pmatrix} W_T$$

and

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} q_t(\theta_0) = \begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^{T} (X_t - (c_{10} - a_0)) \\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T} ((X_t - \overline{X})^2 - \frac{c_{10}}{2}(c_{10} - a_0)) \end{pmatrix} \tag{15}$$

Hereafter, consider the case in which the process under H_A is a geometric diffusion with lognormal invariant density, so that $\log X_t$ is an Ornstein-Uhlenbeck process. That is, under H_A,

$$d \log X_t = -\theta \log X_t dt + \sigma dW_t, \quad \theta > 0,$$

so that $\log X_t \sim N(0, \frac{\sigma^2}{2\theta})$, X_t is a lognormal random variable with $E(X_t) = \exp(\sigma^2/4\theta)$, and $Var(X_t) = \exp(\sigma^2/\theta) - \exp(\sigma^2/2\theta)$.

We begin by showing that A1 and B1-B7 are satisfied. A1 is satisfied under both H_0 and H_A, as in both cases X_t has a stationary invariant density, gamma under the null, lognormal under the alternative, and is geometric ergodic if the following sign conditions hold: $c_1 - a > 0$, and $c_1 > 0$ under H_0 and $\theta > 0$ and $0 < \sigma^2 < \infty$ under H_A. B1 and B5 trivially hold, as there is no need to simulate the paths when constructing the GMM estimates, and B2 holds by the same argument used in Lemma 3.1 of Corradi (1999). B3 holds because $E(f_1) = c_1^\dagger - a^\dagger$ and $E(f_2) = \frac{c_1^\dagger}{2}(c_1^\dagger - a^\dagger)$, where $a^\dagger = a_0$ and $c_1^\dagger = c_{10}$ under H_0. In addition, B4 is satisfied as $\frac{1}{T}\sum f_t$ satisfies a strong law of large numbers and (a, c_1) is uniquely identifiable. It also follows that B6 is satisfied as $\frac{1}{\sqrt{2T\log\log T}}\sum(f_t - E(f))$ satisfies a law of the iterated logarithm. Finally, B7 is trivially satisfied, as we do not need to simulate the paths when constructing GMM estimates under the alternative. In particular, under H_A :

$$\widehat{c}_1 \overset{a.s.}{\to} \frac{\exp(\sigma^2/\theta) - \exp(\sigma^2/2\theta)}{\exp(4\sigma^2/\theta)} = c_1^\dagger$$

and

$$\widehat{a} \overset{a.s.}{\to} \exp(4\sigma^2/\theta) - \frac{\exp(\sigma^2/\theta) - \exp(\sigma^2/2\theta)}{\exp(4\sigma^2/\theta)} = a^\dagger,$$

so that $\sqrt{T}\begin{pmatrix} \widehat{a} - a^\dagger \\ \widehat{c}_1 - c_1^\dagger \end{pmatrix} = J_T^{-1}\begin{pmatrix} 1 & \frac{\widehat{c}_1}{2} \\ -1 & -\widehat{c}_1 + \frac{\widehat{a}}{2} \end{pmatrix} W_T \begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^T (X_t - (c_1^\dagger - a^\dagger)) \\ \frac{1}{\sqrt{T}}\sum_{t=1}^T ((X_t - \overline{X})^2 - \frac{c_1^\dagger}{2}(c_1^\dagger - a^\dagger)) \end{pmatrix}$.

Thus, Proposition 5 holds, so that S_T^2 ($S_{R,T}^2$) and $Z_{R,T}^2(\omega)$ ($\widetilde{Z}_T^2(\omega)$) can be constructed and used as described above.

Before completing this section, it is worth noting that although the process defined under H_0 is very restrictive, and does not correspond to those versions of the Cox, Ingersol and Ross (1985) model frequently estimated in practice, it provides us with a convenient form of the square root process with which to illustrate the application of our diffusion specification test. In addition, consider the simple case where $c_1 = 3$, and $a = -3$ (one of the cases considered in the experiments reported on in the next subsection). Corresponding to the largest sample size used in our Monte Carlo experiments, we generated 1000 observations according to this model. A histogram of these observations is given in Panel 1 of Figure 1. In addition, daily and monthly observations on the 3-month Treasury Bill rate were downloaded from the St. Louis Federal Reserve (FRED) Database. As the daily data were available starting in 02/01/1962, we constructed histograms for these two series starting at that date. These are reported in Panels 2 and 3 of Figure 1. It is interesting that although the historical data are generally 'smoother' than those generated under H_0, they have similar minima, maxima, means, medians, and standard deviations when compared with the simulated data. Thus, even the restrictive example considered here for illustrative purposes is not too far removed from reality.

6.2 Monte Carlo Evidence

In this section the results of a series of Monte Carlo experiments are reported. Corresponding to the example provided in Section 6.1, we examine the following hypotheses:

H_0 : $X(t)$ is a diffusion process solution to the following stochastic differential equation:

$$dX(t) = ((c_1 - a) - X(t))dt + \sqrt{c_1 X(t)}dW(t),\ c_1 > 0,\ \text{and } c_1 - a > 0, \tag{16}$$

H_A : The log of $X(t)$ is an Ornstein-Uhlenbeck process, so that:

$$d\log X(t) = -\theta_1 \log X(t)dt + \sigma dW(t), \quad \theta_1 > 0,$$

Under H_A, data are generated as outlined in Section 4, by simulating paths for the process

$$dY(t) = -\theta Y_t dt + \sigma dW(t), \tag{17}$$

and then forming $X_t = \exp Y_t$. Under H_0 data are generated using an Euler scheme to approximate solution paths to (16). Sample paths generated under H_0 and H_A respectively, are then used to examine the finite sample rejection frequencies of the S_T^2 ($S_{R,T}^2$) statistics, using critical values constructed via use of bootstrap $Z_{R,T}^2(\omega)$ ($\widetilde{Z}_T^2(\omega)$) statistics.[18] Under H_A we set $\sigma = 1$ and $\theta = \{0.3, 0.6, 0.9\}$, while under H_0, rejection frequencies are tabulated for $(a, c_1) = \{(-2, 2), (-3, 3), (-4, 4)\}$.[19]

In all experiments, parameters are estimated using GMM, given the moment conditions implied under H_0, the S_T^2 ($S_{R,T}^2$) test statistics are constructed using the CDF of a gamma random variable, h is set equal to T^{-1}, and samples of $T = \{500, 1000\}$ observations are used. In addition, the pseudo time series, X_t^*, are generated using the bootstrap procedure described in Section 3, with $p = 1/R^{0.25}$ or $p = 1/T^{0.25}$, depending on whether testing strategy (I) or (II) is used, respectively, and $R = \{0.8, 0.9, 0.95\}$. In all cases the integration interval U has been set equal to $[0, 15]$, statistics are formed based on uniform grids of $1,000$ points in U, and critical values are set equal to the

[18]In addition to examining the finite sample performance of S_T^2 and $S_{R,T}^2$, we also examine the properties of $|S_T| = \int_U |S_T(u)|\pi(u)du$, $|S_{R,T}| = \int_U |S_{R,T}(u)|\pi(u)du$, $\sup_{u\in U}|S_T(u)|$, $\sup_{u\in U}|S_{R,T}(u)|$.

[19]The three parameterizations considered under H_0 can be expressed in terms of $\alpha_1$and α_2 as: $(\alpha_1,\alpha_2) = \{(4, 1), (4, 3/2), (4, 2)\}$. The shapes of the densities with these parameterizations are given in Johnson, Kotz and Balakrishnan (1994 pp. 341). Results qualitatively similar to those reported in Table 1 were also found for $(\alpha_1,\alpha_2) = \{(10/3, 3/2), (14/3, 3/2)\}$, although they are not reported here.

90^{th} percentile of the bootstrap distribution. Finally, we tried $B = \{100, 200, 500\}$. As results are qualitatively the same for all three values of B, we report only the findings for $B = 100$.

Results based on data generated according to H_0 are collected in Table 1, while those based on data generated according the H_A are in Table 2. In Part 1 of Table 1, we report findings for the case in which test statistics are computed using T observations, and bootstrap statistics contain the adjustment for parameter estimation error and are computed using R observations. Recalling that the "nominal" rejection rate is 10%, note that for samples of 500 observations the empirical rejection rate is between 12% and 16%, while for samples of 1000 observations it is between 11% and 14%. These results are extremely robust to the choice of parametrization. Note also that the faster the rate of growth of R, the closer is the empirical rejection rate to the nominal one. This is not surprising at all. In fact, in this example recall that

$$\frac{1}{\sqrt{R}}\sum_{t=1}^{R} q_t(\widehat{\theta}_T) = \begin{pmatrix} \frac{1}{\sqrt{R}}\sum_{t=1}^{R}(X_t - (\widehat{c}_1 - \widehat{a})) \\ \frac{1}{\sqrt{R}}\sum_{t=1}^{R}((X_t - \overline{X})^2 - \frac{\widehat{c}_1}{2}(\widehat{c}_1 - \widehat{a})) \end{pmatrix}$$

and

$$\frac{1}{\sqrt{R}}\sum_{t=1}^{R} q_t^*(\widehat{\theta}_T) = \begin{pmatrix} \frac{1}{\sqrt{R}}\sum_{t=1}^{R}(X_t^* - (\widehat{c}_1 - \widehat{a})) \\ \frac{1}{\sqrt{R}}\sum_{t=1}^{R}((X_t^* - \overline{X})^2 - \frac{\widehat{c}_1}{2}(\widehat{c}_1 - \widehat{a})) \end{pmatrix}.$$

Thus, $\frac{1}{\sqrt{R}}\sum_{t=1}^{R}(q_t^*(\widehat{\theta}_T) - q_t(\widehat{\theta}_T))$ does not depend on the estimated parameters, and so ideally we could have taken $R = T$ in this case. On the other hand, it is interesting that even for $R = T^{0.8}$, rejection rates are quite close to 10%. Out of the three statistics considered, the supremum statistic exhibits the best finite sample performance. In Part 2 of Table 1, we report findings for the case in which test statistics are computed using R observations, and bootstrap statistics do not contain the adjustment for parameter estimation error and are computed using T observations. In this case, empirical rejection rates are usually smaller than 10%, with some (although little) improvement when the sample size is increased from 500 to 1000 observations. It is worth noting that in this case, it is preferrable to let R grow at a relatively slow rate. In fact, for $R = T^{0.8}$ and 1000 observations, rejection rates are much higher than in the other cases, with values between 16% and 17%. This pattern is intuitively sensible. We compute critical values which neglect the effect of parameter estimation error. However, the contribution of these omitted terms in the actual statistic is of probability order $\sqrt{R/T}$, so that the slower is the rate of growth of R, relative to T, the faster the effect of parameter estimation error vanishes.

Table 2 reports rejection frequencies under the H_A.[20] Results here generally agree with the assessment made based on examining the results in Table 1 that finite sample test performance is improved when critical values which take parameter estimation error into account are used. For example, note that in Panel 1, rejection rates are always above 86%, even with samples of only 500 observations, while rates are as low as 40% when parameter estimation error is not accounted for. Additionally, rejection rates are not robust to model parameterization when parameter estimation error is not accounted for. Again, this is not surprising, as in Panel 1 the statistic diverges at rate T or $\sqrt{T}$ (depending on the which version of the statistic is considered), while in the other case the statistic diverges at rate R or $\sqrt{R}$. Also not surprisingly, the supremum statistics displays the best performance.

7 Concluding Remarks

In this paper we have outlined a bootstrap procedure which accounts for parameter estimation error and data dependence. This facillitates the construction of parametric specification tests of diffusion processes, for example. In addition, we show that application of the bootstrap leads to straightforward generalizations of the conditional Kolmogorov test of Andrews (1997) and the conditional mean test of Whang (2000) to the case of dependent observations.

In an illustration, the diffusion specification test proposed here is applied to the problem of selecting between two alternative continuous time diffusion models, and it is seen, via a series of Monte Carlo experiments, that accounting for parameter estimation error is important, and that the test performs very well, even with samples of only 500 observations.

[20] As above, Part 1 reports findings for S_T^2, $|S_T|$ and $\sup_u |S_T(u)|$ statistics, with critical values computed using $Z_{RT}^2(\omega)$, $|Z_{R,T}(\omega)|$ and $\sup_u |Z_{R,T}(\omega, u)|$, while Part 2 reports findings for $S_{R,T}^2$, $|S_{R,T}|$ and $\sup_u |S_{R,T}(u)|$ statistics, with critical values computed using $\widetilde{Z}_T^2(\omega)$, $|\widetilde{Z}_T(\omega)|$ and $\sup_u |\widetilde{Z}_T(\omega, u)|$.

8 Appendix

Proof of Theorem 1: **(i)** We first show convergence in distribution for any given $u \in U$, then we show convergence of the finite dimensional distributions and finally stochastic equicontinuity over U, this will ensure that $S_T(.)$ weakly converges to Z, and the desired result then follows from the continuous mapping theorem. Given A1, the skeleton $X_1, X_2, \ldots, X_T$ is a strictly stationary strong mixing sequence with mixing coefficients decaying at a geometric rate. Given A2, we can write

$$\begin{aligned} S_T(u) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big((1\{X_t \leq u\} - F(u,\theta_0)) - (F(u,\widehat{\theta}_T) - F(u,\theta_0))\Big) \\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(1\{X_t \leq u\} - F(u,\theta_0)) - \nabla_\theta F(u,\overline{\theta})'\sqrt{T}(\widehat{\theta}_T - \theta_0) \\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(1\{X_t \leq u\} - F(u,\theta_0)) - \nabla_\theta F(u,\overline{\theta})'A_T(\overline{\theta}_T)\frac{1}{\sqrt{T}}\sum_{t=1}^{T} q(X_t,\theta_0) \\ &= I_{1T}(u) + I_{2T}(u), \end{aligned}$$

where $\overline{\theta}_T \in (\widehat{\theta}_T, \theta_0)$. Recalling A1, A3 and A4, by the central limit theorem for strong mixing sequences,

$$\begin{pmatrix} I_{1T}(u) \\ I_{2T}(u) \end{pmatrix} \xrightarrow{d} N(0, \begin{pmatrix} V_1(u) & C(u) \\ C(u) & V_2(u) \end{pmatrix},$$

where

$$V_1(u) = E(\sum_{s=-\infty}^{\infty}(1\{X_0 \leq u\} - F(u,\theta_0))(1\{X_s \leq u\} - F(u,\theta_0))),$$

$$V_2(u) = E(\sum_{s=-\infty}^{\infty}(\nabla_\theta F(u,\theta_0)'A(\theta_0)q(X_0,\theta_0))(q(X_s,\theta_0)'A(\theta_0)\nabla_\theta F(u,\theta_0))),$$

$$C(u) = -E(\sum_{s=-\infty}^{\infty}(1\{X_0 \leq u\} - F(u,\theta_0))(\nabla_\theta F(u',\theta_0)'A(\theta_0)q(X_s,\theta_0))).$$

Thus, $S_T(u) \xrightarrow{d} N(0, K(u,u))$, where $K(u,u) = V_1(u) + V_2(u) + 2C(u)$. A straightforward application of the Cramer Wold device ensures that

$$\begin{pmatrix} S_T(u) \\ S_T(u') \end{pmatrix} \xrightarrow{d} N(0, \begin{pmatrix} K(u,u) & K(u,u') \\ K(u,u') & K(u',u') \end{pmatrix},$$

where $K(u,u')$ is as defined in (3). As U is compact in $\Re$, (and so totally bounded) in order to show weak convergence, we need to show that $S_T(u)$ is stochastically equicontinuous on U,(see e.g.

Pollard 1990, section 10), that is

$$\lim_{T\to\infty}\sup \Pr\left(\sup_{u':|u-u'|<\delta}|S_T(u)-S_T(u')|>\varepsilon\right)=0 \tag{18}$$

Now (18) will follow if we can show that

$$\lim_{T\to\infty}\sup \Pr\left(\sup_{u':|u-u'|<\delta}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left((1\{X_t\le u\}-F(u,\theta_0))-(1\{X_t\le u'\}-F(u',\theta_0))\right)\right|>\varepsilon/2\right)=0 \tag{19}$$

$$\lim_{T\to\infty}\sup \Pr\left(\sup_{u':|u-u'|<\delta}\left|(\nabla_\theta F(u,\overline{\theta}_T)'-\nabla_\theta F(u',\overline{\theta}_T)')\sqrt{T}(\widehat{\theta}_T-\theta_0)\right|>\varepsilon/2\right)=0 \tag{20}$$

with $\overline{\theta}_T\in(\widehat{\theta}_T,\theta_0)$. We begin by considering (20); by a mean value expansion, we have,

$$\begin{aligned}&\sup_{u':|u-u'|<\delta}\left|\sum_{j=1}^{k}(\nabla_{u,\theta}F(\overline{u},\overline{\theta})_j\sqrt{T}(\widehat{\theta}_{T,j}-\theta_{0,j})(u-u')\right|\\ \le\;& k\max_{j=1,\dots k}\sup_{u\times\theta\in U\times\Theta}|\nabla_{u,\theta}F(u,\theta)_j||\sqrt{T}(\widehat{\theta}_{T,j}-\theta_{0,j})|\sup_{u:|u-u'|<\delta}|u-u'|\end{aligned}$$

Now

$$k\lim_{T\to\infty}\sup \Pr\left(\left|\sqrt{T}(\widehat{\theta}_{T,j}-\theta_{0,j})\right|>\frac{\varepsilon/2}{\delta\sup_{u\times\theta\in U\times\Theta}|\nabla_{u,\theta}F(u,\theta)_j|}\right)=0,$$

given that $\sqrt{T}(\widehat{\theta}_T-\theta_0)=O_p(1)$, and A2 ensures that $\nabla_{\theta,u}F(u,\theta)'$ is jointly continuous on $\Theta\times U$, and so $\sup_{u\times\theta\in U\times\Theta}|\nabla_{u,\theta}F(u,\theta)_j|\le C$. It remains to show (19).

Let $m_t(u)=1\{X_t\le u\}-F(u,\theta_0)$ and note that,

$$\sup_{t\le T,T\ge 1}\sup_{u\in U}|m_t(u)|=1 \tag{21}$$

Without loss of generality set $u<u'$,

$$\begin{aligned}&\sup_{t\le T,T\ge1}\left(E\left(\sup_{u':|u-u'|<\delta}(|m_t(u)-m_t(u')|^p)\right)\right)^{1/p},\ p\ge2\\ &\le\sup_{u':|u-u'|<\delta}|F(u,\theta_0)-F(u',\theta_0)|+\left(E\sup_{u':|u-u'|<\delta}(1\{u\le X_t\le u'\})^p\right)^{1/p}\\ &\le\sup_{u\times U}|\nabla_u F(u,\theta_0)|\delta+\left(\sup_{u\in U}\int_u^{u+\delta}f(x)dx\right)^{1/p}\le 2C\delta,\end{aligned} \tag{22}$$

given A2. Stochastic equicontinuity then follows by Philipp (1982) (see (i)-(iii) in example 2(a) in Andrews (1993)). In fact (i) is satisfied given the geometric ergodicity of the skeleton, (ii) is ensured by (21) and as shown by Andrews (1993, p.201), (iii) is implied by (22). Given stochastic equicontinuity and convergence of the finite dimensional distributions, it follows that $S_T(\cdot) \Rightarrow Z(\cdot)$ where Z is the Gaussian process with the covariance kernel defined in (3). The statement then follows from the continuous mapping theorem.

(ii) Given A1 and A3, X_t is a strictly stationary, strong mixing process with coefficients decaying at a geometric rate. Suppose that X_t has a CDF, $G(u, \beta_0)$, with $\mu\{u : |G(u, \beta_0) - F(u, \theta^\dagger)| > 0\}$, where μ is the Lebesgue measure on U. This is ensured by the fact that the initial conditions, the drift, and the variance terms uniquely determine the invariant density and so the associated CDF. Thus,

$$S_T(u) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Big((1\{X_t \leq u\} - G(u, \beta_0)) - \sqrt{T}(F(u, \widehat{\theta}_T) - G(u, \beta_0)) \Big) .$$

The first term on the RHS of the above expression satisfies a central limit theorem and so is $O_p(1)$. With regard to the second term, note that a mean value expansion yields that,

$$\sqrt{T}(F(u, \widehat{\theta}_T) - G(u, \beta_0)) = \sqrt{T}(F(u, \theta^\dagger) - G(u, \beta_0)) - \nabla_\theta F(u, \overline{\theta}_T)' A_T(\overline{\theta}_T) \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (q(X_t, \theta^\dagger)$$

with $\overline{\theta}_T \in (\theta^\dagger, \widehat{\theta}_T)$. The first term on the right hand side diverges at rate $\sqrt{T}$ for all u in a subset of positive Lebesgue measure. The second term converges in distribution.

The statement in (ii) then follows.

Proof of Corollary 2: (i)

$$\begin{aligned} S_{R,T}(u) &= \frac{1}{\sqrt{R}} \sum_{t=1}^{R} \Big((1\{X_t \leq u\} - F(u, \theta_0)) - (F(u, \widehat{\theta}_T) - F(u, \theta_0)) \Big) \\ &= \frac{1}{\sqrt{R}} \sum_{t=1}^{R} (1\{X_t \leq u\} - F(u, \theta_0)) - \nabla_\theta F(u, \overline{\theta})' \sqrt{R}(\widehat{\theta}_T - \theta_0) \\ &= \frac{1}{\sqrt{R}} \sum_{t=1}^{R} (1\{X_t \leq u\} - F(u, \theta_0)) + o_p(1), \end{aligned}$$

as $\sup_{\theta \times u \in \Theta \times U} |\nabla_{\theta,u} F(u, \theta)'_j| = O(1)$[21] for $j = 1, \ldots, k$, and $\sqrt{R}(\widehat{\theta}_T - \theta_0) = O_p(\frac{\sqrt{R}}{\sqrt{T}}) = o_p(1)$.

(ii) Follows by the same argument used in part (ii) of the proof of Theorem 1.

[21] $\nabla_{\theta,u} F(u, \theta)'_j$ denotes the $j^{th}-$element of $\nabla_{\theta,u} F(u, \theta)'$.

Proof of Theorem 3: Hereafter, let E_*, Var_* and Cov_* denote the expectation, variance and covariance, with respect to the probability measure governining the pseudo time series X_t^*, P_*, conditional on the sample. The proof of Theorem 3 relies on the following Lemma. For notational simplicity we confine our attention to the case where $q_t(\widehat{\theta}_T) = q(X_t, \widehat{\theta}_T)$, but the same argument applies to the case of $q_t(\widehat{\theta}_T) = q(X_t, \ldots, X_{t-m}, \widehat{\theta}_T)$, provided that we construct the pseudo time series z_t^* in the way described in Section 3, above equation 9.

Lemma A1: $\forall u \in U$,

(ia) $E_*(\frac{1}{R}\sum_{t=1}^{R} 1\{X_t^* \leq u\}|X_1, \ldots X_R) = \frac{1}{R}\sum_{t=1}^{R} 1\{X_t \leq u\}$,

(ib) $E_*(\frac{1}{R}\sum_{t=1}^{R} q_t(X_t^*, \widehat{\theta}_T)|X_1, \ldots X_R) = \frac{1}{R}\sum_{t=1}^{R} q_t(X_t, \widehat{\theta}_T)$,

(iia) $Var_*(\frac{1}{\sqrt{R}}\sum_{t=1}^{R} 1\{X_t^* \leq u\}|X_1, \ldots X_R) = C1_0(u) + 2\frac{1}{R}\sum_{j=1}^{R}(1-\frac{j}{R})(1-p)^j C1_j(u)$,

(iib) $Var_*(\frac{1}{\sqrt{R}}\sum_{t=1}^{R} q_t(X_t^*, \widehat{\theta}_T)|X_1, \ldots X_R) = C2_0(u) + 2\frac{1}{R}\sum_{j=1}^{R}(1-\frac{j}{R})(1-p)^j C2_j(u)$, and

(iic) $Cov_*(\frac{1}{\sqrt{R}}\sum_{t=1}^{R} 1\{X_t^* \leq u\}, \frac{1}{\sqrt{R}}\sum_{t=1}^{R} q_t(X_t^*, \widehat{\theta}_T)|X_1, \ldots X_R) = C3_0(u) + 2\frac{1}{R}\sum_{j=1}^{R}(1-\frac{j}{R})(1-p)^j C3_j(u)$, where

$$\begin{aligned}
C1_i(u) &= \frac{1}{R}\sum_{j=1}^{R-i}(1\{X_j \leq u\}1\{X_{j+i} \leq u\}) - \left(\frac{1}{R}\sum_{t=1}^{R} 1\{X_t \leq u\}\right)^2 \\
C2_i(u) &= \frac{1}{R}\sum_{j=1}^{R-i}\left(q(X_j, \widehat{\theta}_T)q(X_{j+i}, \widehat{\theta}_T)\right) - \left(\frac{1}{R}\sum_{t=1}^{R} q(X_t, \widehat{\theta}_T)\right)^2 \\
C3_i(u) &= \frac{1}{R}\sum_{j=1}^{R-i}\left(q(X_j, \widehat{\theta}_T)1\{X_{j+i} \leq u\}\right) - \left(\frac{1}{R}\sum_{t=1}^{R} 1\{X_t \leq u\}\right)\left(\frac{1}{R}\sum_{t=1}^{R} q(X_t, \widehat{\theta}_T)\right)
\end{aligned}$$

Proof of Lemma A1: (ia) Recall that I_i is the initial value of block i (and I_i is a discrete uniform random variable on $1, 2, \ldots, R, \forall i$). Also, L_i denote the length of block i (and has a geometric distribution with parameter $p = p_R, \forall i$). Thus,

$$\begin{aligned}
E_*(1\{X_1^* &\leq u\}|X_{1,}\ldots X_R) = 1\{X_1 \leq u\}P_*(I_1 = 1) + 1\{X_2 \leq u\}P_*(I_1 = 2) \\
+ \ldots 1\{X_R &\leq u\}P_*(I_1 = R) = \frac{1}{R}\sum_{t=1}^{R} 1\{X_t \leq u\}.
\end{aligned}$$

It follows that given the stationarity of the pseudo time series, X_t^*, $E_*(\frac{1}{R}\sum_{t=1}^{R} 1\{X_t^* \leq u\}|X_1, \ldots, X_R) = E_*(1\{X_1^* \leq u\}|X_{1,}\ldots, X_R) = \frac{1}{R}\sum_{t=1}^{R} 1\{X_t \leq u\}$. This proves (ia).

(ib): Follows by the same argument used in (ia).

(iia) Note that

$$
\begin{aligned}
E_*(1\{X_1^* &\leq u\}1\{X_{1+i}^* \leq u\}|X_1,\ldots,X_R) \\
&= E_*(1\{X_1^* \leq u\}1\{X_{1+i}^* \leq u\}|X_1,\ldots,X_R,L_1>i)P_*(L_1>i) \\
+E_*(1\{X_1^* &\leq u\}1\{X_{1+i}^* \leq u\}|X_1,\ldots,X_R,L_1\leq i)P_*(L_1\leq i)
\end{aligned}
$$

By further noting that $P_*(L_1 > i) = \sum_{m=i+1}^{\infty}(1-p)^{m-1}p = \frac{p(1-p)^i}{1-(1-p)} = (1-p)^i$, the right hand side of the above expression can be written as

$$
\begin{aligned}
(1\{X_1 &\leq u\}1\{X_{1+i} \leq u\})P_*(I_1 = 1)(1-p)^i + \\
\ldots + (1\{X_{R-i} &\leq u\}1\{X_R \leq u\})P_*(I_1 = R-i)(1-p)^i \\
+((1\{X_1 &\leq u\})^2)P_*(I_1 = 1)P_*(I_{1+i} = 1)(1-(1-p)^i) + \ldots \\
+\ldots + (1\{X_1 &\leq u\}1\{X_R \leq u)P_*(I_1 = 1)P_*(I_{1+i} = R)(1-(1-p)^i) \\
+\ldots + ((1\{X_R &\leq u\})^2 P_*(I_1 = R)P_*(I_{1+i} = R)(1-(1-p)^i) \\
&= \frac{1}{R}\sum_{j=1}^{R-i} 1\{X_j \leq u\}1\{X_{j+i} \leq u\}(1-p)^i - \left(\frac{1}{R}\sum_{t=1}^{R} 1\{X_t \leq u\}\right)^2 (1-(1-p)^i).
\end{aligned}
$$

Let

$$
C1_i(u) = \frac{1}{R}\sum_{j=1}^{R-i}(1\{X_j \leq u\}1\{X_{j+i} \leq u\}) - \left(\frac{1}{R}\sum_{t=1}^{R} 1\{X_t \leq u\}\right)^2.
$$

Then,

$$
Var_*(\frac{1}{\sqrt{R}}\sum_{t=1}^{R} 1\{X_t^* \leq u\}|X_1,\ldots X_R) = C1_0(u) + 2\frac{1}{R}\sum_{j=1}^{R}(1-\frac{j}{R})(1-p)^j C1_j(u).
$$

(iib),(iic) Follow by the same argument used in (iia).

Proof of Theorem 3-(Cont.): Hereafter, $\xrightarrow{d}_*$denotes convergence in distribution with respct to the probability law governing the psuedo time series, conditional on the sample. Given A1 and A4, by Theorem 2 in PR[22] we have that $\forall u$,

$$
\left(\begin{array}{c}
\frac{1}{\sqrt{R}}\sum_{t=1}^{R}\left(1\{X_t^*(\omega) \leq u\} - 1\{X_t \leq u\}\right) \\
-\left(\nabla_\theta F(u,\widehat{\theta}_T)\right)' \frac{1}{\sqrt{R}}\sum_{t=1}^{R}\left(A_T(\widehat{\theta}_T)(q(X_t^*(\omega),\widehat{\theta}_T) - q(X_t,\widehat{\theta}_T))\right)
\end{array}\right)
$$
$$
\xrightarrow{d}_* N\left(0, \left(\begin{array}{cc} V_{11}(\omega) & V_{12}(\omega) \\ V_{12}(\omega) & V_{22}(\omega)\end{array}\right)\right),
$$

[22]Note that geometric ergodicity and A4 imply the satisfaction of condition (8) in PR.

where from Lemma A1,

$$
\begin{aligned}
V_{11}(\omega) &= \lim_{R\to\infty} Var_*(\frac{1}{\sqrt{R}}\sum_{t=1}^{R}(1\{X_t^*\leq u\}|X_1,\dots X_R)) \\
&= \lim_{R\to\infty}(C1_0(u)+2\frac{1}{R}\sum_{j=1}^{R}(1-\frac{j}{R})(1-p)^j C1_j(u)),
\end{aligned}
$$

$$
\begin{aligned}
V_{22}(\omega) &= \lim_{R\to\infty} Var_*\left(\frac{1}{\sqrt{R}}\sum_{t=1}^{R} q_t(X_t^*,\widehat{\theta}_T)|X_1,\dots X_R\right) \\
&= \lim_{R\to\infty}(C2_0(u)+2\frac{1}{R}\sum_{j=1}^{R}(1-\frac{j}{R})(1-p)^j C2_j(u)),
\end{aligned}
$$

$$
\begin{aligned}
V_{12}(\omega) &= \lim_{R\to\infty} Cov_*(\frac{1}{\sqrt{R}}\sum_{t=1}^{R}1\{X_t^*\leq u\},\frac{1}{\sqrt{R}}\sum_{t=1}^{R} q_t(X_t^*,\widehat{\theta}_T)|X_1,\dots X_R) \\
&= \lim_{R\to\infty}(C3_0(u)+2\frac{1}{R}\sum_{j=1}^{R}(1-\frac{j}{R})(1-p)^j C3_j(u)),
\end{aligned}
$$

where $C1_i(u), C2_i(u)$ and $C3_i(u)$ are defined in Lemma A1. Let

$$
- ^*_R = \left\{ \begin{array}{c}
\omega : |Var_*\left(\frac{1}{\sqrt{R}}\sum_{t=1}^{R}\left(1\{X_t^*\leq u\}\right)|X_1,\dots X_R\right) - Var\left(\frac{1}{\sqrt{R}}\sum_{t=1}^{R}1\{X_t\leq u\}\right)| < \epsilon \\
|Var_*\left(\frac{1}{\sqrt{R}}\sum_{t=1}^{R} q_t(X_t^*,\widehat{\theta}_T)|X_1,\dots X_R\right) - Var\left(\frac{1}{\sqrt{R}}\sum_{t=1}^{R} q_t(X_t,\widehat{\theta}_T)\right)| < \epsilon \\
|Cov_*(\frac{1}{\sqrt{R}}\sum_{t=1}^{R}1\{X_t^*\leq u\},\frac{1}{\sqrt{R}}\sum_{t=1}^{R} q_t(X_t^*,\widehat{\theta}_T)|X_1,\dots X_R) \\
-Cov(\frac{1}{\sqrt{R}}\sum_{t=1}^{R}1\{X_t\leq u\},\frac{1}{\sqrt{R}}\sum_{t=1}^{R} q_t(X_t,\widehat{\theta}_T))| < \epsilon
\end{array}\right\}.
$$

Given A3, $\sqrt{R}(\widehat{\theta}_T-\theta^*) = o_{a.s.}(1)$, and from Theorem 1 in PR we have that $\lim_{R\to\infty} P(- ^*_R) = 1$. Furthermore, given A3, $A_T^*(\theta)$ satisfies a uniform strong law of large numbers and $\sup_{\theta\times u\in\Theta\times U}|\nabla^2_\theta F(u,\theta)'_j| = O(1)$, $j = 1,\dots,k$. This completes the proof of the statement for any given $u \in U$.

By a straightforward application of the Cramer-Wold device, it also follows that $(Z_{R,T}(u), Z_{R,T}(u'))$ has the same limiting distribution under both H_0 and H_A. Also, this distribution corresponds with that of $(S_T(u), S_T(u'))$, under the null. We now need to show the convergence, conditional on the sample, of $Z_{R,T}(.)$ as a process. As we have already shown the convergence of the finite dimensional distributions, we need to show that $Z_{T,R}(u,\omega)$ is stochastically equicontinuous on $u \in U$, $P-\omega$, that is conditional on the sample and for all sample but a set with probablity measure converging to zero. Recall that P_* and P denote the probability law of the pseudo times series conditional on the sample and the probability law of the sample, respectively. In order to show that

$$\limsup_{R\to\infty} P_* \left(\sup_{u':|u-u'|<\delta} |Z_{R,T}(u,\omega) - Z_{R,T}(u',\omega)| > \varepsilon \right) = 0$$

$\forall \omega \in -_{AR}$, with $P(-_{AR}) \to 1$, it suffices to show that

$$\sup_{u':|u-u'|<\delta} \left| \frac{1}{\sqrt{R}} \sum_{t=1}^{R} (1\{X_t \leq u\} - 1\{X_t \leq u'\}) \right| \leq \varepsilon/3 \tag{23}$$

$\forall \omega \in -_{BR}$, with $P(-_{BR}) \to 1$,

$$\limsup_{R\to\infty} P_* \left(\sup_{u':|u-u'|<\delta} \left| \frac{1}{\sqrt{R}} \sum_{t=1}^{R} (1\{X_t^*(\omega) \leq u\} - 1\{X_t^*(\omega) \leq u'\}) \right| > \varepsilon/3 \right) = 0 \tag{24}$$

$\forall \omega \in -_{CR}$, with $P(-_{CR}) \to 1$, and

$$\limsup_{R\to\infty} P_* \left(\sup_{u':|u-u'|<\delta} \left| (\nabla_\theta F(u,\widehat{\theta}_T)' - \nabla_\theta F(u,\widehat{\theta}_T)') A_T(\widehat{\theta}_T) \frac{1}{\sqrt{R}} \sum_{t=1}^{R} (q_t(X_t^*(\omega),\widehat{\theta}_T) - q_t(X_t,\widehat{\theta}_T)) \right| > \varepsilon/3 \right) \tag{25}$$

$\forall \omega \in -_{DR}$, with $P(-_{DR}) \to 1$, and set $-_{AR} = -_{BR} \cap -_{CR} \cap -_{DR}$.

We have shown in the proof of Theorem 1 that $\frac{1}{\sqrt{R}} \sum_{t=1}^{R} (1\{X_t \leq u\}$ is stochastic equicontinuous on U, thus (23) follows immediately. Now $\sup_{u\in U} 1\{X_t^*(\omega) \leq u\} = 1$, without loss of generality set $u < u'$,

$$\begin{aligned}
&(\sup_{t\leq T, T>1} E_*(\sup_{u':|u-u'|<\delta} \left| 1\{X_t^*(\omega) \leq u\} - 1\{X_t^*(\omega) \leq u'\} \right|^p / X_1 \ldots X_R))^{1/p} \\
= &\left(\sup_{u\in U} \left| \frac{1}{R} \sum_{t=1}^{R} (u \leq X_t \leq u+\delta) \right| \right)^{1/p}
\end{aligned}$$

and the term on the right hand side as $R \to \infty$ is less than or equal to $\varepsilon/3$ $\forall \omega \in -_{CR}$, with $P(-_{CR}) \to 1$, because of the law of large numbers. This shows (24). Finally as for (25), using a mean value expansion note that,

$$\begin{aligned}
&\limsup_{R\to\infty} P_* \left(\sup_{u\in U} \left| \nabla_{u,\theta} F(u,\widehat{\theta}_T)' A_T(\widehat{\theta}_T) \frac{1}{\sqrt{R}} \sum_{t=1}^{R} (q_t(X_t^*(\omega),\widehat{\theta}_T) - q_t(X_t,\widehat{\theta}_T))\delta \right| > \varepsilon/3 \right) \\
\leq\ &\limsup_{R\to\infty} P_* \left(\sup_{u\times\theta\in U\times\Theta} \left| \nabla_{u,\theta} F(u,\theta)' A_T(\theta) \frac{1}{\sqrt{R}} \sum_{t=1}^{R} (q_t(X_t^*(\omega),\widehat{\theta}_T) - q_t(X_t,\widehat{\theta}_T)) \right| > \frac{\varepsilon}{3\delta} \right) = 0
\end{aligned}$$

$\forall \omega \in -_{DR}$, with $P(-_{DR}) \to 1$, given A2, A3 and recalling that $\frac{1}{\sqrt{R}} \sum_{t=1}^{R} (q_t(X_t^*(\omega),\widehat{\theta}_T) - q_t(X_t,\widehat{\theta}_T))$ converges in distribution conditional on the sample and for all sample but a subset with probability measure converging to zero. This shows (25), the desired result then follows from the continuous mapping theorem.

Proof of Corollary 4: It follows immediately, given that $\sqrt{R}(\widehat{\theta}_T - \theta^\dagger) = O_p(\sqrt{R}/\sqrt{T}) = o_p(1)$.

Proof of Proposition 5:

$$\sup_{\theta\in\Theta} |(G_T(\theta) - G_\infty(\theta)) - (E(g_1) - E(g_1^\theta))|$$

$$\leq \left| \frac{1}{T}\sum_{t=1}^{T}(g_t - E(g_1)) \right| + \sup_{\theta\in\Theta} \left| \frac{1}{T}\sum_{t=1}^{T}(g_t^\theta - E(g_1^\theta)) \right|.$$

A1 and B1-B3 ensure that both the first and the second terms on the RHS of the above expression are $o_{a.s.}(1)$. Given B4, by the same argument used in the proof of Theorem 1 in Duffie and Singleton (1993), it follows that $\widehat{\theta}_T - \theta^\dagger = o_{a.s.}(1)$. By a mean value expansion around $\theta^\dagger$,

$$\sqrt{T}(\widehat{\theta}_T - \theta^\dagger) = J_T^{-1}\nabla_\theta G_T(\widehat{\theta}_T) W_T \sqrt{T} G_T(\theta^\dagger)$$

and $J_T = \nabla_\theta G_T(\widehat{\theta}_T)' W_T \nabla_\theta G_T(\overline{\theta}_T)$, with $\overline{\theta}_T \in (\widehat{\theta}_T, \theta^\dagger)$. Under the null, $\theta^\dagger = \theta_0$, and

$$\sqrt{T}G_T(\theta^\dagger) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(g_t - E(g_1)) - \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(g_t^{\theta^\dagger} - E(g_1^{\theta^\dagger})) + \sqrt{T}(E(g_1) - E(g_1^{\theta^\dagger})).$$

By Pardoux and Talay (1985), the last term on the RHS is $o(1)$ provided $h\sqrt{T} \to 0$ as $T \to \infty$. It then follows that $\sqrt{T}G_T(\theta_0) \xrightarrow{d} N(0, 2\Sigma_0)$. As J_T satisfies a uniform strong law of large numbers, given B1-B3, A3 holds under both hypotheses.

Proof of Theorem 6: Given the differentiability of F with respect to θ in the interior of Θ,

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

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

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

Given Assumption C1(i),(ii),(iii)(a-b) 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 the Cramer Wold device. We now need to show stochastic equicontinuity, the desired result then follows from the continuous mapping theorem. In order to show that

$$\lim\sup_{T\to\infty} \Pr\left(\sup_{\substack{u':|u-u'|<\delta \\ v':|v-v'|<\delta}} |KS_T(u,v) - KS_T(u',v')| > \varepsilon\right) = 0$$

it suffices to show that

$$\lim\sup_{T\to\infty} \Pr\left(\sup_{\substack{u':|u-u'|<\delta \\ v':|v-v'|<\delta}} \left|\frac{1}{T^{1/2}}\sum_{t=1}^{T}\left(m_t(u,v) - m_t(u',v')\right)\right| > \varepsilon/2\right) = 0 \tag{26}$$

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

$$\begin{aligned}
&\lim\sup_{T\to\infty} \Pr\left(\sup_{\substack{u':|u-u'|<\delta \\ v':|v-v'|<\delta}} \left|\frac{1}{T}\sum_{t=1}^{T}\left(\nabla_\theta F(u|X_t,\overline{\theta})'1\{X_t \le v\} - \nabla_\theta F(u'|X_t,\overline{\theta})'1\{X_t \le v'\}\right)\sqrt{T}(\widehat{\theta}_T - \theta_0)\right| > \frac{\varepsilon}{2}\right) \\
&= \quad 0
\end{aligned} \tag{27}$$

We begin by showing (26). First,

$$\sup_{\substack{u':|u-u'|<\delta \\ v':|v-v'|<\delta}} |m_t(u,v)| = 1 \tag{28}$$

Now

$$\sup_{t\le T, T>1}\left(E\left(\sup_{\substack{u':|u-u'|<\delta \\ v':|v-v'|<\delta}} |m_t(u,v) - m_t(u',v')|^p\right)\right)^{1/p} \tag{29}$$

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

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

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

$$\leq (\sup_{u\in U}\left|\int_u^{u+\delta}f_y(s)ds\right|)^{1/p}+\left|E\left(\sup_{u\in U}\nabla_u F(u,\theta_0)\right)\right|^{1/p}\delta+(\sup_{u\in U}\left|\int_u^{u+\delta}f_x(s)ds\right|)^{1/p}\leq C\delta \quad (30)$$

where f_x, f_y denote the marginal densities of X and y respectively. (26) then follows by Philipp (1982) (see example 2(a) in Andrews (1993)). In fact (i) is satisfied given the C(i), (ii) is ensured by (28) and as shown by Andrews (1993, p.201), (iii) is implied by the fact that (29) is majorized by (30). As for (27),

$$\frac{1}{T}\sum_{t=1}^{T}\left(\nabla_\theta F(u|X_t,\overline{\theta})'1\{X_t\leq v\}-\nabla_\theta F(u'|X_t,\overline{\theta})'1\{X_t\leq v'\}\right)$$

$$=\frac{1}{T}\sum_{t=1}^{T}(\nabla_\theta F(u|X_t,\overline{\theta})'-\nabla_\theta F(u'|X_t,\overline{\theta})')1\{X_t\leq v\} \quad (31)$$

$$+\frac{1}{T}\sum_{t=1}^{T}\left(\nabla_\theta F(u'|X_t,\overline{\theta})\right)'(1\{X_t\leq v\}-1\{X_t\leq v'\}) \quad (32)$$

it suffices to show that $\sup_{u'|u-u'|<\delta}$ of (31) and $\sup_{v'|v-v'|<\delta}$ of (32) also converge to zero in probability. The summands in (31) 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 C(ii). This ensures that $\sup_{u'|u-u'|<\delta}$ of (31) converges to zero in probability. Finally the summands in (32) satisfy conditions TSE and DM in Andrews (1992), in fact

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

as $\frac{1}{T}\sum_{t=1}^{T}(\nabla_\theta F(u|X_t,\overline{\theta})'=O_p(1)$ uniformly in $U\times\Theta$, thus satisfying TSE, furthermore DM is trivially satisfied because of C(ii). This ensures that $\sup_{v'|v-v'|<\delta}$ of (32) also converges to zero in probability.

Proof of Theorem 7: Hereafter, by conditioning on the data, we mean conditioning on $y_1, \ldots, y_R$, $X_1, \ldots, X_R$. By the same argument used in Theorem 3,

$$E_*\left(\frac{1}{T}\sum_{t=1}^{T} z_t^*(u,v)|\text{ data}\right) = \frac{1}{T}\sum_{t=1}^{T} z_t(u,v),$$

where the right hand side term above satisfies a uniform strong law of large numbers given C1(i)-(iii). Also,

$$Var_*(\frac{1}{R^{1/2}}\sum_{t=1}^{R} z_t^*(u,v)|\text{ data}) = C_0(u,v) + \frac{2}{R}\sum_{j=1}^{R-1}(1-\frac{j}{R})(1-p)^j C_j(u,v),$$

where

$$C_j(u,v) = \frac{1}{T}\sum_{t=1}^{R-j} z_t(u,v) z_{t+j}(u,v) - \left(\frac{1}{T}\sum_{t=1}^{R} z_t(u,v)\right)^2.$$

Let

$$-_{1R} = \left\{\ \omega : |Var_*\left(\tfrac{1}{\sqrt{R}}\textstyle\sum_{t=1}^{R} z_t^*(u,v)|\text{ data}\right) - Var\left(\tfrac{1}{\sqrt{R}}\textstyle\sum_{t=1}^{R} z_t(u,v)\right)| < \epsilon\ \right\}.$$

From Theorem 1 in PR, it follows that $P(-_{1R}) \to 1$ as $R \to \infty$. Thus, $\forall \omega \in -_{1R}$, and given A1 and A4, by Theorem 2 in PR, the law of $\frac{1}{\sqrt{R}}\sum_{t=1}^{R}(z_t^*(\omega,u,v) - z_t(u,v))$, as $R, T \to \infty$, and with $p_R = R^{-\gamma}$, $0 < \gamma < 1$, converges to the law of $\frac{1}{\sqrt{R}}\sum_{t=1}^{R}(z_t(u,v) - E(z_t(u,v)))$ which coincides with the law of $KS_T(u,v)$ under the null hypothesis. We now need to show that $KS_{R,T}^*(\omega,u,v)$ is P^*-stochastic equicontinuous $\forall \omega$, except for a subset of probability measure approaching zero, that is, we need to show that

$$\limsup_{T\to\infty} P_*\left(\sup_{\substack{u:|u-u'|<\delta \\ v:|v-v'|<\delta}} |KS_{R,T}^*(\omega,u,v) - KS_{R,T}^*(\omega,u',v')| > \varepsilon\right) = 0 \tag{33}$$

$\forall \omega \in -_{AR}$, with $P(-_{AR}) \to 1$ as $R \to \infty$. In order to show (33), it suffices to show (34), (35) and (36) below.

$$\sup_{\substack{u:|u-u'|<\delta \\ v:|v-v'|<\delta}} \left| \frac{1}{R^{1/2}}\sum_{t=1}^{R}\left(\begin{array}{c} \left(1\{y_t \le u\} - F(u|X_t,\widehat{\theta}_T)\right) 1\{X_t \le v\} \\ -\left(1\{y_t \le u'\} - F(u'|X_t,\widehat{\theta}_T)\right) 1\{X_t \le v'\} \end{array}\right)\right| \le \frac{\varepsilon}{3} \tag{34}$$

$\forall \omega \in -_{BR}$, with $P(-_{BR}) \to 1$ as $R \to \infty$,

$$\begin{aligned} &\limsup_{T\to\infty} P_*\left(\sup_{\substack{u:|u-u'|<\delta \\ v:|v-v'|<\delta}} \left| \frac{1}{R^{1/2}}\sum_{t=1}^{R}\left(\begin{array}{c} \left(1\{y_t^*(\omega) \le u\} - F(u|X_t^*(\omega),\widehat{\theta}_T)\right) 1\{X_t^*(\omega) \le v\} \\ -\left(1\{y_t^*(\omega) \le u'\} - F(u'|X_t^*(\omega),\widehat{\theta}_T)\right) 1\{X_t^*(\omega) \le v'\} \end{array}\right)\right| > \frac{\varepsilon}{3}\right) \\ &= 0 \end{aligned} \tag{35}$$

$\forall\omega \in -_{CR}$, with $P(-_{CR}) \to 1$ as $R \to \infty$,

$$\limsup_{T\to\infty} P_*\left(\sup_{\substack{u:|u-u'|<\delta \\ v:|v-v'|<\delta}}\left|\begin{array}{c}\left(\frac{1}{T}\sum_{t=1}^{T}\left(\nabla_\theta F(u|X_t,\widehat{\theta}_T)'1\{X_t\le v\} - \nabla_\theta F(u'|X_t,\widehat{\theta}_T)'1\{X_t\le v'\}\right)\right) \\ \times A_T(\widehat{\theta}_T)\left(\frac{1}{R^{1/2}}\sum_{t=1}^{R}(q(y_t^*(\omega),X_t^*(\omega),\widehat{\theta}_T)-q(y_t,X_t,\widehat{\theta}_T))\right)\end{array}\right| > \frac{\varepsilon}{3}\right) = 0 \quad (36)$$

$\forall\omega \in -_{DR}$, with $P(-_{DR}) \to 1$ as $R \to \infty$, and set $-_{AR} = -_{BR} \cap -_{CR} \cap -_{DR}$.

Recalling that $\widehat{\theta}_T - \theta^\dagger = o_{a.s.}(1)$, by Theorem 6 we know that $\left(1\{y_t \le u\} - F(u|X_t,\widehat{\theta}_T)\right)1\{X_t \le v\}$ is equicontinuous on $U \times V$ $\forall\omega \in -_{BR}$, with $P(-_{BR}) \to 1$ as $R \to \infty$, thus (34) follows immediately. Now, conditional on the sample and for all sample but a set of probability measure converging to zero, $\frac{1}{R^{1/2}}\sum_{t=1}^{R}(q(y_t^*(\omega),X_t^*(\omega),\widehat{\theta}_T)-q(y_t,X_t,\widehat{\theta}_T))$ is bounded in P_*-probability, as it converges in ditribution. Also, $\frac{1}{T}\sum_{t=1}^{T}\nabla_\theta F(u|X_t,\widehat{\theta}_T)'1\{X_t \le v\}$ is $P-$stochastic equicontinuous on $U \times V$ by the same argument used in the proof of Theorem 6. Thus (36) follows. It remains to show (35). Now,

$$\sup_{\substack{u:|u-u'|<\delta \\ v:|v-v'|<\delta}}\left|\left(1\{y_t^*(\omega)\le u\} - F(u|X_t^*(\omega),\widehat{\theta}_T)\right)1\{X_t^*(\omega)\le v\}\right| \le 1$$

without loss of generality let $u < u'$ and $v < v'$,

$$\left(E_*\sup_{\substack{u:|u-u'|<\delta \\ v:|v-v'|<\delta}}\left(\left|\begin{array}{c}\left(1\{y_t^*(\omega)\le u\} - F(u|X_t^*(\omega),\widehat{\theta}_T)\right)1\{X_t^*(\omega)\le v\} \\ -\left(1\{y_t^*(\omega)\le u'\} - F(u'|X_t^*(\omega),\widehat{\theta}_T)\right)1\{X_t^*(\omega)\le v'\}\end{array}\right|^p / \text{ data}\right)\right)^{1/p}$$

$$\le \left(E_*\sup_{\substack{u:|u-u'|<\delta \\ v:|v-v'|<\delta}}\left(\left|\left(1\{y_t^*(\omega)\le u\} - 1\{y_t^*(\omega)\le u'\right)1\{X_t^*(\omega)\le v\}\right|^p / \text{ data}\right)\right)^{1/p}$$

$$+\left(E_*\sup_{\substack{u:|u-u'|<\delta \\ v:|v-v'|<\delta}}\left(\left|\left(1\{X_t^*(\omega)\le v\} - 1\{X_t^*(\omega)\le v'\right)1\{y_t^*(\omega)\le u'\}\right|^p / \text{ data}\right)\right)^{1/p}$$

$$+\left(E_*\sup_{\substack{u:|u-u'|<\delta \\ v:|v-v'|<\delta}}\left(\left|\left(F(u|X_t^*(\omega),\widehat{\theta}_T) - F(u'|X_t^*(\omega),\widehat{\theta}_T)\right)1\{X_t^*(\omega)\le v\}\right|^p / \text{ data}\right)\right)^{1/p}$$

$$+\left(E_*\sup_{\substack{u:|u-u'|<\delta \\ v:|v-v'|<\delta}}\left(\left|F(u|X_t^*(\omega),\widehat{\theta}_T)(1\{X_t^*(\omega)\le v\} - 1\{X_t^*(\omega)\le v'\})\right|^p / \text{ data}\right)\right)^{1/p}$$

$$\leq \left(E_* \sup_{\substack{u:|u-u'|<\delta \\ v:|v-v'|<\delta}} \left(\left|1\{u \leq y_t^*(\omega) \leq u'\}\right|^p / \text{ data}\right)\right)^{1/p} + \left(E_* \sup_{\substack{u:|u-u'|<\delta \\ v:|v-v'|<\delta}} \left(\left|1\{v \leq X_t^*(\omega) \leq v'\}\right|^p / \text{ data}\right)\right)^{1/p}$$

$$+ \left(E_* \sup_{u \times \theta \in U \times \Theta} |\nabla_u F(u|X_t^*, \theta)|^p / \text{ data}\right)^{1/p} \delta$$

$$+ \left(E_* \sup_{u \times \theta \in U \times \Theta} |F(u|X_t^*, \theta)|^{2p} / \text{ data}\right)^{1/2p} \left(E_* \sup_{\substack{u:|u-u'|<\delta \\ v:|v-v'|<\delta}} \left(\left|1\{v \leq X_t^*(\omega) \leq v'\}\right|^{2p} / \text{ data}\right)\right)^{1/2p}$$

$$\leq \left(\sup_{u \in U} \left|\frac{1}{R}\sum_{t=1}^{R} 1\{u \leq y_t \leq u+\delta\}\right|\right)^{1/p} + \left(\sup_{v \in V} \left|\frac{1}{R}\sum_{t=1}^{R} 1\{v \leq X_t \leq v+\delta\}\right|\right)^{1/p}$$

$$+ C^{1/p}\delta + C^{1/2p}\left(\sup_{v \in V} \left|\frac{1}{R}\sum_{t=1}^{R} 1\{v \leq X_t \leq v+\delta\}\right|\right)^{1/2p} \leq \frac{\varepsilon}{3}$$

$\forall \omega \in \text{-}_{CR}$, with $P(\text{-}_{CR}) \to 1$ as $R \to \infty$. This shows (33). The desired result then follows by the continuous mapping theorem.

Proof of Theorem 8: The only difference with the proof of Theorem 6 is due to the fact that y_t and $g(X_t, \widehat{\theta}_T)$ are unbounded sequences and so we cannot show stochastic equicontinuity using the same arguments as above. Let $m_t(v) = (y_t - g(X_t, \widehat{\theta}_T))1\{X_t \leq v\}$, we shall show that for $p > 2$,

$$\sup_{v \in V} E(|m_t(v)|^p) \leq \infty, \ p > 2 \tag{37}$$

and

$$(E(|m_t(v) - m_t(v')|^p))^{1/p} \leq B|v - v'|^{\lambda}, \ B < \infty, \ \lambda > 0 \tag{38}$$

As shown by Hansen (1996, proof of Theorem 1), (37) and (38) ensure that the conditions for stochastic equicontinuity of Doukhan, Massart and Rio (1995) are satisfied. The left hand side of (37) is majorized by

$$\sup_{\theta \in \Theta} \left(E(y_t - g(X_t, \theta))^{2p}\right)^{1/2p} \sup_{v \in V} \left(E(1\{X_t \leq v\})\right)^{1/2p} < \infty,$$

given C$'$(ii). As for (38), let $v < v'$ and note that it is majorized by

$$\sup_{\theta \in \Theta} \left(E(y_t - g(X_t, \theta))^{2p}\right)^{1/2p} \left(E(1\{X_t \leq v\})\right)^{1/2p}$$

$$\leq C \left(\int_v^{v'} f_x(s) ds\right)^{1/2p} \leq C|v - v'|^{1/2p}, \ C < \infty$$

Proof of Theorem 9: Also in this case, the only difference with the proof of Theorem 7 is due to the fact that y_t and $g(X_t, \widehat{\theta}_T)$ are unbounded sequences and so we cannot show stochastic equicontinuity using the same arguments as above. In order to show that

$$\lim_{T\to\infty} \sup P_* \left(\sup_{v:|v-v'|<\delta} |K^*_{R,T}(\omega, v) - K^*_{R,T}(\omega, v')| > \varepsilon \right) = 0 \tag{39}$$

$\forall \omega \in \text{-}_{AR}$, with $P(\text{-}_{AR}) \to 1$ as $R \to \infty$, we need to show that

$$\sup_{v:|v-v'|<\delta} \left| \frac{1}{\sqrt{R}} \sum_{t=1}^{R} (y_t - g(X_t, \widehat{\theta}_T))(1\{X_t \le v\} - 1\{X_t \le v'\}) \right| \le \frac{\varepsilon}{3} \tag{40}$$

$\forall \omega \in \text{-}_{BR}$, with $P(\text{-}_{BR}) \to 1$ as $R \to \infty$,

$$\lim_{T\to\infty} \sup P_* \left(\sup_{v:|v-v'|<\delta} \left| \frac{1}{\sqrt{R}} \sum_{t=1}^{R} (y^*_t(\omega) - g(X^*_t(\omega), \widehat{\theta}_T))(1\{X_t(\omega) \le v\} - 1\{X_t(\omega) \le v'\}) \right| > \frac{\epsilon}{3} \right) = 0 \tag{41}$$

$\forall \omega \in \text{-}_{CR}$, with $P(\text{-}_{CR}) \to 1$ as $R \to \infty$,

$$\lim_{T\to\infty} \sup P_* \left(\sup_{v:|v-v'|<\delta} \left| \begin{array}{l} \left(\frac{1}{T}\sum_{t=1}^{T} \nabla_\theta g(X_t, \overline{\theta}_T)(1\{X_t(\omega) \le v\} - 1\{X_t(\omega) \le v'\})\right) \\ \times A_T(\overline{\theta}_T) \frac{1}{\sqrt{R}} \sum_{t=1}^{R} (q(y^*_t, X^*_t(\omega), \widehat{\theta}_T) - q(y^*_t, X^*_t(\omega), \widehat{\theta}_T)) \end{array} \right| > \frac{\epsilon}{3} \right) = 0 \tag{42}$$

$\forall \omega \in \text{-}_{DR}$, with $P(\text{-}_{DR}) \to 1$ as $R \to \infty$, and set $\text{-}_{AR} = \text{-}_{BR} \cap \text{-}_{CR} \cap \text{-}_{DR}$.

Now (40) follows straightforwardly from the $P-$stochastic equicontinuity of $\frac{1}{\sqrt{R}} \sum_{t=1}^{R} (y_t - g(X_t, \widehat{\theta}_T))1\{X_t \le v\}$, which has been proved above. As for (41),

$$\sup_{v\in V} E_*(|(y_t - g(X_t, \widehat{\theta}_T))1\{X_t \le v\}|^p / \text{ data })^{1/p}$$

$$\le \sup_{v\in V} \left(\frac{1}{R} \sum_{t=1}^{R} (y_t - g(X_t, \widehat{\theta}))1\{X_t \le v\} \right)$$

which is bounded for all samples but a set probability measure approaching zero because of the uniform law of large numbers. Also, for $v < v'$,

$$E_* \left(|(y^*_t(\omega) - g(X^*_t(\omega), \widehat{\theta}_T))(1\{X^*_t(\omega) \le v\} - 1\{X^*_t(\omega) \le v\})|^p / \text{ data} \right)^{1/p}$$

$$\le E_* \left(|(y^*_t(\omega) - g(X^*_t(\omega), \widehat{\theta}_T))|^{2p} / \text{ data} \right)^{1/2p} E_* \left(|1\{v \le X^*_t(\omega) \le v'\}| / \text{ data} \right)^{1/2p}$$

$$\le \left[\sup_{\theta\in\Theta} \left(\frac{1}{R} \sum_{t=1}^{R} (y_t - g(X_t, \theta)) \right)^{2p} \right]^{1/2p} \left[\left(\frac{1}{R} \sum_{t=1}^{R} 1\{v \le X_t \le v'\} \right) \right]^{1/2p}$$

$$\le B \left(\int_v^{v'} f_x(s) ds \right)^{1/2p} \le B|v - v'|^{1/2p} \; \forall \omega \in \text{-}_{CR}, \; P(\text{-}_{CR}) \to 1 \text{ as } R \to \infty$$

as the term in the first square brackets satisfy a uniform law of large numbers (under C$'$(i)(ii)) and the term in the second square bracket has been shown to be $P-$stochastic equicontinuous on V. This completes the proof of (41). Finally as for (42), $\frac{1}{\sqrt{R}}\sum_{t=1}^{R}(q(y_t^*, X_t^*(\omega), \widehat{\theta}_T) - q(y_t^*, X_t^*(\omega), \widehat{\theta}_T))$ is $O_{P^*}(1)$ for all samples but a set with probability measure approaching zero, because it converges in distribution. Also $\frac{1}{T}\sum_{t=1}^{T}\nabla_\theta g(X_t, \overline{\theta}_T)(1\{X_t(\omega) \leq v\} - 1\{X_t(\omega) \leq v'\})$ can be shown to be $P-$stochastic equicontinuous by the same argument used in the proof of Theorem 6, given C$'$(ii). Thus (39) follows.

9 References

Aït-Sahalia, Y., (1996), Testing Continuous Time Models of the Spot Interest Rate, *Review of Financial Studies,* 9, 385-426.

Altissimo, F. and V. Corradi, (2000), Bounds for Inference in the Presence of Nuisance Parameters Unidentified Under the Null, Working Paper, University of Exeter.

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.

Bai, J., (1998), Testing Parametric Conditional Distributions of Dynamic Models, Working Paper, Boston College.

Beran, R. and P.W. Millar, (1989), A Stochastic Minimum Distance Test for Multivariate Parametric Models, *Annals of Statistic,* 17, 125-140.

Bierens, H. and W. Ploberger, (1997), A Stochastic Minimum Distance Test for Multivariate Parametric Models, *Annals of Statistic,* 17, 125-140.

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

Corradi, V., (1999), Deciding between $I(0)$ and $I(1)$ via FLIL-based Bounds, *Econometric Theory,* 15, 643-663.

Corradi, V. and N.R. Swanson, (2000), A Consistent Test for Out of Sample Nonlinear Predictive Ability, Working Paper, Texas A&M University.

Cox, J.C., J.E. Ingersoll and S.A. Ross, (1985), A Theory of the Term Structure of Interest Rates, *Econometrica,* 53, 385-407.

Doukhan, P., P. Massart and E. Rio, (1995), Invariance Principles for Absolutely Regular Empirical Processes, *Annales de l'Institute Henri Poincare: Probabiltes et Statistique,* 31, 393-427.

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

Dridi, R., (1999), Simulated Asymptotic Least Squares Theory, Working Paper, London School of

Economics.

Duffie, D. and K. Singleton, (1993), Simulated Moment Estimation of Markov Models of Asset Prices, *Econometrica,* 61, 929-952.

Fournie, E., (1993), Un Test de Type Kolmogorov-Smirnov pour Processus de Diffusion Ergodiques, W.P. 1696, INRIA.

Gallant, A.R. and G. Tauchen, (1996), Which Moments to Match, *Econometric Theory,* 12, 657-681.

Gard, T.C., (1988), *Introduction to Stochastic Differential Equations,* Marcel Dekker, New York.

Gourieroux, C., A. Monfort and E. Renault, (1993), Indirect Inference, *Journal of Applied Econometrics,* 8, 85-118.

Hall, P., (1992), *The Bootstrap and Egeworth 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.

Inoue, A., (1998), Testing for Distributional Changes in Times Series, *Econometric Theory,* forthcoming.

Inoue, A., (1999), A Conditional Goodness of Fit Test for Time Series, Working Paper, North Carolina State University.

Johnson, N.L., S. Kotz and N. Balakrishnan, (1994), *Continuous Univariate Distributions I,* Wiley, New York.

Karlin, S. and H.M. Taylor, (1981), *A Second Course in Stochastic Processes,* Academic Press, San Diego.

Kolmogorov, A.N., (1933), Sulla Determinazione Empirica di una Legge di Distribuzione, *Giornale dell'Istituto degli Attuari,* 4, 83-91.

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

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.

Pardoux, E. and D. Talay, (1985), Discretization and Simulation of Stochastic Differential Equations, *Acta Applicandae Mathematicae,* 3, 23-47.

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

Philipp, W., (1982), Invariance Principles for Sums of Mixing Random Elements and the Multivariate Empirical Process, *Colloquia Mathematica Societatis Janos Bolyai,* 36, 843-873.

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

Pollard, D., (1990), *Empirical Processes: Theory and Applications,* CBMS, Conference Series in Probability and Statistics, v.2, Hayward, CA, Institute of Mathematical Statistics.

Radulovic D., (1996), The Bootstrap for Empirical Processes Based on Stationary Observations, *Stochastic Processes and their Applications,* 65, 259-281.

Sin, C.Y. and H. White, (1996), Information Criteria for Selecting Possibly Misspecified Parametric Models, *Journal of Econometrics,* 71, 207-225.

Smirnov, N., (1939), On the Estimation of the Discrepancy Between Empirical Curves of Distribution for Two Independent Samples, *Bulletin Mathematique de l'Universite' de Moscou,* 2, fasc. 2.

Stramer, O. and R.L. Tweedie, (1997) Geometric and Subgeometric Convergence of Diffusions with Given Stationary Distribution and their Discretization, Manuscript, University of Iowa and Colorado State University.

Thompson, S.B., (2000), Specification Tests for Continuous Time Models, Working Paper, Harvard University.

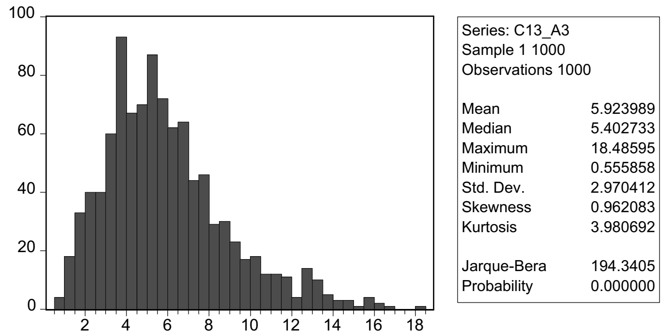
Whang, Y.J., (2000), Consistent Bootstrap Tests of Parametric Regression Functions, *Journal of Econometrics,* 98, 27-46.

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

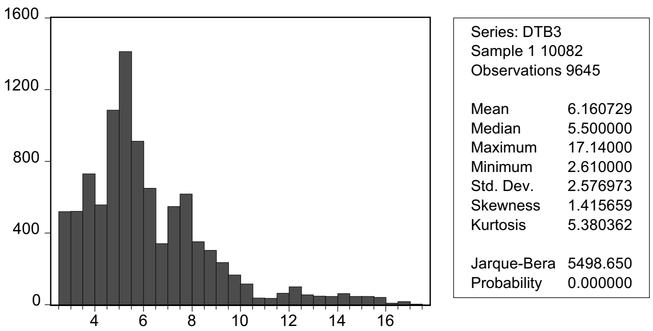
Wong, E., (1964), The Construction of a Class of Stationary Markov Processes, in *Sixteenth Symposia in Applied Mathematics: Stochastic Processes in Mathematical Physics and Engineering,* ed. by R. Bellman, American Mathematical Society, Providence, RI.

Figure 1: Simulated and Actual Data

Panel 1: 1000 Observations Generated According to $dX(t) = (6 - X(t))dt + \sqrt{3X(t)}dW(t)$



Panel 2: Monthly 3-mo Treasury Bill Data for the Period 2/1962-9/2000 (Ann. %)



Panel 3: Daily 3-mo Treasury Bill Data for the Period 02/01/1962-09/22/2000 (Ann. %)

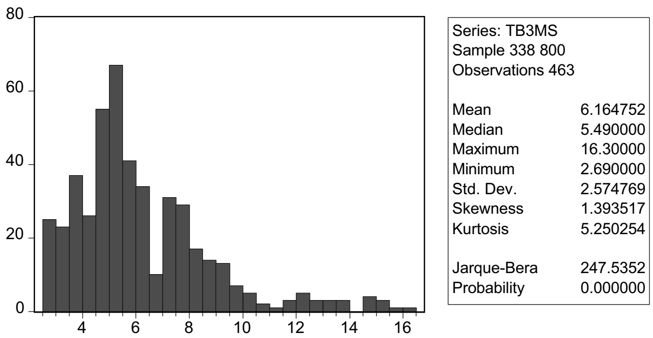


Table 1: Monte Carlo Rejection Frequencies: Data Generated Under H_0 *

Part 1: Parameter Estimation Error Accounted For

R	Sample (T) = 500			Sample (T) = 1000		
	S_T^2	$\lvert S_T \rvert$	$\sup_{u\in U} \lvert S_T(u) \rvert$	S_T^2	$\lvert S_T \rvert$	$\sup_{u\in U} \lvert S_T(u) \rvert$
Panel a: $c_1 = 2$, $a = -2$						
$T^{0.80}$	0.164	0.131	0.127	0.150	0.131	0.116
$T^{0.90}$	0.141	0.131	0.127	0.131	0.125	0.110
$T^{0.95}$	0.142	0.130	0.125	0.129	0.129	0.113
Panel b: $c_1 = 3$, $a = -3$						
$T^{0.80}$	0.139	0.134	0.127	0.132	0.121	0.113
$T^{0.90}$	0.137	0.132	0.128	0.122	0.114	0.113
$T^{0.95}$	0.129	0.127	0.128	0.119	0.123	0.113
Panel c: $c_1 = 4$, $a = -4$						
$T^{0.80}$	0.137	0.129	0.134	0.136	0.138	0.127
$T^{0.90}$	0.135	0.127	0.130	0.123	0.126	0.128
$T^{0.95}$	0.134	0.129	0.115	0.124	0.125	0.126

Part 2: Parameter Estimation Error Not Accounted For

R	Sample (T) = 500			Sample (T) = 1000		
	$S_{R,T}^2$	$\lvert S_{R,T} \rvert$	$\sup_{u\in U} \lvert S_{R,T}(u) \rvert$	$S_{R,T}^2$	$\lvert S_{R,T} \rvert$	$\sup_{u\in U} \lvert S_{R,T}(u) \rvert$
Panel a: $c_1 = 2$, $a = -2$						
$T^{0.80}$	0.153	0.138	0.141	0.174	0.177	0.166
$T^{0.90}$	0.066	0.068	0.084	0.089	0.091	0.093
$T^{0.95}$	0.021	0.022	0.037	0.033	0.028	0.034
Panel b: $c_1 = 3$, $a = -3$						
$T^{0.80}$	0.146	0.138	0.141	0.172	0.171	0.166
$T^{0.90}$	0.065	0.071	0.092	0.089	0.091	0.097
$T^{0.95}$	0.025	0.022	0.039	0.031	0.028	0.039
Panel c: $c_1 = 4$, $a = -4$						
$T^{0.80}$	0.147	0.136	0.138	0.171	0.174	0.162
$T^{0.90}$	0.073	0.073	0.089	0.093	0.091	0.092
$T^{0.95}$	0.025	0.022	0.034	0.033	0.029	0.038

* Notes: All entries are rejection frequencies of the null hypothesis. Critical values are set equal to the 90^{th} percentile of the bootstrap distribution. Results are based on 1000 Monte Carlo simulations, with 100 bootstrap statistics constructed at each iteration in order to obtain the bootstrap distribution. See above for further details.

Table 2: Monte Carlo Rejection Frequencies: Data Generated Under H_A *

Part 1: Parameter Estimation Error Accounted For						
R	Sample (T) = 500			Sample (T) = 1000		
	S_T^2	$\|S_T\|$	$\sup_{u \in U} \|S_T(u)\|$	S_T^2	$\|S_T\|$	$\sup_{u \in U} \|S_T(u)\|$
Panel a: $\theta_1 = 0.3$						
$T^{0.80}$	0.879	0.888	0.975	0.939	0.940	0.982
$T^{0.90}$	0.914	0.929	0.994	0.970	0.970	0.992
$T^{0.95}$	0.934	0.950	0.998	0.992	0.994	0.996
Panel b: $\theta_1 = 0.6$						
$T^{0.80}$	0.864	0.931	0.987	0.935	0.972	0.994
$T^{0.90}$	0.880	0.967	0.997	0.938	0.986	0.999
$T^{0.95}$	0.881	0.967	0.999	0.951	0.993	1.000
Panel c: $\theta_1 = 0.9$						
$T^{0.80}$	0.891	0.953	0.970	0.967	0.989	0.996
$T^{0.90}$	0.928	0.982	0.996	0.990	0.999	0.999
$T^{0.95}$	0.944	0.988	0.995	0.995	1.000	1.000
Part 2: Parameter Estimation Error Not Accounted For						
R	Sample (T) = 500			Sample (T) = 1000		
	$S_{R,T}^2$	$\|S_{R,T}\|$	$\sup_{u \in U} \|S_{R,T}(u)\|$	$S_{R,T}^2$	$\|S_{R,T}\|$	$\sup_{u \in U} \|S_{R,T}(u)\|$
Panel a: $\theta_1 = 0.3$						
$T^{0.80}$	0.919	0.942	0.968	0.992	0.997	1.000
$T^{0.90}$	0.952	0.967	0.993	0.999	1.000	1.000
$T^{0.95}$	0.967	0.982	0.996	1.000	1.000	1.000
Panel b: $\theta_1 = 0.6$						
$T^{0.80}$	0.560	0.720	0.853	0.762	0.933	0.988
$T^{0.90}$	0.675	0.878	0.961	0.939	0.996	0.999
$T^{0.95}$	0.732	0.917	0.977	0.982	1.000	1.000
Panel c: $\theta_1 = 0.9$						
$T^{0.80}$	0.444	0.556	0.604	0.661	0.803	0.891
$T^{0.90}$	0.596	0.721	0.819	0.890	0.973	0.989
$T^{0.95}$	0.689	0.821	0.896	0.966	0.995	0.999

* See notes to Table 1.