Corradi, Valentina; Swanson, Norman R.

Working Paper

Bootstrap Specification Tests for Diffusion Processes

Working Paper, No. 2003-21

Provided in Cooperation with:
Department of Economics, Rutgers University

Suggested Citation: Corradi, Valentina; Swanson, Norman R. (2003) : Bootstrap Specification Tests for Diffusion Processes, Working Paper, No. 2003-21, Rutgers University, Department of Economics, New Brunswick, NJ

This Version is available at:
https://hdl.handle.net/10419/23177

Bootstrap Specification Tests for Diffusion Processes*

Valentina Corradi
Queen Mary, University of London

Norman R. Swanson
Rutgers University

June 2003

Abstract

This paper introduces bootstrap specification tests for diffusion processes. In the one-dimensional case, the proposed test is closest to the nonparametric 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. However, we compare cumulative distribution functions, while Aït-Sahalia compares densities. In the multidimensional and/or multifactor case, the proposed test is based on the comparison of the empirical CDF of the actual data and the empirical CDF of the simulated data. The limiting distributions of both tests are functionals of zero mean Gaussian processes with covariance kernels that reflect data dependence and parameter estimation error (PEE). In order to obtain asymptotically valid critical values for the tests, we use an empirical process version of the block bootstrap which properly accounts for the contribution of PEE. An example based on a simple version of the Cox, Ingersol and Ross (1985) square root process is outlined and related Monte Carlo experiments are carried out. These experiments suggest that the test has good finite sample properties, even for samples as small as 400 observations when tests are formed using critical values constructed with as few as 100 bootstrap replications.

JEL classification: C12, C22.
Keywords: block bootstrap, diffusion process, multifactor model, parameter estimation error, specification test, stochastic volatility.

*Valentina Corradi, Department of Economics, Queen Mary, University of London, Mile End Road, London E1 4NS, UK, v.corradi@qmul.ac.uk. Norman R. Swanson, Department of Economics, Rutgers University, 75 Hamilton Street, New Brunswick, NJ 08901-1248, USA, nswanson@econ.rutgers.edu. Parts of this paper were written while the second author was visiting the University of California, San Diego. We wish to thank Ron Gallant, the associate editor, two anonymous referees, Filippo Altissimo, Marcelo Fernandes, Silvia Goncalves, Soren Johansen, Oliver Linton, Allan Timmerman, Paolo Zaffaroni, and seminar participants at Ente Einaudi-Roma, Queen Mary, University of London, Université Libre Bruxelles, European Universitary Institute, University of Washington, Purdue University, Rutgers University, ESRC-UK econometrics Group in Bristol, the 2000 Winter Meetings of the Econometric Society, and the 2001 Meeting of the European Econometric Society for useful comments on earlier versions of this paper. In addition, special thanks are due to Antonio Mele and Bent Sorensen for providing numerous very helpful suggestions. Corradi gratefully acknowledges financial support from ESRC, grant code R000230006.

1 Introduction

This paper introduces two bootstrap specification tests for diffusion processes. In the one-dimensional case, note that the invariant density associated with the diffusion is implied by the specification of the drift and variance terms. Therefore, we can construct a Kolomogorov type test, based on comparison of the empirical cumulative distribution function and the cumulative distribution function (CDF) implied by the specification of the drift and the variance, under the null model. This test, which is the first of our two tests, is closest to the nonparametric 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, although our test is based on a comparison of CDFs, while Aït-Sahalia's is based on a comparison of densities. Thus, his approach requires the use of a nonparametric density estimator (and hence the choice of the bandwidth parameter) and is characterized by a nonparametric rate, while our test has a parametric rate. In the case of either multidimensional diffusions or multifactor models characterized by stochastic volatility, the functional form of the invariant density of the return(s) is no longer guaranteed to be given in closed form, upon joint specification of the drift and variance terms. However, in this case we can still compare empirical distributions, and our second test is thus based on the comparison of the empirical distribution of the actual data and the empirical distribution of the (model) simulated data.

It should be noted that tests based on the comparison of CDFs have no power against *i.i.d.* alternatives that are generated by the same marginal density as that implied under H_0. However, this feature of our tests is not particularly relevant in the context of highly dependent financial data, for example. Nevertheless, it would in principle be interesting to construct specification tests for diffusions based on the specification of the transition density. The main difficulty with this is that knowledge of the drift and variance terms of a diffusion does not in turn generally imply knowledge of the transition density. Indeed, if the functional form for the transition density were known, we could test the hypothesis of correct specification of a diffusion via the generalized cross-spectrum approach of Hong (2001), (see also Hong and Li (2002), and Hong, Li and Zhao (2002) for application to testing financial models), or via the robust periodogram approach of Thompson (2002). In the case in which the transition density is unknown, a test can be constructed by comparing the kernel (conditional) density estimator of the actual and simulated data, as in Altissimo, Fornari

and Mele (2002).[1] Alternatively, closed form approximation of the transition density (and hence the likelihood function) is proposed by Aït-Sahalia (1999, 2002), who also provides conditions under which the argmax of the approximated likelihood is asymptotically equivalent to the "true" maximum likelihood estimator. Whether the same approach can be used to obtain an approximation of the conditional distribution to be used in an implementation of a conditional Kolmogorov test, along the lines of Andrews (1997), is left to future research. A possible justification for comparing marginal distributions instead of conditional distributions is the case of subordinated diffusions (see e.g. Conley, Hansen, Luttmer and Scheinkman (1997)), where the subordinated process may have the same stationary density as the underlying diffusion, but not the same transition density.

As mentioned above, the test that we propose for the one-dimensional case is based on comparison of the empirical CDF of the data and the distribution implied by the drift and the variance of the diffusion, evaluated at the estimated parameters.[2] In the multidimensional and/or multifactor case, our test is based on comparison of the empirical distribution of the data and the empirical distribution of data simulated using estimated parameters. In both cases, parameters are estimated via the simulated generalized method of moments (SGMM) approach of Duffie and Singleton (1993), using as many moment conditions as parameters (exact identification). As is common with these types of tests, the limiting distributions are functionals of zero mean Gaussian processes with covariance kernels that reflect both the contribution of parameter estimation error (PEE) as well as the time series nature of the data. Thus, the limiting distributions are not nuisance parameters free and critical values cannot be tabulated. Our approach is to provide valid asymptotic critical values via an extension of the empirical process version of the block bootstrap which properly captures the contribution of PEE, for the case where parameters are estimated via SGMM. Of note in this context is that when the simulation error is negligible (i.e. the simulated sample grows faster than the historical sample), and given exact identification, the results developed by Goncalves and White

[1] Unfortunately, in this context there is not a well defined conditional empirical distribution. Interestingly, Thompson (2002) suggests an ingenious device (based on the use of an Euler scheme) for approximating 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 his approach in conjunction with standard techniques.

[2] A Kolmogorov-Smirnov test for diffusion processes has previously been suggested by Fournie (1993), for the case in which the continuous trajectories are observed.

(2002) for QMLE estimators extend to SGMM estimators.[3]

The potential usefulness of our proposed bootstrap based tests is examined via a series of Monte Carlo experiments in the context of testing the goodness of fit of a square root diffusion process, which is specified using a simplified version of the Cox, Ingersoll and Ross (1985) model, under the null hypothesis. Under the alternative, logged data are generated according to an Ornstein-Uhlenbeck process, so that the data are lognormal. For samples of 400, 800, and 1200 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 nominal values, and rejection rates under the alternative are generally high.[4]

The rest of the paper is organized as follows. In Section 2, we outline the specification test and analyze its asymptotic behavior, for the case of one-dimensional diffusions. Section 3 outlines the bootstrap procedure which we propose, and establishes its asymptotic validity. Section 4 discusses extention to multifactor models, and Section 5 contains the results from our Monte Carlo experiments. Concluding remarks are contained in Section 6. All proofs are collected in an appendix.

2 One-Dimensional Diffusion Specification Test

In this section we outline our test for the joint correct specification of the drift and variance terms in one-dimensional stationary-ergodic diffusion processes. Let $X(t)$, $t \geq 0$, be a diffusion process solution to the following stochastic differential equation:

$$dX(t) = b(X(t), \theta_0)dt + \sigma(X(t), \theta_0)dW(t), \tag{1}$$

where $\theta_0 \in \Theta$, $\Theta \subset \Re^k$, and Θ is a compact set.

It is known that the drift and variance terms ($b(\cdot)$ and $\sigma^2(\cdot)$, respectively) uniquely determine the stationary density, say $f(x, \theta_0)$, associated with the invariant probability measure of the above

[3]The issue of PEE is addressed by Thompson (2002) by providing upper bounds, which are valid for the case of efficiently estimated parameters. Another approach is that of Hong and Li (2002), who perform out of sample tests, with the estimation period growing faster than the prediction period; so that the contribution of PEE vanishes asymptotically.

[4]It is worth noting that the joint problem of simulating paths and simulating bootstrap replications make this Monte Carlo study rather computationally intensive, and we are not aware of other simulation studies which analyze the performance of bootstrap tests for diffusion processes.

diffusion process (see e.g. Karlin and Taylor (1981), pp. 241). In particular,

$$f(x,\theta_0) = \frac{c(\theta_0)}{\sigma^2(x,\theta_0)} \exp\left(\int^x \frac{2b(v,\theta_0)}{\sigma^2(v,\theta_0)} dv\right), \tag{2}$$

where $c(\theta)$ is a constant ensuring that the density integrates to one. 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 $\widehat{\theta}_{T,S,h}$, which is based on the skeleton of the observed data as well as on a (model) simulated path.[5] 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 associated with the underlying diffusion, and let $F(u,\theta_0)$ be the CDF associated with the density in (2). We consider the following hypotheses:

$$H_0 : F_0(u) = F(u,\theta_0), \text{ for all } u \in U \tag{3}$$

versus

$$H_A : F_0(u) \neq F(u,\theta_0), \text{ for some } u \in U, \text{ with nonzero Lebesgue measure.} \tag{4}$$

Now, note the a null hypothesis of joint correct specification of the drift and the variance terms implies H_0 in (3). However, the reverse does not hold, as H_0 does not necessarily imply the correct specification of the diffusion process (see equation (2)). In fact, we cannot rule out the possibility that even though X_t is not a skeleton of the solution to the stochastic differential equation in (1), we still have that $\Pr(X_t \leq u) = F(u,\theta_0)$. In order to test H_0 versus H_A, consider the following test statistic:[6]

$$V^2_{T,S,h} = \int_U V^2_{T,S,h}(u)\pi(u)du, \tag{5}$$

where

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

[5] In the case in which the moment conditions can be written in closed form, we have that $\widehat{\theta}_{T,S,h} = \widehat{\theta}_T$, as S is the sample length of the simulated path used in estimation of θ_0, and h is the discretization parameter used in the application of Euler and/or Milstein approximation schemes, for example (see below for further details).

[6] In Monte Carlo experiments that are reported on below, we also examine the finite sample properties of two different version of the test statistic, namely $|V_{T,S,h}| = \int_U |V_{T,S,h}(u)|\pi(u)du$, and $|V_{T,S,h}|^{\sup} = \sup_{u\in U} |V_{T,S,h}(u)|$.

with U the compact interval defined below (see Assumption 2) and $\int_U \pi(u)du = 1$. Further, $\widehat{\theta}_{T,S,h}$ is the simulated GMM (SGMM) estimator, defined to be:

$$\begin{aligned}\widehat{\theta}_{T,S,h} &= \arg\min_{\theta\in\Theta}\left(\frac{1}{T}\sum_{t=1}^{T}g(X_t) - \frac{1}{S}\sum_{t=1}^{S}g(X_{t,h}^{\theta})\right)' W_T \left(\frac{1}{T}\sum_{t=1}^{T}g(X_t) - \frac{1}{S}\sum_{t=1}^{S}g(X_{t,h}^{\theta})\right) \\ &= \arg\min_{\theta\in\Theta} G_{T,S,h}(\theta)' W_T G_{T,S,h}(\theta), \end{aligned} \tag{6}$$

where g denotes a vector of p moment conditions, $\Theta \subset R^p$ (so that we have as many moment conditions as parameters), and $X_{t,h}^{\theta} = X_{[Nth/S]}^{\theta}$, with $S = Nh$ (S denotes simulation path length and h is the discretization interval). Finally, W_T is the inverse of a heteroskedasticity and autocorrelation (HAC) robust covariance matrix estimator. That is:

$$W_T^{-1} = \frac{1}{T}\sum_{\tau=-l_T}^{l_T} w_\tau \sum_{t=\tau+1+l_T}^{T-l_T}\left(g(X_t) - \frac{1}{T}\sum_{t=1}^{T}g(X_t)\right)\left(g(X_{t-\tau}) - \frac{1}{T}\sum_{t=1}^{T}g(X_t)\right)', \tag{7}$$

where $w_\tau = 1 - \tau/(l_T + 1)$. In order to construct simulated estimators, we require simulated paths, under the null diffusion. If we use a Milstein scheme (see e.g. Pardoux and Talay (1985)), then

$$\begin{aligned} X_{kh}^{\theta} - X_{(k-1)h}^{\theta} &= b(X_{(k-1)h}^{\theta},\theta)h + \sigma(X_{(k-1)h}^{\theta},\theta)\epsilon_{kh} - \frac{1}{2}\sigma(X_{(k-1)h}^{\theta},\theta)'\sigma(X_{(k-1)h}^{\theta},\theta)h \\ &\quad + \frac{1}{2}\sigma(X_{(k-1)h}^{\theta},\theta)'\sigma(X_{(k-1)h}^{\theta},\theta)\epsilon_{kh}^2 \end{aligned} \tag{8}$$

where $\epsilon_{kh} \overset{iid}{\sim} N(0,h)$, $k = 1,\ldots,N$, $Nh = S$, and σ' is the derivative with respect to the first argument. Also define, the pseudo true value:

$$\theta^{\dagger} = \arg\min_{\theta\in\Theta} G_{\infty}(\theta)' W_0 G_{\infty}(\theta),$$

where $G_{\infty}(\theta)' W_0 G_{\infty}(\theta) = p\lim_{T,S\to\infty,h\to 0} G_{T,S,h}(\theta)' W_T G_{T,S,h}(\theta)$, and $\theta^{\dagger} = \theta_0$ under the null. The reason why we limit our attention to the exactly identified case is that this ensures that $G_{\infty}(\theta^{\dagger}) = 0$, even when the model used to simulate the diffusion is misspecified, in the sense of differing from the underlying DGP.[7]

A complete treatment of the asymptotic behavior of (non simulated) GMM in the joint presence of overidentification and misspecification is provided by Hall and Inoue (2003), who show that in

[7]First order conditions imply that

$$\nabla_{\theta} G_{\infty}(\theta^{\dagger})' W^{\dagger} G_{\infty}(\theta^{\dagger}) = 0.$$

However, in the case for which the number of parameters and the number of moment conditions is the same, $\nabla_{\theta} G_{\infty}(\theta^{\dagger})' W^{\dagger}$ is invertible, and so the first order conditions also imply that $G_{\infty}(\theta^{\dagger}) = 0$.

the case of a HAC-type weighting matrix, the rate of convergence depends on the lag truncation parameter. For the sake of simplicity, we focus on SGMM estimators in this paper. However, in the case of correct specification, we could equally rely on Indirect Inference (II: Gourieroux, Monfort and Renault (1993)) and Efficient Method of Moments (EMM: Gallant and Tauchen (1996)).[8] In addition to simulation based methods, approximate maximum likelihood estimators have recently received considerable attention. For example, Aït-Sahalia (1999, 2002) suggests a closed form approximation of the likelihood function, and Altissimo, Mele and Fornari (2003) propose an estimator based on the minimization of the distance between a kernel density estimator constructed using the actual data and one constructed using simulated data, in which case they provide conditions under which such estimators are asymptotically equivalent to maximum likelihood estimators. Nevertheless, it is not immediate to see what the properties of the Ait-Sahalia and Altissimo et al. estimators are in the misspecified case. Finally, nonparametric estimation of multidimensional diffusions via spectral decomposition of the generator function is studied by Chen, Hansen and Scheinkman (2000).

The following assumptions are used in the sequel.

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): $b(\cdot)$ and $\sigma(\cdot)$, as defined in (1), are twice continuously differentiable. Also, b, b', σ, and σ' are Lipschitz, with Lipschitz constant independent of θ.

Assumption A3 (A3): $F(u, \theta)$ is twice continuously differentiable in the interior of $\Theta \times U$, where Θ and U are compact subsets of $\Re^p$ 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 A4 (A4): For any fixed h and $\forall \theta \in \Theta$, X^θ_{kh} is geometrically ergodic and strictly stationary.[9]

Assumption A5 (A5): $W_T \xrightarrow{a.s.} W_0 = \sum_0^{-1}$, where,

[8] A unified framework for simulation based estimators, which nests SGMM, II and EMM, is provided in Dridi (1999).

[9] 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 inherited by the simulated paths. This is in general the case for the Euler or the Milstein scheme, whenever the drift grows at most at a linear rate and the drift and variance terms are not "too big".

$\sum_0 = \sum_{j=-\infty}^{\infty} E\left((g(X_1) - E(g(X_1)))(g(X_{1+j}) - E(g(X_{1+j})))'\right)$.

Assumption A6 (A6): $\forall\theta \in \Theta$, and for all h, $||g(X_{t,h}^{\theta})||_{2+\delta} < C < \infty$, $g(X_{t,h}^{\theta})$ is Lipschitz, uniformly on Θ, $\theta \to E(g(X_{t,h}^{\theta}))$ is continuous, and $g(X_t), g(X_{t,h}^{\theta}), \nabla_\theta X_{t,h}^{\theta}$ are $2r-$dominated (the last two also on Θ) for $r > 3/2$.[10]

Assumption A7 (A7): Unique identifiability: $G_\infty(\theta^\dagger)'W_0G_\infty(\theta^\dagger) < G_\infty(\theta)'W_0G_\infty(\theta)$, $\forall\theta \neq \theta^\dagger$.

Assumption A8 (A8): (i) $\widehat{\theta}_{T,S,h}$ and $\theta^\dagger$ are in the interior of Θ; (ii) $g(X_t^{\theta})$ is twice continuously differentiable in the interior of Θ; and (iii) $D^\dagger = E(\partial g_1^{\theta}/\partial\theta|_{\theta=\theta^\dagger})$ exists and is of full rank, p.

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. A3 imposes very mild smoothness requirements on the cumulative distribution function under the null, and is thus easily verified. A4-A7 ensure consistency and asymptotic normality of $\widehat{\theta}_{T,S,h}$, under both hypotheses.

Theorem 1: Let A1-A8 hold. As $T, S \to \infty$, $h \to 0$, $T/S \to 0$, and $Th^2 \to 0$:
(i) Under H_0,

$$V_{T,S,h}^2 \Rightarrow \int_U Z^2(u)\pi(u)du,$$

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

$$\begin{aligned} K(u,u') &= E(\sum_{s=-\infty}^{\infty}(1\{X_1 \leq u\} - F(u,\theta_0))(1\{X_s \leq u\} - F(u,\theta_0))) \\ &\quad +\nabla_\theta F(u,\theta_0)'D^{0\prime}W_0D^0\nabla_\theta F(u,\theta_0))) - 2\nabla_\theta F(u,\theta_0)'(D^{0\prime}W_0D^0)^{-1}D^{0\prime}W^0 \\ &\quad \times \sum_{s=-\infty}^{\infty} E\left((g(X_s) - E(g(X_1)))(1\{X_1 \leq u\} - F(u,\theta_0))\right). \end{aligned} \tag{9}$$

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

$$\lim_{T\to\infty} \Pr(\frac{1}{T}V_{T,S,h}^2 > \varepsilon) = 1.$$

As the estimated parameters are $\sqrt{T}$ consistent, PEE does not vanish asymptotically, but instead enters into the asymptotic covariance kernel (the last two lines in (9) summarize the contribution of PEE to the kernel). Note that in the statement of the theorem above, we require that the simulated sample size grows at a faster rate than the historical sample. We can relax this requirement and still get convergence to a functional of a Gaussian process, although the covariance kernel would be slightly different. However, in order to establish validity of the block bootstrap

[10]Let $g(X_{t,h}^{\theta})_i$ be the $i-th$ element of $g(X_{t,h}^{\theta})$. We require $\sup_{\theta\in\Theta}|g(X_{t,h}^{\theta})_i| \leq D_t$, with $\sup_t E(D_t)^{2r} < \infty$.

under SGMM, we require $T/S \to 0$. Of further note is that by varying the interval of integration, U, over which the specification test is constructed, one can assess the 'goodness' of specification over various regions of the distribution.

3 Bootstrap Critical Values

The limiting distribution of $V^2_{T,S,h}$ is a functional of a Gaussian process with a covariance kernel that reflects both PEE and the time series nature of the data. Thus, critical values cannot be tabulated. In the present context, valid asymptotic critical values can be obtained in three ways. First, one can use the conditional p-value approach of Corradi and Swanson (2002), which extends Hansen's (1996) and Inoue's (2001) results to the case of non vanishing PEE. Second, one can use the subsampling method of Politis, Romano and Wolf (1999). Third, one can use an appropriate block bootstrap procedure. A drawback of the first two approaches is that the simulated (or subsample based) critical values diverge at rate l (where l plays the role of the blocksize length or denotes the subsample size) under the alternative. Thus, we choose to use the third approach.[11] In order to show the first order validity of the block bootstrap in our context, we derive the limiting distribution of appropriately formed bootstrap statistics and show that they coincide with the limiting distribution in Theorem 1. Then, a test with correct asymptotic size and unit asymptotic power can be obtained by comparing the value of the original statistic with bootstrapped critical values.

In the presence of dependent observations, but no PEE, valid bootstrap critical values are straightforwardly provided by an empirical version of the Künsch (1989) block bootstrap (see e.g. Buhlmann (1994), Naik-Nimbalkar and Rajarshi (1994) or Peligrad (1998)).[12] However, in the

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

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

present context we need a bootstrap procedure which properly mimic the contribution of PEE to the covariance kernel. Goncalves and White (2002) show the first order validity of the block bootstrap for QMLE (or $m-$estimators), for the case of dependent and heterogeneous observations. In the sequel we show that, in the stationary case, their results are valid also for SGMM estimators in the exact identification case and when the simulated series sample size grows faster than the historical sample size. While a formal proof is provided in the appendix, it is worthwhile to also give an intuitive explanation of our result. First, if $T/S \to 0$, simulation error is negligible, SGMM is asymptotically equivalent to GMM, and consequently we do not need to bootstrap the simulated series. Second, in the exactly identified case, the bootstrap sample always satisfies the moment conditions, therefore GMM estimators can be treated the same way that QMLE estimators are treated.

In order to implement the appropriate bootstrap statistic, we proceed as follows. At each replication, draw b blocks (with replacement) of length l from the sample X_t where $T = lb$. Thus, the first block is equal to $X_{i+1}, ..., X_{i+l}$, for some $i = 0, 1, ...T - l$, with probability $1/(T - l + 1)$, the second block is equal to $X_{i+1}, ..., X_{i+l}$, for some i, with probability $1/(T - l + 1)$, and so on for all blocks. More formally, let I_k, $k = 1, ..., b$ be *i.i.d.* discrete uniform random variables on $[0, 1, ..., T-l]$, and let $T = bl$. Then, the resampled series, is such that $X_1^*, X_2^*, ..., X_l^*, X_{l+1}^*, ..., X_T^* = X_{I_1+1}, X_{I_1+2}, ..., X_{I_1+l}, X_{I_2}, ..., X_{I_b+l}$, and so a resampled series consists of b blocks that are discrete *i.i.d.* uniform random variables, conditional on the sample. Now, define the simulated SGMM estimator as:

$$\begin{aligned} \widehat{\theta}^*_{T,S,h} &= \arg\min_{\theta\in\Theta} \left(\frac{1}{T}\sum_{t=1}^{T} g(X_t^*) - \frac{1}{S}\sum_{t=1}^{S} g(X_{t,h}^{\theta})\right)' W_T \left(\frac{1}{T}\sum_{t=1}^{T} g(X_t^*) - \frac{1}{S}\sum_{t=1}^{S} g(X_{t,h}^{\theta})\right) \\ &= \arg\min_{\theta\in\Theta} G^*_{T,S,h}(\theta)' W_T G^*_{T,S,h}(\theta), \end{aligned} \tag{10}$$

and note that we do not need to resample the simulated series. The reason is that, as $T/S \to \infty$, simulation error vanishes asymptotically and so there is no need to mimic its contribution to the covariance kernel. Finally, define the bootstrap statistic as:

$$V^{2*}_{T,S,h} = \int_U V^{2*}_{T,S,h}(u)\pi(u)du, \tag{11}$$

smaller first order variance. In addition, the overlapping block boostrap is more efficient than the non overlapping block bootstrap.

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

$$V^*_{T,S,h}(u) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left((1\{X^*_t \leq u\} - 1\{X_t \leq u\}) - \left(F(u, \widehat{\theta}^*_{T,S,h}) - F(u, \widehat{\theta}_{T,S,h}) \right) \right). \qquad (12)$$

The validity of the suggested bootstrap procedure is stated in the subsequent theorem.

Theorem 2: Let A1-A7 hold. As $T, S \to \infty$, $h \to 0$, $T/S \to 0$, $Th^2 \to 0$, $l \to \infty$ and $l/\sqrt{T} \to 0$:

$$P\left(\omega : \sup_{v \in \Re} \left| P^* \left(\int_U V^{2*}_{T,S,h}(u)\pi(u)du \leq v \right) \right. \right.$$

$$\left. \left. - P\left(\int_U \left(V_{T,S,h}(u) - \sqrt{T}E\left(1\{X_t \leq u\} - F(u, \theta^\dagger) \right) \right)^2 \pi(u)du \leq v \right) \right| > \varepsilon \right) \to 0,$$

where P^* denotes the probability law of the resampled series, conditional on the sample.

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

4 Extension to Multifactor Diffusion Models

In the case of multidimensional and/or multifactor diffusion models, knowledge of the drift and of the diffusion matrix no longer implies knowledge of the invariant density. Thus, the test suggested in Section 2 cannot be generalized in a straightforward manner to the multidimensional case. This is quite a severe limitation, as many applications, such as term structure analysis, require multifactor models (see e.g. Day and Singleton (2000,2002)). Therefore, in this section we outline a new test which is based on comparison of the empirical CDF of historical data and the empirical CDF of (model) simulated data, where data are simulated using estimators constructed via application of SGMM. A related test based on the comparison of historical data and simulated data is discussed in Corradi and Swanson (CS: 2003), in the context of real business cycle model evaluation. The test introduced below differs from the CS tests in two main respects. Namely: (i) we are interested in continuous time processes and thus we account for discretization error (CS consider only discrete time models); and (ii) parameters estimators are constructed using both actual and simulated data (in CS, parameters are either calibrated or estimated using only historical data).

In providing an extension to the multidimensional case of the above test of one-dimensional models, a first difficulty lies in the choice of the discrete approximation scheme. In particular, the diffusion process $X(t)$ can be expressed as a function of the driving Brownian motion $W(t)$, in the one-dimensional case. However, in the multidimensional case, where we have $X(t) \in R^p$, note that $X(t)$ cannot in general be expressed as a function of the p driving Brownian motions, but is instead a function of $(W_j(t), \int_0^t W_j(s)dW_i(s))$, $i,j = 1,...p$ (see e.g. Pardoux and Talay (1985), p.30-32). For this reason, simple approximation schemes like the Euler or the Milstein schemes, which do not involve approximation of stochastic integrals, may not be adequate in the multidimensional case. One situation in which the Milstein scheme does straigthforwardly generalize to the multidimensional case is when the diffusion matrix is commutative. Let $\Sigma(X) = \begin{pmatrix} \sigma_1(X) & . & . & \sigma_P(X) \end{pmatrix}$, where $\sigma_i(X)$ is a $p \times 1$ vector, for $i = 1,...,p$. If for all $i,j = 1,...,p$,

$$\begin{pmatrix} \frac{\partial \sigma_j(X)}{\partial X_1} & . & . & \frac{\partial \sigma_j(X)}{\partial X_p} \end{pmatrix} \sigma_i(X) = \begin{pmatrix} \frac{\partial \sigma_i(X)}{\partial X_1} & . & . & \frac{\partial \sigma_i(X)}{\partial X_p} \end{pmatrix} \sigma_j(X),$$

then $\Sigma(X)$ is commutative. It is immediate to see that almost all of the most frequently used stochastic volatility (SV) models violate the commutativity property. The intuitive reason for this is that both the variance of the observable asset and the variance of the (unobservable) variance

process depend only on the volatility process. In this situation, more "sophisticated" approximation schemes are necessary.

In the remainder of this section, we specialize to the case of two-factor stochastic volatility models. Extension to general multidimensional and multifactor models follows directly. Consider:

$$\begin{pmatrix} dX(t) \\ dV(t) \end{pmatrix} = \begin{pmatrix} b_1(X(t),\theta) \\ b_2(V(t),\theta) \end{pmatrix} dt + \begin{pmatrix} \sigma_{11}(V(t),\theta) \\ 0 \end{pmatrix} dW_1(t) + \begin{pmatrix} \sigma_{12}(V(t),\theta) \\ \sigma_{22}(V(t),\theta) \end{pmatrix} dW_2(t), \tag{13}$$

where $W_{1,t}$ and $W_{2,t}$ are independent standard Brownian motions. It is immediate to see that the diffusion in (13) violates the commutativity property. Also, note that most of the popular SV models, such as the square-root model of Heston (1993), the GARCH diffusion model of Nelson (1990), the lognormal model of Hull and White (1987) and the eigenfunction models of Meddahi (2001) can be written as (13) above.

Let

$$b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}, \ \sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ 0 & \sigma_{22} \end{pmatrix}, \tag{14}$$

and define the following generalized Milstein scheme (see eq. (3.3), p.346 in Kloeden and Platen (1999)):

$$\begin{aligned} X^\theta_{(k+1)h} &= X^\theta_{kh} + \widetilde{b}_1(X^\theta_{kh},\theta)h + \sigma_{11}(V^\theta_{kh},\theta)\epsilon_{1,(k+1)h} + \sigma_{12}(V^\theta_{kh},\theta)\epsilon_{2,(k+1)h} \\ &\quad + \frac{1}{2}\sigma_{22}(V^\theta_{kh},\theta)\frac{\partial\sigma_{12}(V^\theta_{kh},\theta)}{\partial V}\epsilon^2_{2,(k+1)h} \\ &\quad + \sigma_{22}(V^\theta_{kh},\theta)\frac{\partial\sigma_{11}(V^\theta_{kh},\theta)}{\partial V}\int_{kh}^{(k+1)h}\left(\int_{kh}^{s} dW_{1,\tau}\right)dW_{2,s} \end{aligned} \tag{15}$$

$$\begin{aligned} V^\theta_{(k+1)h} &= V^\theta_{kh} + \widetilde{b}_2(V^\theta_{kh},\theta)h + \sigma_{22}(V^\theta_{kh},\theta)\epsilon_{2,(k+1)h} \\ &\quad + \frac{1}{2}\sigma_{22}(V^\theta_{kh},\theta)\frac{\partial\sigma_{22}(V^\theta_{kh},\theta)}{\partial V}\epsilon^2_{2,(k+1)h} \end{aligned} \tag{16}$$

where $h^{-1/2}\epsilon_{i,kh} \sim N(0,1)$, $i = 1,2$, $E(\epsilon_{1,kh}\epsilon_{2,mh}) = 0$ for all k and m, and

$$\widetilde{b}(V,\theta) = \begin{pmatrix} \widetilde{b}_1(V,\theta) \\ \widetilde{b}_2(V,\theta) \end{pmatrix} = \begin{pmatrix} b_1(V,\theta) - \frac{1}{2}\sigma_{22}(V,\theta)\frac{\partial\sigma_{12}(V,\theta)}{\partial V} \\ b_2(V,\theta) - \frac{1}{2}\sigma_{22}(V,\theta)\frac{\partial\sigma_{22}(V,\theta)}{\partial V} \end{pmatrix}.$$

The last terms on the RHS of (15) involves stochastic integrals and cannot be explicitly computed. However, they can be approximated, up to an error of order $o(h)$ by (see eq. (3.7), p.347 in Kloeden and Platen (1999)):

$$\int_{kh}^{(k+1)h}\left(\int_{kh}^{s} dW_{1,\tau}\right)dW_{2,s} \approx h\left(\frac{1}{2}\xi_1\xi_2 + \sqrt{\rho_p}\left(\mu_{1,p}\xi_2 - \mu_{2,p}\xi_1\right)\right)$$

$$+\frac{h}{2\pi}\sum_{r=1}^{p}\frac{1}{r}\left(\varsigma_{1,r}\left(\sqrt{2}\xi_2+\eta_{2,r}\right)-\varsigma_{2,r}\left(\sqrt{2}\xi_1+\eta_{1,r}\right)\right),$$

where for $j=1,2$, $\xi_j,\mu_{j,p},\varsigma_{j,r},\eta_{j,r}$ are $iidN(0,1)$ and $\rho_p=\frac{1}{12}-\frac{1}{2\pi^2}\sum_{r=1}^{p}\frac{1}{r^2}$, and p is such that as $h\to 0$, $p\to\infty$. Now, define,

$$Z_{T,S,h}(u)=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(1\{X_t\le u\}-\frac{1}{S}\sum_{t=1}^{S}1\{X_{t,h}^{\widehat{\theta}_{T,S,h}}\le u\}\right), \tag{17}$$

where $X_{t,h}^{\theta}=X_{[Nth/S]}^{\theta}$, $S=Nh$, $X_{t,h}^{\theta}$ is generated using (15) and (16), and $\widehat{\theta}_{T,S,h}$ is as defined in equation (6), except that $X_{t,h}^{\theta}$ is simulated using (15) and (16).

The theorem below requires a strengthened version of Assumption A2 above. Namely:

Assumption A2′ (A2′): $b(\cdot)$ and $\sigma(\cdot)$ (as defined in (13) and (14)) and $\sigma_{ij}(V,\theta)\frac{\partial\sigma_{kl}(V,\theta)}{\partial V}$ are twice continuously differentiable, Lipschitz, with Lipschitz constant independent of θ, and grow at most at a linear rate, uniformly in Θ, for $i,j,k,\iota=1,2$.

Theorem 3: Let A1,A2′ and A3-A8 hold, with $X_{t,h}^{\theta}$ generated according to (15) and (16). As $T,S\to\infty$, $h\to 0$, $T/S\to 0$, $T^2/S\to\infty$, $Sh\to 0$:

(i) Under H_0,

$$Z_{T,S,h}^2\Rightarrow\int_U Z^2(u)\pi(u)du,$$

where Z is a Gaussian process with covariance kernel given by $K(u,u')$ equal to:

$$K(u,u')=E(\sum_{s=-\infty}^{\infty}(1\{X_1\le u\}-F(u,\theta_0))(1\{X_s\le u\}-F(u,\theta_0)))+\mu_f'(\theta_0)D^{0\prime}W_0D^0\mu_f(\theta_0)$$

$$-2\mu_f'(\theta_0)'(D^{0\prime}W_0D^0)^{-1}D^{0\prime}W^0\sum_{s=-\infty}^{\infty}E\left((g(X_s)-E(g(X_1))(1\{X_1\le u\}-F(u,\theta_0))\right), \tag{18}$$

where $\mu_f(\theta_0)=E\left(f(u,\theta_0)\nabla_\theta X_{t,h}^{\theta_0}\right)$.

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

$$\lim_{T\to\infty}\Pr(\frac{1}{T}Z_{T,S,h}^2>\varepsilon)=1.$$

Analogous to that in Theorem 1, the limiting distribution in Theorem 3 is a functional of a zero mean Gaussian process, with a covariance kernel that reflects both PEE and the dependence of the data. In fact, the only difference between the covariance kernel in (9) and that in (18) lies in the contribution of PEE. Note also that we require $Sh\to 0$ instead of $Th^2\to 0$, which is a stronger

requirement on the discrete approximation stepsize, h. Further, we require S to grow at a rate faster than T, but slower than T^2.

The relevant bootstrap statistic is:

$$Z^{2*}_{T,S,h}(u) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left((1\{X^*_t \leq u\} - 1\{X_t \leq u\}) - \frac{1}{S} \sum_{t=1}^{S} \left(1\{X^{\widehat{\theta}^*_{T,S,h}}_{t,h} \leq u\} - 1\{X^{\widehat{\theta}_{T,S,h}}_{t,h} \leq u\} \right) \right), \tag{19}$$

where $X^{\widehat{\theta}^*_{T,S,h}}_{t,h}$ and $X^{\widehat{\theta}_{T,S,h}}_{t,h}$ are constructed as in (15), using $\widehat{\theta}^*_{T,S,h}$ and $\widehat{\theta}_{T,S,h}$ respectively, and where $\widehat{\theta}^*_{T,S,h}$ and $\widehat{\theta}_{T,S,h}$ are defined in (6) and (10). The following theorem then holds.

Theorem 4: Let A1,A2′ and A3-A7 hold, with $X^{\theta}_{t,h}$ generated according to (15) and (16). As $T, S \to \infty$, $h \to 0$, $T/S \to 0$, $T^2/S \to \infty$, $Sh \to 0$, $l \to \infty$ and $l/\sqrt{T} \to 0$:

$$P\left(\omega : \sup_{v \in \Re} \left| P^* \left(\int_U Z^{2*}_{T,S,h}(u)\pi(u)du \leq v \right) \right. \right.$$

$$\left. \left. - P\left(\int_U \left(Z^*_{T,S,h}(u) - \sqrt{T}E\left(1\{X_t \leq u\} - F(u,\theta^\dagger) \right) \right)^2 \pi(u)du \leq v \right) \right| > \varepsilon \right) \to 0.$$

We can thus proceed using the approach discussed below the statement of Theorem 2.

5 Monte Carlo Evidence

In this section we discuss implementation of the diffusion specification test and of the bootstrap procedure outlined in Sections 2 and 3, for the case in which we want to test the null hypothesis that the CDF is a gamma distribution. In particular consider the diffusion below, which is a simplified version of the square root process discussed in Cox, Ingersoll and Ross (1985):[13]

$$dX(t) = ((c_1 - a) - X(t))dt + \sqrt{c_1 X(t)}dW(t), \; c_1 > 0, \text{ and } c_1 - a > 0. \tag{20}$$

From Wong (1964, pp. 264-265), it is known that the stationary density associated with $X(t)$ belongs to the linear exponential (or Pearson) family. The process has a non-central chi-squared

[13]Note that this diffusion is obtained by defining the drift and the variance as in equations (10) and (11) in Wong (1964), setting $\beta = 1$, $c = 0$, $d = c_1$, $e = 0$ in his equation (10), and setting $a = -1$, $b = -a$ in his equation (11). The resulting diffusion is indeed a square root process.

transition density, and the invariant density is a gamma. The invariant density can be written as[14]:

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

It follows that,

$$F(u;a,c_1)=\frac{\int_0^u(\frac{c_1}{2})^{-2(1-a/c_1)}x^{2(1-a/c_1)-1}\exp(-x/(\frac{c_1}{2}))dx}{\Gamma(2(1-a/c_1))}. \tag{21}$$

For the CDF in (21), the first two moments are known (the mean is (c_1-a) and the variance is $\frac{c_1}{2}(c_1-a)$). For details, see Johnson, Kotz and Balakrishnan ((1994), ch. 17). Thus, we can use (non simulated) GMM to estimate the parameters of the CDF. Also, as we have two moment conditions and two parameters (exact identification), there is some $(a^\dagger,c_1^\dagger)$ for which the moment conditions are satisfied under both the null and the alternative hypotheses. Along these lines, define

$$(\widehat{a},\ \widehat{c}_1)'=\arg\min_{a,c_1\in\Theta} G'(\theta)W_TG(\theta), \tag{22}$$

where

$$G'(\theta)=\Big(\ \Big(\tfrac{1}{T}\textstyle\sum_{t=1}^T X_t-(c_1-a)\Big),\ \ \Big(\tfrac{1}{T}\textstyle\sum_{t=1}^T(X_t-\overline{X})^2-\tfrac{c_1}{2}(c_1-a)\Big)\ \Big).$$

and

$$W_T^{-1}=\frac{1}{T}\sum_{t=1}^T f_tf_t'+\frac{1}{T}\sum_{\tau=1}^{l_T}w_\tau\sum_{t=\tau+t}^T(f_tf_{t-\tau}'+f_{t-\tau}f_t')$$

with $f_t=(f_{1t},f_{2t})'$, $f_{1t}=(X_t-\overline{X})$, $f_{2t}=(X_t-\overline{X})^2$, $\overline{f}=\frac{1}{T}\sum_{t=1}^T f_t$, and $w_\tau=1-\frac{\tau}{l_T+1}$. Further, assume that we are interested in testing the null hypothesis that the CDF is as defined in (21). Finally, note 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 the above model, consider the simple case where $c_1=3$, and $a=-3$ (one of the cases considered in the experiments reported on below). 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

[14]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 $\alpha_2=c_1/2$.

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.

Given the discussion in Section 2 above, the statistic used in our simulation exercise is:[15]

$$V_{T,S,h}^2 = V_T^2 = \int_U V_T^2(u)\pi(u)du,$$

where

$$V_T(u) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(1\{X_t \leq u\} - \frac{\int_0^u (\frac{\widehat{c}_1}{2})^{-2(1-\widehat{a}/\widehat{c}_1)} x^{2(1-\widehat{a}/\widehat{c}_1)-1}\exp(-x/(\frac{\widehat{c}_1}{2}))dx}{\Gamma(2(1-\widehat{a}/\widehat{c}_1))}\right),$$

and $\widehat{a}$ and $\widehat{c}_1$ are defined as in (22). The bootstrap statistic is:

$$V_{T,S,h}^{2*} = V_T^{2*} = \int_U V_T^{2*}(u)\pi(u)du,$$

where

$$\begin{aligned} V_T^*(u) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(1\{X_t^* \leq u\} - \frac{\int_0^u (\frac{\widehat{c}_1^*}{2})^{-2(1-\widehat{a}^*/\widehat{c}_1^*)} x^{2(1-\widehat{a}^*/\widehat{c}_1^*)-1}\exp(-x/(\frac{\widehat{c}_1^*}{2}))dx}{\Gamma(2(1-\widehat{a}^*/\widehat{c}_1^*))}\right) \\ &\quad -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(1\{X_t \leq u\} - \frac{\int_0^u (\frac{\widehat{c}_1}{2})^{-2(1-\widehat{a}/\widehat{c}_1)} x^{2(1-\widehat{a}/\widehat{c}_1)-1}\exp(-x/(\frac{\widehat{c}_1}{2}))dx}{\Gamma(2(1-\widehat{a}/\widehat{c}_1))}\right). \end{aligned}$$

In the above expression, X_t^* denotes the pseudo time series, resampled according to the moving block scheme described in Section 3. Also, $\widehat{a}^*$ and $\widehat{c}_1^*$ are defined as:

$$(\widehat{a}^*,\ \widehat{c}_1^*)' = \arg\min_{a,c_1\in\Theta} G^{*\prime}(\theta)W_T G^*(\theta), \tag{23}$$

where

$$G^{*\prime}(\theta) = \left(\ \left(\tfrac{1}{T}\textstyle\sum_{t=1}^{T} X_t^* - (c_1 - a)\right),\ \left(\tfrac{1}{T}\textstyle\sum_{t=1}^{T}(X_t^* - \overline{X}^*)^2 - \tfrac{c_1}{2}(c_1 - a)\right)\ \right).$$

Given the exact identification of the parameters in this example, the bootstrap series satisfies the moment conditions (in the limit) and so no recentering term is required.

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

In our test level experiments, we simulate data according to (20) by using the discrete Milstein approximation scheme given in equation (8).[16] It follows that discrete samples are generated according to:

$$b(X^{\theta}_{(k-1)h},\theta)h = ((c_1 - a) - X_{(k-1)h})h,\ \sigma(X^{\theta}_{(k-1)h},\theta) = \sqrt{c_1 X_{(k-1)h}},\ \sigma(X^{\theta}_{(k-1)h},\theta)' = \frac{1}{2}(X_{(k-1)h})^{-1/2}\sqrt{c_1}. \tag{24}$$

Rejection frequencies are tabulated for $(a, c_1) = \{(-2,2),(-3,3),(-4,4)\}$.[17] For power experiments we generate log of $X(t)$ as an Ornstein-Uhlenbeck process. Namely, we define:

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

so that $X(t)$ has a lognormal CDF. In practice we simulate the approximate discrete trajectories of $Y_{kh} = \log(X_{kh})$ using the floowing Milstein scheme:

$$Y_{kh} = -\theta_1 Y_{(k-1)h}h + \sigma\epsilon_{kh},\ \sqrt{h}\epsilon_{kh} \sim N(0,1),\ X_{kh} = \exp(Y_{kh}).$$

Note that under both hypotheses the DGP is characterized by a diffusion term which is locally, but not globally Lipschitz. In the power experiments, we set $\sigma^2 = \{0.1, 0.5, 1.0\}$ and $\theta = \{0.3, 0.6, 0.9\}$. Summarizing, the null is that X_t has a gamma CDF, while the alternative is that X_t does not have a gamma CDF. In all experiments, parameters are estimated using GMM, given the moment conditions implied under H_0, and V_T^2, $|V_T|$, and $|V_T|^{\sup}$ test statistics are constructed. Further, h is set equal to T^{-1}, samples of $T = \{400, 800, 1200\}$ observations are generated, block lengths of $l = \{2,4,5,8,10,20,25,40,50\}$ are tried, the integration interval U is set equal to $[0,15]$, statistics are formed based on uniform grids of 50 points in U, and critical values are set equal to the 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 Tables 2-4. Recalling that the "nominal" rejection rate is 10%, note that for samples of 400 observations, empirical rejection rates range from around 8% to

[16]Note that in (24) below, paths are only simulated at the "true" parameter values. Thus, we can suppress the dependence of parameters on the simulation trajectory.

[17]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.

15%, while for samples of 1200 observations, the range is approximately 10% to 13%. Interestingly, results seem to be quite robust to the choice of blocklength, l. From Tables 2-4, we see that empirical rejection frequencies under the alternative are also somewhat robust to the choice of blocklength, although finite sample power in some cases increases rather substantially with sample size (see e.g. Table 2, Panels a and b). However, when the variability of the process is $\sigma^2 = 0.5$ or $\sigma^2 = 1.0$, finite sample power is above 0.85 for all sample sizes and cases considered (see Tables 3 and 4).

6 Concluding Remarks

In this paper we have outlined two parametric specification tests for diffusion processes. In the one-dimensional case, we outline a test that is based on a comparison of the empirical distribution of the skeleton of the diffusion and the CDF implied by the specification of the drift and the variance terms, with the latter evaluated at estimated (simulated GMM) parameters. In the multidimensional and/or multifactor case, we outline a related test that is based on a comparison of the empirical distributions of the actual and simulated data. In both cases, the limiting distribution is a functional of Gaussian process with a covariance kernel that reflects data dependence and parameter estimation error. In order to obtain asymptotically valid critical values for the tests, we propose an extension of the empirical version of the block bootstrap which properly captures the contribution of parameter estimation error.

In an illustration, the one-dimensional test is applied to the problem of selecting between two alternative continuous time diffusion models, and a limited number of Monte Carlo experiments are carried out, with results suggesting that the tests hold some promise, even when applied to samples of data as small as 400 observations.

7 Appendix

The proof of Theorem 1 requires the following Lemma.

Lemma A1: Let A1-A2, A4-A8 hold. If as $T, S \to \infty$, $h \to 0$, $T/S \to 0$ and $h^2 T \to 0$, then

$$\sqrt{T}\left(\widehat{\theta}_{T,S,h} - \theta^{\dagger}\right) \xrightarrow{d} N(0, D^{\dagger\prime} W_0 D^{\dagger}),$$

where W_0 and $D^{\dagger}$ are defined in assumptions A5 and A8 respectively.

Proof of Lemma A1: Note that, because of the first order conditions, $\nabla_\theta G_{T,S,h}(\widehat{\theta}_{T,S,h})' W_T G_{T,S,h}(\widehat{\theta}_{T,S,h}) = 0$, then via a mean value expansion around $\theta^{\dagger}$,

$$0 = \nabla_\theta G_{T,S,h}(\widehat{\theta}_{T,S,h})' W_T G_{T,S,h}(\theta^{\dagger}) + \nabla_\theta G_{T,S,h}(\widehat{\theta}_{T,S,h})' W_T \nabla_\theta G_{T,S,h}(\overline{\theta}_{T,S,h}) \left(\widehat{\theta}_{T,S,h} - \theta^{\dagger}\right),$$

where $\overline{\theta}_{T,S,h} \in \left(\widehat{\theta}_{T,S,h}, \theta^{\dagger}\right)$. Thus,

$$\sqrt{T}\left(\widehat{\theta}_{T,S,h} - \theta^{\dagger}\right) = \left(-\nabla_\theta G_{T,S,h}(\widehat{\theta}_{T,S,h})' W_T \nabla_\theta G_{T,S,h}(\overline{\theta}_{T,S,h})\right)^{-1} \nabla_\theta G_{T,S,h}(\widehat{\theta}_{T,S,h})' W_T \sqrt{T} G_{T,S,h}(\theta^{\dagger}). \tag{25}$$

We begin by showing that as $T, S \to \infty$, $h \to 0$, $T/S \to 0$ and $h^2 T \to 0$,

$$\sqrt{T} G_{T,S,h}(\theta^{\dagger}) \xrightarrow{d} N(0, W_0^{-1}). \tag{26}$$

Now, given A1 and A5,

$$\begin{aligned} \sqrt{T} G_{T,S,h}(\theta^{\dagger}) &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (g(X_t) - E(g(X_1))) - \frac{\sqrt{T}}{S} \sum_{t=1}^{S} \left(g(X_{t,h}^{\theta^{\dagger}}) - E(g(X_{1,h}^{\theta^{\dagger}}))\right) \\ &\quad - \sqrt{T}\left(E(g(X_{1,h}^{\theta^{\dagger}})) - E(g(X_1^{\theta^{\dagger}}))\right) - \sqrt{T}\left(E(g(X_1^{\theta^{\dagger}})) - E(g(X_1))\right), \end{aligned} \tag{27}$$

we need to show that the second, third and fourth term on the RHS of (27) approach zero in probability, as $T, S \to \infty$, $h \to 0$, $T/S \to 0$ and $h^2 T \to 0$. By the first order conditions, $\nabla_\theta G_\infty(\theta^{\dagger})' W^{\dagger} G_\infty(\theta^{\dagger}) = 0$ and so, as $\left(\nabla_\theta G_\infty(\theta^{\dagger})' W^{\dagger}\right)^{-1}$ exists, $G_\infty(\theta^{\dagger}) = 0$, where, given A1 and A3,

$$0 = G_\infty(\theta^{\dagger}) = \text{plim}_{S\to\infty, h\to 0} \frac{1}{S} \sum_{t=1}^{S} g(X_{t,h}^{\theta^{\dagger}}) - \text{plim}_{T\to\infty} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} g(X_t) = E(g(X_1^{\theta^{\dagger}})) - E(g(X_1)),$$

thus the last term on the RHS of (27) is zero. Now, given A2 and recalling that $hT^2 \to 0$, the second last term on the RHS of (27) is $o_P(1)$, as a straightforward consequence of Theorem 2.3 in Pardoux and Talay (1985). Given A1, A2, A4 and A6, $\frac{1}{\sqrt{S}} \sum_{t=1}^{S} \left(g(X_{t,h}^{\theta^{\dagger}}) - E(g(X_{1,h}^{\theta^{\dagger}}))\right)$ is bounded

in probability, as it satisfies a central limit theorem, and so for $T = o(S)$ it vanishes to zero. Given A5, the statement in (26) follows straightforwardly. Finally, given A2, and A4-A8, via the uniform law of large numbers for mixing process, as $T, S \to \infty$ and $h \to 0$,

$$\sup_{\theta \in \Theta} |G_{T,S,h}(\theta)' W_T G_{T,S,h}(\theta) - G_\infty' W_0 G_\infty| \xrightarrow{\Pr} 0,$$

and so given unique identifiability, A7, $\left(\widehat{\theta}_{T,S,h} - \theta^\dagger\right) \xrightarrow{\Pr} 0$. Also, given A2, and A4-A7,

$$\sup_{\theta \in \Theta} |\nabla_\theta G_{T,S,h}(\theta)' W_T \nabla G_{T,S,h}(\theta) - D(\theta)' W_0 D(\theta)| \xrightarrow{\Pr} 0,$$

and so, given that $\left(\widehat{\theta}_{T,S,h} - \theta^\dagger\right) \xrightarrow{\Pr} 0$,

$$\nabla_\theta G_{T,S,h}(\widehat{\theta}_{T,S,h})' W_T \nabla G_{T,S,h}(\overline{\theta}_{T,S,h}) \xrightarrow{\Pr} D^{\dagger\prime} W_0 D^\dagger,$$

where $D(\theta^\dagger) = D^\dagger$. The statement in the Lemma then follows.

Proof of Theorem 1: **(i)** Recall that $\theta^\dagger = \theta_0$, under the null. 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 $V_{T,S,h}(.)$ 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

$$\begin{aligned} V_{T,S,h}(u) &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left((1\{X_t \leq u\} - F(u, \theta_0)) - (F(u, \widehat{\theta}_{T,S,h}) - F(u, \theta_0)) \right) \\ &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (1\{X_t \leq u\} - F(u, \theta_0)) - \nabla_\theta F(u, \overline{\theta}_{T,S,h})' \sqrt{T} (\widehat{\theta}_{T,S,h} - \theta_0) \\ &= I_{1T}(u) + I_{2T}(u), \end{aligned}$$

where $\overline{\theta}_{T,S,h} \in (\widehat{\theta}_{T,S,h}, \theta_0)$. Recalling A1, A4-A8, 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, given Lemma A1,

$$V_1(u) = E(\sum_{s=-\infty}^{\infty} (1\{X_1 \leq u\} - F(u, \theta_0))(1\{X_s \leq u\} - F(u, \theta_0))),$$

$$V_2(u) = \nabla_\theta F(u,\theta_0)' D^{0\prime} W_0 D^0 \nabla_\theta F(u,\theta_0))),$$

$$C(u) = -\nabla_\theta F(u,\theta_0)'(D^{0\prime}W_0D^0)^{-1}D^{0\prime}W_0 \sum_{s=-\infty}^{\infty} E\left((g(X_s) - E(g(X_1))(1\{X_1 \le u\} - F(u,\theta_0))\right).$$

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

$$\begin{pmatrix} V_{T,S,h}(u) \\ V_{T,S,h}(u') \end{pmatrix} \xrightarrow{d} N(0, \begin{pmatrix} K(u,u) & K(u,u') \\ K(u,u') & K(u',u') \end{pmatrix},$$

where $K(u,u')$ is as defined in (9). As U is compact in $\Re$, (and so totally bounded) in order to show weak convergence, we need to show that $V_{T,S,h}(u)$ is stochastically equicontinuous on U,(see e.g. Pollard 1990, section 10), that is

$$\text{limsup}_{T,S\to\infty,h\to 0} \Pr\left(\sup_u \sup_{u':|u-u'|<\delta} |V_{T,S,h}(u) - V_{T,S,h}(u')| > \varepsilon\right) = 0 \tag{28}$$

Now (28) will follow if we can show that

$$\text{limsup}_{T\to\infty} \Pr\left(\sup_u \sup_{u':|u-u'|<\delta} \left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left((1\{X_t \le u\} - F(u,\theta_0)) - (1\{X_t \le u'\} - F(u',\theta_0))\right)\right| > \varepsilon/2\right) = 0 \tag{29}$$

$$\text{limsup}_{T,S\to\infty,h\to 0} \Pr\left(\sup_u \sup_{u':|u-u'|<\delta} \left|(\nabla_\theta F(u,\overline{\theta}_T)' - \nabla_\theta F(u',\overline{\theta}_{T,S,h})')\sqrt{T}(\widehat{\theta}_{T,S,h} - \theta_0)\right| > \varepsilon/2\right) = 0 \tag{30}$$

with $\overline{\theta}_{T,S,h} \in (\widehat{\theta}_{T,S,h}, \theta_0)$. We begin by considering (30); by a mean value expansion, we have,

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

Now

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

given that $\sqrt{T}(\widehat{\theta}_{T,S,h} - \theta_0) = O_p(1)$, and A3 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 (29).

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

$$\sup_{t\le T, T\ge 1} \sup_{u\in U} |m_t(u)| = 1 \tag{31}$$

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

$$\sup_{t\leq T, T\geq 1}\left(E\left(\sup_{u}\sup_{u':|u-u'|<\delta}(|m_t(u)-m_t(u')|^p)\right)\right)^{1/p},\ p\geq 2$$

$$\leq \sup_{u}\sup_{u':|u-u'|<\delta}|F(u,\theta_0)-F(u',\theta_0)|+\left(E\left(\sup_{u}\sup_{u':|u-u'|<\delta}(1\{u\leq X_t\leq u'\})^p\right)\right)^{1/p}$$

$$\leq \sup_{u\times U}|\nabla_u F(u,\theta_0)|\delta+\left(\sup_{u\in U}\int_u^{u+\delta}f(x)dx\right)^{1/p}\leq 2C\delta, \tag{32}$$

given A3. 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 (31) and as shown by Andrews (1993, pp. 201), (iii) is implied by (32). Given stochastic equicontinuity and convergence of the finite dimensional distributions, it follows that $V_{T,S,h}(\cdot)\Rightarrow Z(\cdot)$ where Z is the Gaussian process with the covariance kernel defined in (9). The statement then follows from the continuous mapping theorem.

(ii)

$$S_{T,S,h}(u)=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left((1\{X_t\leq u\}-F_0(u))-\sqrt{T}(F(u,\widehat{\theta}_{T,S,h})-F_0(u))\right).$$

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, it diverges at rate $\sqrt{T}$ for all u in a subset of positive Lebesgue measure. The statement in (ii) then follows.

Hereafter, P^* denotes the probability law of the resampled series X_t^*, and d^* denotes P^*-convergence in distribution, conditional on the sample. Also, with the notation $o_{P^*}(1)$ $\Pr-P$, we mean a term which approaches zero in P^*-probability, conditional on the sample and for all sample, but a subset of measure approaching zero.

The proof of Theorem 2 requires the following Lemma.

Lemma A2: Let A1-A2, A4-A8 hold. If as $T,S\to\infty$, $h\to 0$, $T/S\to 0$, $h^2T\to 0$, $l\to\infty$, and $l/\sqrt{T}\to 0$, then

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

Proof of Lemma A2: Via a mean value expansion around $\theta^\dagger$,

$$0=\nabla_\theta G^*_{T,S,h}(\widehat{\theta}^*_{T,S,h})'W_T G^*_{T,S,h}(\theta^\dagger)+\nabla_\theta G^*_{T,S,h}(\widehat{\theta}^*_{T,S,h})'W_T\nabla_\theta G^*_{T,S,h}(\overline{\theta}^*_{T,S,h})\left(\widehat{\theta}^*_{T,S,h}-\theta^\dagger\right),$$

where $\overline{\theta}^*_{T,S,h} \in \left(\widehat{\theta}^*_{T,S,h}, \theta^\dagger\right)$. Thus,

$$\sqrt{T}\left(\widehat{\theta}^*_{T,S,h} - \theta^\dagger\right) = \left(-\nabla_\theta G^*_{T,S,h}(\widehat{\theta}^*_{T,S,h})' W_T \nabla_\theta G^*_{T,S,h}(\overline{\theta}^*_{T,S,h})\right)^{-1} \nabla_\theta G^*_{T,S,h}(\widehat{\theta}^*_{T,S,h})' W_T \sqrt{T} G^*_{T,S,h}(\theta^\dagger),$$

and so given (25),

$$\begin{aligned}
&\sqrt{T}\left(\widehat{\theta}^*_{T,S,h} - \widehat{\theta}_{T,S,h}\right) \\
= \;& \left(-\nabla_\theta G^*_{T,S,h}(\widehat{\theta}^*_{T,S,h})' W_T \nabla_\theta G^*_{T,S,h}(\overline{\theta}^*_{T,S,h})\right)^{-1} \nabla_\theta G^*_{T,S,h}(\widehat{\theta}^*_{T,S,h})' W_T \sqrt{T}\left(G^*_{T,S,h}(\theta^\dagger) - G_{T,S,h}(\theta^\dagger)\right) \\
&+ \Big(\left(-\nabla_\theta G^*_{T,S,h}(\widehat{\theta}^*_{T,S,h})' W_T \nabla_\theta G^*_{T,S,h}(\overline{\theta}^*_{T,S,h})\right)^{-1} \nabla_\theta G^*_{T,S,h}(\widehat{\theta}^*_{T,S,h})' W_T \\
&- \left(-\nabla_\theta G_{T,S,h}(\widehat{\theta}_{T,S,h})' W_T \nabla_\theta G_{T,S,h}(\overline{\theta}_{T,S,h})\right)^{-1} \nabla_\theta G_{T,S,h}(\widehat{\theta}_{T,S,h})' W_T \Big) \sqrt{T} G_{T,S,h}(\theta^\dagger).
\end{aligned}$$

We begin by showing that $\sqrt{T}\left(G^*_{T,S,h}(\theta^\dagger) - G_{T,S,h}(\theta^\dagger)\right)$ has the same limiting distribution as $\sqrt{T}G_{T,S,h}(\theta^\dagger)$, conditionally on the samples, and for all samples except a set of probability measure approaching zero. Note that,

$$\sqrt{T}\left(G^*_{T,S,h}(\theta^\dagger) - G_\infty(\theta^\dagger)\right) = \sqrt{T}\left(G^*_{T,S,h}(\theta^\dagger) - G_{T,S,h}(\theta^\dagger)\right) + \sqrt{T}\left(G_{T,S,h}(\theta^\dagger) - G_\infty(\theta^\dagger)\right),$$

where for $h^2 T \to 0$, the last term on the RHS above is $o_P(1)$, by the same argument used in the proof of Lemma A1. Also, given A1-A4 and given that $T/S \to 0$ (i.e. the simulation error is vanishing), by the same argument as in the proof of Lemma A1,

$$\sqrt{T}\left(G^*_{T,S,h}(\theta^\dagger) - G_{T,S,h}(\theta^\dagger)\right) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(g(X^*_t) - g(X_t)\right) + o_{P^*}(1). \tag{33}$$

Thus, given A1 and A6, by Theorem 3.5 in Künsch (1989), the RHS of (33) has the same limiting distribution as $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(g(X_t) - E(g(X_1))\right)$, conditionally on the samples, and for all samples except a set of probability measure approaching zero. Therefore, $\sqrt{T}\left(G^*_{T,S,h}(\theta^\dagger) - G_{T,S,h}(\theta^\dagger)\right) \xrightarrow{d^*} N(0, W_0)$, $\Pr - P$, where d^* denotes P^*-convergence in distribution. We now need to show that

$$\widehat{\theta}^*_{T,S,h} - \theta^* = o_{P^*}(1), \Pr - P. \tag{34}$$

First, note that

$$G^*_{T,S,h}(\theta) - G_\infty(\theta) = \left(G^*_{T,S,h}(\theta) - G_{T,S,h}(\theta)\right) + \left(G_{T,S,h}(\theta) - G_\infty(\theta)\right) = o_{P^*}(1) + o(1) \;\Pr - P, \tag{35}$$

uniformly in θ. In fact, as we do note resample the simulated series, given A1, A4, A6, the first term on the RHS of (35) is $o_{P^*}(1) + o(1)$ $\Pr - P$ because of Lemma A2 in Goncalves and White

(2002), the second is $o_P(1)$ by the same argument as in the proof of Lemma A1. Therefore, given A5,

$$\sup_{\theta\in\Theta}\left|G^*_{T,S,h}(\theta)'W_TG^*_{T,S,h}(\theta)-G_\infty'W_0G_\infty\right|\xrightarrow{\Pr^*}0,\ \Pr-P,$$

and, given A7, (34) follows. By a similar argument, given (34), it also follows that,

$$\Big(\left(-\nabla_\theta G^*_{T,S,h}(\widehat{\theta}^*_{T,S,h})'W_T\nabla_\theta G^*_{T,S,h}(\overline{\theta}^*_{T,S,h})\right)^{-1}\nabla_\theta G^*_{T,S,h}(\widehat{\theta}^*_{T,S,h})'W_T$$
$$-\left(-\nabla_\theta G_{T,S,h}(\widehat{\theta}_{T,S,h})'W_T\nabla_\theta G_{T,S,h}(\overline{\theta}_{T,S,h})\right)^{-1}\nabla_\theta G_{T,S,h}(\widehat{\theta}_{T,S,h})'W_T\Big)\xrightarrow{\Pr^*}0,\ \Pr-P,$$

and, as $\sqrt{T}G_{T,S,h}(\theta^\dagger)=O_P(1)$, the statement in the lemma follows.

Proof of Theorem 2:

$$V^*_{T,S,h}(u)=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(1\{X^*_t\le u\}-1\{X_t\le u\})-\nabla_\theta F(u,\overline{\theta}^*_{T,S,h})\sqrt{T}\left(\widehat{\theta}^*_{T,S,h}-\widehat{\theta}_{T,S,h}\right). \tag{36}$$

Now, given A1 and A3, by the empirical process version of the block bootstrap of Naik-Nimbalkar and Rajarshi (1994, Theorem 2.1), $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(1\{X^*_t\le u\}-1\{X_t\le u\})$ has the same limiting distribution, as a process over U, as $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(1\{X_t\le u\}-F(u,\theta^\dagger)\right)$. Also, given A3 and recalling (34), $\nabla_\theta F(u,\overline{\theta}^*_{T,S,h})-\nabla_\theta F(u,\theta^\dagger)=o_{P^*}(1)$ $\Pr-P$. The statement in the theorem then follows from Lemma A2.

Proof of Theorem 3: **(i)** Recall that $\theta^\dagger=\theta_0$, under the null. Now, for any given $u\in U$,

$$\begin{aligned}
Z_{T,S,h}(u) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}1\{X_t\le u\}-\frac{\sqrt{T}}{S}\sum_{t=1}^{S}1\left\{X_t^{\theta_0}\le u-\left(X_{t,h}^{\widehat{\theta}_{T,S,h}}-X_t^{\theta_0}\right)\right\}\\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(1\{X_t\le u\}-F(u,\theta_0))-\frac{\sqrt{T}}{S}\sum_{t=1}^{S}\left(1\left\{X_t^{\theta_0}\le u-\left(X_{t,h}^{\widehat{\theta}_{T,S,h}}-X_t^{\theta_0}\right)\right\}\right.\\
&\quad \left.-F\left(u-\left(X_{t,h}^{\widehat{\theta}_{T,S,h}}-X_t^{\theta_0}\right),\theta_0\right)\right)\\
&\quad -\frac{\sqrt{T}}{S}\sum_{t=1}^{S}\left(F\left(u-\left(X_{t,h}^{\widehat{\theta}_{T,S,h}}-X_t^{\theta_0}\right),\theta_0\right)-F(u,\theta_0)\right).
\end{aligned}$$

We need to show that,

$$\begin{aligned}
&\frac{\sqrt{T}}{S}\sum_{t=1}^{S}\left(1\left\{X_t^{\theta_0}\le u-\left(X_{t,h}^{\widehat{\theta}_{T,S,h}}-X_t^{\theta_0}\right)\right\}-F\left(u-\left(X_{t,h}^{\widehat{\theta}_{T,S,h}}-X_t^{\theta_0}\right),\theta_0\right)\right)\\
=\ &\frac{\sqrt{T}}{S}\sum_{t=1}^{S}\left(1\left\{X_t^{\theta_0}\le u\right\}-F(u,\theta_0)\right)+o_P(1),
\end{aligned} \tag{37}$$

where the $o_P(1)$ term holds uniformly in u. In order to show (37), we first need to show that,

$$\sup_{t\leq S}\left|X_{t,h}^{\widehat{\theta}_{T,S,h}}-X_t^{\theta_0}\right|=o_P(1). \tag{38}$$

First, note that

$$\sup_{t\leq S}\left|X_{t,h}^{\widehat{\theta}_{T,S,h}}-X_t^{\theta_0}\right|\leq\sup_{t\leq S}\left|X_{t,h}^{\theta_0}-X_t^{\theta_0}\right|+\sup_{t\leq S}\left|X_{t,h}^{\widehat{\theta}_{T,S,h}}-X_{t,h}^{\theta_0}\right|. \tag{39}$$

As for the first term on the RHS of (39),

$$\begin{aligned}\Pr\left(\sup_{t\leq S}\left|X_{t,h}^{\theta_0}-X_t^{\theta_0}\right|>\varepsilon\right) &\leq \sum_{t=1}^{S}\Pr\left(\left|X_{t,h}^{\theta_0}-X_t^{\theta_0}\right|>\varepsilon\right)\leq\frac{1}{\varepsilon^2}\sum_{t=1}^{S}E\left(\left|X_{t,h}^{\theta_0}-X_t^{\theta_0}\right|^2\right)\\ &= \frac{S}{\varepsilon^2}E\left(\left|X_{1,h}^{\theta_0}-X_1^{\theta_0}\right|^2\right)\leq C\frac{Sh}{\varepsilon^2}\to 0 \text{ as } Sh\to 0,\end{aligned} \tag{40}$$

where the second last equality on the RHS of (40) follows from A1 (stationarity-ergodicity of the discrete approximation for all h), while the last inequality, given A2′ and (15),(16) follows from Theorem 10.3.5 in Kloeden and Platen (1999). As for the second term on the RHS of (39),

$$\sup_{t\leq S}\left|X_{t,h}^{\widehat{\theta}_{T,S,h}}-X_{t,h}^{\theta_0}\right|\leq\sup_{t\leq S}\left|\nabla'_\theta X_{t,h}^{\overline{\theta}_{T,S,h}}\left(\widehat{\theta}_{T,S,h}-\theta_0\right)\right|,$$

where, $\overline{\theta}_{T,S,h}\in\left(\widehat{\theta}_{T,S,h},\theta_0\right)$, and, given Lemma A1, $\left(\widehat{\theta}_{T,S,h}-\theta_0\right)=O_P(T^{-1/2})$. Further, given the domination condition in A6,

$$\begin{aligned}\Pr\left(\sup_{t\leq S}T^{-1/2}\left|\nabla'_\theta X_{t,h}^{\overline{\theta}_{T,S,h}}\right|>\varepsilon\right)&\leq\sum_{t=1}^{S}\Pr\left(T^{-1/2}\left|\nabla'_\theta X_{t,h}^{\overline{\theta}_{T,S,h}}\right|>\varepsilon\right)\\ &\leq\sum_{t=1}^{S}\Pr\left(T^{-1/2}D_t>\varepsilon\right)\leq\frac{1}{\varepsilon^4}\frac{S}{T^2}E(D_t^4)\to 0 \text{ if } \frac{S}{T^2}\to 0.\end{aligned}$$

Then, (37) follows because of the stochastic equicontinuity in u of $\frac{1}{\sqrt{S}}\sum_{t=1}^{S}\left(1\left\{X_t^{\theta_0}\leq u\right\}-F(u,\theta_0)\right)$. Thus,

$$\begin{aligned}Z_{T,S,h}(u) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(1\{X_t\leq u\}-F(u,\theta_0))+\frac{\sqrt{T}}{S}\sum_{t=1}^{S}\left(1\left\{X_t^{\theta_0}\leq u\right\}-F(u,\theta_0)\right)\\ &\quad-\frac{\sqrt{T}}{S}\sum_{t=1}^{S}\left(F\left(u-\left(X_{t,h}^{\widehat{\theta}_{T,S,h}}-X_t^{\theta_0}\right),\theta_0\right)-F(u,\theta_0)\right)+o_P(1),\end{aligned} \tag{41}$$

with the $o_P(1)$ term holding uniformly in u. The second term on the RHS of (41) is $o_P(1)$, for $T/S\to 0$, by the same argument as that used in the proof of Theorem 1. As for the third term on the RHS of (41), it can be written as:

$$-\frac{1}{S}\sum_{t=1}^{S}f\left(u-\left(X_{t,h}^{\overline{\theta}_{T,S,h}}-X_t^{\theta_0}\right),\theta_0\right)\nabla'_\theta X_{t,h}^{\overline{\theta}_{T,S,h}}\sqrt{T}\left(\widehat{\theta}_{T,S,h}-\theta_0\right),$$

where f denotes the derivative of F with respect to its argument. Given A3, and recalling (38), and Lemma A1, $\frac{1}{S}\sum_{t=1}^{S} f\left(u-\left(X_{t,h}^{\overline{\theta}_{T,S,h}}-X_t^{\theta_0}\right),\theta_0\right)\nabla_\theta X_{t,h}^{\overline{\theta}_{T,S,h}} = E\left(f(u,\theta_0)\nabla'_\theta X_{t,h}^{\theta_0}\right)+o_P(1) = \mu'_f(\theta_0+o_P(1)$. Thus,

$$Z_{T,S,h}(u)=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(1\{X_t\le u\}-F(u,\theta_0)\right)-\mu'_f(\theta_0)\sqrt{T}\left(\widehat{\theta}_{T,S,h}-\theta_0\right).$$

The covariance kernel is thus as given in (18), given Lemma A1, and the final result is an immediate consequence of the continuous mapping theorem.

(ii) By the same argument as that used in part **(i)**,

$$\begin{aligned}Z_{T,S,h}(u) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(1\{X_t\le u\}-F_0(u)\right)-\frac{\sqrt{T}}{S}\sum_{t=1}^{S}\left(F\left(u-\left(X_{t,h}^{\widehat{\theta}_{T,S,h}}-X_t^{\theta_0}\right),\theta_0\right)-F(u,\theta_0)\right)\\ &\quad -\sqrt{T}\left(F(u,\theta_0)-F_0(u)\right)+o_P(1).\end{aligned}$$

The result then follows by the same arguments as those used in the proof of Theorem 1.

Proof of Theorem 4:

$$\begin{aligned}Z^*_{T,S,h}(u) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(1\{X_t^*\le u\}-1\{X_t\le u\}\right)-\frac{\sqrt{T}}{S}\sum_{t=1}^{S}\left(1\{X_{t,h}^{\widehat{\theta}^*_{T,S,h}}\le u\}-1\{X_{t,h}^{\widehat{\theta}_{T,S,h}}\le u\}\right)\\ &= I_{1,T}-I_{2,T,S,h}.\end{aligned}$$

Now,

$$\begin{aligned}I_{2,T,S,h} &= \frac{\sqrt{T}}{S}\sum_{t=1}^{S}\left(1\left\{X_t^{\theta_0}\le u-\left(X_{t,h}^{\widehat{\theta}^*_{T,S,h}}-X_t^{\theta_0}\right)\right\}-F\left(u-\left(X_{t,h}^{\widehat{\theta}^*_{T,S,h}}-X_t^{\theta_0}\right),\theta_0\right)\right)\\ &\quad -\frac{\sqrt{T}}{S}\sum_{t=1}^{S}\left(1\left\{X_t^{\theta_0}\le u-\left(X_{t,h}^{\widehat{\theta}_{T,S,h}}-X_t^{\theta_0}\right)\right\}-F\left(u-\left(X_{t,h}^{\widehat{\theta}_{T,S,h}}-X_t^{\theta_0}\right),\theta_0\right)\right)\\ &\quad +\frac{\sqrt{T}}{S}\sum_{t=1}^{S}\left(F\left(u-\left(X_{t,h}^{\widehat{\theta}^*_{T,S,h}}-X_t^{\theta_0}\right),\theta_0\right)-F\left(u-\left(X_{t,h}^{\widehat{\theta}_{T,S,h}}-X_t^{\theta_0}\right),\theta_0\right)\right). \quad (42)\end{aligned}$$

By a similar argument to that used in the proof of Theorem 3, and recalling that $T/S\to 0$, the first two terms on the RHS of (42) are $o_P(1)$, uniformly in u. As for the last term on the RHS of (42), it can be written as:

$$\begin{aligned}&\frac{1}{S}\sum_{t=1}^{S} f\left(u-\left(X_{t,h}^{\overline{\theta}^*_{T,S,h}}-X_t^{\theta_0}\right),\theta_0\right)\nabla'_\theta X_{t,h}^{\overline{\theta}^*_{T,S,h}}\sqrt{T}\left(\widehat{\theta}^*_{T,S,h}-\theta_0\right)\\ &-\frac{1}{S}\sum_{t=1}^{S} f\left(u-\left(X_{t,h}^{\overline{\theta}_{T,S,h}}-X_t^{\theta_0}\right),\theta_0\right)\nabla'_\theta X_{t,h}^{\overline{\theta}_{T,S,h}}\sqrt{T}\left(\widehat{\theta}_{T,S,h}-\theta_0\right)\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{S}\sum_{t=1}^{S} f\left(u - \left(X_{t,h}^{\overline{\theta}^*_{T,S,h}} - X_t^{\theta_0}\right), \theta_0\right) \nabla'_\theta X_{t,h}^{\overline{\theta}^*_{T,S,h}} \sqrt{T}\left(\widehat{\theta}^*_{T,S,h} - \widehat{\theta}_{T,S,h}\right) \\
&\quad + \frac{1}{S}\sum_{t=1}^{S}\left(f\left(u - \left(X_{t,h}^{\overline{\theta}^*_{T,S,h}} - X_t^{\theta_0}\right), \theta_0\right) \nabla'_\theta X_{t,h}^{\overline{\theta}^*_{T,S,h}} - \sum_{t=1}^{S} f\left(u - \left(X_{t,h}^{\overline{\theta}_{T,S,h}} - X_t^{\theta_0}\right), \theta_0\right) \nabla'_\theta X_{t,h}^{\overline{\theta}_{T,S,h}}\right) \\
&\quad \times \sqrt{T}\left(\widehat{\theta}_{T,S,h} - \theta_0\right) \\
&= \frac{1}{S}\sum_{t=1}^{S} f\left(u - \left(X_{t,h}^{\overline{\theta}^*_{T,S,h}} - X_t^{\theta_0}\right), \theta_0\right) \nabla'_\theta X_{t,h}^{\overline{\theta}^*_{T,S,h}} \sqrt{T}\left(\widehat{\theta}^*_{T,S,h} - \widehat{\theta}_{T,S,h}\right) + o_{P^*}(1), \ \Pr - P,
\end{aligned}
$$

the desired result then follows from the same arguments as those used in the proof of Theorem 2.

8 References

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

Aït-Sahalia, Y., (1999), Transition Densities for Interest Rate and Others Nonlinear Diffusions, *Journal of Finance*, LIV, 1361-1395.

Aït-Sahalia, Y., (2002), Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed Form Approximation Approach, *Econometrica,* 70, 223-262.

Altissimo, F., F. Fornari and Mele, (2002), Testing the Closeness of Conditional Densities by Simulated Nonparametric Methods, Working Paper, London School of Economics.

Altissimo, F., A. Mele and F. Fornari, (2003), Simulated Nonparametric Estimation of Continuous Time Models of Asset Prices and Returns, Working Paper, London School of Economics.

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.

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

Chen, X., L.P. Hansen, and J. Scheinkman, (2000), Principal Components and the Long Run, Working Paper, New York University.

Conley, T.G., E.G.J. Luttmer, L.P. Hansen and J.A. Scheinkman (1997), Short-Term Interest Rates as Subordinated Diffusions, *Review of Financial Studies*, 10, 525-577.

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

Corradi, V. and N.R. Swanson, (2003), Evaluation of Dynamic Stochastic General Equilibrium Models Bases on Distributional Comparison of Simulated and Historical Data, Working Paper, Rutgers University and Queen Mary, University of London.

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

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.

Gonçalves, S. and H. White, (2002), Maximum Likelihood and the Bootstrap for Dynamic Nonlinear Models, Working Paper, CRDE, Cirano, University of Montreal.

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.

Hall, A.R., and A. Inoue, (2003), The Large Sample Behavior of the Generalized Method of Moments Estimator in Misspecified Models, *Journal of Econometrics,* 361-394.

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

Heston, S.L., (1993), A Closed Form Solution for Option with Stochastic Volatility with Applications to Bond and Currency Options, *Review of Financial Studies*, 6, 327-344.

Hong, Y., (2001), Evaluation of Out of Sample Probability Density Forecasts with Applications to S&P 500 Stock Prices, Working Paper, Cornell University.

Hong, Y.M., and H. Li, (2002) Nonparametric Specification Testing for Continuous Time Models with Applications to Term Structure Interest Rates, Working Paper, Cornell University.

Hong, Y.M., H. Li, and F. Zhao, (2002) Out of Sample Performance of Spot Interest Rate Models, Working Paper, Cornell University.

Hll, J., and A. White, (1987), The Pricing of Options on Assets with Stochastic Volatility, *Journal of Finance*, XLII, 281-300.

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

Inoue, A., and M. Shintani, (2001), Bootstrapping GMM Estimators for Time Series, Working Paper, North Carolina State University and Vanderbilt 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.

Kloeden, P.E., and E. Platen, (1999), *Numerical Solution of Stochastic Differential Equations*, Springer and Verlag, New York.

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

Meddahi, N., (2001), An Eigenfunction Approach for Volatility Modeling, Working Paper, University of Montreal.

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.

Nelson, D.B., (1990), ARCH as Diffusion Approximations, *Journal of Econometrics*, 45, 7-38.

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, (1994a), The Stationary Bootstrap, *Journal of the American Statistical Association,* 89, 1303-1313.

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

Politis, D.N., J.P. Romano, and M. Wolf, (1999), *Subsampling*, Springer and Verlag, New York.

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.

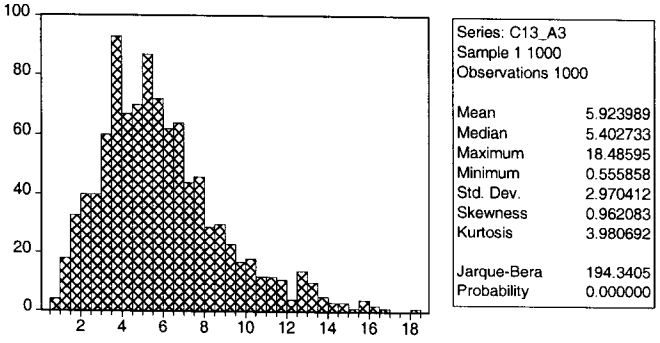
Stramer, O. and R.L. Tweedie, (1997) Geometric and Subgeometric Convergence of Diffusions with Given Stationary Distribution and their Discretization, Working Paper, University of Iowa and Colorado State University.

Thompson, S.B., (2002), Evaluating the Goodness of Fit of Conditional Distributions with an Application to the Affine Term Structure, Working Paper, Harvard University.

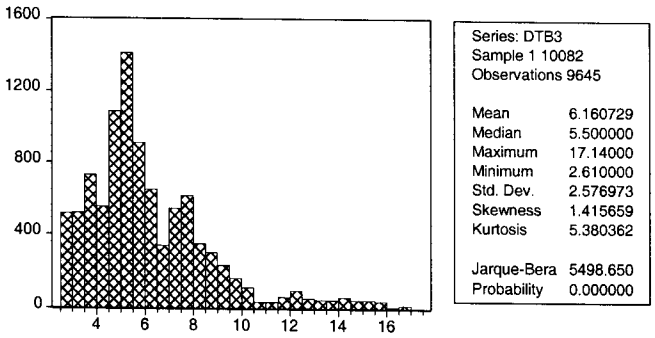
Wong, E., (1964), The Construction of a Class of Stationary Markov Processes, in *Sixteenth Symposia in Applied Mathematics: Stochastic Processes in Mathematical Physics and Engineering*, ed. by R. Bellman, American Mathematical Society, Providence, RI.

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. %)

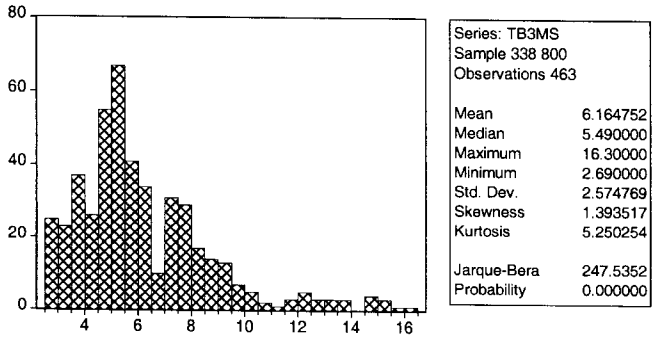


Table 1: Diffusion Specification Test Rejection Frequencies: Empirical Level*
Data Generated Under H_0 With (c_1, a)=$\{(2,-2),(3,-3),(4,-4)\}$

l	T=400 observations			T=800 observations			T=1200 observations		
	V_T^2	$\lvert V_T\rvert$	$\lvert V_T\rvert^{sup}$	V_T^2	$\lvert V_T\rvert$	$\lvert V_T\rvert^{sup}$	V_T^2	$\lvert V_T\rvert$	$\lvert V_T\rvert^{sup}$
Panel a: $c_1 = 2$, $a = -2$									
2	0.086	0.096	0.092	0.114	0.122	0.124	0.130	0.144	0.132
3	0.088	0.092	0.090	0.096	0.106	0.118	0.112	0.136	0.118
5	0.080	0.096	0.098	0.100	0.114	0.114	0.132	0.134	0.118
8	0.084	0.098	0.090	0.104	0.114	0.114	0.112	0.124	0.126
10	0.078	0.086	0.088	0.106	0.110	0.130	0.122	0.128	0.122
20	0.086	0.096	0.088	0.096	0.096	0.112	0.116	0.126	0.128
25	0.086	0.096	0.092	0.112	0.108	0.104	0.118	0.122	0.126
40	0.094	0.100	0.094	0.096	0.106	0.114	0.114	0.124	0.118
50	0.096	0.108	0.112	0.098	0.106	0.112	0.128	0.118	0.130
Panel b: $c_1 = 3$, $a = -3$									
2	0.140	0.134	0.128	0.148	0.142	0.118	0.116	0.122	0.094
3	0.124	0.124	0.128	0.130	0.134	0.130	0.108	0.120	0.094
5	0.144	0.144	0.126	0.134	0.136	0.122	0.112	0.112	0.108
8	0.130	0.122	0.120	0.132	0.126	0.116	0.108	0.120	0.102
10	0.136	0.134	0.132	0.148	0.142	0.122	0.110	0.110	0.104
20	0.126	0.124	0.132	0.146	0.142	0.118	0.114	0.110	0.100
25	0.134	0.134	0.138	0.140	0.134	0.136	0.112	0.114	0.096
40	0.128	0.128	0.134	0.138	0.142	0.116	0.112	0.102	0.102
50	0.144	0.146	0.146	0.144	0.148	0.126	0.112	0.108	0.096
Panel c: $c_1 = 4$, $a = -4$									
2	0.120	0.124	0.122	0.114	0.132	0.112	0.108	0.114	0.112
3	0.110	0.124	0.114	0.120	0.122	0.110	0.108	0.108	0.110
5	0.120	0.130	0.122	0.106	0.104	0.106	0.112	0.108	0.106
8	0.114	0.122	0.120	0.122	0.122	0.102	0.106	0.102	0.106
10	0.128	0.126	0.130	0.112	0.116	0.112	0.108	0.112	0.112
20	0.136	0.138	0.116	0.130	0.130	0.112	0.104	0.116	0.104
25	0.148	0.150	0.124	0.128	0.132	0.112	0.112	0.116	0.120
40	0.138	0.140	0.124	0.142	0.140	0.116	0.114	0.124	0.112
50	0.148	0.156	0.156	0.128	0.146	0.120	0.104	0.118	0.106

(*) Notes: The first column of numerical entries are block lengths used in the construction of bootstrap critical values. All other numerical entries are rejection frequencies based on application of the one-dimensional diffusion specification test discussed above. Results for three versions of the test, denoted V_T^2, $|V_T|$, and $|V_T|^{sup}$ are reported. Critical values are set equal to the 90^{th} percentile of the empirical distribution of the bootstrap statistics, and all empirical distributions are constructed using 100 bootstrap replications. All results are based on 1000 Monte Carlo simulations (see above for further details).

Table 2: Diffusion Specification Test Rejection Frequencies: Empirical Power I*
Data Generated Under H_A With $\sigma^2 = 0.1$

l	T=400 observations			T=800 observations			T=1200 observations		
	V_T^2	$\|V_T\|$	$\|V_T\|^{sup}$	V_T^2	$\|V_T\|$	$\|V_T\|^{sup}$	V_T^2	$\|V_T\|$	$\|V_T\|^{sup}$
Panel a: $\theta = 0.3$									
2	0.543	0.563	0.473	0.783	0.820	0.727	0.893	0.903	0.850
3	0.480	0.503	0.457	0.750	0.783	0.710	0.877	0.887	0.837
5	0.483	0.490	0.433	0.750	0.787	0.700	0.870	0.880	0.827
8	0.463	0.480	0.423	0.720	0.760	0.677	0.860	0.863	0.830
10	0.450	0.483	0.397	0.720	0.747	0.700	0.850	0.867	0.823
20	0.460	0.447	0.427	0.723	0.753	0.677	0.847	0.873	0.790
25	0.437	0.443	0.417	0.727	0.740	0.667	0.847	0.870	0.820
40	0.467	0.477	0.427	0.707	0.727	0.683	0.847	0.867	0.827
50	0.450	0.450	0.413	0.717	0.727	0.670	0.857	0.870	0.827
Panel b: $\theta = 0.6$									
2	0.333	0.363	0.290	0.570	0.587	0.483	0.723	0.763	0.630
3	0.340	0.353	0.287	0.567	0.583	0.467	0.717	0.747	0.623
5	0.337	0.353	0.260	0.560	0.593	0.453	0.710	0.750	0.617
8	0.340	0.360	0.290	0.560	0.587	0.450	0.717	0.743	0.623
10	0.343	0.357	0.280	0.560	0.607	0.443	0.710	0.757	0.613
20	0.330	0.357	0.283	0.540	0.590	0.440	0.720	0.740	0.620
25	0.327	0.357	0.283	0.547	0.577	0.427	0.703	0.743	0.607
40	0.350	0.350	0.313	0.533	0.573	0.447	0.717	0.750	0.613
50	0.353	0.363	0.330	0.537	0.583	0.457	0.707	0.753	0.627
Panel c: $\theta = 0.9$									
2	0.213	0.243	0.193	0.377	0.460	0.297	0.510	0.603	0.360
3	0.223	0.253	0.183	0.383	0.453	0.273	0.513	0.583	0.360
5	0.233	0.247	0.197	0.390	0.480	0.297	0.523	0.587	0.363
8	0.237	0.253	0.200	0.380	0.453	0.283	0.497	0.587	0.370
10	0.230	0.250	0.193	0.367	0.440	0.270	0.503	0.580	0.367
20	0.223	0.257	0.193	0.363	0.470	0.283	0.490	0.573	0.363
25	0.243	0.270	0.217	0.390	0.480	0.277	0.477	0.590	0.387
40	0.267	0.280	0.233	0.370	0.467	0.300	0.503	0.590	0.387
50	0.260	0.273	0.237	0.393	0.477	0.313	0.497	0.580	0.373

(*) Notes: See notes to Table 1.

Table 3: Diffusion Specification Test Rejection Frequencies: Empirical Power II*
Data Generated Under H_A With $\sigma^2 = 0.5$

l	T=400 observations			T=800 observations			T=1200 observations		
	V_T^2	$\lvert V_T\rvert$	$\lvert V_T\rvert^{sup}$	V_T^2	$\lvert V_T\rvert$	$\lvert V_T\rvert^{sup}$	V_T^2	$\lvert V_T\rvert$	$\lvert V_T\rvert^{sup}$
Panel a: $\theta = 0.3$									
2	0.953	0.873	0.980	0.990	0.957	1.000	0.993	0.977	1.000
3	0.950	0.843	0.977	0.997	0.963	1.000	0.993	0.967	1.000
5	0.943	0.857	0.980	0.987	0.960	1.000	0.990	0.973	1.000
8	0.933	0.853	0.977	0.990	0.940	1.000	0.990	0.960	1.000
10	0.950	0.850	0.967	0.987	0.937	1.000	0.990	0.950	1.000
20	0.937	0.833	0.970	0.990	0.947	1.000	0.990	0.957	1.000
25	0.920	0.837	0.957	0.983	0.947	1.000	0.993	0.960	1.000
40	0.910	0.833	0.957	0.987	0.943	1.000	0.990	0.947	1.000
50	0.907	0.820	0.943	0.993	0.943	1.000	0.990	0.963	1.000
Panel b: $\theta = 0.6$									
2	0.937	0.883	0.937	1.000	0.987	1.000	1.000	1.000	1.000
3	0.917	0.860	0.910	0.997	0.987	1.000	1.000	1.000	1.000
5	0.923	0.857	0.913	1.000	0.990	1.000	1.000	1.000	1.000
8	0.917	0.853	0.897	0.993	0.990	1.000	1.000	1.000	1.000
10	0.923	0.857	0.900	0.997	0.993	1.000	1.000	1.000	1.000
20	0.903	0.847	0.897	1.000	0.997	1.000	1.000	1.000	1.000
25	0.890	0.837	0.910	1.000	0.990	1.000	1.000	1.000	1.000
40	0.913	0.820	0.890	0.993	0.990	1.000	1.000	0.997	1.000
50	0.900	0.830	0.900	0.993	0.993	1.000	1.000	1.000	1.000
Panel c: $\theta = 0.9$									
2	0.870	0.857	0.723	0.983	0.980	0.950	0.997	0.997	0.993
3	0.867	0.853	0.737	0.980	0.980	0.953	0.993	0.997	0.990
5	0.857	0.853	0.723	0.987	0.983	0.957	0.997	0.997	0.993
8	0.857	0.850	0.713	0.987	0.980	0.957	0.997	0.997	0.990
10	0.863	0.850	0.743	0.980	0.973	0.963	0.993	0.997	0.993
20	0.850	0.840	0.723	0.983	0.980	0.957	0.997	0.993	0.993
25	0.843	0.827	0.717	0.987	0.977	0.960	0.997	0.997	0.993
40	0.830	0.827	0.720	0.983	0.980	0.953	0.997	0.993	0.993
50	0.823	0.830	0.723	0.973	0.973	0.950	0.997	0.993	0.997

(*) Notes: See notes to Table 1.

Table 4: Diffusion Specification Test Rejection Frequencies: Empirical Power III*
Data Generated Under H_A With $\sigma^2 = 1.0$

l	T=400 observations			T=800 observations			T=1200 observations		
	V_T^2	$\|V_T\|$	$\|V_T\|^{sup}$	V_T^2	$\|V_T\|$	$\|V_T\|^{sup}$	V_T^2	$\|V_T\|$	$\|V_T\|^{sup}$
Panel a: $\theta = 0.3$									
2	0.987	0.973	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	0.977	0.967	0.997	0.997	1.000	1.000	1.000	1.000	1.000
5	0.983	0.967	1.000	1.000	1.000	1.000	1.000	1.000	1.000
8	0.967	0.953	1.000	1.000	1.000	1.000	1.000	1.000	1.000
10	0.967	0.953	0.997	1.000	1.000	1.000	1.000	1.000	1.000
20	0.947	0.933	0.990	1.000	1.000	1.000	1.000	1.000	1.000
25	0.960	0.933	0.997	1.000	1.000	1.000	1.000	1.000	1.000
40	0.933	0.920	0.987	0.997	0.997	1.000	1.000	1.000	1.000
50	0.937	0.907	0.987	0.997	0.993	1.000	1.000	0.997	1.000
Panel b: $\theta = 0.6$									
2	0.987	0.890	0.997	0.997	0.960	1.000	0.997	0.973	1.000
3	0.977	0.887	1.000	0.993	0.957	1.000	0.993	0.957	1.000
5	0.980	0.890	0.997	0.990	0.970	1.000	0.997	0.957	1.000
8	0.983	0.900	0.993	0.990	0.943	1.000	0.993	0.963	1.000
10	0.987	0.893	1.000	0.993	0.957	1.000	0.997	0.967	1.000
20	0.973	0.883	0.993	0.993	0.953	1.000	0.997	0.970	1.000
25	0.983	0.890	0.997	0.997	0.960	1.000	0.993	0.967	1.000
40	0.983	0.890	0.997	0.993	0.960	1.000	0.993	0.967	1.000
50	0.990	0.887	0.997	0.993	0.963	1.000	0.990	0.970	1.000
Panel c: $\theta = 0.9$									
2	0.987	0.903	0.990	0.993	0.977	1.000	1.000	0.997	1.000
3	0.973	0.927	0.990	1.000	0.980	1.000	1.000	0.997	1.000
5	0.983	0.910	0.993	0.997	0.980	1.000	1.000	0.993	1.000
8	0.973	0.903	0.993	1.000	0.977	1.000	1.000	0.993	1.000
10	0.973	0.903	0.993	1.000	0.990	1.000	1.000	1.000	1.000
20	0.953	0.913	0.987	0.997	0.980	1.000	0.997	0.993	1.000
25	0.960	0.893	0.990	0.997	0.990	1.000	0.997	0.993	1.000
40	0.970	0.897	0.990	1.000	0.987	1.000	1.000	0.997	1.000
50	0.970	0.910	0.997	0.997	0.983	1.000	1.000	0.990	1.000

(*) Notes: See notes to Table 1.