Volume 161 Issue 2 April 2011 ISSN 0304-4076

ELSEVIER

JOURNAL OF Econometrics

CONTENTS

CONTENTS CONTINUED ON BACK COVER

Journal of Econometrics 161 (2011) 304–324

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Predictive density construction and accuracy testing with multiple possibly misspecified diffusion models☆

Valentina Corradi[a,*], Norman R. Swanson[b]

[a] Department of Economics, University of Warwick, Coventry CV4 7AL, UK

[b] Department of Economics, Rutgers University, 75 Hamilton Street, New Brunswick, NJ 08901, USA

ARTICLE INFO

Article history:
Received 19 April 2009
Received in revised form
8 July 2010
Accepted 23 December 2010
Available online 29 December 2010

JEL classification:
C22
C51

Keywords:
Block bootstrap
Diffusion processes
Jumps
Nonparametric simulated quasi maximum likelihood
Parameter estimation error
Recursive estimation
Stochastic volatility

ABSTRACT

This paper develops tests for comparing the accuracy of predictive densities derived from (possibly misspecified) diffusion models. In particular, we first outline a simple simulation-based framework for constructing predictive densities for one-factor and stochastic volatility models. We then construct tests that are in the spirit of Diebold and Mariano (1995) and White (2000). In order to establish the asymptotic properties of our tests, we also develop a recursive variant of the nonparametric simulated maximum likelihood estimator of Fermanian and Salanié (2004). In an empirical illustration, the predictive densities from several models of the one-month federal funds rates are compared.

1. Introduction

Correct specification of models describing dynamics of financial assets is crucial for everything from pricing bonds and derivative assets to designing appropriate hedging strategies. Hence, it is of little surprise that there has been considerable attention given to the issue of testing for the correct specification of diffusion models. In this paper, we do not construct specification tests in the usual sense, but instead assume that all models are (possibly) misspecified and outline a simulation-based methodology for comparing the accuracy of predictive densities based on alternative models.

To place this paper in the correct historical context, note that a first generation of specification testing papers, initiated by the work of Aït-Sahalia (1996), compares the marginal densities implied by hypothesized null models with nonparametric estimates thereof, for the case of one-factor models (see also Pritsker (1998) and Jiang (1998)). While one-factor models may in some cases provide a reasonable representation for short-term interest rates, there is a somewhat widespread consensus that stock returns and term structures are better modeled using multifactor diffusions. To take this into account, Corradi and Swanson (2005a) outline a test for comparing the cumulative distribution (marginal or joint) implied by a hypothesized null model with the corresponding empirical distribution. Their test can be used in the context of multidimensional and/or multifactor models. Needless to say, tests based on the comparison of marginal distributions have no power against *iid* alternatives with the same marginal, while tests based on the comparison of joint distributions do not suffer from this problem. Nevertheless, correct specification of the joint distribution is not equivalent to

☆ Corradi gratefully acknowledges ESRC grant RES-000-23-0006 and RES-062-23-0311, and Swanson acknowledges financial support from a Rutgers University Research Council grant. We would like to thank the co-editor, Ron Gallant, a referee, Federico Bandi, Marine Carrasco, Javier Hidalgo, Antonio Mele, Andrew Patton, Eric Renault, John Rust and the seminar participants at the annual joint CORE, ECARES, and KU Leuven econometrics workshop, the 2007 Summer Meeting of the Econometric Society, the Marseille conference in honor of Russell Davidson, as well as faculty at the following universities: University of Montreal, LSE, Michigan State University, New York University, University of Chicago GBS, and the University of Maryland, for their useful comments on earlier versions of this paper. Additionally, we would like to thank Lili Cai for excellent research assistance.

* Corresponding author.
E-mail addresses: v.corradi@warwick.ac.uk (V. Corradi), nswanson@econ.rutgers.edu (N.R. Swanson).

0304-4076/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2010.12.009

that of the conditional; and hence focus in the literature now centers on comparing conditional distributions. When considering conditional distributions, a key difficulty that arises stems from the fact that knowledge of the drift and variance terms of a diffusion process does not in turn imply knowledge of the transition density, in general. Indeed, if the functional form of the transition density were known, one could test the hypothesis of correct specification of a diffusion via the probability integral transform approach of Diebold et al. (1998); the cross-spectrum approach of Hong (2001), Hong et al. (2002) and Hong and Li (2005); the martingalization-type Kolmogorov test of Bai (2003); or via the normality transformation approaches of Bontemps and Meddahi (2005) and Duan (2003). Furthermore, for the case in which the transition density is unknown, tests could be constructed by comparing the kernel (conditional) density estimator of the actual and simulated data, as in Altissimo and Mele (2009) and Thompson (2008); by comparing the conditional distribution of the simulated and of the historical data, as in Bhardwaj et al. (2008); or by using the approaches of Aït-Sahalia (2002) and Aït-Sahalia et al. (2009), where closed form approximations of conditional densities under the null are compared with data-driven kernel density estimates.

All of the papers cited above deal with testing for the correct specification of a given diffusion model. Nevertheless, and as alluded to above, we believe that all models are probably best viewed as approximations of reality and, thus, are likely to be misspecified. Therefore, we focus on choosing the "best" model from amongst (multiple) misspecified alternatives. Moreover, the "best" model is selected by constructing tests that compare both predictive densities and/or predictive conditional confidence intervals associated with alternative models.

Our approach is to measure accuracy using a distributional generalization of mean square error, as defined in Corradi and Swanson (2005b). Namely, let $F_k^\tau(u|X_t,\theta_k^\dagger)$ be the distribution of $X_{t+\tau}$ given X_t, evaluated at u, implied by diffusion model k, and let $F_0^\tau(u|X_t,\theta_0)$ be the distribution associated with the underlying and unknown "true" model. Now, choose model k over model 1, say, if $E((F_k^\tau(u|X_t,\theta_k^\dagger) - F_0^\tau(u|X_t,\theta_0))^2) < E((F_1^\tau(u|X_t,\theta_1^\dagger) - F_0^\tau(u|X_t,\theta_0))^2)$. Our tests can be viewed as distributional generalizations of both Diebold and Mariano (1995) and White (2000). Note that if we knew $F_k^\tau(u|X_t,\theta_k^\dagger)$ in closed form, then we could proceed as in Corradi and Swanson (2006a,b). However, the functional form of the model implied conditional distribution is unknown in closed form, in general, and hence we rely on a simulation-based approach to facilitate testing. As is customary in the out-of-sample evaluation literature, the sample of T observations is split into two subsamples, such that $T = R + P$, where only the last P observations are used for predictive evaluation. We first simulate $P - \tau$ τ-step ahead paths, using $X_R, \ldots, X_{R+P-\tau}$ as starting values. Then, a scaled difference between the conditional distribution, estimated with historical as well as simulated data, is used to construct our test statistic. One complication that arises in this setup is that for the case of stochastic volatility (SV) models, the initial value of the volatility process is unobserved. To overcome this problem, it suffices to simulate the process using different random initial values for the volatility process. Thereafter, one simply constructs the empirical distribution of the asset price process for any given initial value of the volatility process and takes an average over the latter. This integrates out the effect of the volatility initial value.

The limiting distributions of the suggested statistics are shown to be (functionals of) Gaussian processes with covariance kernels that reflect the contribution of recursive parameter estimation error. In order to provide asymptotically (first-order) valid critical values, we introduce a new bootstrap procedure that mimics the contribution of parameter estimation error in a recursive setting. This is achieved by establishing consistency and asymptotic normality of nonparametric simulated quasi maximum likelihood (NPSQML) estimators of (possibly misspecified) diffusion models, in a recursive setting, and by establishing the first-order validity of their bootstrap analogs.

Of final note is that we test the same null hypothesis as Corradi and Swanson (2006a), and we estimate empirical conditional distributions using both historical and simulated data, as in Bhardwaj et al. (2008). However, there are many differences between those papers and this one. Five such differences are the following. First, we show the asymptotic equivalence of recursive NPSQMLE (Nonparametric Simulated Quasi Maximum Likelihood Estimators) and recursive QMLE. Second, we show the asymptotic equivalence of recursively estimated NPSQMLE and recursive QMLE for partially unobservable multidimensional diffusions (e.g. for stochastic volatility models). This extends in a non-trivial manner the NPSQMLE of Fermanian and Salanié (2004). Third, we establish the first order validity of bootstrap critical values for recursive NPSQMLE, in the case of both observable and partially unobservable diffusions. To the best of our knowledge, there are no available results on bootstrapping NPSQMLE. Fourth, we allow for jumps in the return process, and we recursively estimate the intensity and the parameters of the jump size density. Finally, we develop Diebold–Mariano type Reality Check tests for cases where (a) the CDF is not known in closed form, and (b) data are generated by partially unobservable jump diffusion processes.

The rest of the paper is organized as follows. In Section 2, we define the setup. Section 3 outlines the testing procedure for choosing between $m \geqslant 2$ models and establishes the asymptotic properties thereof. In Section 4, we develop a recursive version of the NPSQML estimator of Fermanian and Salanié (2004) and outline conditions under which asymptotic equivalence between NPSQML and the corresponding recursive QMLE obtains. An empirical illustration is provided in Section 5, in which various models of the effective federal funds rate are compared. All proofs are collected in an Appendix. Hereafter, let P^* denote the probability law governing the resampled series, conditional on the (entire) sample, let E^* and Var^* denote the mean and variance operators associated with P^*. Further, let $o_P^*(1)$ Pr $-P$ denote a term converging to zero in P^*-probability, conditional on the sample except a subset of probability measure approaching zero. Finally, let $O_P^*(1)$ Pr $-P$ denote a term which is bounded in P^*-probability, conditional on the sample, and for all samples except a subset with probability measure approaching zero.

2. Set-up

First, consider m one factor jump diffusion models. Namely, for $k = 1, \ldots, m$ consider[1]:

$$X(t_-) = \int_0^t b_k(X(s_-),\theta_k^\dagger)\mathrm{d}s - \lambda_k t \int_Y y\phi_k(y)\mathrm{d}y + \int_0^t \sigma_k(X(s_-),\theta_k^\dagger)\mathrm{d}W(s) + \sum_{j=1}^{J_{k,t}} y_{k,j},$$

where $J_{k,t}$ is a Poisson process with intensity parameter λ_k, λ_k finite, and the jump size, $y_{k,j}$, is *iid* with marginal distribution given by ϕ_k. Both $J_{k,t}$ and $y_{k,j}$ are assumed to be independent of the driving Brownian motion, $W(t)$. Also, note that $\int_Y y\phi_k(y)\mathrm{d}y$ denotes the mean jump size under model k, hereafter denoted by $\mu_{y,k}$. The case of no jumps corresponds to $J_{k,t} = 0$ for all t, and $\lambda_k = 0$. Note that over a unit time interval, there are on average λ_k

[1] Hereafter, $X(t_-)$ denotes the *cadlag* (right continuous with left limit) for $t \in \mathfrak{R}^+$, while X_t denotes the discrete skeleton for $t = 1, 2, \ldots$.

jumps; so that over the time span $[0, t]$, there are on average $\lambda_k t$ jumps. The dynamics of $X(t_-)$ is then given by:

$$dX(t) = (b_k(X(t_-), \theta_k^\dagger) - \lambda_k \mu_{y,k})dt + \sigma_k(X(t_-), \theta_k^\dagger)dW(t) + \int_Y yp(dy, dt), \quad (1)$$

where $p(dy, dt)$ is a random Poisson measure giving point mass at y if a jump occurs in the interval dt. Hereafter, let $\vartheta_k = (\theta_k, \lambda_k, \mu_{y,k})$. If model k is correctly specified, then $b_k(X(t_-), \theta_k^\dagger) = b_0(X(t_-), \theta_0)$, $\sigma_k(X(t_-), \theta_k^\dagger) = \sigma_0(X(t_-), \theta_0)$, $\lambda_k = \lambda_0$, and $\phi_k = \phi_0$. Now, let $F_k^\tau(u|X_t, \vartheta_k^\dagger) = P^\tau_{\vartheta_k^\dagger}(X_{t+\tau} \leq u|X_t)$ (i.e., $F_k^\tau(u|X_t, \vartheta_k^\dagger)$ defines the conditional distribution of $X_{t+\tau}$, given X_t, and evaluated at u, under the probability law generated by model k). Analogously, define $F_0^\tau(u|X_t, \vartheta_0) = P^\tau_{\vartheta_0}(X_{t+\tau} \leq u|X_t)$ to be the "true" conditional distribution. We measure model accuracy in terms of a distributional analog of mean square error. In particular, model 1 is defined to be more accurate than model k if:

$$E(((F_1^\tau(u_2|X_t, \vartheta_1^\dagger) - F_1^\tau(u_1|X_t, \vartheta_1^\dagger)) - (F_0^\tau(u_2|X_t, \vartheta_0) - F_0^\tau(u_1|X_t, \vartheta_0)))^2) < E(((F_k^\tau(u_2^\tau|X_t, \vartheta_k^\dagger) - F_k^\tau(u_1^\tau|X_t, \vartheta_k^\dagger)) - (F_0^\tau(u_2|X_t, \vartheta_0) - F_0^\tau(u_1|X_t, \vartheta_0)))^2).$$

This measure defines a norm and implies a standard goodness of fit measure (see, for example, Corradi and Swanson 2005b). Recalling that $E(1\{u_1 \leq X_{t+\tau} \leq u_2\}|X_t) = F_0^\tau(u_2|X_t, \vartheta_0) - F_0^\tau(u_1|X_t, \vartheta_0)$, it is straightforward to construct a sequence of $P - \tau\tau$-step ahead prediction errors under model k as $1\{u_1 \leq X_{t+\tau} \leq u_2\} - (F_k^\tau(u_2|X_t, \widehat{\vartheta}_{k,t,N,h}) - F_k^\tau(u_1|X_t, \widehat{\vartheta}_{k,t,N,h}))$, for $t = R, \ldots, R + P - \tau$, where $\widehat{\vartheta}_{k,t,N,h}$ is an estimator of $\vartheta_k^\dagger$ computed using all observations up to time t, $P + R = T$, N is the number of simulation paths used in estimation, and h is the discretization interval. Hence, prediction errors should be constructed as follows. Simulate $P - \tau$ paths of length τ, using $X_{R+1}, \ldots, X_{R+P-\tau}$ as starting values and using the recursively estimated parameters, $\widehat{\vartheta}_{k,t,N,h}, t = R, \ldots, R + P - \tau$. Then, construct the empirical distribution of the series simulated under model k. Then, test statistics are constructed relying on the fact that, under some regularity conditions, as discussed in Bhardwaj et al. (2008):

$$\frac{1}{N}\sum_{i=1}^{N} 1\{u_1 \leq X_{k,t+\tau,i}^{\widehat{\vartheta}_{k,t,N,h}}(X_t) \leq u_2\} \xrightarrow{pr} F_{X_{k,t+\tau}^{\vartheta_k^\dagger}(X_t)}(u_2) - F_{X_{k,t+\tau}^{\vartheta_k^\dagger}(X_t)}(u_1), \quad t = R, \ldots, T - \tau, \quad (2)$$

where $F_{X_{k,t+\tau}^{\vartheta_k^\dagger}(X_t)}(u)$ is the marginal distribution of $X_{t+\tau}^{\vartheta_k^\dagger}(X_t)$ implied by k model (i.e., by the model used to simulate the series), conditional on the (simulation) starting value X_t. Furthermore, the marginal distribution of $X_{t+\tau}^{\vartheta_k^\dagger}(X_t)$ is the distribution of $X_{t+\tau}$ under model k, conditional on the values observed at time t. Thus, $F_{X_{k,t+\tau}^{\vartheta_k^\dagger}(X_t)}(u) = F_k^\tau(u|X_t, \vartheta_k^\dagger)$. In the above expression, $X_{k,t+\tau,i}^{\widehat{\vartheta}_{k,t,N,h}}(X_t)$ is generated according to a Milstein scheme, where

$$\begin{aligned} X_{(q+1)h}^{\widehat{\vartheta}_{k,t,N,h}} - X_{qh}^{\widehat{\vartheta}_{k,t,N,h}} &= b_k(X_{qh}^{\widehat{\vartheta}_{k,t,N,h}}, \widehat{\theta}_{k,t,N,h})h \\ &+ \sigma_k(X_{qh}^{\widehat{\vartheta}_{k,t,N,h}}, \widehat{\theta}_{k,t,N,h})\epsilon_{(q+1)h} - \frac{1}{2}\sigma_k'(X_{qh}^{\widehat{\vartheta}_{k,t,N,h}}, \widehat{\theta}_{k,t,N,h}) \\ &\times \sigma_k(X_{qh}^{\widehat{\vartheta}_{k,t,N,h}}, \widehat{\theta}_{k,t,N,h})h + \frac{1}{2}\sigma_k(X_{qh}^{\widehat{\vartheta}_{k,t,N,h}}, \widehat{\theta}_{k,t,N,h}) \\ &\times \sigma_k'(X_{qh}^{\widehat{\vartheta}_{k,t,N,h}}, \widehat{\theta}_{k,t,N,h})\epsilon^2_{(q+1)h} - \widehat{\lambda}_k\widehat{\mu}_{y,k}h \\ &+ \sum_{j=1}^{\mathcal{J}_k} y_{k,j}1\{qh \leq \mathcal{U}_j \leq (q+1)h\}, \end{aligned} \quad (3)$$

with $\epsilon_{qh} \overset{iid}{\sim} N(0, h)$, $q = 1, \ldots, Q$; and where σ' is the derivative of $\sigma(\cdot)$ with respect to its first argument. Additionally, the argument X_t in $X_{k,t+\tau,i}^{\widehat{\vartheta}_{k,t,N,h}}(X_t)$ denotes that the starting value for the simulation is X_t. Note that the last term on the right-hand side (RHS) of (3) is nonzero whenever we have one (or more) jump realization(s) in the interval $[(q-1)h, qh]$. Moreover, as neither the intensity nor the jump size is state dependent, the jump component can be simulated without any discretization error, as follows. Begin by making a draw from a Poisson distribution with intensity parameter $\widehat{\lambda}_k\tau$, say $\mathcal{J}_k$. This gives a realization for the number of jumps over the simulation time span. Then, draw $\mathcal{J}_k$ uniform random variables over $[0, \tau]$, and sort them in ascending order so that $\mathcal{U}_1 \leq \mathcal{U}_2 \leq \cdots \leq \mathcal{U}_{\mathcal{J}_k}$. These provide realizations for the $\mathcal{J}_k$ jump times. Then, make $\mathcal{J}_k$ independent draws from ϕ_k, say $y_{k,1}, \ldots, y_{k,\mathcal{J}_k}$. An important feature of this simulation procedure is that to generate $X_{k,t+\tau,i}^{\widehat{\vartheta}_{k,t,N,h}}(X_t)$, $i = 1, \ldots, N$, for $t = R, \ldots, T - \tau$, we must use (for each t) the same set of randomly drawn errors as well as the same draws for numbers of jumps, jump times and jump sizes. Thus, only the starting value used to initialize the simulations changes. More precisely, the errors used in simulation are defined to be $\epsilon_{qh,i} \overset{iid}{\sim} N(0, h)$, with $Qh = \tau$, $i = 1, \ldots, N$.

Now, proceed by constructing $X_{k,R+\tau,i}^{\widehat{\vartheta}_{k,R,N,h}}(X_R), \ldots, X_{k,T,i}^{\widehat{\vartheta}_{k,T-\tau,N,h}}(X_{T-\tau})$, where $i = 1, \ldots, N$. This yields an $N \times (P - \tau + 1)$ matrix of simulated values. The key feature of this setup is that it enables the comparison of simulated values $X_{k,R+j+\tau,i}^{\widehat{\vartheta}_{k,R+j,N,h}}(X_{R+j})$ with actual values that are τ periods ahead (i.e., $X_{R+j+\tau}$), for $j = 1, \ldots, P - \tau + 1$. In this manner, we are able to propose tests for simulation based on *ex-ante* predictive density comparison.

Turning now to the case of SV models, whenever both intensity and jump size are non-state dependent, a jump component can be simulated and added to either the return and/or the volatility process in the same manner as above. Therefore, for the sake of simplicity, we consider SV models without jumps in the sequel. Extension to general multidimensional and multifactor models both with and without jumps follows directly. Finally, note that as we are considering the case of no jumps, parameters and estimators will be denoted by θ instead of ϑ. Consider model k, $k = 1, \ldots, m$, defined as follows:

$$\begin{pmatrix} dX(t) \\ dV(t) \end{pmatrix} = \begin{pmatrix} b_{1,k}(X(t), \theta_k^\dagger) \\ b_{2,k}(V(t), \theta_k^\dagger) \end{pmatrix} dt + \begin{pmatrix} \sigma_{11,k}(V(t), \theta_k^\dagger) \\ 0 \end{pmatrix} dW_1(t) + \begin{pmatrix} \sigma_{12,k}(V(t), \theta_k^\dagger) \\ \sigma_{22,k}(V(t), \theta_k^\dagger) \end{pmatrix} dW_2(t), \quad (4)$$

where $W_{1,t}$ and $W_{2,t}$ are independent standard Brownian motions. Following a generalized Milstein scheme (see, for example, Eq. (3.3), pp. 346 in Kloeden and Platen 1999), for models $k = 1, 2, \ldots, m$, and for $\widehat{\theta}_{k,t,N,S,h}$ an estimator of $\theta_k^\dagger$:

$$\begin{aligned} X_{(q+1)h}^{\widehat{\theta}_{k,t,N,S,h}} &= X_{qh}^{\widehat{\theta}_{k,t,N,S,h}} + \widetilde{b}_{1,k}(X_{qh}^{\widehat{\theta}_{k,t,N,S,h}}, \widehat{\theta}_{k,t,N,S,h})h \\ &+ \sigma_{11,k}(V_{qh}^{\widehat{\theta}_{k,t,N,S,h}}, \widehat{\theta}_{k,t,N,S,h})\epsilon_{1,(q+1)h} \\ &+ \sigma_{12,k}(V_{qh}^{\widehat{\theta}_{k,t,N,S,h}}, \theta_k)\epsilon_{2,(q+1)h} \\ &+ \frac{1}{2}\sigma_{22,k}(V_{qh}^{\widehat{\theta}_{k,t,N,S,h}}, \theta_i) \\ &\times \frac{\partial\sigma_{12,k}(V_{qh}^{\widehat{\theta}_{k,t,N,S,h}}, \theta_k)}{\partial V}\epsilon^2_{2,(q+1)h} \\ &+ \sigma_{22,k}(V_{qh}^{\widehat{\theta}_{k,t,N,S,h}}, \theta_k)\frac{\partial\sigma_{11,k}(V_{qh}^{\widehat{\theta}_{k,t,N,S,h}}, \theta_k)}{\partial V} \\ &\times \int_{qh}^{(q+1)h}\left(\int_{qh}^{s} dW_{1,\tau}\right) dW_{2,s} \end{aligned} \quad (5)$$

$$V_{(q+1)h}^{\widehat{\theta}_{k,t,N,S,h}} = V_{qh}^{\widehat{\theta}_{k,t,N,S,h}} + \widetilde{b}_{2,k}(V_{qh}^{\widehat{\theta}_{k,t,N,S,h}}, \theta_k)h + \sigma_{22,k}(V_{qh}^{\widehat{\theta}_{k,t,N,S,h}}, \theta_k)\epsilon_{2,(q+1)h} + \frac{1}{2}\sigma_{22,k}(V_{qh}^{\widehat{\theta}_{k,t,N,S,h}}, \theta_k) \times \frac{\partial \sigma_{22}(V_{qh}^{\widehat{\theta}_{k,t,N,S,h}}, \theta_k)}{\partial V}\epsilon_{2,(q+1)h}^2 \quad (6)$$

where $h^{-1/2}\epsilon_{i,qh} \sim N(0,1)$, $i = 1, 2$, $E(\epsilon_{1,qh}\epsilon_{2,q'h}) = 0$ for all $q \neq q'$, and

$$\widetilde{b}_k(V, \widehat{\theta}_{k,t,N,S,h}) = \begin{pmatrix} \widetilde{b}_{1,k}(V, \widehat{\theta}_{k,t,N,S,h}) \\ \widetilde{b}_{2,k}(V, \widehat{\theta}_{k,t,N,S,h}) \end{pmatrix} = \begin{pmatrix} b_{1,k}(V, \widehat{\theta}_{k,t,N,S,h}) - \frac{1}{2}\sigma_{22,k}(V, \widehat{\theta}_{k,t,N,S,h})\frac{\partial \sigma_{12,k}(V, \widehat{\theta}_{k,t,N,S,h})}{\partial V} \\ b_{2,k}(V, \widehat{\theta}_{k,t,N,S,h}) - \frac{1}{2}\sigma_{22,k}(V, \widehat{\theta}_{k,t,N,S,h})\frac{\partial \sigma_{22,k}(V, \widehat{\theta}_{k,t,N,S,h})}{\partial V} \end{pmatrix}.$$

The last terms on the RHS of (5) involve stochastic integrals and cannot be explicitly computed. However, they can be approximated, up to an error of order $o(h)$ by (see, for example, Eq. (3.7), pp. 347 in Kloeden and Platen 1999):

$$\int_{qh}^{(q+1)h}\left(\int_{qh}^{s} dW_{1,\tau}\right) dW_{2,s} \approx h\left(\frac{1}{2}\xi_1\xi_2 + \sqrt{\rho_p}(\mu_{1,p}\xi_2 - \mu_{2,p}\xi_1)\right) + \frac{h}{2\pi}\sum_{r=1}^{p}\frac{1}{r}(\varsigma_{1,r}(\sqrt{2}\xi_2 + \eta_{2,r}) - \varsigma_{2,r}(\sqrt{2}\xi_1 + \eta_{1,r})),$$

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

In order to simulate paths for SV models, proceed as follows:

Step 1. Using the schemes in (5) and (6), simulate $(P-\tau+1)\times S\times N$ paths of length τ, setting the initial values for the observable state variable equal to the initial value X_t, $t = R, \ldots, R+P-\tau$, and for each X_t, using the S different starting values for volatility (i.e., $V_j^{\widehat{\theta}_{k,t,N,S,h}}, j = 1, \ldots, S$). Thus, there are S paths rather than one, for each starting value of X_t. For any initial value X_t and $V_j^{\widehat{\theta}_{k,t,N,S,h}}$, $t = R+1, \ldots, R+P-\tau$ and $j = 1, \ldots, S$, generate N independent paths of length τ. Also, keep the simulated randomness (i.e., $\epsilon_{1,qh}, \epsilon_{2,qh}, \int_{qh}^{(q+1)h}(\int_{qh}^{s} dW_{1,\tau})dW_{2,s}$) constant across the different starting values for the unobservable and observable state variables. Now, define $X_{k,t+\tau,i,j}^{\widehat{\theta}_{k,t,N,S,h}}(X_t, V_j^{\widehat{\theta}_{k,t,N,S,h}})$ to be the τ-step ahead value for the return series simulated (under model k), at replication i, $i = 1, \ldots, N$, using initial values X_t and $V_j^{\widehat{\theta}_{k,t,N,S,h}}$.

Step 2. Construct an estimator of $F_{X_{k,t+\tau}^{\theta_k^\dagger}(X_t)}(u_2) - F_{X_{k,t+\tau}^{\theta_k^\dagger}(X_t)}(u_1)$ using $\frac{1}{NS}\sum_{j=1}^{S}\sum_{i=1}^{N} 1\{u_1 \leq X_{k,t+\tau,i,j}^{\widehat{\theta}_{k,t,N,S,h}}(X_t, V_{k,t,j}^{\widehat{\theta}_{k,t,N,S,h}}) \leq u_2\}$, where $V_{k,t,j}^{\widehat{\theta}_{k,t,N,S,h}}$ denotes the value of volatility at time t and at simulation j, simulated under model k, using parameters $\widehat{\theta}_{k,t,N,S,h}$.

The asymptotic results in the sequel require the following assumptions.

Assumption A1. (i) $X(t)$, $t \in \Re^+$, is a strictly stationary, geometric ergodic β-mixing diffusion; and (ii) $\int_Y y^p \phi_k(y)dy < \infty$ for some $p > 2$.

Assumption A2. For $k = 1, \ldots, m$, $b_k(\cdot, \theta^\dagger)$ and $\sigma_k(\cdot, \theta^\dagger)$, as defined in (1), are twice continuously differentiable. Also, $b_k(\cdot, \cdot), b_k(\cdot, \cdot)', \sigma_k(\cdot, \cdot)$, and $\sigma_k(\cdot, \cdot)'$ are Lipschitz in the first argument, with Lipschitz constant independent of θ_k, where $b_k(\cdot, \cdot)'$ and $\sigma_k(\cdot, \cdot)'$ denote derivatives with respect to the first argument of the function.

Assumption A2′. For $j, i = 1, 2$, let $b_{j,k}(\cdot, \cdot)$ and $\sigma_{i,j,k}(\cdot, \cdot)$ (as defined in (4)) and $\sigma_{ll',k}(V, \theta_k)\frac{\partial \sigma_{k\iota}(V,\theta_k)}{\partial V}$ be twice continuously differentiable, Lipschitz in the first argument, with a Lipschitz constant independent of θ_k, and assume that these terms grow at most at a linear rate, uniformly in Θ_k, for $l, l', j, \iota = 1, 2$ and $k = 1, \ldots, m$.

Assumption A3. For $k = 1, \ldots, m$: (i) for any fixed h and $\vartheta_k \in \Theta_k$, Θ_k compact set in $\mathcal{R}^{d_k}$, $X_{qh}^{\vartheta_k}$ is geometrically ergodic and β-mixing; (ii) $X_{k,t+\tau,i}^{\vartheta_k}$ is continuously differentiable in the interior of Θ_k, for $i = 1, \ldots, N$; and (iii) $\nabla_{\vartheta_k}X_{k,t+\tau,i}^{\vartheta_k}$ is r-dominated in Θ_k, uniformly in i for $r > 4$.

Assumption A4. For each model $k = 1, \ldots, m$ the parameters $\widehat{\vartheta}_{k,t,N,h}$ admit the following expansion:

$$\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}(\widehat{\vartheta}_{k,t,N,h} - \vartheta_k^\dagger) = A_k^\dagger\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\psi_{k,t,N,h}(\vartheta_k^\dagger) + o_p(1)$$

and as $P, R, N \to \infty$ and $h \to 0$,

$$\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\psi_{k,t,N,h}(\vartheta_k^\dagger) \xrightarrow{d} N(0, V_k^\dagger),$$

where $V_k^\dagger = \lim_{T,R,N,h^{-1}\to\infty} Var(\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\psi_{k,t,N,h}(\vartheta_k^\dagger))$.

Assumption A4′. For each model $k = 1, \ldots, m$ the parameters $\widehat{\vartheta}_{k,t,N,S,h}$ admit the following expansion:

$$\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}(\widehat{\vartheta}_{k,t,N,S,h} - \vartheta_k^\dagger) = A_k^\dagger\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\psi_{k,t,N,S,h}(\vartheta_k^\dagger) + o_p(1)$$

and as $P, R, N, S \to \infty$ and $h \to 0$,

$$\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\psi_{k,t,N,S,h}(\vartheta_k^\dagger) \xrightarrow{d} N(0, V_k^\dagger),$$

where $V_k^\dagger = \lim_{T,R,N,S,h^{-1}\to\infty} Var(\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\psi_{k,t,N,S,h}(\vartheta_k^\dagger))$.

Assumption A1(i) requires the diffusion, $X(t)$, to be geometrically ergodic and β-mixing. In the case of no jumps, conditions for (geometric) β-mixing for (multivariate) diffusions that can be relatively easily verified are provided by Meyn and Tweedie (1993). Such conditions also suffice for the case of jump diffusions, when both the intensity parameters and the jump sizes are independent of the state of the system. Recently, Masuda (2007) has extended the conditions for β-mixing to the case of jump diffusions in which the intensity parameter is constant, but the size of the jumps is state dependent.

Assumptions A4 and A4′ require that the recursively estimated parameters are $\sqrt{P}$-consistent and asymptotically normal, regardless of whether or not the underlying model is misspecified. As outlined in detail in Section 4, a key point here is that $E(\psi_{k,t,N,h}(\theta_k^\dagger))$ and $E(\psi_{k,t,N,S,h}(\theta_k^\dagger))$ are $o(P^{-1/2})$, regardless of misspecification. We shall show that NPSQMLE satisfies this requirement. Needless to say, in some cases the transition density is known in closed form and can be used to obtain QML estimators. For example, if the drift and variance terms as well as the intensity of the jump process have affine structures, then there is no need to rely on simulation methods and parameters can be estimated via the use of the conditional empirical characteristic function (see, for example, Singleton 2001).

3. Test statistics

3.1. One factor models

First, consider comparing the predictive accuracy of two possibly misspecified diffusion models. The hypotheses of interest are:

$$H_0 : E_X((F_{X^{\vartheta_1^\dagger}_{1,t+\tau}(X_t)}(u_2) - F_{X^{\vartheta_1^\dagger}_{1,t+\tau}(X_t)}(u_1)) - (F_0^\tau(u_2|X_t) - F_0^\tau(u_1|X_t)))^2 - E_X((F_{X^{\vartheta_k^\dagger}_{k,t+\tau}(X_t)}(u_2) - F_{X^{\vartheta_k^\dagger}_{k,t+\tau}(X_t)}(u_1)) - (F_0^\tau(u_2|X_t) - F_0^\tau(u_1|X_t)))^2 = 0$$

H_A : negation of H_0.

Notice that the hypotheses are stated using a particular interval (i.e., $(u_1, u_2) \in U \times U$) so that the objective is evaluation of predictive densities for a given range of values. The test statistic is:

$$D_{k,P,N}(u_1,u_2) = \frac{1}{\sqrt{P}}\sum_{t=R}^{T-\tau}\left(\left[\frac{1}{N}\sum_{i=1}^{N}1\{u_1 \le X^{\widehat{\vartheta}_{1,t,N,h}}_{1,t+\tau,i}(X_t) \le u_2\} - 1\{u_1 \le X_{t+\tau} \le u_2\}\right]^2 - \left[\frac{1}{N}\sum_{i=1}^{N}1\{u_1 \le X^{\widehat{\vartheta}_{k,t,N,h}}_{k,t+\tau,i}(X_t) \le u_2\} - 1\{u_1 \le X_{t+\tau} \le u_2\}\right]^2\right).$$

Theorem 1. *Let Assumptions A1–A4 hold. Also, assume that models 1 and k are nonnested. If as $P, R, N \to \infty, h \to 0, P/N \to 0, h^2P \to 0$, and $P/R \to \pi$, where $0 < \pi < \infty$, then:* (i) *Under H_0, $D_{k,P,N}(u_1,u_2) \xrightarrow{d} N(0, W_k(u_1,u_2))$, where $W_k(u_1,u_2)$ is defined in the Appendix.* (ii) *Under H_A,* $\Pr\left(\frac{1}{\sqrt{P}}\left|D_{k,P,N}(u_1,u_2)\right| > \varepsilon\right) \to 1$.

Note that $W_k(u_1, u_2)$ reflects the contribution of the recursive parameter estimation error. The intuitive argument underlying the proof to Theorem 1 is the following. Note that:

$$\frac{1}{N}\sum_{i=1}^{N}1\{X^{\widehat{\vartheta}_{k,t,N,h}}_{k,t+\tau,i}(X_t) \le u\} = \frac{1}{N}\sum_{i=1}^{N}1\{X^{\vartheta_k^\dagger}_{k,t+\tau,i}(X_t) \le u\} + E(f_{X^{\vartheta_k^\dagger}_{k,t+\tau,i}(X_t)}(u)\nabla_{\vartheta_k}X^{\vartheta_k^\dagger}_{k,t+\tau,i}(X_t))\frac{1}{\sqrt{P}}\sum_{t=R}^{T}(\widehat{\vartheta}_{k,t,N,h} - \vartheta^\dagger) + o_P(1)$$
$$= F_{X^{\vartheta_k^\dagger}_{k,t+\tau}(X_t)}(u) + E(f_{X^{\vartheta_k^\dagger}_{k,t+\tau,i}(X_t)}(u)\nabla_{\vartheta_k}X^{\vartheta_k^\dagger}_{k,t+\tau,i}(X_t)) \times \frac{1}{\sqrt{P}}\sum_{t=R}^{T}(\widehat{\vartheta}_{k,t,N,h} - \vartheta^\dagger) + o_P(1) + o_N(1),$$

where $o_N(1)$ denotes terms approaching zero, as $N \to \infty$. The statement follows by the same argument used in the case in which the closed form of the conditional distribution is known. Note that as $N/P \to \infty$, we can neglect the contribution of simulation error in the asymptotic covariance matrix. Finally, it is easy to see that if $P/R \to \pi = 0$, then the contribution of parameter estimation error vanishes.

In some circumstances, one may be interested in comparing one (benchmark) model against multiple competing models. In this case, the null hypothesis is that no model can outperform the benchmark model. More specifically, setting model 1 as the benchmark, the hypotheses of interest are:

$$H_0' : \max_{k=2,\ldots,m}(E_X((F_{X^{\vartheta_1^\dagger}_{1,t+\tau}(X_t)}(u_2) - F_{X^{\vartheta_1^\dagger}_{1,t+\tau}(X_t)}(u_1)) - (F_0(u_2|X_t) - F_0(u_1|X_t)))^2 - E_X((F_{X^{\vartheta_k^\dagger}_{k,t+\tau}(X_t)}(u_2) - F_{X^{\vartheta_k^\dagger}_{k,t+\tau}(X_t)}(u_1)) - (F_0(u_2|X_t) - F_0(u_1|X_t)))^2) \le 0$$

H_A' : negation of H_0'.

The statistic for testing these hypotheses is:

$$D^{\text{Max}}_{k,P,N}(u_1,u_2) = \max_{k=2,\ldots,m} D_{k,P,N}(u_1,u_2).$$

Corollary 1. *Let Assumptions A1–A4 hold. Also, assume that models 1 and k are nonnested for at least one $k = 2, \ldots, m$. If as $P, R, N \to \infty, h \to 0, P/N \to 0, h^2P \to 0$, and $P/R \to \pi$, where $0 < \pi < \infty$, then:*

$$\max_{k=2,\ldots,m}(D_{k,P,N}(u_1,u_2) - \mu_k(u_1,u_2)) \xrightarrow{d} \max_{k=2,\ldots,m} Z_k(u_1,u_2),$$

where, with an abuse of notation, $\mu_k(u_1,u_2) = \mu_1(u_1,u_2) - \mu_k(u_1,u_2)$, and

$$\mu_j(u_1,u_2) = E(((F_{X^{\vartheta_j^\dagger}_{j,t+\tau}(X_t)}(u_2) - F_{X^{\vartheta_j^\dagger}_{j,t+\tau}(X_t)}(u_1)) - (F_0(u_2|X_t) - F_0(u_1|X_t)))^2),$$

for $j = 1, \ldots, m$, and where $(Z_1(u_1,u_2), \ldots, Z_m(u_1,u_2))$ is an m-dimensional Gaussian random variable for which the associated covariance matrix has kk element given by $W_k(u_1,u_2)$, as in Theorem 1(i).

Critical values for these tests can be obtained using a recursive version of the block bootstrap. When forming block bootstrap samples in the recursive case, observations at the beginning of the sample are used more frequently than observations at the end of the sample. This introduces a location bias to the usual block bootstrap, as under standard resampling with replacement, all blocks from the original sample have the same probability of being selected. Also, the bias term varies across samples and can be either positive or negative, depending on the specific sample. A first-order valid bootstrap procedure for non-simulation-based *m*- estimators constructed using a recursive estimation scheme is outlined in Corradi and Swanson (2007). Here we extend the results of Corradi and Swanson (2007) by establishing asymptotic results for cases in which simulation-based estimators are bootstrapped in a recursive setting.

In order to carry out the bootstrap, begin by resampling b blocks of length l from the full sample, with $lb = T$. For any given τ, it is necessary to jointly resample $X_t, X_{t+1}, \ldots, X_{t+\tau}$. More precisely, let $Z^{t,\tau} = (X_t, X_{t+1}, \ldots, X_{t+\tau})$, $t = 1, \ldots, T - \tau$. Now, resample b overlapping blocks of length l from $Z^{t,\tau}$. This yields $Z^{t,*} = (X_t^*, X_{t+1}^*, \ldots, X_{t+\tau}^*)$, $t = 1, \ldots, T - \tau$. Use these data to construct $\widehat{\vartheta}^*_{k,t,N,h}$. Recall that N is the number of simulated series used to estimate the parameters. Note that as we assume $N/P \to \infty$, simulation error vanishes and there is no need to resample the simulated series. We proceed by assuming the first-order asymptotic validity of the bootstrap estimator, as outlined in the following assumption (in Section 4 we shall provide primitive conditions under which NPSQMLE and satisfies this assumption).

Assumption A5. As $P, R, N \to \infty$ and $h \to 0$, for $k = 1, \ldots, m$:

$$P\left(\omega : \sup_{v\in\Re^{\varrho}}\left|P_T^*\left(\frac{1}{\sqrt{P}}\sum_{t=R}^{T}(\widehat{\vartheta}^*_{k,t,N,h} - \widehat{\vartheta}_{k,t,N,h}) \le v\right) - P\left(\frac{1}{\sqrt{P}}\sum_{t=R}^{T}(\widehat{\vartheta}_{k,t,N,h} - \vartheta_k^\dagger) \le v\right)\right| > \varepsilon\right) \to 0.$$

It can be seen immediately that A5 ensures that $\frac{1}{\sqrt{P}}\sum_{t=R}^{T}(\widehat{\vartheta}^*_{k,t,N,h}-\widehat{\vartheta}_{k,t,N,h})$ has the same limiting distribution as $\frac{1}{\sqrt{P}}\sum_{t=R}^{T}(\widehat{\vartheta}_{k,t,N,h}-\vartheta_k^\dagger)$, conditional on sample, and for all samples except a set with probability measure approaching zero. Given this assumption, the appropriate bootstrap statistic is:

$$D^*_{k,P,N}(u_1,u_2)=\frac{1}{\sqrt{P}}\sum_{t=R}^{T-\tau}\Bigg\{\Bigg(\Bigg[\frac{1}{N}\sum_{i=1}^{N}1\{u_1\le X_{1,t+\tau,i}^{\widehat{\vartheta}^*_{1,t,N,h}}(X_t^*)\le u_2\} - 1\{u_1\le X^*_{t+\tau}\le u_2\}\Bigg]^2 - \Bigg(\frac{1}{T}\sum_{j=1}^{T}\Bigg[\frac{1}{N}\sum_{i=1}^{N}1\{u_1\le X_{1,t+\tau,i}^{\widehat{\vartheta}_{1,t,N,h}}(X_j)\le u_2\} - 1\{u_1\le X_{j+\tau}\le u_2\}\Bigg]^2\Bigg)\Bigg) - \Bigg(\Bigg[\frac{1}{N}\sum_{i=1}^{N}1\{u_1\le X_{k,t+\tau,i}^{\widehat{\vartheta}^*_{k,t,N,h}}(X_t^*)\le u_2\} - 1\{u_1\le X^*_{t+\tau}\le u_2\}\Bigg]^2 - \Bigg(\frac{1}{T}\sum_{j=1}^{T}\Bigg[\frac{1}{N}\sum_{i=1}^{N}1\{u_1\le X_{k,t+\tau,i}^{\widehat{\vartheta}_{k,t,N,h}}(X_j)\le u_2\} - 1\{u_1\le X_{j+\tau}\le u_2\}\Bigg]^2\Bigg)\Bigg)\Bigg\}.$$

Note that each bootstrap term is recentered around the (full) sample mean. This is necessary because the bootstrap statistic is constructed using the last P resampled observations, which in turn have been resampled from the full sample. In particular, this is necessary regardless of the ratio, P/R. Thus, even if $P/R\to 0$, so that there is no need to mimic the parameter estimation error (and hence the above statistic can be constructed using $\widehat{\vartheta}_{k,t,N,h}$ instead of $\widehat{\vartheta}^*_{k,t,N,h}$), it remains the case that recentering of all bootstrap terms around the (full) sample mean is necessary.

Theorem 2. *Let Assumptions A1–A5 hold. Also, assume that models 1 and k are nonnested. If as $P,R,N\to\infty$, $h\to 0$, $P/N\to 0$, $h^2P\to 0$, $l\to\infty$, $l/T^{1/4}\to 0$, and $P/R\to\pi$, where $0<\pi<\infty$, then:*

$$P\left(\omega:\sup_{v\in\Re^{\varrho}}|P_T^*(D^*_{k,P,N}(u_1,u_2)\le v) - P(D_{k,P,N}(u_1,u_2)-\mu_k(u_1,u_2)\le v)|>\varepsilon\right)\to 0.$$

Corollary 2. *Let Assumptions A1–A5 hold. Also, assume that at least one model is nonnested with model 1. If as $P,R,N\to\infty$, $h\to 0$, $P/N\to 0$, $h^2P\to 0$, $l\to\infty$, $l/T^{1/4}\to 0$, and $P/R\to\pi$, where $0<\pi<\infty$, then:*

$$P\left(\omega:\sup_{v\in\Re^{\varrho}}\left|P_T^*\left(\max_{k=2,\ldots,m}D^*_{k,P,N}(u_1,u_2)\le v\right) - P\left(\max_{k=2,\ldots,m}(D_{k,P,N}(u_1,u_2)-\mu_k(u_1,u_2))\le v\right)\right|>\varepsilon\right)\to 0.$$

The above results suggest proceeding in the following manner. For any bootstrap replication, compute the bootstrap statistic (i.e. $D^*_{k,P,N}(u_1,u_2)$ or $\max_{k=2,\ldots,m}D^*_{k,P,N}(u_1,u_2)$). Perform B bootstrap replications (B large) and compute the percentiles of the empirical distribution of the B bootstrap statistics. Reject H_0, if $D_{k,P,N}(u_1,u_2)$ is less than the $(\alpha/2)$th-percentile or greater than the $(1-\alpha/2)$th-percentile of the bootstrap empirical distribution. This provides a test with asymptotic size α and unit asymptotic power. Furthermore, reject H_0' if $\max_{k=2,\ldots,m}D_{k,P,N}(u_1,u_2)$ is greater than the $(1-\alpha)$th-percentile of the bootstrap empirical distribution. Whenever $\mu_1(u_1,u_2)=\mu_k(u_1,u_2)$, for $k=2,\ldots,m$ (i.e., when all competitors are as good as the benchmark), then the asymptotic size is α. However, whenever $\mu_k(u_1,u_2)>\mu_1(u_1,u_2)$ for some k, the bootstrap critical values define upper bounds, and the inference drawn on them is conservative.

3.2. Stochastic volatility models

The test statistic for comparing two models is:

$$DV_{k,P,S,N}(u_1,u_2)=\frac{1}{\sqrt{P}}\sum_{t=R}^{T-\tau}\left(\left(\frac{1}{NS}\sum_{j=1}^{S}\sum_{i=1}^{N}1\{u_1\le X_{1,t+\tau,i,j}^{\widehat{\theta}_{1,t,N,S,h}}(X_t,V_{1,j}^{\widehat{\theta}_{1,t,N,S,h}})\le u_2\} - 1\{u_1\le X_{t+\tau}\le u_2\}\right)^2 - \left(\frac{1}{NS}\sum_{j=1}^{S}\sum_{i=1}^{N}1\{u_1\le X_{k,t+\tau,i,j}^{\widehat{\theta}_{k,t,N,S,h}}(X_t,V_{k,j}^{\widehat{\theta}_{k,t,N,S,h}})\le u_2\} - 1\{u_1\le X_{t+\tau}\le u_2\}\right)^2\right),$$

and the bootstrap test statistic is:

$$DV^*_{k,P,S,N}(u_1,u_2)=\frac{1}{\sqrt{P}}\sum_{t=R}^{T-\tau}\Bigg\{\Bigg(\Bigg[\frac{1}{NS}\sum_{j=1}^{S}\sum_{i=1}^{N}1\{u_1\le X_{1,t+\tau,i,j}^{\widehat{\theta}^*_{1,t,N,S,h}}(X_t^*,V_{1,j}^{\widehat{\theta}^*_{1,t,N,S,h}})\le u_2\} - 1\{u_1\le X^*_{t+\tau}\le u_2\}\Bigg]^2 - \Bigg(\frac{1}{T}\sum_{l=1}^{T}\Bigg[\frac{1}{NS}\sum_{j=1}^{S}\sum_{i=1}^{N}1\{u_1\le X_{1,t+\tau,i,j}^{\widehat{\theta}_{1,t,N,S,h}}(X_l,V_{1,j}^{\widehat{\theta}_{1,t,N,S,h}})\le u_2\} - 1\{u_1\le X_{l+\tau}\le u_2\}\Bigg]^2\Bigg)\Bigg) - \Bigg(\Bigg[\frac{1}{NS}\sum_{j=1}^{S}\sum_{i=1}^{N}1\{u_1\le X_{k,t+\tau,i,j}^{\widehat{\theta}^*_{k,t,N,S,h}}(X_t^*,V_{k,j}^{\widehat{\theta}^*_{k,t,N,S,h}})\le u_2\} - 1\{u_1\le X^*_{t+\tau}\le u_2\}\Bigg]^2 - \Bigg(\frac{1}{T}\sum_{l=1}^{T}\Bigg[\frac{1}{NS}\sum_{j=1}^{S}\sum_{i=1}^{N}1\{u_1\le X_{k,t+\tau,i,j}^{\widehat{\theta}_{k,t,N,S,h}}(X_l,V_{k,j}^{\widehat{\theta}_{k,t,N,S,h}})\le u_2\} - 1\{u_1\le X_{l+\tau}\le u_2\}\Bigg]^2\Bigg)\Bigg)\Bigg\}.$$

Note that we do not need to resample the volatility process, although volatility is simulated under both $\widehat{\theta}_{k,t,N,S,h}$ and $\widehat{\theta}^*_{k,t,N,S,h}$, for $k = 1, \ldots, m$.

Also, $\max_{k=2,\ldots,m} DV_{k,P,N}(u_1, u_2)$ and $\max_{k=2,\ldots,m} DV^*_{k,P,N}(u_1, u_2)$ are defined analogous to their one-factor counterparts.

Assumption A5′. As $P, R, N, S \to \infty$ and $h \to 0$, for $k = 1, \ldots, m$:

$$P\left(\omega : \sup_{v\in\Re^{\varrho}} \left| P_T^*\left(\frac{1}{\sqrt{P}}\sum_{t=R}^{T}\left(\widehat{\theta}^*_{k,t,N,S,h} - \widehat{\theta}_{k,t,N,S,h}\right) \le v\right) - P\left(\frac{1}{\sqrt{P}}\sum_{t=R}^{T}\left(\widehat{\theta}_{k,t,N,S,h} - \theta_k^{\dagger}\right) \le v\right)\right| > \varepsilon\right) \to 0.$$

Theorem 3. *Let Assumptions A1, A2′, A3, and A4′ hold. Also, assume that models 1 and k are nonnested. If as $P, R, S, N \to \infty, h \to 0, P/N \to 0, P/S \to 0, h^2P \to 0$, and $P/R \to \pi$, where $0 < \pi < \infty$, then:* (i) *under H_0, $DV_{k,P,N,S}(u_1, u_2) \xrightarrow{d} N(0, \widetilde{W}_k(u_1, u_2))$, where $\widetilde{W}_k(u_1, u_2)$ has the same format as $W_k(u_1, u_2)$ in the statement of Theorem 1*(i). *Also,*

$$\max_{k=2,\ldots,m}(DV_{k,P,N,S}(u_1, u_2) - \mu(u_1, u_2)) \xrightarrow{d} \max_{k=2,\ldots,m} Z_k(u_1, u_2),$$

where $\mu(u_1, u_2)$ and $Z_k(u_1, u_2)$ are defined as in the statement of Corollary 1; and (ii) *under H_A, for $k = 2, \ldots, m$,* $\Pr(\frac{1}{\sqrt{P}}|DV_{k,P,N,S}(u_1, u_2)| > \varepsilon) \to 1$.

Theorem 4. *Let Assumptions A1, A2′, A3, and A4′–A5′ hold. Also, assume that models 1 and k are nonnested. If as $P, R, S, N \to \infty, h \to 0, P/N \to 0, P/S \to 0, h^2P \to 0, l \to \infty, l/T^{1/4} \to 0$, and $P/R \to \pi$, where $0 < \pi < \infty$, then:*

$$P\left(\omega : \sup_{v\in\Re^{\varrho}} |P_T^*(DV^*_{k,P,N,S}(u_1, u_2) \le v) - P(DV_{k,P,N,S}(u_1, u_2) - \mu_k(u_1, u_2) \le v)| > \varepsilon\right) \to 0,$$

and

$$P\left(\omega : \sup_{v\in\Re^{\varrho}} \left|P_T^*\left(\max_{k=2,\ldots,m} DV^*_{k,P,N,S}(u_1, u_2) \le v\right) - P\left(\max_{k=2,\ldots,m}(DV_{k,P,N,S}(u_1, u_2) - \mu(u_1, u_2)) \le v\right)\right| > \varepsilon\right) \to 0,$$

where $\mu_k(u_1, u_2)$ is defined as in the statement of Corollary 1.

3.3. Further extensions

3.3.1. Out of sample specification tests

In this paper the focus is on the comparison of out of sample predictive accuracy of possible misspecified diffusion models, when the conditional distribution is not known in closed form. One may wonder whether the tools developed here can also allow for the construction of out of sample specification tests, based on recursively estimated parameters. As we briefly outline below, this is indeed possible.

As mentioned in the introduction, specification tests for the conditional distribution of a diffusion, when its closed form is unknown, have also been recently suggested by Aït-Sahalia et al. (2009) and by Bhardwaj et al. (2008). The former test is based on the integrated mean square error of the difference between a local polynomial estimator of the conditional CDF and an exact approximation, based on a Hermite expansion, of the parametric CDF under the null, constructed with historical data. The latter test is based on the comparison of empirical CDFs constructed with historical and simulated data respectively. Both tests require estimated parameters, the former to construct the approximated CDF under the null, and the latter to simulate data. In both cases, parameters are estimated using the full sample, and tests are thus in-sample. Hereafter, to facilitate comparison, set $\tau = 1$.

Aït-Sahalia, Fan and Peng's statistic, Eq. (4.4) in their paper, is based on

$$h_T^{1/2}\sum_{t=1}^{T-1}(\widehat{F}_{T,h_T}(X_{t+1}|X_t) - F(X_{t+1}|X_t; \widehat{\vartheta}_T))^2 w(X_{t+1}, X_t), \tag{7}$$

where $\widehat{F}_{T,h_T}(X_{t+1}|X_t)$ is a local polynomial estimator of the conditional distribution constructed using the full sample, and evaluated at (X_{t+1}, X_t), h_T is a bandwidth parameter, $F(X_{t+1}|X_t; \widehat{\vartheta}_T)$ is an exact approximation of the CDF under the null, evaluated at $\widehat{\vartheta}_T$, and $w(X_{t+1}, X_t)$ is a weighting function. It is straightforward to construct an out of sample version of their test based on recursively estimated parameters. Namely, consider:

$$h_T^{1/2}\sum_{t=R}^{T-1}(\widehat{F}_{t,h_T}(X_{t+1}|X_t) - F(X_{t+1}|X_t; \widehat{\vartheta}_t))^2 w(X_{t+1}, X_t), \tag{8}$$

where $\widehat{F}_{t,h_T}(X_{t+1}|X_t)$ is a recursive local polynomial estimator constructed using observations up to time t, and $\widehat{\vartheta}_t$ are recursively estimated parameters. As the statistic in (8) converges at a nonparametric rate (see their Theorem 3), the contribution of parameter estimation error is always negligible, regardless of the estimation scheme. On the other hand, the asymptotic bias terms may be affected by nonparametric recursive estimation.

Turning to Bhardwaj et al. (2008), their statistic given in their Eqs. (10)–(11) can be written as $V_T = \sup_{u\times v\in U\times V}|V_T(u, v)|$, with

$$V_T(u, v) = \frac{1}{\sqrt{T-1}} \times \sum_{t=1}^{T-1}\left(\frac{1}{N}\sum_{s=1}^{N} 1\{X_{s,t+1}^{\widehat{\vartheta}_{T,N,h}}(X_t) \le u\} - 1\{X_{t+1} \le u\}\right)1\{X_t \le v\},$$

where $X_{s,t+1}^{\widehat{\vartheta}_{T,N,h}}$ is the process simulated using $\widehat{\vartheta}_{T,N,h}$ and with initial value X_t. It is immediate to see that an out of sample version of this test can be written as $V_P = \sup_{u\times v\in U\times V}|V_P(u, v)|$, with

$$V_P(u, v) = \frac{1}{\sqrt{P-1}} \times \sum_{t=R}^{T-1}\left(\frac{1}{N}\sum_{s=1}^{N} 1\{X_{s,t+1}^{\widehat{\vartheta}_{t,N,h}}(X_t) \le u\} - 1\{X_{t+1} \le u\}\right)1\{X_t \le v\},$$

where $\widehat{\vartheta}_{t,N,h}$ are recursively estimated parameters. Thus, the limiting distribution of V_P follows by combining the proof of Theorem 3 in Bhardwaj et al. (2008) with that of Theorem 1 in this paper. Validity of bootstrap critical values and extension to stochastic volatility models also follow using the tools developed in this paper.

3.3.2. Choice of intervals

Thus far, our focus has centered on comparing models over a specific conditional interval. Needless to say, one may choose different models for different intervals. However, if one is interested in comparing predictive accuracy over multiple intervals, one can construct weighted versions of $\max_{k=2,\ldots,m} D_{k,P,N}(u_1, u_2)$ and $\max_{k=2,\ldots,m} DV_{k,P,S,N}(u_1, u_2)$. For notational simplicity, hereafter we limit our attention to the one-factor case, but extension to stochastic volatility models follows by a similar argument. More precisely, let $(-\infty, u_1], (u_1, u_2], \ldots, (u_{J-1}, \infty)$ be a partition of the support of the variable to be predicted, and define

$$D^J_{k,P,N} = \max_{k=2,\ldots,m} \sum_{j=0}^{J-1} D_{k,P,N}(u_j, u_{j+1}) w(u_j, u_{j+1}),$$

where $u_0 = -\infty$, $u_J = \infty$, and $\sum_{j=0}^{J-1} w(u_j, u_{j+1}) = 1$. Of course, $D^J_{k,P,N}$ is not independent of the bounds of the interval, and in fact it depends on the number of intervals considered and on their relative length. Moreover, J should be finite and not too large, otherwise one will be left with fewer observations than needed to construct reliable estimates. Intuitively, this is the price we pay for using statistics that converge at a parametric rate.

Alternatively, if interest lies in approximation of the entire conditional distribution, one can consider the one-sided interval $(-\infty, u]$, and construct the following statistic,

$$D_{k,P,N} = \int_U D_{k,P,N}(u)\phi(u)\mathrm{d}u,$$

where $\phi(u) \geq 0$ and $\int \phi(u)\mathrm{d}u = 1$. If we restrict U to be a compact set in $\mathcal{R}$, say $U = [\underline{u}, \overline{u}]$, then it is not difficult to show that $D_{k,P,N}(u)$ is stochastic equicontinuous on U, and under the conditions of Theorem 1, it follows that $\int_U D_{k,P,N}(u)\phi(u)\mathrm{d}u \xrightarrow{d} \int_U Z_k(u)\phi(u)\mathrm{d}u$, where $Z_k(u)$ is a Gaussian process with covariance kernel $W_k(u)$, as defined in the Appendix. The asymptotic validity of the bootstrap critical values can be established as in Theorem 2, provided the bootstrap counterpart of $D_{k,P,N}(u)$, $D^*_{k,P,N}(u)$, is also stochastic equicontinuous on U.

Finally, one may prefer to minimize the integrated mean square error between the model and the true conditional density. In this case, the null hypothesis is:

$$H_0' : \max_{k=2,\ldots,m} \int_U \int_V [(f_1(u|v;\vartheta_1^\dagger) - f_0(u|v))^2 - (f_k(u|v;\vartheta_k^\dagger) - f_0(u|v))^2] f_0(u,v)\mathrm{d}u\mathrm{d}v \leq 0,$$

where $f_k(u|v;\vartheta_k^\dagger)$ is the conditional density implied by model k, $f_0(u|v)$ is the true conditional density. A test statistic for H_0', requires estimators of $f_1(u|v;\vartheta_1^\dagger)$, $f_k(u|v;\vartheta_k^\dagger)$, and $f_0(u|v)$.

Given recursively estimated parameters $\widehat{\vartheta}_{k,t,N,h}$, one should thus generate a sample of length N, $X_{qh}^{\widehat{\vartheta}_{k,t,N,h}}$ $q = 1, \ldots, Q$ where $Qh = N$, and sample the simulated data at the same frequency as the historical data, in order to get a discrete sample, $X_i^{\widehat{\vartheta}_{t,N,h}}$, $i = 1, \ldots, N$. As $N/T \to \infty$, the initial value effect is negligible. Then, construct a kernel estimator of the conditional density, using both simulated and historical data,

$$\widehat{f}_{k,N}^{\widehat{\vartheta}_{k,t,N,h}}(X_{t+\tau}|X_t) = \frac{\frac{1}{Nh_N^2}\sum_{i=1}^{N-\tau} K\left(\frac{X_{i+\tau}^{\widehat{\vartheta}_{k,t,N,h}} - X_{t+\tau}}{h_N}\right) K\left(\frac{X_i^{\widehat{\vartheta}_{k,t,N,h}} - X_t}{h_N}\right)}{\frac{1}{Nh_N}\sum_{i=1}^{N-\tau} K\left(\frac{X_i^{\widehat{\vartheta}_{k,t,N,h}} - X_t}{h_N}\right)}$$

and

$$\widehat{f}_t(X_{t+\tau}|X_t) = \frac{\frac{1}{Th_T^2}\sum_{j=1}^{t} K\left(\frac{X_{j+\tau} - X_{t+\tau}}{h_T}\right) K\left(\frac{X_j - X_t}{h_T}\right)}{\frac{1}{Th_T}\sum_{j=1}^{t} K\left(\frac{X_j - X_t}{h_T}\right)},$$

$$t = R, \ldots, T-\tau$$

where h_N and h_T are bandwidth parameters. Then, we can test H_0' via a statistics based on:

$$h_T^{1/2}\sum_{t=R}^{T-\tau}[(\widehat{f}_{1,N}^{\widehat{\vartheta}_{1,t,N,h}}(X_{t+\tau}|X_t) - \widehat{f}_t(X_{t+\tau}|X_t))^2 - (\widehat{f}_{k,N}^{\widehat{\vartheta}_{k,t,N,h}}(X_{t+\tau}|X_t) - \widehat{f}_t(X_{t+\tau}|X_t))^2] w(X_{t+\tau}, X_t),$$

with $w(X_{t+\tau}, X_t)$ denoting a proper trimming function.

The study of the asymptotic properties of the statistic above is left to future research.

4. Recursive nonparametric simulated quasi maximum likelihood estimators

In this section we develop a recursive version of the nonparametric simulated (quasi) maximum likelihood (NPSQML) estimator of Fermanian and Salanié (2004) and outline conditions under which asymptotic equivalence between the NPSQML estimator and the corresponding recursive QML estimator obtains, hence ensuring that A4 and A4′ hold. Analogous results are also established for the bootstrap counterpart of the recursive NPSQML estimators.

A previous version of this paper contains results analogous to those reported in this section for the case of exactly identified simulated generalized methods of estimators of Duffie and Singleton (1993).[2]

4.1. One factor models

The idea underlying the nonparametric simulated maximum likelihood estimator of Fermanian and Salanié (2004) is to replace the unknown conditional density with a kernel estimator, constructed using simulated data. Fermanian and Salanié (2004) focus on the case of exogenous conditioning variables, while Kristensen and Shin (2008) consider extensions to (fully observed) Markov models. In the sequel, we extend the estimator of Fermanian and Salanié (2004) and Kristensen and Shin (2008) to the recursive estimation case. In a subsequent section, we outline a bootstrap version of the estimator and establish the first-order validity thereof.

Hereafter, let $f_k(X_t|X_{t-1},\vartheta_k^\dagger)$ be the conditional density implied by model k. If we knew f_k in closed form, we could just estimate $\vartheta_{t,k}^\dagger$ recursively, using standard QML as[3]:

$$\widehat{\vartheta}_{t,k} = \arg\max_{\vartheta_k \in \Theta_k} \frac{1}{t}\sum_{j=2}^{t} \ln f_k(X_t|X_{t-1},\vartheta_k), \quad t = R+1, \ldots, R+P.$$

Now, define:

$$\vartheta_k^\dagger = \arg\max_{\vartheta_k \in \Theta_k} E(\ln f_k(X_t|X_{t-1},\vartheta_k)). \tag{9}$$

Following Kristensen and Shin (2008), generate $T-1$ paths of length one for each simulation replication, using $X_1, \ldots, X_{T-1}$ as starting values and hence construct $X_{k,t,j}^{\vartheta}(X_{t-1})$, for $t = 2, \ldots, T-$

[2] See http://econweb.rutgers.edu/nswanson/papers.htm. We conjecture that one could establish the asymptotic properties of recursive versions and bootstrap analogs for all other simulation-based estimators, such as indirect inference (Gourieroux et al., 1993; Dridi et al., 2007), an efficient method of moment (Gallant and Tauchen, 1996) and simulated GMM with a continuum of moment conditions (Carrasco et al., 2007). We leave this to future research.

[3] Note that as model k is, in general, misspecified, $\sum_{t=1}^{T-1} \ln f_k(X_t|X_{t-1},\vartheta_k)$ is a quasi-likelihood and $\nabla_{\theta_k} \ln f_k(X_t|X_{t-1},\vartheta_k^\dagger)$ is not necessarily a martingale difference sequence.

$1, j = 1, \ldots, N$. Note that we keep the N random draws fixed across different initial values. Then, define the following estimator of the conditional density:

$$\widehat{f}_{k,N,h}(X_t|X_{t-1}, \vartheta_k) = \frac{1}{N\xi_N} \sum_{i=1}^{N} K\left(\frac{X^{\vartheta_k}_{t,i,h}(X_{t-1}) - X_t}{\xi_N}\right).$$

Further, define the recursive NPSQML estimator as follows:

$$\widehat{\vartheta}_{k,t,N,h} = \arg\max_{\vartheta_k \in \Theta_k} \frac{1}{t} \sum_{s=2}^{t} \ln \widehat{f}_{k,N,h}(X_s|X_{s-1}, \vartheta_k) \times \tau_N(\widehat{f}_{k,N,h}(X_s|X_{s-1}, \vartheta_k)), \quad t \geq R,$$

where the trimming function $\tau_N(\widehat{f}_{k,N,h}(X_t|X_{t-1}, \vartheta_k))$ is a positive and increasing function, such that $\tau_N(\widehat{f}_{k,N,h}(X_t, X_{t-1}, \vartheta_k)) = 0$, if $\widehat{f}_{k,N,h}(X_t, X_{t-1}, \vartheta_k) < \xi_N^{\delta}$, and $\tau_N(\widehat{f}_{k,N,h}(X_t, X_{t-1}, \vartheta_k)) = 1$, if $\widehat{f}_{k,N,h}(X_t, X_{t-1}, \vartheta_k) > 2\xi_N^{\delta}$, for some $\delta > 0$.[4] The reason for the trimming parameter is that when the log density is close to zero, the derivative tends to infinity and so even very tiny simulation errors have a large impact on the likelihood. Our result in this subsection requires the following additional assumptions.

Assumption A3′. For $k = 1, \ldots, m$: (i) $X_i^{\vartheta_k}(x)$ and $X_{i,h}^{\vartheta_k}(x)$ are geometrically ergodic and β-mixing, (ii) $\frac{\partial X_i^{\vartheta_k}(x)}{\partial \vartheta_k}$, $\frac{\partial X_i^{\vartheta_k}(x)}{\partial x}$, $\frac{\partial^2 X_i^{\vartheta_k}(x)}{\partial \vartheta_k \partial \vartheta_k'}$, $\frac{\partial^2 X_i^{\vartheta_k}(x)}{\partial \vartheta_k \partial x}$ and $\frac{\partial X_{i,h}^{\vartheta_k}(x)}{\partial \vartheta_k}$, $\frac{\partial X_{i,h}^{\vartheta_k}(x)}{\partial x}$, $\frac{\partial^2 X_{i,h}^{\vartheta_k}(x)}{\partial \vartheta_k \partial \vartheta_k'}$, $\frac{\partial^2 X_{i,h}^{\vartheta_k}(x)}{\partial \vartheta_k \partial x}$ are r-dominated on Θ_k and on $X^{T,a} = \{x : x \leq T^a\}$ for $r > 4$ and $a > 1$.

Assumption A6. Let $\mathcal{N}_{\vartheta_k^\dagger}$ be a neighborhood of $\vartheta_k^\dagger$, $E(\sup_{\vartheta_k \in \mathcal{N}_{\vartheta_k^\dagger}} \| \frac{\partial \ln f_k(X_t|X_{t-1}, \vartheta_k)}{\partial \vartheta_k} \|^r) < \infty$, $E(\sup_{\vartheta_k \in \mathcal{N}_{\vartheta_k^\dagger}} \| \frac{\partial X_i^{\vartheta_k}(X_{t-1})}{\partial \vartheta_k} \|^r) < \infty$, $E(\sup_{\vartheta_k \in \mathcal{N}_{\vartheta_k^\dagger}} \| \frac{\partial X_{i,h}^{\vartheta_k}(X_{t-1})}{\partial \vartheta_k} \|^r) < \infty$, for $k = 1, \ldots, m$ and for $r > 4$.

Assumption A7. For $k = 1, \ldots, m$: (i) $\vartheta_k^\dagger$ is uniquely identified (i.e. $E(\ln f_k(X_t|X_{t-1}, \vartheta_k)) < E(\ln f_k(X_t|X_{t-1}, \vartheta_k^\dagger))$ for any $\vartheta_k \neq \vartheta_k^\dagger$); (ii) $\widehat{\vartheta}_{k,t,N,h}$ and $\vartheta_k^\dagger$ are in the interior of Θ_k, (iii) $f_k(x|x_{-1}, \vartheta_k)$ is $s+1$-continuously differentiable on the interior of Θ_k, $f_k(x|x_{-1}, \vartheta_k)$, $\nabla_x^s f_k(x|x_{-1}, \vartheta_k)$, $\nabla_x^s \nabla_{\vartheta_k} f_k(x|x_{-1}, \vartheta_k)$ are bounded on $\mathcal{R} \times \mathcal{R} \times \Theta_k$, for $s \geq 2$; (iii) the elements of $\nabla_{\vartheta_k} f_k(X_t|X_{t-1}, \vartheta_k)$, $\nabla^2_{\vartheta_k} f_k(X_t|X_{t-1}, \vartheta_k)$, $\nabla_{\vartheta_k} \ln f_k(X_t|X_{t-1}, \vartheta_k)$ and $\nabla^2_{\vartheta_k} \ln f_k(X_t|X_{t-1}, \vartheta_k)$ are r-dominated on Θ_k, with $r > 4$; and (iv) $E(-\nabla^2_{\vartheta_k} \ln f_k(\vartheta_k))$ is positive definite, uniformly on Θ_k.

Assumption A8. The kernel, K, is a symmetric, nonnegative, continuous function with bounded support $[-\Delta, \Delta]$, s-time differentiable on the interior of its support and satisfies: $\int K(u)du = 1$, $\int u^{s-1}K(u)(u)du = 0$, $s \geq 2$. Let $K^{(j)}$ be the j-th derivative of the kernel. Then, $K^{(j)}(-\Delta) = K^{(j)}(\Delta) = 0$, for $j = 1, \ldots, s$, $s \geq 2$.

Theorem 5. *Let Assumptions A1–A2, A3′, and A6–A8 hold. Let $T = R + P$, $P/R \to \pi$, where $0 < \pi < \infty$ and let $N = T^a$ $a > 1$. As $T, P, N \to \infty$, (a) $T^{\frac{1}{2}} |\ln \xi_N|^{\frac{r}{r-1}} \Pr(\inf_{\vartheta \in \mathcal{N}_{\vartheta_k^\dagger}} f(X_j|X_{j-1}, \vartheta) \leq 2\xi_N^{\delta}) \to 0$, (b) $T^{1/2}\xi_N^{s-\delta} |\ln \xi_N| \to 0$, (c) $T^{(1-a)}\xi_N^{-4-2\delta}(\ln \xi_N^2) \ln T^a \to 0$, (d) $T^{1/2}\xi_N^{-(\delta+3)} h |\ln \xi_N^{\delta}| \to 0$. Then, for $k = 1, \ldots, m$: (i) $\sup_{t \geq R}(\widehat{\vartheta}_{k,t,N,h} - \vartheta_k^\dagger) \xrightarrow{p} 0$ and (ii) $\frac{1}{\sqrt{P}} \sum_{t=R}^{T} (\widehat{\vartheta}_{k,t,N,h} - \vartheta_k^\dagger) \xrightarrow{d} N(0, 2\Pi A_k^\dagger V_k^\dagger A_k^\dagger)$, where $A_k^\dagger = E(-\nabla^2_{\theta_k} \ln f_k(X_t|X_{t-1}, \vartheta_k^\dagger))$, $V_k^\dagger = \sum_{i=-\infty}^{\infty} E(\nabla_{\theta_k} \ln f_k(X_2|X_1, \vartheta_k^\dagger) \nabla_{\vartheta_k} \ln f_k(X_{2+i}|X_{1+i}, \vartheta_k^\dagger)')$ and $\Pi = 1 - \pi^{-1}\ln(1+\pi)$.*

As $0 < \pi < \infty$, P grows at the same rate as T, for sake of simplicity, we have stated the rate conditions (a)–(d) in terms of T, instead of a combination of T and P. Note that if we simulate the process using the Euler scheme, instead of the Milstein scheme, the rate condition in (d) should be strengthened to $T^{1/2}\xi_N^{-(d+3)}h^{1/2}|\ln \xi_N^{\delta}| \to 0$.

From Theorem 5 it can be seen immediately that the NPSQML estimator satisfies Assumption A4 and is asymptotically equivalent to the unfeasible QML estimator, which is constructed by maximizing the likelihood of model k. An interesting alternative to nonparametric simulated maximum likelihood estimator has recently been suggested by Altissimo and Mele (2009). Their estimator is based on the minimization of a properly weighted distance between kernel conditional density estimators based on historical and simulated data. For fully observable systems, it is asymptotically equivalent to the maximum likelihood estimator.

Under the rate conditions in Theorem 5, the contribution of simulation error is asymptotically negligible, and thus there is no need to resample the simulated observations. In particular, let $Z^{t,*} = (X_t^*, X_{t+1}^*, \ldots, X_{t+\tau}^*)$, $t = 1, \ldots, T-\tau$ be as outlined in Section 3. For each simulation replication, generate $T-1$ paths of length one, using as starting values $X_1^*, \ldots, X_{T-1}^*$, and so obtaining $X^{\vartheta_k}_{k,t,j}(X^*_{t-1})$, for $t = 2, \ldots, T-1, j = 1, \ldots, N$. Further, let:

$$\widehat{f}^*_{k,N,h}(X_t^*|X_{t-1}^*, \vartheta_k) = \frac{1}{N\xi_N} \sum_{j=1}^{N} K\left(\frac{X^{\vartheta_k}_{t,j,h}(X^*_{t-1}) - X^*_t}{\xi_N}\right).$$

Now, for $t = R, \ldots, R+P-1$, define:

$$\begin{aligned}
\widehat{\vartheta}^*_{k,t,N,h} = \arg\max_{\vartheta_k \in \Theta_k} \frac{1}{t} \sum_{l=2}^{t} \Bigg(& \ln \widehat{f}_{k,N,h}(X_l^*|X_{l-1}^*, \vartheta_k) \\
& \times \tau_N(\widehat{f}_{k,N,h}(X_l^*|X_{l-1}^*, \vartheta_k)) \\
& - \vartheta_k' \Bigg(\frac{1}{T} \sum_{l'=2}^{T} \frac{\nabla_{\vartheta_k} \widehat{f}_{k,N,h}(X_{l'} \mid X_{l'-1}, \vartheta_k)}{\widehat{f}_{k,N,h}(X_{l'}|X_{l'-1}, \vartheta_k)} \Bigg|_{\vartheta_k = \widehat{\vartheta}_{k,t,N,h}} \\
& \times \tau_N(\widehat{f}_{k,N,h}(X_{l'}|X_{l'-1}, \widehat{\vartheta}_{k,t,N,h})) + \tau_N'(\widehat{f}_{k,N,h}(X_{l'}|X_{l'-1}, \vartheta_k)) \\
& \times \nabla_{\vartheta_k} \widehat{f}_{k,N,h}(X_{l'}|X_{l'-1}, \vartheta_k)|_{\vartheta_k = \widehat{\vartheta}_{k,t,N,h}} \\
& \times \ln \widehat{f}_{k,N,h}(X_{l'}|X_{l'-1}, \widehat{\vartheta}_{k,t,N,h}) \Bigg) \Bigg),
\end{aligned}$$

where $\tau_N'(\cdot)$ denotes the derivative of $\tau_N(\cdot)$ with respect to its argument. Note that each term in the simulated likelihood is recentered around the (full) sample mean of the score, evaluated at $\widehat{\vartheta}_{k,t,N,h}$. This ensures that the bootstrap score has mean zero, conditional on the sample. The recentering term requires computation of $\nabla_{\theta_k} \widehat{f}_{k,N,h}(X_{l'}|X_{l'-1}, \widehat{\vartheta}_{k,t,N,h})$, which is not known in closed form. Nevertheless, it can be computed numerically, by simply taking the numerical derivative of the simulated likelihood.

Theorem 6. *Let Assumptions A1–A2, A3′, and A6–A8 hold. Let $T = R + P$, $P/R \to \pi$, where $0 < \pi < \infty$ and let $N = T^a$ $a > 1$. As $T, N, l \to \infty$, $l/T^{1/4} \to 0$, and (a) $T^{\frac{1}{2}} |\ln \xi_N|^{\frac{r}{r-1}} \Pr(\inf_{\vartheta \in \mathcal{N}_{\vartheta_k^\dagger}}$*

[4] As an example of a trimming function, Fermanian and Salanié (2004) suggest using:

$$\tau_N(x) = \frac{4(x - a_N)^3}{a_N^3} - \frac{3(x - a_N)4}{a_N^4},$$

for $a_N \leq x \leq 2a_N$.

$f(X_j|X_{j-1},\vartheta) \le 2\xi_N^\delta) \to 0$, (b) $T^{1/2}\xi_N^{s-\delta}|\ln\xi_N| \to 0$, (c) $T^{(1-a)}\xi_N^{-4-2\delta}(\ln\xi_N^2)\ln T^a \to 0$, (d) $T^{1/2}\xi_N^{-(\delta+3)}h|\ln\xi_N^\delta| \to 0$. *Then, for* $k=1,\ldots,m$:

$$P\left(\omega : \sup_{v\in\Re^\varrho}\left|P_T^*\left(\frac{1}{\sqrt{P}}\sum_{t=R}^{T}(\widehat{\vartheta}^*_{k,t,N,h}-\widehat{\vartheta}_{k,t,N,h})\le v\right) - P\left(\frac{1}{\sqrt{P}}\sum_{t=R}^{T}(\widehat{\vartheta}_{k,t,N,h}-\vartheta_k^\dagger)\le v\right)\right|>\varepsilon\right)\to 0,$$

where P_T^* denotes the probability law of the resampled series, conditional on the (entire) sample.

Thus, $\frac{1}{\sqrt{P}}\sum_{t=R}^{T}(\widehat{\vartheta}^*_{k,t,N,h}-\widehat{\vartheta}_{k,t,N,h})$ has the same limiting distribution as $\frac{1}{\sqrt{P}}\sum_{t=R}^{T}(\widehat{\vartheta}_{k,t,N,h}-\vartheta_k^\dagger)$, conditional on the sample, and for all samples except a set with probability measure approaching zero, and A5 is satisfied by the bootstrap NPSQML estimator.

4.2. Stochastic volatility models

Since volatility is not observable, we cannot proceed as in the single factor case. Instead, let $V_s^{\theta_k}$ be generated according to (4), setting $qh=s$, $q=1,\ldots,1/h$, and $s=1,\ldots,S$. For each model $k=1,\ldots,m$, and at each simulation replication, $i=1,\ldots,N$, generate S paths of length one, using X_{t-1} as the starting value for the observable, and using S different starting values for the unobservable volatility (i.e., $V_s^{\theta_k}$, $s=1,\ldots,S$). Thus, for any $t=1,\ldots,T-1$, and for any set i, $i=1,\ldots,N$ of random errors $\epsilon_{1,t+(q+1)h,i}$ and $\epsilon_{2,t+(q+1)h,i}$, $q=1,\ldots,1/h$, generate S different values for the observable at time $t+1$, each of them corresponding to a different starting value for the unobservable. Note that it is important to use the same set of random errors $\epsilon_{1,t+(q+1)h,i}$ and $\epsilon_{2,t+(q+1)h,i}$ across different initial values for volatility. Using (5) and (6), generate $X_{t,i}^{\theta_k}(X_t,V_s^{\theta_k})$ for $t=2,\ldots,T$, $i=1,\ldots,N$ and $s=1,\ldots,S$. Now, define:

$$\widehat{f}_{k,N,S,h}(X_t|X_{t-1},\theta_k) = \frac{1}{S}\sum_{s=1}^{S}\frac{1}{N\xi_N}\times\sum_{i=1}^{N}K\left(\frac{X_{t,i,h}^{\theta_k}(X_{t-1},V_s^{\theta_k})-X_t}{\xi_N}\right),$$

and note that by averaging over the initial values for the unobservable volatility, its effect is integrated out. Finally, define:

$$\widehat{\theta}_{k,t,N,S,h} = \arg\min_{\theta_k\in\Theta_k}\frac{1}{t}\sum_{l=2}^{t}\ln\widehat{f}_{k,N,S,h}(X_l|X_{l-1},\theta_k)\times\tau_N(\widehat{f}_{k,N,S,h}(X_l|X_{l-1},\theta_k)),\quad t\ge R.$$

Before establishing the asymptotic properties of $\widehat{\theta}_{k,t,N,S,h}$, we need another assumption.

Assumption A9. Let $\mathcal{N}_{\theta_k^\dagger}$ be a neighborhood of $\theta_k^\dagger$, $E(\sup_{\theta_k\in\mathcal{N}_{\theta_k^\dagger}}\|\frac{\partial X_i^{\theta_k}(X_{t-1},V_j^{\theta_k})}{\partial\theta_k}\|^r)<\infty$, $E(\sup_{\theta_k\in\mathcal{N}_{\theta_k^\dagger}}\|\frac{\partial X_{i,h}^{\theta_k}(X_{t-1},V_j^{\theta_k})}{\partial\theta_k}\|^r)<\infty$, for $k=1,\ldots,m$ and for $r>4$.

Theorem 7. *Let Assumptions A1, A2′–A3′, and A6–A9 hold. Let $T=R+P$, $P/R\to\pi$, where $0<\pi<\infty$, and let $N=T^a$, $a>1$. As $T,N,S\to\infty$,*
(a) $T^{\frac12}|\ln\xi_N|^{\frac{r}{r-1}}\Pr(\inf_{\vartheta\in\mathcal{N}_{\vartheta_k^\dagger}}f(X_j|X_{j-1},\vartheta)\le 2\xi_N^\delta)\to 0$, (b) $T^{1/2}\xi_N^{s-\delta}|\ln\xi_N|\to 0$, (c) $T^{(1-a)}\xi_N^{-4-2\delta}(\ln\xi_N^2)\ln T^a\to 0$, (d) $T^{1/2}\xi_N^{-(\delta+3)}h|\ln\xi_N^\delta|\to 0$, (e) $T^{1/2}S^{-1/2}\xi_N^{-2(1+\delta)}\to 0$. *Then for* $k=1,\ldots,m$:

(i) $\sup_{t\ge R}(\widehat{\theta}_{k,t,N,S,h}-\theta_k^\dagger)\xrightarrow{p}0$ *and* (ii) $\frac{1}{\sqrt{P}}\sum_{t=R}^{T}(\widehat{\theta}_{k,t,N,S,h}-\theta_k^\dagger)\xrightarrow{d}N(0,2\Pi A_k^\dagger V_k^\dagger A_k^\dagger)$, *where* $A_k^\dagger=E(-\nabla^2_{\theta_k}\ln f_k(X_t|X_{t-1},\theta_k^\dagger))$, $V_k^\dagger=\sum_{i=-\infty}^{\infty}E(\nabla_{\theta_k}\ln f_k(X_2|X_1,\theta_k^\dagger)\nabla_{\theta_k}\ln f_k(X_{2+i}|X_{1+i},\theta_k^\dagger)')$, *and* $\Pi=1-\pi^{-1}\ln(1+\pi)$.

Note that in this case, X_t is no longer Markov (i.e., X_t and V_t are jointly Markovian, but X_t is not). Therefore, even in the case in which model k is the true data generating process, the joint likelihood cannot be expressed as the product of the conditional and marginal distributions. Thus, $\widehat{\theta}_{k,t,N,S,h}$ is necessarily a QML estimator. Furthermore, note that $\nabla_{\theta_k}\ln f(X_t|X_{t-1},\theta_k^\dagger)$ is no longer a martingale difference sequence; therefore, we need to use HAC robust covariance matrix estimators, regardless of whether k is the "correct" model or not.

Note that for the bootstrap counterpart of $\widehat{\theta}_{k,t,N,S,h}$, since $S/T\to\infty$ and $N/T\to\infty$, the contribution of simulation error is asymptotically negligible. Hence, there is no need to resample the simulated observations or the simulated initial values for volatility. Define:

$$\widehat{f}_{k,N,S,h}(X_t^*|X_{t-1}^*,\theta_k) = \frac{1}{S}\sum_{s=1}^{S}\frac{1}{N\xi_N}\times\sum_{i=1}^{N}K\left(\frac{X_{t,i}^{\theta_k}(X_{t-1}^*,V_{s'-1}^{\theta_k})-X_t^*}{\xi_N}\right).$$

Now, for $t=R,\ldots,R+P-1$, define:

$$\begin{aligned}\widehat{\theta}^*_{k,t,N,S,h} = \arg\max_{\theta_k\in\Theta_k}\frac{1}{t}\sum_{l=2}^{t}\Bigg(&\ln\widehat{f}_{k,N,S,h}(X_l^*|X_{l-1}^*,\theta_k)\\ &\times\tau_N(\widehat{f}_{k,N,S,h}(X_l^*|X_{l-1}^*,\theta_k))\\ &-\theta_k'\Bigg(\frac{1}{T}\sum_{l'=2}^{T}\frac{\nabla_{\theta_k}\widehat{f}_{k,N,S,h}(X_{l'}|X_{l'-1},\theta_k)}{\widehat{f}_{k,N,h}(X_{l'}^*|X_{l'-1}^*,\theta_k)}\Bigg|_{\widehat{\theta}_{k,t,N,h}}\\ &\times\tau_N(\widehat{f}_{k,N,S,h}(X_{l'}|X_{l'-1},\widehat{\theta}_{k,t,N,S,h}))\\ &+\tau_N'(\widehat{f}_{k,N,Sh}(X_{l'}|X_{l'-1},\widehat{\theta}_{k,t,N,S,h}))\\ &\times\nabla_{\vartheta_k}\widehat{f}_{k,N,S,h}(X_{l'}|X_{l'-1},_{\widehat{\theta}_{k,t,N,h}})\\ &\times\ln\widehat{f}_{k,N,S,h}(X_{l'}|X_{l'-1},\widehat{\theta}_{k,t,N,h})\Bigg)\Bigg),\end{aligned}$$

where $\tau_N'(\cdot)$ denotes the derivative with respect to its argument. We have:

Theorem 8. *Let Assumptions A1, A2′–A3′, and A6–A9 hold. Let $T=R+P$, $P/R\to\pi$, where $0<\pi<\infty$, and let $N=T^a$, $a>1$. As $T,N,S,l\to\infty$, $l/T^{1/4}\to 0$, and* (a) $T^{\frac12}|\ln\xi_N|^{\frac{r}{r-1}}\Pr(\inf_{\vartheta\in\mathcal{N}_{\vartheta_k^\dagger}}f(X_j|X_{j-1},\vartheta)\le 2\xi_N^\delta)\to 0$, (b) $T^{1/2}\xi_N^{s-\delta}|\ln\xi_N|\to 0$, (c) $T^{(1-a)}\xi_N^{-4-2\delta}(\ln\xi_N^2)\ln T^a\to 0$, (d) $T^{1/2}\xi_N^{-(\delta+3)}h|\ln\xi_N^\delta|\to 0$, (e) $T^{1/2}S^{-1/2}\xi_N^{-2(1+\delta)}\to 0$. *Then, for* $k=1,\ldots,m$:

$$P\left(\omega : \sup_{v\in\Re^\varrho}\left|P_T^*\left(\frac{1}{\sqrt{P}}\sum_{t=R}^{T}(\widehat{\theta}^*_{k,t,N,S,h}-\widehat{\theta}_{k,t,N,S,h})\le v\right) - P\left(\frac{1}{\sqrt{P}}\sum_{t=R}^{T}(\widehat{\theta}_{k,t,N,S,h}-\vartheta^\dagger)\le v\right)\right|>\varepsilon\right)\to 0,$$

where P_T^* denotes the probability law of the resampled series, conditional on the (entire) sample.

Table 1
Predictive density model selection test results sample period January 6, 1989–December 31, 1998 (CIR model is the benchmark, bootstrap block length = 5).

τ	u_1, u_2	$DV^{\mathrm{Max}}_{k,P,S,N}(u_1,u_2)$	$PDMSFE_{CIR}$	$PDMSFE_{SV}$	$PDMSFE_{SVJ}$	5% CV	10% CV	15% CV	20% CV
1	$\overline{X} \pm 0.5\sigma_X$	2.82927*	5.66205	3.62009	2.83278	1.76793	1.65848	1.59048	1.53149
	$\overline{X} \pm \sigma_X$	1.31996	1.58636	0.3691	0.2664	1.78705	1.64695	1.57157	1.5188
2	$\overline{X} \pm 0.5\sigma_X$	1.57134*	4.13194	2.62781	2.56061	0.95374	0.85015	0.81374	0.77364
	$\overline{X} \pm \sigma_X$	0.53925	0.85434	0.34105	0.31509	0.88404	0.8354	0.7433	0.67953
3	$\overline{X} \pm 0.5\sigma_X$	0.80223*	4.26257	3.87959	3.46034	0.23338	0.20535	0.19317	0.16539
	$\overline{X} \pm \sigma_X$	1.19189*	1.82012	0.93572	0.62823	0.48909	0.40461	0.36703	0.30468
4	$\overline{X} \pm 0.5\sigma_X$	1.23058*	4.32896	3.82788	3.09838	0.34424	0.28591	0.22947	0.21701
	$\overline{X} \pm \sigma_X$	0.48079*	1.02194	0.76792	0.54115	0.32672	0.28204	0.22131	0.20073
5	$\overline{X} \pm 0.5\sigma_X$	−0.00077	3.71976	3.72053	3.97788	0.25028	0.2032	0.17763	0.16541
	$\overline{X} \pm \sigma_X$	0.18502	1.09725	1.01962	0.91223	0.2864	0.2164	0.19567	0.14872
6	$\overline{X} \pm 0.5\sigma_X$	1.52213*	4.949	3.83724	3.42687	0.11366	0.08187	0.07064	0.05948
	$\overline{X} \pm \sigma_X$	0.58406*	1.63659	1.05253	1.18955	0.16156	0.12362	0.11468	0.10462
12	$\overline{X} \pm 0.5\sigma_X$	0.56293*	4.58393	4.37846	4.021	0.03752	0.03085	0.02742	0.01931
	$\overline{X} \pm \sigma_X$	0.41295*	1.30048	1.5585	0.88753	0.02381	0.01912	0.01574	0.01425

Notes: Numerical entries in the table are test statistics, predictive density type *PDMSFE*s (see Section 7 for further discussion), and associated bootstrap critical values, constructed using intervals given in the second column of the table, and for predictive horizons, $\tau = 1, 2, 3, 4, 5, 6, 12$. Starred entries denote rejection of the null hypothesis that the CIR model yields predictive densities at least as accurate as the competitor SV and SVJ models. Weekly data are used in all estimations, and the sample period across which predictive densities are constructed is $T/2$, where T is the sample size. Predictive densities are constructed using simulations of length $S = 10T$. Empirical bootstrap distributions are constructed using 100 bootstrap replications, and critical values are reported for the 95th, 90th, 85th, and 80th percentiles of the bootstrap distribution. $\overline{X}$ and σ_X are the mean and variance of an initial sample of data used in the first in-sample estimation, prior to the construction of the first predictive density (i.e., using $T/2$ observations). Finally, the predictive density type "mean square forecast errors" (*MSFE*s) reported in the fourth through sixth columns of the table are defined above and reported entries are multiplied by $P^{1/2}$, where $P = T/2$ is the *ex-ante* prediction period.

Table 2
Predictive density model selection test results sample period January 6, 1989–December 31, 1998 (CIR model is the benchmark, bootstrap block length = 10).

τ	u_1, u_2	$DV^{\mathrm{Max}}_{k,P,S,N}(u_1,u_2)$	$PDMSFE_{CIR}$	$PDMSFE_{SV}$	$PDMSFE_{SVJ}$	5% CV	10% CV	15% CV	20% CV
1	$\overline{X} \pm 0.5\sigma_X$	2.82927*	5.66205	3.62009	2.83278	2.00777	1.87189	1.79275	1.74894
	$\overline{X} \pm \sigma_X$	1.31996	1.58636	0.3691	0.2664	2.04287	1.94914	1.92829	1.82353
2	$\overline{X} \pm 0.5\sigma_X$	1.57134*	4.13194	2.62781	2.56061	1.20729	1.12574	1.09287	1.01652
	$\overline{X} \pm \sigma_X$	0.53925	0.85434	0.34105	0.31509	1.18983	1.12383	1.02568	0.93639
3	$\overline{X} \pm 0.5\sigma_X$	0.80223*	4.26257	3.87959	3.46034	0.30797	0.26336	0.23572	0.21822
	$\overline{X} \pm \sigma_X$	1.19189*	1.82012	0.93572	0.62823	0.72656	0.61716	0.5816	0.5347
4	$\overline{X} \pm 0.5\sigma_X$	1.23058*	4.32896	3.82788	3.09838	0.39022	0.31387	0.28829	0.27063
	$\overline{X} \pm \sigma_X$	0.48079*	1.02194	0.76792	0.54115	0.52736	0.45501	0.41484	0.37745
5	$\overline{X} \pm 0.5\sigma_X$	−0.00077	3.71976	3.72053	3.97788	0.20617	0.18285	0.16524	0.13619
	$\overline{X} \pm \sigma_X$	0.18502	1.09725	1.01962	0.91223	0.36255	0.29925	0.2721	0.22753
6	$\overline{X} \pm 0.5\sigma_X$	1.52213*	4.949	3.83724	3.42687	0.11792	0.10103	0.08588	0.08082
	$\overline{X} \pm \sigma_X$	0.58406*	1.63659	1.05253	1.18955	0.1695	0.14107	0.12773	0.09614
12	$\overline{X} \pm 0.5\sigma_X$	0.56293*	4.58393	4.37846	4.021	0.05866	0.04347	0.03611	0.03507
	$\overline{X} \pm \sigma_X$	0.41295*	1.30048	1.5585	0.88753	0.03615	0.03183	0.02711	0.02122

Notes: see Table 1.

5. Empirical illustration: choosing between CIR, SV, and SVJ models

In this section, we choose between Cox–Ingersoll–Ross (CIR), stochastic volatility (SV) and stochastic volatility with jumps (SVJ) models by comparing the models' predictive performance across two different sample periods. Our primary objective is to illustrate the implementation of our test statistics and our secondary objective is to assess whether the choice of model is impacted by the choice of sample period. There are many precedents in the empirical literature suggesting that evaluation of subsample robustness is an important issue when evaluating models. For example, see Bandi and Reno (2008), who compare their semiparametric estimates of a jump diffusion for S&P500 returns to a less general affine model estimated by Eraker et al. (2003). In their analysis, the alternative models are rather similar, but they use different sample periods and different variance filtering methods. In our example, we use the same estimation method for different models across different estimation periods. In particular, we consider two samples of weekly data, one from January 6, 1989–December 31, 1998 (526 observations) and one from January 8, 1999–April 30, 2008 (491 observations), chosen arbitrarily. The variable that we model is the effective (or market) federal funds rate (i.e., the interbank interest rate), measured at the close.

In our analysis, we use the three models implemented in Bhardwaj et al. (2008). Other than considering similar models, our empirical illustration is quite different from theirs. Namely, they report on *in-sample* Kolmogorov type consistent specification tests for individual models, while we report the model selection type test statistics and related forecast error measures discussed in this paper. More specifically, we jointly compare the *out-of-sample* predictive accuracy of various models using recursively estimated models and recursively constructed predictive densities. The three models that we examine are:

CIR: $\mathrm{d}X(t) = \kappa_1(\alpha_1 - X(t))\mathrm{d}t + \gamma_1\sqrt{X(t)}\mathrm{d}W_1(t)$, where $\kappa_1 > 0$, $\gamma_1 > 0$ and $2\kappa_1\alpha_1 \geq \gamma_1^2$,

SV: $\mathrm{d}X(t) = \kappa_2(\alpha_2 - X(t))\mathrm{d}t + \sqrt{V(t)}\mathrm{d}W_r(t)$, and $\mathrm{d}V(t) = \kappa_3(\alpha_3 - V(t))\mathrm{d}t + \gamma_2\sqrt{V(t)}\mathrm{d}W_v(t)$, where $W_r(t)$ and $W_v(t)$ are independent Brownian motions, and where $\kappa_2 > 0$, $\kappa_3 > 0$, $\gamma_2 > 0$, and $2\kappa_3\alpha_3 \geq \gamma_2^2$.

SVJ: $\mathrm{d}X(t) = \kappa_4(\alpha_4 - X(t))\mathrm{d}t + \sqrt{V(t)}\mathrm{d}W_{rJ}(t) + J_u\mathrm{d}q_u - J_d\mathrm{d}q_d$, and $\mathrm{d}V(t) = \kappa_5(\alpha_5 - V(t))\mathrm{d}t + \gamma_3\sqrt{V(t)}\mathrm{d}W_{vJ}(t)$, where $W_{rJ}(t)$ and $W_{vJ}(t)$ are independent Brownian motions, and where $\kappa_4 > 0$, $\kappa_5 > 0$, $\gamma_3 > 0$, and $2\kappa_5\alpha_5 \geq \gamma_3^2$. Further q_u and q_d are Poisson processes with jump intensity λ_u and λ_d, and are independent of the Brownian motions $W_1(t)$ and $W_2(t)$. Jump sizes are *iid* and are controlled by jump magnitudes $\zeta_u, \zeta_d > 0$, which are drawn from

Table 3
Predictive density model selection test results sample period January 8, 1999–April 30, 2008 (CIR model is the benchmark, bootstrap block length = 5).

τ	u_1, u_2	$DV^{Max}_{k,P,S,N}(u_1, u_2)$	$PDMSFE_{CIR}$	$PDMSFE_{SV}$	$PDMSFE_{SVJ}$	5% CV	10% CV	15% CV	20% CV
1	$\overline{X} \pm 0.5\sigma_X$	3.36528*	3.93191	0.56663	2.35979	2.4573	2.31001	2.17511	2.05169
	$\overline{X} \pm \sigma_X$	0.39113	0.39172	0.00059	0.13535	2.09495	1.99902	1.93683	1.84544
2	$\overline{X} \pm 0.5\sigma_X$	1.8218*	2.32377	0.50197	2.04596	1.82588	1.71781	1.64691	1.55461
	$\overline{X} \pm \sigma_X$	0.59514	0.60979	0.01464	0.26331	2.182	2.09447	1.99572	1.93641
3	$\overline{X} \pm 0.5\sigma_X$	1.2709	1.86856	0.59766	2.29788	1.47533	1.33248	1.19701	1.11857
	$\overline{X} \pm \sigma_X$	0.97425	1.04645	0.0722	0.46272	1.98624	1.77604	1.71385	1.63308
4	$\overline{X} \pm 0.5\sigma_X$	1.33461*	1.86611	0.5315	2.50816	1.18714	1.03895	0.92443	0.74572
	$\overline{X} \pm \sigma_X$	0.59446	0.78217	0.18771	0.23341	1.44947	1.31151	1.23566	1.18198
5	$\overline{X} \pm 0.5\sigma_X$	1.55731*	1.92318	0.36586	2.3208	0.94807	0.72157	0.63611	0.56305
	$\overline{X} \pm \sigma_X$	0.62454*	0.92698	0.30244	0.42899	1.12818	0.91251	0.81989	0.69776
6	$\overline{X} \pm 0.5\sigma_X$	1.07981	1.5355	0.45569	2.23224	0.90627	0.81358	0.58599	0.49386
	$\overline{X} \pm \sigma_X$	1.0877*	1.3928	0.39654	0.3051	1.11448	0.88946	0.69749	0.57532
12	$\overline{X} \pm 0.5\sigma_X$	1.06647*	1.72738	0.66091	2.59892	0.96992	0.7709	0.65347	0.54271
	$\overline{X} \pm \sigma_X$	0.74472*	0.9282	0.43853	0.18348	0.93258	0.73613	0.59269	0.4251

Notes: see Table 1.

Table 4
Predictive density model selection test results sample period January 8, 1999–April 30, 2008 (CIR model is the benchmark, bootstrap block length = 10).

τ	u_1, u_2	$DV^{Max}_{k,P,S,N}(u_1, u_2)$	$PDMSFE_{CIR}$	$PDMSFE_{SV}$	$PDMSFE_{SVJ}$	5% CV	10% CV	15% CV	20% CV
1	$\overline{X} \pm 0.5\sigma_X$	3.36528*	3.93191	0.56663	2.35979	3.22922	2.79456	2.66332	2.49582
	$\overline{X} \pm \sigma_X$	0.39113	0.39172	0.00059	0.13535	2.49945	2.30575	2.18381	2.15431
2	$\overline{X} \pm 0.5\sigma_X$	1.8218	2.32377	0.50197	2.04596	2.97083	2.41921	2.29894	2.2163
	$\overline{X} \pm \sigma_X$	0.59514	0.60979	0.01464	0.26331	2.82514	2.67829	2.64444	2.55817
3	$\overline{X} \pm 0.5\sigma_X$	1.2709	1.86856	0.59766	2.29788	2.51858	2.25422	2.06351	1.93476
	$\overline{X} \pm \sigma_X$	0.97425	1.04645	0.0722	0.46272	2.98617	2.8359	2.75257	2.59837
4	$\overline{X} \pm 0.5\sigma_X$	1.33461	1.86611	0.5315	2.50816	2.14655	1.91697	1.73401	1.59074
	$\overline{X} \pm \sigma_X$	0.59446	0.78217	0.18771	0.23341	2.72152	2.56512	2.49455	2.37684
5	$\overline{X} \pm 0.5\sigma_X$	1.55731	1.92318	0.36586	2.3208	1.9112	1.80572	1.4376	1.33975
	$\overline{X} \pm \sigma_X$	0.62454	0.92698	0.30244	0.42899	2.57883	2.30651	2.14454	1.96686
6	$\overline{X} \pm 0.5\sigma_X$	1.07981	1.5355	0.45569	2.23224	2.11693	1.64939	1.47409	1.34432
	$\overline{X} \pm \sigma_X$	1.0877	1.3928	0.39654	0.3051	2.37199	2.08945	1.83042	1.71404
12	$\overline{X} \pm 0.5\sigma_X$	1.06647*	1.72738	0.66091	2.59892	1.36719	1.00359	0.8389	0.57706
	$\overline{X} \pm \sigma_X$	0.74472	0.9282	0.43853	0.18348	1.77444	0.98574	0.75872	0.54984

Notes: see Table 1.

exponential distributions, with densities: $f(J_u) = \frac{1}{\zeta_u}\exp(-\frac{J_u}{\zeta_u})$ and $f(J_d) = \frac{1}{\zeta_d}\exp(-\frac{J_d}{\zeta_d})$. Here, λ_u is the probability of a jump up, $\Pr(dq_u(t) = 1) = \lambda_u$, and jump up size is controlled by J_u; while λ_d and J_d control jump down intensity and size. Note that the case of Poisson jumps with constant intensity and jump size with exponential density is covered by the assumptions stated in the previous sections.

Note that the CIR model is neither nesting the SVJ and the SV models nor is nested in either of them. On the other hand, SV is clearly nested in SVJ.

The tests that we construct are $D^{Max}_{k,P,N}(u_1, u_2)$ and $DV^{Max}_{k,P,S,N}(u_1, u_2)$. In our tables, we also report the so-called "predictive density" mean square forecast error (*PDMSFE*) terms in these statistics, which are constructed using the following formulae:

$$\frac{1}{P}\sum_{t=R}^{T-\tau}\left(\frac{1}{NS}\sum_{j=1}^{S}\sum_{i=1}^{N}1\{u_1 \le X^{\widehat{\theta}_{1,t,N,S,h}}_{1,t+\tau,i,j}(X_t, V^{\widehat{\theta}_{1,t,N,S,h}}_{1,j}) \le u_2\} - 1\{u_1 \le X_{t+\tau} \le u_2\}\right)^2$$

and

$$\frac{1}{P}\sum_{t=R}^{T-\tau}\left(\frac{1}{N}\sum_{i=1}^{N}1\{u_1 \le X^{\widehat{\vartheta}_{1,t,N,h}}_{1,t+\tau,i}(X_t) \le u_2\} - 1\{u_1 \le X_{t+\tau} \le u_2\}\right)^2,$$

depending upon whether we are predicting using one factor or SV models. We define the CIR model to be our "benchmark", against which the other models are compared. Thus, both competitors are neither nested in nor nesting the benchmark. For the estimation of parameters as well as the construction of predictive densities, data were generated using the Milstein scheme discussed above, with $h = 1/T$, where T is the sample size. The jump component in our SVJ model was simulated without any error because of the constancy of the intensity parameter. The three models fall in the class of affine diffusions. Therefore, it is possible to compute parameter estimates using the conditional characteristic function (see Singleton (2001) for the CIR model, Jiang and Knight (2002) for the SV model, and Chacko and Viceira (2003) for the SVJ model). We leave analysis of the predictive accuracy of the models discussed herein under different estimation methods to future research. All parameters are estimated recursively, all empirical bootstrap distributions are constructed using 500 bootstrap replications, and critical values are reported for the 95th,90th,85th, and 80th percentiles of the relevant bootstrap empirical distributions. For the bootstrap, block lengths of 5 and 10 are reported on. Additionally, we set $S = 1000$, and for model SV and SVJ we set $N = S$. Tests were carried out based on the construction of τ*-step* ahead predictive densities and associated confidence intervals, for $\tau = \{1, 2, 3, 4, 5, 6, 12\}$. We set (u_1, u_2) equal to $\overline{X} \pm 0.5\sigma_X$, and $\overline{X} \pm \sigma_X$, where $\overline{X}$ and σ_X are the mean and variance of an initial sample of data.

Test statistic values, *PDMSFE*s, and bootstrap critical values are reported for various u_1, u_2 combinations, forecast horizons, and bootstrap block lengths in Tables 1–4. The first two tables report

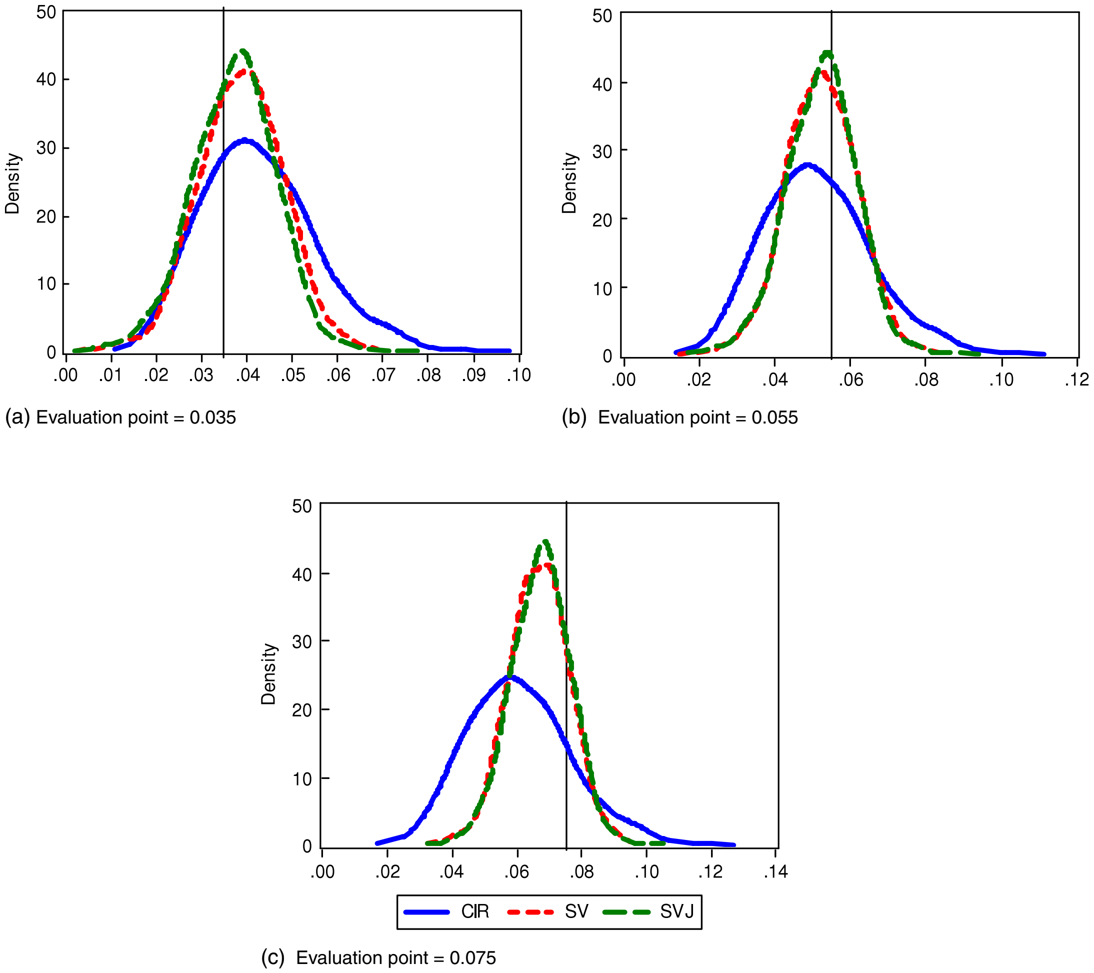

Fig. 1. Predictive densities for CIR, SV and SVJ models — 01:1989–12:1998.

results for the sample period January 6, 1989–December 31, 1998, while Tables 3 and 4 report results for the sample period January 8, 1999–April 30, 2008. Interestingly, a number of very clear-cut conclusions emerge. In particular, *PDMSFE*s are lower for the SVJ model in 12 of 14 cases in Table 1. Moreover, in the two cases where SVJ is not "*PDMSFE*-best", there is little to choose between the *PDMSFEs* of the different models. Perhaps not surprisingly, then, the null hypothesis that the CIR model yields predictive densities at least as accurate as the two competitor models is rejected in almost all cases, at a 95% level of confidence. (Starred entries in the tables denote rejection using CVs equal to the 95th percentile of the empirical bootstrap distributions.) Notice also that although bootstrap CVs increase in magnitude when a longer block length is used (see Table 2), the number of rejections of the null hypothesis remains the same, suggesting that our findings, thus far, are somewhat robust to bootstrap block length.

Turning now to Table 3, note that it is now the SV model that yields the "*PDMSFE*-best" predictive densities in all but two cases. Moreover, in the two cases that SV does not "win", the SVJ model "wins", albeit with only marginally lower *PDMSFE*s. However, significant rejection of the null only occurs in 8 of 14 cases based on the more recent sample of data used in construction of the statistics reported in Tables 3 and 4, rather than 10 cases, as in Tables 1 and 2. Moreover, when the block length is increased from 5 to 10, the number of rejections of the null deceases almost to zero (see Table 4). Thus, while the point *PDMSFE* is lower in 12 of 14 cases, it is more difficult to discern a statistically significant difference between the SV and the CIR model when using data from 1999–2008. Two points are worth mentioning in this regard. First, in Tables 3 and 4, the absolute magnitude of the SV *PDMSFEs* are actually substantively lower than those for the CIR model, when comparing CIR and SV models, just as they were when comparing CIR and SVJ models in Tables 1 and 2, suggesting that the reduction in rejections when increasing the block length in Table 4 may be due in part to size bias in the case of the longer block length. Second, and more important, regardless of the above findings, it is very clear that the selection of *PDMSFE*-best model is indeed dependent upon the sample period used to construct predictive densities. While the one factor model generally performs worse than the other two models, whether or not jumps improve model performance depends on the sample period being investigated. Thus, different sample periods do not result in the same model being chosen, which is not surprising, given that the extant empirical evidence concerning which model to use when examining interest rates is rather mixed.[5]

In Figs. 1 and 2, predictive densities are plotted for various evaluation points given a particular set of recursively estimated

[5] Note that SV is nested by SVJ. However, it neither nests or is nested by *CIR*, and hence the "nestedness" assumptions made in the statements of the theorems and corollaries above are not violated. Additionally, note that one might be tempted to think that if there is a model that outperforms *CIR*, this should be SVJ, as SVJ nests SV. However, as we are performing true *ex-ante* prediction experiments using predictive densities, this is clearly not the case; more parsimonious models may perform better, particularly if they are "better approximations" of the true underlying DGP.

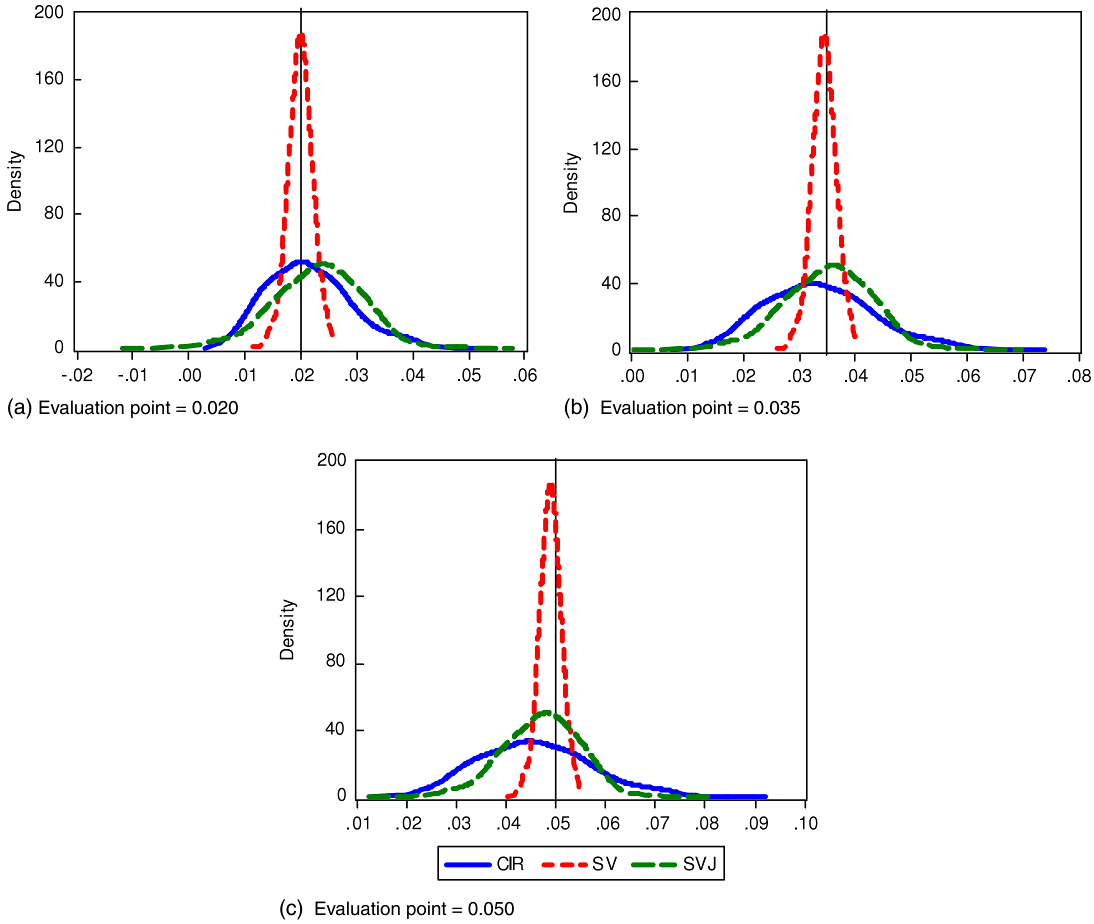

Fig. 2. Predictive densities for CIR, SV and SVJ models — 01:1999–04:2008.

parameters (chosen to illustrate the variety of predictive densities that arise, in practical applications). Evaluation points are chosen to be equal to the mean of the data and various points around the mean. Fig. 1 reports densities for our first sample period and Fig. 2 for our second sample period. Notice that a model yielding a density centered around the evaluation point is preferred, assuming that it yields predictions with equal or less dispersion than its competitor model. Interestingly, in Fig. 1 it is quite apparent that the *SVJ* model is preferred, although none of the models is particularly well centered for evaluation points not equal to the mean of 0.055. In Fig. 2, where results are reported for the second sample period, the models are well-centered around the evaluation point, even for points that are relatively distant from the mean (see Fig. 1(a) and (c)). Moreover, in this particular set of plots, the SV model is clearly dominant, as it yields densities that are better centered and exhibit much less dispersion.

Appendix

Proof of Theorem 1. (i) We begin by analyzing the term in the test statistic that is associated with model 1. Without loss of generality and for the sake of brevity, set $u_1 = -\infty$ and $u_2 = u$. Consider:

$$\frac{1}{\sqrt{P}}\sum_{t=R}^{T-\tau}\left(\frac{1}{N}\sum_{i=1}^{N}1\{X_{1,t+\tau,i}^{\widehat{\vartheta}_{1,t,N,h}}(X_t)\le u\}-1\{X_{t+\tau}\le u\}\right)^2$$

$$=\frac{1}{\sqrt{P}}\sum_{t=R}^{T-\tau}\left(\frac{1}{N}\sum_{i=1}^{N}1\{X_{1,t+\tau,i}^{\vartheta_1^\dagger}(X_t)\le u\}-1\{X_{t+\tau}\le u\}\right)^2$$

$$+\frac{1}{\sqrt{P}}\sum_{t=R}^{T-\tau}\left(\frac{1}{N}\sum_{i=1}^{N}(1\{X_{1,t+\tau,i}^{\widehat{\vartheta}_{1,t,N,h}}(X_t)\le u\} - 1\{X_{1,t+\tau,i}^{\vartheta_1^\dagger}(X_t)\le u\})\right)^2$$

$$+\frac{2}{\sqrt{P}}\sum_{t=R}^{T-\tau}\left[\left(\frac{1}{N}\sum_{i=1}^{N}1\{X_{1,t+\tau,i}^{\vartheta_1^\dagger}(X_t)\le u\}-1\{X_{t+\tau}\le u\}\right)\times\left(\frac{1}{N}\sum_{i=1}^{N}(1\{X_{1,t+\tau,i}^{\widehat{\vartheta}_{1,t,N,h}}(X_t)\le u\}-1\{X_{1,t+\tau,i}^{\vartheta_1^\dagger}(X_t)\le u\})\right)\right]$$

$$=I_{P,N,h}+II_{P,N,h}+III_{P,N,h}. \qquad (10)$$

Now,

$$I_{P,N,h}=\frac{1}{\sqrt{P}}\sum_{t=R}^{T-\tau}((1\{X_{t+\tau}\le u\}-F_0(u|X_t))+(F_0(u|X_t)-F_{X_{1,t+\tau}^{\vartheta_1^\dagger}}(u|X_t)))^2+o_p(1),$$

as $E(\frac{1}{N}\sum_{i=1}^{N}1\{X_{1,t+\tau,i}^{\vartheta_1^\dagger}(X_t)\le u\}-F_{X_{1,t+\tau}^{\vartheta_1^\dagger}}(u|X_t))=0$; and for $N/P\to\infty$,

$$\frac{1}{\sqrt{P}}\sum_{t=R}^{T-\tau}\left|\frac{1}{N}\sum_{i=1}^{N}1\{X_{1,t+\tau,i}^{\vartheta_1^\dagger}(X_t)\le u\}-F_{X_{1,t+\tau}^{\vartheta_1^\dagger}}(u|X_t)\right|=o_p(1).$$

Letting $\mu_{F_1} = E(F_{X_{1,t+\tau}^{\vartheta_1^\dagger}}(u|X_t) - 1\{X_{t+\tau} \le u\})$,

$$III_{P,N,h} = \frac{1}{\sqrt{P}} \sum_{t=R}^{T-\tau} \left[(F_{X_{1,t+\tau}^{\vartheta_1^\dagger}}(u|X_t) - 1\{X_{t+\tau} \le u\}) \times \left(\frac{1}{N} \sum_{i=1}^{N} (1\{X_{1,t+\tau,i}^{\widehat{\vartheta}_{1,t,N,h}}(X_t) \le u\} - 1\{X_{1,t+\tau,i}^{\vartheta_1^\dagger}(X_t) \le u\}) \right) \right] + o_P(1)$$

$$= \frac{\mu_{F_1}}{\sqrt{P}} \sum_{t=R}^{T-\tau} \left(\frac{1}{N} \sum_{i=1}^{N} (1\{X_{1,t+\tau,i}^{\widehat{\vartheta}_{1,t,N,h}}(X_t) \le u\} - 1\{X_{1,t+\tau,i}^{\vartheta_1^\dagger}(X_t) \le u\}) \right) + o_P(1)$$

$$= \frac{\mu_{F_1}}{\sqrt{P}} \sum_{t=R}^{T-\tau} \left(\left(\frac{1}{N} \sum_{i=1}^{N} 1\{X_{1,t+\tau,i}^{\vartheta_1^\dagger}(X_t) \le u - (X_{1,t+\tau,i}^{\widehat{\vartheta}_{1,t,N,h}}(X_t) - X_{1,t+\tau,i}^{\vartheta_1^\dagger}(X_t))\} - F_{X_{1,t+\tau}^{\vartheta_1^\dagger}}((u - (X_{1,t+\tau,i}^{\widehat{\vartheta}_{1,t,N,h}}(X_t) - X_{1,t+\tau,i}^{\vartheta_1^\dagger}(X_t)))|X_t) \right) - \left(\frac{1}{N} \sum_{i=1}^{N} 1\{X_{1,t+\tau,i}^{\vartheta_1^\dagger}(X_t) \le u\} - F_{X_{1,t+\tau}^{\vartheta_1^\dagger}}(u|X_t) \right) \right)$$
$$+ \frac{\mu_{F_1}}{\sqrt{P}} \sum_{t=R}^{T-\tau} \frac{1}{N} \sum_{i=1}^{N} (F_{X_{1,t+\tau}^{\vartheta_1^\dagger}}((u - (X_{1,t+\tau,i}^{\widehat{\vartheta}_{1,t,N,h}}(X_t) - X_{1,t+\tau,i}^{\vartheta_1^\dagger}(X_t)))|X_t) - F_{X_{1,t+\tau}^{\vartheta_1^\dagger}}(u|X_t)) + o_p(1). \quad (11)$$

By arguments similar to those used in the proof of Proposition 1 in Corradi and Swanson (2005b), the first term of the last equality on the RHS of (11) is $o_P(1)$. Now, by taking a mean value expansion around $\vartheta_1^\dagger$, it is easy to see that the second term of the last equality on the RHS of (11) can be written as:

$$\frac{1}{\sqrt{P}} \sum_{t=R}^{T-\tau} \left(\frac{1}{N} \sum_{i=1}^{N} f_1((u - (X_{1,t+\tau,i}^{\overline{\vartheta}_{1,t,N,h}}(X_t) - X_{1,t+\tau,i}^{\vartheta_1^\dagger}(X_t)))|X_t) \times \nabla_{\theta_1} X_{1,t+\tau,i}^{\overline{\vartheta}_{1,t,N,h}}(X_t) \right) (\widehat{\vartheta}_{1,t,N,h} - \vartheta_1^\dagger), \quad (12)$$

where $f_1(\cdot|X_t)$ denotes the conditional density under model 1.

Finally, $II_{P,N,h}$ is $o_P(1)$, given that it is of smaller order than the other two terms on the RHS of (10). By treating model k in the same manner as model 1, we have that,

$$D_{k,P,N}(u) = \frac{1}{\sqrt{P}} \sum_{t=R}^{T-\tau} ((F_{X_{1,t+\tau}^{\vartheta_1^\dagger}}(u|X_t) - F_0(u|X_t))^2 - (F_{X_{k,t+\tau}^{\vartheta_k^\dagger}}(u|X_t) - F_0(u|X_t))^2)$$
$$+ \frac{1}{\sqrt{P}} \sum_{t=R}^{T-\tau} ((F_0(u|X_t) - 1\{X_{t+\tau} \le u\})(F_{X_{1,t+\tau}^{\vartheta_1^\dagger}}(u|X_t) - F_{X_{k,t+\tau}^{\vartheta_k^\dagger}}(u|X_t))) + \mu_{F_1} \frac{2}{\sqrt{P}} \sum_{t=R}^{T-\tau} \left(\frac{1}{N} \sum_{i=1}^{N} f_1((u - (X_{1,t+\tau,i}^{\overline{\vartheta}_{1,t,N,h}}(X_t) - X_{1,t+\tau,i}^{\vartheta_1^\dagger}(X_t)))|X_t) \times \nabla_{\theta_1} X_{1,t+\tau,i}^{\overline{\vartheta}_{1,t,N,h}}(X_t)' \right) (\widehat{\vartheta}_{1,t,N,h} - \vartheta_1^\dagger)$$
$$- \mu_{F_k} \frac{2}{\sqrt{P}} \sum_{t=R}^{T-\tau} \left(\frac{1}{N} \sum_{i=1}^{N} f_k((u - (X_{k,t+\tau,i}^{\overline{\vartheta}_{k,t,N,h}}(X_t) - X_{k,t+\tau,i}^{\vartheta_k^\dagger}(X_t)))|X_t) \nabla_{\theta_k} X_{k,t+\tau,i}^{\overline{\vartheta}_{k,t,N,h}}(X_t)' \right) \times (\widehat{\vartheta}_{k,t,N,h} - \vartheta_k^\dagger) + o_P(1).$$

Now, let $\mu_{f_k,\theta_k^\dagger}(u)' = E_X(f_{X_{1,t+\tau}^{\vartheta_k^\dagger}}(u|X_t) E_N(\nabla_{\theta_k} X_{k,t+\tau,i}^{\overline{\vartheta}_{k,t,N,h}}(X_t))')$, where E_X denotes expectation with respect to the probability measure governing the data and E_N denotes expectation with respect to the probability measure governing the simulated data. Thus, given Assumption A4:

$$D_{k,P,N}(u) = \frac{1}{\sqrt{P}} \sum_{t=R}^{T-\tau} ((F_{X_{1,t+\tau}^{\vartheta_1^\dagger}}(u|X_t) - F_0(u|X_t))^2 - (F_{X_{k,t+\tau}^{\vartheta_k^\dagger}}(u|X_t) - F_0(u|X_t))^2)$$
$$+ \frac{1}{\sqrt{P}} \sum_{t=R}^{T-\tau} ((F_0(u|X_t) - 1\{X_{t+\tau} \le u\}) \times (F_{X_{1,t+\tau}^{\vartheta_1^\dagger}}(u|X_t) - F_{X_{k,t+\tau}^{\vartheta_k^\dagger}}(u|X_t)))$$
$$+ \mu_{F_1} \mu_{f_1,\vartheta_1^\dagger}(u) A_1^\dagger \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \psi_{1,t,N,h}(\vartheta_1^\dagger) - \mu_{F_k} \mu_{f_k,\vartheta_k^\dagger}(u) A_k^\dagger \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \psi_{k,t,N,h}(\vartheta_k^\dagger) + o_P(1). \quad (13)$$

It then follows that $D_{k,P,N}(u) \xrightarrow{d} N(0, W_k(u))$, where

$$W_k(u) = C(u) + V(u) + 2CV(u) + P_{11}(u) + P_{kk}(u) - P_{1k}(u) + P_1C(u) - P_kC(u) + P_1V(u) - P_kV(u),$$

and where, recalling A4,

$$C(u) = \sum_{j=0}^{\infty} E(((F_{X_{1,1+\tau}^{\vartheta_1^\dagger}}(u|X_1) - F_0^\tau(u|X_1))^2 - (F_{X_{k,1+\tau}^{\vartheta_k^\dagger}}(u|X_1) - F_0^\tau(u|X_1))^2)((F_{X_{1,1+j+\tau}^{\vartheta_1^\dagger}}(u|X_{1+j}) - F_0^\tau(u|X_{1+j}))^2 - (F_{X_{k,1+j+\tau}^{\vartheta_k^\dagger}}(u|X_{1+j}) - F_0^\tau(u|X_{1+j}))^2))$$

$$V(u) = \sum_{j=0}^{\infty} E(((F_0^\tau(u|X_1) - 1\{X_{1+\tau} \le u\})(F_{X_{1,1+\tau}^{\vartheta_1^\dagger}}(u|X_1) - F_{X_{k,1+\tau}^{\vartheta_k^\dagger}}(u|X_1)))((F_0^\tau(u|X_{1+j}) - 1\{X_{1+j+\tau} \le u\})(F_{X_{1,1+j+\tau}^{\vartheta_1^\dagger}}(u|X_{1+j}) - F_{X_{k,1+j+\tau}^{\vartheta_k^\dagger}}(u|X_{1+j}))))$$

$$\begin{aligned}CV(u) &= \sum_{j=0}^{\infty} E(((F_{X_{1,1+\tau}^{\vartheta_1^\dagger}}(u|X_1) - F_0(u|X_1))^2 \\ &\quad - (F_{X_{k,1+\tau}^{\vartheta_k^\dagger}}(u|X_1) - F_0(u|X_1))^2)((F_0^\tau(u|X_{1+j}) \\ &\quad - 1\{X_{1+j+\tau} \le u\})(F_{X_{1,1+j+\tau}^{\vartheta_1^\dagger}}(u|X_{1+j}) \\ &\quad - F_{X_{k,1+j+\tau}^{\vartheta_k^\dagger}}(u|X_{1+j}))))\end{aligned}$$

$$P_{11}(u) = 4\Pi \mu_{F_1}^2(u)\mu'_{f_1,\vartheta_1^\dagger}(u)(A_1^\dagger V_1^\dagger A_1^\dagger)\mu_{f_1,\vartheta_1^\dagger}$$

$$\begin{aligned}P_{1k}(u) &= 8\Pi \mu_{F_1}(u)\mu_{f_1,\vartheta_1^\dagger}(u)' A_1^\dagger \\ &\quad \times \sum_{j=0}^{\infty} E(\psi_{1,1}(\theta_1^\dagger)\psi_{k,1+j}(\theta_k^\dagger)')A_k^{\dagger\prime}\mu_{f_k,\vartheta_k^\dagger}(u)\mu_{F_k},\end{aligned}$$

$$\begin{aligned}P_1C(u) &= 4\Pi \mu_{F_1}(u)\mu_{f_1,\vartheta_1^\dagger}(u)A_1^\dagger \sum_{j=0}^{\infty} E(\psi_{1,1}(\theta_1^\dagger) \\ &\quad \times ((F_{X_{1,1+j+\tau}^{\vartheta_1^\dagger}}(u|X_{1+j}) - F_0(u|X_{1+j}))^2 \\ &\quad - (F_{X_{k,1+j+\tau}^{\vartheta_k^\dagger}}(u|X_{1+j}) - F_0(u|X_{1+j}))^2)),\end{aligned}$$

and

$$\begin{aligned}P_1V_1(u) &= 4\Pi \mu_{F_1}(u)\mu_{f_1,\theta_1^\dagger}(u)' A_1^\dagger \\ &\quad \times \sum_{j=0}^{\infty} E(\psi_{1,1}(\theta_1^\dagger)((F_0^\tau(u|X_{1+j}) \\ &\quad - 1\{X_{1+j+\tau} \le u\})(F_{X_{1,1+j+\tau}^{\theta_1^\dagger}}(u|X_{1+j}) \\ &\quad - F_{X_{k,1+j+\tau}^{\theta_k^\dagger}}(u|X_{1+j})))),\end{aligned}$$

(ii) Under H_A, $\frac{1}{\sqrt{P}}\sum_{t=R}^{T-\tau}((F_{X_{1,t+\tau}^{\vartheta_1^\dagger}}(u|X_t) - F_0(u|X_t))^2 - (F_{X_{k,t+\tau}^{\vartheta_k^\dagger}}(u|X_t) - F_0(u|X_t))^2)$ diverges at rate $\sqrt{P}$. This drives the statistic to either plus or minus infinity. □

Proof of Corollary 1. For any given k, the limiting distribution of $D_{k,P,N}(u_1,u_2) - \mu_k(u_1,u_2)$ follows from inspection of Theorem 1(i). Also, by the Cramer–Wold device,

$$\begin{aligned}&((D_{2,P,N}(u_1,u_2) - \mu_2(u_1,u_2)), \ldots, \\ &(D_{m,P,N}(u_1,u_2) - \mu_m(u_1,u_2)))\end{aligned} \quad (14)$$

converges to a $(m-1)$-dimensional mean zero Gaussian random variable with covariance matrix that has kk element given by $W_k(u_1,u_2)$, as defined in the statement of Theorem 1(i). The statement in the corollary then follows as a straightforward consequence of the Cramer–Wold device and the continuous mapping theorem. Theorem 1 considers the case of two nonnested models. We then need to study what happens if a subset of models is instead nested with the benchmark. For notational simplicity, hereafter let $u_1 = -\infty$ and $u_2 = u$. Suppose that models $2, \ldots, m/2$ are nested with model 1, while models $m/2+1, \ldots, m$ are nonnested with 1, so that under the null, for $j = 2, \ldots, m/2$, $F_{X_{1,t+\tau}^{\vartheta_1^\dagger}}(u|X_t) = F_{X_{j,t+\tau}^{\vartheta_j^\dagger}}(u|X_t)$ *a.s.*, and then the first two terms on the RHS of (13) are almost sure zero. Further, if the nested models are also correctly specified, then $F_{X_{1,t+\tau}^{\vartheta_1^\dagger}}(u|X_t) = F_{X_{j,t+\tau}^{\vartheta_j^\dagger}}(u|X_t) = F_0(u|X_t)$ *a.s.* implying $\mu_{F_1} = \mu_{F_j} = 0$, and so the last two terms on RHS of (13) are also zero. Hence, if model j and 1 are nested, and correctly specified, $D_{j,P,N}(u) = o_p(1)$ and $\mu_j(u) = 0$. On the other hand, if the nested models are not correctly specified for the conditional distribution, then $F_{X_{1,t+\tau}^{\vartheta_1^\dagger}}(u|X_t) = F_{X_{j,t+\tau}^{\vartheta_j^\dagger}}(u|X_t)$ *a.s.*, but they are different from $F_0(u|X_t)$ over a set of strictly positive probability, and so $\mu_{F_1} = \mu_{F_j}$ but they are both different from zero, so that the last two terms on RHS of (13) do not vanish. Concluding, if models $j = 2, \ldots, m/2$ are nested with the benchmark, and correctly specified, then $D_{k,P,N}^{\text{Max}}(u_1,u_2) = \max_{k=2,\ldots,m} D_{k,P,N}(u_1,u_2) = \max_{k=m/2+1,\ldots,m} D_{k,P,N}(u_1,u_2) + o_p(1)$. As the asymptotic covariance of (14) has to be positive semidefinite, it suffices that at least one competing model is not nested with the benchmark. □

Proof of Theorem 2. As before, set $u_1 = -\infty$ and $u_2 = u$. We begin by analyzing the term in the test statistic that is associated with model 1, which can be written as:

$$\begin{aligned}&= \frac{1}{\sqrt{P}}\sum_{t=R}^{T-\tau}\left(\left[\frac{1}{N}\sum_{i=1}^{N} 1\{X_{1,t+\tau,i}^{\widehat{\vartheta}_{1,t,N,h}}(X_t^*) \le u\} - 1\{X_{t+\tau}^* \le u\}\right]^2\right. \\ &\quad \left. - \left(\frac{1}{T}\sum_{j=1}^{T}\left[\frac{1}{N}\sum_{i=1}^{N} 1\{X_{1,t+\tau,i}^{\widehat{\vartheta}_{1,t,N,h}}(X_j) \le u\} - 1\{X_{j+\tau} \le u\}\right]^2\right)\right) \\ &\quad + \frac{1}{\sqrt{P}}\sum_{t=R}^{T-\tau}\left[\frac{1}{N}\sum_{i=1}^{N}(1\{X_{1,t+\tau,i}^{\widehat{\vartheta}^*_{1,t,N,h}}(X_t^*) \le u\}\right. \\ &\quad \left. - 1\{X_{1,t+\tau,i}^{\widehat{\vartheta}_{1,t,N,h}}(X_t^*) \le u\})\right]^2 \\ &\quad + 2\frac{1}{\sqrt{P}}\sum_{t=R}^{T-\tau}\left[\left(\frac{1}{N}\sum_{i=1}^{N} 1\{X_{1,t+\tau,i}^{\widehat{\vartheta}_{1,t,N,h}}(X_t^*) \le u\} - 1\{X_{t+\tau}^* \le u\}\right)\right. \\ &\quad \left. \times \left(\frac{1}{N}\sum_{i=1}^{N}\left(1\{X_{1,t+\tau,i}^{\widehat{\vartheta}^*_{1,t,N,h}}(X_t^*) \le u\} - 1\{X_{1,t+\tau,i}^{\widehat{\vartheta}_{1,t,N,h}}(X_t^*) \le u\}\right)\right)\right].\end{aligned} \quad (15)$$

First, note that:

$$\begin{aligned}&= \frac{1}{\sqrt{P}}\sum_{t=R}^{T-\tau}E^*\left(\left[\frac{1}{N}\sum_{i=1}^{N} 1\{X_{1,t+\tau,i}^{\widehat{\vartheta}_{1,t,N,h}}(X_t^*) \le u\} - 1\{X_{t+\tau}^* \le u\}\right]\right)^2 \\ &= \frac{1}{\sqrt{P}}\sum_{t=R}^{T-\tau}\left(\frac{1}{T}\sum_{j=1}^{T}\left[\frac{1}{N}\sum_{i=1}^{N} 1\{X_{1,t+\tau,i}^{\widehat{\vartheta}_{1,t,N,h}}(X_j) \le u\}\right.\right. \\ &\quad \left.\left. - 1\{X_{j+\tau} \le u_2\}\right]^2\right) + O(l/P^{1/2}) \Pr -P.\end{aligned}$$

Also, by the same arguments as those used in the proof of Theorem 4 in Bhardwaj et al. (2008),

$$\begin{aligned}&\text{Var}^*\left(\frac{1}{\sqrt{P}}\sum_{t=R}^{T-\tau}\left[\frac{1}{N}\sum_{i=1}^{N} 1\{X_{1,t+\tau,i}^{\widehat{\vartheta}_{1,t,N,h}}(X_t^*) \le u\} - 1\{X_{t+\tau}^* \le u\}\right]^2\right) \\ &= Var\left(\frac{1}{\sqrt{P}}\sum_{t=R}^{T-\tau}\left[\frac{1}{N}\sum_{i=1}^{N} 1\{X_{1,t+\tau,i}^{\vartheta_1^\dagger}(X) \le u\}\right.\right. \\ &\quad \left.\left. - 1\{X_{t+\tau} \le u\}\right]^2\right) + O(l/P^{1/2}) \Pr -P.\end{aligned}$$

Thus, from Theorem 3.5 in Künsch (1989), it follows that the first term on the RHS of the last equality in (15) has the same limiting distribution as:

$$\begin{aligned}&\frac{1}{\sqrt{P}}\sum_{t=R}^{T-\tau}\left(\left(\frac{1}{N}\sum_{i=1}^{N} 1\{X_{1,t+\tau,i}^{\vartheta_1^\dagger}(X_t) \le u\} - 1\{X_{t+\tau} \le u\}\right)^2\right. \\ &\quad \left. - E\left(\frac{1}{N}\sum_{i=1}^{N} 1\{X_{1,t+\tau,i}^{\vartheta_1^\dagger}(X_t) \le u\} - 1\{X_{t+\tau} \le u\}\right)^2\right).\end{aligned}$$

Now, $\frac{1}{N}\sum_{i=1}^{N} 1\{X^{\vartheta_1^\dagger}_{1,t+\tau,i}(X_t) \le u\} - F_{X^{\vartheta_1^\dagger}_{1,t+\tau}}(u|X_t) = O_N(N^{-1/2})$, and as $N/P \to \infty$, the third term on the RHS of (15) can be written as:

$$2\mu_{F_1}\frac{1}{\sqrt{P}}\sum_{t=R}^{T-\tau}\left(\frac{1}{N}\sum_{i=1}^{N}(1\{X^{\widehat{\vartheta}^*_{1,t,N,h}}_{1,t+\tau,i}(X_t^*) \le u\} - 1\{X^{\widehat{\vartheta}_{1,t,N,h}}_{1,t+\tau,i}(X_t^*) \le u\})\right) + o_p^*(1) \Pr -P, \quad (16)$$

where $\mu_{F_1} = E(F_{X^{\vartheta_1^\dagger}_{1,t+\tau}}(u|X_t) - 1\{X_{t+\tau} \le u\})$. Now,

$$\frac{1}{\sqrt{P}}\sum_{t=R}^{T-\tau}\left(\frac{1}{N}\sum_{i=1}^{N}(1\{X^{\widehat{\vartheta}^*_{1,t,N,h}}_{1,t+\tau,i}(X_t^*) \le u\} - 1\{X^{\widehat{\vartheta}_{1,t,N,h}}_{1,t+\tau,i}(X_t^*) \le u\})\right)$$
$$= \frac{1}{\sqrt{P}}\sum_{t=R}^{T-\tau}\left(\left(\frac{1}{N}\sum_{i=1}^{N} 1\{X^{\widehat{\vartheta}_{1,t,N,h}}_{1,t+\tau,i}(X_t) \le u - (X^{\widehat{\vartheta}^*_{1,t,N,h}}_{1,t+\tau,i}(X_t) - X^{\widehat{\vartheta}_{1,t,N,h}}_{1,t+\tau,i}(X_t))\} - F_{X^{\widehat{\vartheta}_{1,t,N,h}}_{1,t+\tau}}((u - (X^{\widehat{\vartheta}^*_{1,t,N,h}}_{1,t+\tau,i}(X_t) - X^{\widehat{\vartheta}_{1,t,N,h}}_{1,t+\tau,i}(X_t)))|X_t\right)\right) - \left(\frac{1}{N}\sum_{i=1}^{N} 1\{X^{\widehat{\vartheta}_{1,t,N,h}}_{1,t+\tau,i}(X_t) \le u\} - F_{X^{\widehat{\vartheta}_{1,t,N,h}}_{1,t+\tau}}(u|X_t))\right) + \frac{1}{\sqrt{P}}\sum_{t=R}^{T-\tau}\frac{1}{N}\sum_{i=1}^{N}(F_{X^{\widehat{\vartheta}_{1,t,N,h}}_{1,t+\tau}}((u - (X^{\widehat{\vartheta}^*_{1,t,N,h}}_{1,t+\tau,i}(X_t) - X^{\widehat{\vartheta}_{1,t,N,h}}_{1,t+\tau,i}(X_t)))|X_t) - F_{X^{\widehat{\vartheta}_{1,t,N,h}}_{1,t+\tau}}(u|X_t)). \quad (17)$$

By the same argument as that used in the proof of Theorem 1(i):

$$= \frac{1}{\sqrt{P}}\sum_{t=R}^{T-\tau}\left(\left(\frac{1}{N}\sum_{i=1}^{N} 1\{X^{\widehat{\vartheta}_{1,t,N,h}}_{1,t+\tau,i}(X_t) \le u - (X^{\widehat{\vartheta}^*_{1,t,N,h}}_{1,t+\tau,i}(X_t) - X^{\widehat{\vartheta}_{1,t,N,h}}_{1,t+\tau,i}(X_t))\} - F_{X^{\widehat{\vartheta}_{1,t,N,h}}_{1,t+\tau}}((u - (X^{\widehat{\vartheta}^*_{1,t,N,h}}_{1,t+\tau,i}(X_t) - X^{\widehat{\vartheta}_{1,t,N,h}}_{1,t+\tau,i}(X_t)))|X_t)\right) - \left(\frac{1}{N}\sum_{i=1}^{N} 1\{X^{\widehat{\vartheta}_{1,t,N,h}}_{1,t+\tau,i}(X_t) \le u\} - F_{X^{\widehat{\vartheta}_{1,t,N,h}}_{1,t+\tau}}(u|X_t)\right)\right) = o_{P^*}(1) \Pr -P.$$

Finally, the last term on the RHS of (17), conditional on the sample, and for all samples except a set with probability measure approaching zero, has the same limiting distribution as:

$$\frac{1}{\sqrt{P}}\sum_{t=R}^{T-\tau}\frac{1}{N}\sum_{i=1}^{N}(F_{X^{\vartheta_1^\dagger}_{1,t+\tau}}((u - (X^{\widehat{\vartheta}_{1,t,N,h}}_{1,t+\tau,i}(X_t) - X^{\vartheta_1^\dagger}_{1,t+\tau,i}(X_t)))|X_t) - F_{X^{\vartheta_1^\dagger}_{1,t+\tau}}(u|X_t))$$

and the statement then follows by the same argument as that used in Theorem 1(i). □

Proof of Corollary 2. Given Theorem 2, for the case of nonnested models, the result follows directly upon application of the Cramer–Wold device and the continuous mapping theorem. It remains to show that whenever $D_{j,P,N}(u) = o_p(1)$, because of model j and 1 are nested and correctly specified, then also $D^*_{j,P,N}(u) = o_p^*(1)$, conditional on the sample. Now, given (16), it is immediate to see that whenever $\mu_{F_1} = \mu_{F_j} = 0$, then the bootstrap counterpart of the parameter estimation error is $o_{p^*}(1)$, $\Pr -P$. Now, recalling (15) the bootstrap counterpart of $D_{j,P,N}(u)$ for the case of vanishing parameter estimation error writes as:

$$\widetilde{D}^*_{j,P,N}(u)\frac{1}{\sqrt{P}}\sum_{t=R}^{T-\tau}\left\{\left(\left(\left[\frac{1}{N}\sum_{i=1}^{N}1\{X^{\widehat{\vartheta}_{1,t,N,h}}_{1,t+\tau,i}(X_t^*) \le u\} - 1\{X^*_{t+\tau} \le u\}\right]^2 - \left(\frac{1}{T}\sum_{\iota=1}^{T}\left[\frac{1}{N}\sum_{i=1}^{N}1\{X^{\widehat{\vartheta}_{1,t,N,h}}_{1,t+\tau,i}(X_\iota) \le u\} - 1\{X_{\iota+\tau} \le u\}\right]^2\right)\right) - \left(\left[\frac{1}{N}\sum_{i=1}^{N}1\{X^{\widehat{\vartheta}_{j,t,N,h}}_{j,t+\tau,i}(X_t^*) \le u\} - 1\{X^*_{t+\tau} \le u\}\right]^2 - \left(\frac{1}{T}\sum_{\iota=1}^{T}\left[\frac{1}{N}\sum_{i=1}^{N}1\{X^{\widehat{\vartheta}_{j,t,N,h}}_{j,t+\tau,i}(X_\iota) \le u\} - 1\{X_{\iota+\tau} \le u\}\right]^2\right)\right)\right\}.$$

Now, a straightforward calculation gives that

$$\begin{aligned}\mathrm{var}^*(\widetilde{D}^*_{j,P,N}(u)) = &\frac{1}{P}\sum_{t=R}^{T-\tau}((F_{X^{\widehat{\vartheta}_{1,t,N,h}}_{1,t+\tau}}(u|X_t) - 1\{X_{t+\tau} \le u\})^2 \\ &- \frac{1}{T}\sum_{\iota=1}^{T}(F_{X^{\widehat{\vartheta}_{1,t,N,h}}_{1,t+\tau}}(u|X_\iota) - 1\{X_{\iota+\tau} \le u\})^2)^2 \\ &+ \frac{1}{P}\sum_{t=R}^{T-\tau}((F_{X^{\widehat{\vartheta}_{j,t,N,h}}_{j,t+\tau}}(u|X_t) - 1\{X_{t+\tau} \le u\})^2 \\ &- \frac{1}{T}\sum_{\iota=1}^{T}(F_{X^{\widehat{\vartheta}_{j,t,N,h}}_{j,t+\tau}}(u|X_\iota) - 1\{X_{\iota+\tau} \le u\})^2)^2 \\ &- \frac{2}{P}\sum_{t=R}^{T-\tau}(((F_{X^{\widehat{\vartheta}_{j,t,N,h}}_{j,t+\tau}}(u|X_t) - 1\{X_{t+\tau} \le u\})^2 \\ &- \frac{1}{T}\sum_{\iota=1}^{T}(F_{X^{\widehat{\vartheta}_{j,t,N,h}}_{j,t+\tau}}(u|X_\iota) - 1\{X_{\iota+\tau} \le u\})^2) \\ &\times ((F_{X^{\widehat{\vartheta}_{1,t,N,h}}_{1,t+\tau}}(u|X_t) - 1\{X_{t+\tau} \le u\})^2 \\ &- \frac{1}{T}\sum_{\iota=1}^{T}(F_{X^{\widehat{\vartheta}_{1,t,N,h}}_{1,t+\tau}}(u|X_\iota) - 1\{X_{\iota+\tau} \le u\})^2)) \\ &+ O_p(l/\sqrt{P}) + o_N(1),\end{aligned}$$

which is $o_p(1)$, as $F_{X^{\vartheta_1^\dagger}_{1,t+\tau}}(u|X_t) = F_{X^{\vartheta_j^\dagger}_{j,t+\tau}}(u|X_t) = F_0(u|X_t)$, and $\widehat{\vartheta}_{1,t,N,h}$ and $\widehat{\vartheta}_{j,t,N,h}$ are consistent for $\vartheta_1^\dagger$ and $\vartheta_j^\dagger$, respectively. Hence, $D^*_{j,P,N}(u) = o_p^*(1)$. □

Proof of Theorem 3. We begin by analyzing the term in the test statistic that is associated with model 1. Without loss of generality and for the sake of brevity, we yet again set $u_1 = -\infty$ and $u_2 = u$. Consider:

$$\frac{1}{\sqrt{P}}\sum_{t=R}^{T-\tau}\left(\frac{1}{NS}\sum_{j=1}^{S}\sum_{i=1}^{N}1\{X^{\widehat{\vartheta}_{1,t,N,Sh}}_{1,t+\tau,i,j}(X_t) \le u\} - 1\{X_{t+\tau} \le u\}\right)^2$$
$$= \frac{1}{\sqrt{P}}\sum_{t=R}^{T-\tau}\left(\frac{1}{NS}\sum_{j=1}^{S}\sum_{i=1}^{N}1\{X^{\vartheta_1^\dagger}_{1,t+\tau,i,j}(X_t) \le u\} - 1\{X_{t+\tau} \le u\}\right)^2$$

$$+\frac{1}{\sqrt{P}}\sum_{t=R}^{T-\tau}\left(\frac{1}{NS}\sum_{j=1}^{S}\sum_{i=1}^{N}(1\{X_{1,t+\tau,i,j}^{\widehat{\theta}_{1,t,N,S,h}}(X_t)\le u\}\right.$$
$$\left.-\ 1\{X_{1,t+\tau,i,j}^{\theta_1^\dagger}(X_t)\le u\})\right)^2$$
$$+\frac{2}{\sqrt{P}}\sum_{t=R}^{T-\tau}\left[\left(\frac{1}{NS}\sum_{j=1}^{S}\sum_{i=1}^{N}1\{X_{1,t+\tau,i,j}^{\theta_1^\dagger}(X_t)\le u\}\right.\right.$$
$$\left.-\ 1\{X_{t+\tau}\le u\}\right)\left(\frac{1}{NS}\sum_{j=1}^{S}\sum_{i=1}^{N}(1\{X_{1,t+\tau,i,j}^{\widehat{\vartheta}_{1,t,N,S,h}}(X_t)\le u\}\right.$$
$$\left.\left.-\ 1\{X_{1,t+\tau,i,j}^{\theta_1^\dagger}(X_t)\le u\})\right)\right]$$
$$= I_{P,N,S,h}+II_{P,N,S,h}+III_{P,N,S,h}.$$

The statement follows by the same argument as that used in Theorem 1, as by Proposition 5 in Bhardwaj et al. (2008), for $S/P\to\infty$ and $N/P\to\infty$,

$$\frac{1}{\sqrt{P}}\sum_{t=R}^{T-\tau}\left|\frac{1}{NS}\sum_{j=1}^{S}\sum_{i=1}^{N}1\{X_{1,t+\tau,i,j}^{\theta_1^\dagger}(X_t)\le u\}\right.$$
$$\left.-\ F_{X_{1,t+\tau}^{\vartheta_1^\dagger}}(u|X_t,V_{j,h}^{\theta_1^\dagger})\right|=o_p(1).\quad\Box$$

Proof of Theorem 4. Since $S/T\to\infty$, we do not need to resample the initial value of volatility, and the statement thus follows by the same argument as that used in Theorem 2.

For notational simplicity, in the proof of Theorems 5–8 below, we drop the subscript k, as the arguments used in this proof are the same for all k.

Proof of Theorem 5. Define,

$$\widehat{f}_N(X_t|X_{t-1},\vartheta)=\frac{1}{N\xi_N}\sum_{i=1}^{N}K\left(\frac{X_{t,i}^{\vartheta}(X_{t-1})-X_t}{\xi_N}\right),$$

where $X_{t,i}^{\vartheta}(X_{t-1})$ is the i-th simulated value, when starting the path at X_{t-1}, for the case in which there is no discretization error (i.e. for the case in which we could generate continuous paths), and define:

$$L_{t,h}^N(\vartheta)=\frac{1}{t}\sum_{j=1}^{t}\ln\widehat{f}_{k,N,h}(X_j|X_{j-1},\vartheta)\times\tau_N(\widehat{f}_{k,N,h}(X_j|X_{j-1},\vartheta)),\qquad(18)$$

$$L_t^N(\vartheta)=\frac{1}{t}\sum_{j=1}^{t}\ln\widehat{f}_{k,N}(X_j|X_{j-1},\vartheta)\tau_N(\widehat{f}_{k,N}(X_j|X_{j-1},\vartheta)),\qquad(19)$$

and

$$L_t(\vartheta)=\frac{1}{t}\sum_{j=1}^{t}\ln f(X_j|X_{j-1},\vartheta),\qquad(20)$$

where $L_t(\vartheta)$ is the pseudo true density under P_θ.

We organize the proof into four steps. Steps 1 and 2 suffice for the statement in (i) to hold.

Step 1:

$$\sup_{\vartheta\in\Theta}\sup_{t\ge R}|L_t^N(\vartheta)-L_t(\vartheta)|=o_p(1).$$

Step 2:

$$\sup_{\vartheta\in\Theta}\sup_{t\ge R}|L_{t,h}^N(\vartheta)-L_t^N(\vartheta)|=o_p(1).$$

Step 3:

$$\sup_{\vartheta\in\mathcal{N}_{\vartheta^\dagger}}\frac{1}{\sqrt{P}}\sum_{t=R}^{T}|\nabla_\vartheta L_t^N(\vartheta)-\nabla_\vartheta L_t(\vartheta)|=o_p(1).$$

Step 4:

$$\sup_{\vartheta\in\mathcal{N}_{\vartheta_k^\dagger}}\frac{1}{\sqrt{P}}\sum_{t=R}^{T}|\nabla_\vartheta L_{t,h}^N(\vartheta)-\nabla_\vartheta L_t^N(\vartheta)|=o_p(1).$$

Proof of Steps 1 *and* 3.

We first need to show that our assumptions imply the assumptions in Theorems 1.1 and 1.2 in Fermanian and Salanié (2004), and then we outline which steps in their proofs have to be modified in order to take into account the fact that X_t is β-mixing (instead of *iid*) and the fact that our estimator is recursive. Then, the statements in Steps 1 and 3 will follow directly from their Theorems 1.1 and 1.2. Now, Assumption A8 implies K1, in Fermanian and Salanié (2004). A1(ii)–(iii) and Assumptions A6 and A7 imply L1 and L2, with $\beta=r$, and L3, with $\gamma=\gamma'=r>4$ in Fermanian and Salanié (2004). A3′ implies M1 with $s_0=0$, and M2 with $r_0=s_1=0$ and $p_0=\zeta=r>4$, in Fermanian and Salanié (2004). It remains to check that the rate conditions T1, R1, T2, R2 and R3 in Fermanian and Salanié (2004) are implied by the rate conditions in the statement of the theorem. First, recall that T,R,P grow at the same rate, given $0<\pi<\infty$ and $N=T^a$, $a>1$. Given A1(iii), $\Pr(\sup_t|X_t|>\varepsilon T^a)\le\sum_{t=1}^{T}\Pr(|X_t|>\varepsilon T^a)\le\frac{1}{\varepsilon^r}T^{1-ar}E(|X_t|^r)$, and as $a>1$ and $r>4$, and so (c) in the statement of the theorem implies T2 (and hence T1) in Fermanian and Salanié (2004) for $v=1$ and $\gamma=\gamma'=\zeta=r>4$. Condition (a) corresponds to R3 in Fermanian and Salanié (2004), for $\gamma=r$. Finally, (c) and (b) are equivalent to R2 in Fermanian and Salanié (2004), for $m=1$ and $r_0=0$.

As the proof in Fermanian and Salanié (2004) is based on the rate at which

$$1\{\|X_t,X_{t-1}\|<N\}\sup_{\vartheta\in\Theta}|\ln\widehat{f}_N(X_t|X_{t-1},\vartheta)-\ln f(X_j|X_{j-1},\vartheta)|$$

and

$1\{\|X_t,X_{t-1}\|>N\}\sup_{\vartheta\in\Theta}|\ln\widehat{f}_N(X_t|X_{t-1},\vartheta)|$ approach zero, the fact that we are estimating parameters in a recursive manner plays no role. On the other hand, the *iid* assumption is used in the exponential inequalities in the proof of Lemma 1 and Theorem 1.1 in Fermanian and Salanié (2004). However, given the geometric β-mixing assumption in A1(i), the rate in the exponential (Bernstein and Hoeffding) inequalities is slower than in the *iid* case, only up to a logarithmic term (see e.g. Doukhan, 1995, p. 33–36). Thus, consistency follows from their Theorem 1.1 and asymptotic normality from their Theorem 1.2. Hence, it remains to prove Step 4, as Step 2 follows by the same argument.

Proof of Step 4.

$$\sup_{\vartheta\in\mathcal{N}_{\vartheta_k^\dagger}}\frac{1}{\sqrt{P}}\sum_{t=R}^{T}|\nabla_\vartheta L_{t,h}^N(\vartheta)-\nabla_\vartheta L_t^N(\vartheta)|$$
$$\le\sup_{\vartheta\in\mathcal{N}_{\vartheta_k^\dagger}}\frac{1}{\sqrt{P}}\sum_{t=R}^{T}\left(\left|\frac{1}{t}\sum_{j=1}^{t}\tau_N(\widehat{f}_N(X_j|X_{j-1},\vartheta))\right.\right.$$
$$\times\frac{1}{\widehat{f}_N(X_j|X_{j-1},\vartheta)}$$
$$\left.\times\left(\frac{\partial\widehat{f}_{N,h}(X_j|X_{j-1},\vartheta)}{\partial\vartheta}-\frac{\partial\widehat{f}_N(X_j|X_{j-1},\vartheta)}{\partial\vartheta}\right)\right|$$
$$+\left|\frac{1}{t}\sum_{j=1}^{t}\left(\frac{\tau_N(\widehat{f}_{N,h}(X_j|X_{j-1},\vartheta))}{\widehat{f}_{N,h}(X_j|X_{j-1},\vartheta)}-\frac{\tau_N(\widehat{f}_N(X_j|X_{j-1},\vartheta))}{\widehat{f}_N(X_j|X_{j-1},\vartheta)}\right)\right.$$

$$\times \left.\frac{\partial\widehat{f}_{N,h}(X_j|X_{j-1},\vartheta)}{\partial\vartheta}\right|$$
$$+\left|\frac{1}{t}\sum_{j=1}^{t}\tau_N'(\widehat{f}_{N,h}(X_j|X_{j-1},\vartheta))\frac{\partial\widehat{f}_{N,h}(X_j|X_{j-1},\vartheta)}{\partial\vartheta}\right.$$
$$\times \left.\ln\widehat{f}_{N,h}(X_j|X_{j-1},\vartheta)\right|$$
$$+\left|\frac{1}{t}\sum_{j=1}^{t}\tau_N'(\widehat{f}_{N}(X_j|X_{j-1},\vartheta))\frac{\partial\widehat{f}_{N}(X_j|X_{j-1},\vartheta)}{\partial\vartheta}\right.$$
$$\times \left.\left.\ln\widehat{f}_{N}(X_j|X_{j-1},\vartheta)\right|\right)$$
$$= A_{1,T,N,h}+A_{2,T,N,h}+A_{3,T,N,h}+A_{4,T,N,h}.$$

Now, note that $\overline{X}^{\vartheta}_{j,i,h}(X_{j-1}) \in (X^{\vartheta}_{j,i,h}(X_{j-1}), X^{\vartheta}_{j,i}(X_{j-1}))$, and recall by Theorem 2.3 in Pardoux and Talay (1985) that $E((X^{\vartheta}_{j,i,h}(X_{j-1}) - X^{\vartheta}_{j,i}(X_{j-1}))^2) = O(h^2)$. Thus,

$$A_{1,T,N,h} \le \xi_N^{-\delta}\sqrt{P}\sup_{t\ge R}\sup_{\vartheta\in\mathcal{N}_{\vartheta_k^\dagger}}\left|\frac{1}{t}\sum_{j=1}^{t}\frac{1}{N\xi_N}\right.$$
$$\times\sum_{i=1}^{N}\left(\nabla_\vartheta K\left(\frac{X^{\vartheta}_{j,i,h}(X_{j-1})-X_j}{\xi_N}\right)\right.$$
$$\left.\left.-\ \nabla_\vartheta K\left(\frac{X^{\vartheta}_{t,i}(X_{j-1})-X_j}{\xi_N}\right)\right)\right|$$
$$\le \xi_N^{-(\delta+3)}\sqrt{P}\sup_{t\ge R}\sup_{\vartheta\in\mathcal{N}_{\vartheta_k^\dagger}}\left|\frac{1}{t}\sum_{j=1}^{t}\frac{1}{N}\sum_{i=1}^{N}\nabla^2_\vartheta K\right.$$
$$\times\left.\left(\frac{X^{\vartheta}_{j,i,h}(X_{j-1})-X_j}{\xi_N}\right)\right|_{\overline{X}^{\vartheta}_{j,i,h}(X_{j-1})}$$
$$\times\left.(X^{\vartheta}_{j,i,h}(X_{j-1})-X^{\vartheta}_{j,i}(X_{j-1}))\right|$$
$$= O_p(\sqrt{P}\xi_N^{-(\delta+3)}h), \tag{21}$$

and given A6,

$$A_{2,T,N,h} \le \sqrt{P}\sup_{t\ge R}\sup_{\vartheta\in\mathcal{N}_{\vartheta_k^\dagger}}\left|\frac{1}{t}\right.$$
$$\times\sum_{j=1}^{t}\frac{\tau_N(\widehat{f}_N(X_j|X_{j-1},\vartheta))(\widehat{f}_{N,h}(X_j|X_{j-1},\vartheta)-\widehat{f}_N(X_j|X_{j-1},\vartheta))}{\widehat{f}_N(X_j|X_{j-1},\vartheta)}$$
$$\times\left.\frac{\partial\ln\widehat{f}_{N,h}(X_j|X_{j-1},\vartheta)}{\partial\vartheta}\right|+\sqrt{P}\sup_{t\ge R}\sup_{\vartheta\in\mathcal{N}_{\vartheta_k^\dagger}}\left|\frac{1}{t}\right.$$
$$\times\sum_{j=1}^{t}\left(\frac{\tau_N(\widehat{f}_N(X_j|X_{j-1},\vartheta))-\tau_N(\widehat{f}_{N,h}(X_j|X_{j-1},\vartheta))}{\widehat{f}_N(X_j|X_{j-1},\vartheta)}\right)$$
$$\times\left.\widehat{f}_{N,h}(X_j|X_{j-1},\vartheta)\frac{\partial\ln\widehat{f}_{N,h}(X_j|X_{j-1},\vartheta)}{\partial\vartheta}\right|$$
$$= O_p(\sqrt{P}\xi_N^{-(\delta+3)}h|\ln\xi_N^{\delta}|). \tag{22}$$

Hence, $A_{1,T,N,h}$ and $A_{2,T,N,h}$ are $o_p(1)$ because of (d). Finally, given the rate conditions in (a)–(c), $A_{3,T,N,h}$ and $A_{4,T,N,h}$ are $o_P(1)$, by the same argument as used in the study of the term A4 in the proof of Theorem 2.2 in Fermanian and Salanié (2004). □

Proof of Theorem 6. Define,

$$L^{*N}_{t,h}(\theta) = \frac{1}{t}\sum_{j=1}^{t}\left(\ln\widehat{f}_{N,h}(X_j^*|X_{j-1}^*,\vartheta)\tau_N(\widehat{f}_{N,h}(X_j^*|X_{j-1}^*,\vartheta))\right.$$
$$\left.-\ \vartheta'\frac{1}{T}\sum_{i=1}^{T}\nabla_\vartheta L^N_{i,h}(\widehat{\vartheta}_{t,N,h})\right)$$
$$L_t^*(\theta) = \frac{1}{t}\sum_{j=1}^{t}\left(\ln f(X_j^*|X_{j-1}^*,\vartheta) - \vartheta'\frac{1}{T}\sum_{i=1}^{T}\nabla_\vartheta L_i(\widehat{\vartheta}_t)\right)$$

and let $\widehat{\vartheta}_t^* = \arg\min_{\vartheta\in\Theta} L_t^*(\theta)$. We organize the proof into two steps.

Step 1:

$$\sup_{\vartheta\in\Theta}\sup_{t\ge R}|L^{*N}_{t,h}(\vartheta)-L_t^*(\vartheta)| = o_{p^*}(1).$$

Step 2:

$$\sup_{\vartheta\in\mathcal{N}_{\vartheta^\dagger}}\frac{1}{\sqrt{P}}\sum_{t=R}^{T}|\nabla_\vartheta L^{*N}_{t,h}(\vartheta)-\nabla_\vartheta L_t^*(\vartheta)| = o_p^*(1).$$

Given Steps 1 and 2, the desired outcome follows from Theorem 1 in Corradi and Swanson (2007).

Proof of Step 1. Given the definition of $L^{*N}_{t,h}(\vartheta)$ and $L_t^*(\vartheta)$, and recalling that Θ is a compact set, it suffices to show that:

$$\arg\max_{\vartheta\in\Theta_k}\sup_{t\ge R}\left|\frac{1}{t}\sum_{l=2}^{t}(\ln\widehat{f}_{N,h}(X_l^*|X_{l-1}^*,\vartheta)\tau_N(\widehat{f}_{N,h}(X_l^*|X_{l-1}^*,\vartheta))\right.$$
$$\left.-\ \ln f(X_l^*|X_{l-1}^*,\vartheta))\right|$$
$$= o_{p^*}(1) \tag{23}$$

and

$$\sup_{t\ge R}\frac{1}{T}\sum_{i=1}^{T}(\nabla_\vartheta L^{*N}_{i,h}(\widehat{\vartheta}_{t,N,h})-\nabla_\vartheta L_i^*(\widehat{\vartheta}_t)) = o_{p^*}(1). \tag{24}$$

Now, (23) follows from Steps 1 and 2 in the proof of Theorem 5, given that the only difference is that we evaluate the likelihood at the resampled observations. Note also that (24) is majorized by:

$$\sup_{t\ge R}\left|\frac{1}{T}\sum_{i=1}^{T}(\nabla_\vartheta L^{*N}_{i,h}(\widehat{\vartheta}_{t,N,h})-\nabla_\vartheta L_i^*(\widehat{\vartheta}_{t,N,h}))\right|$$
$$+\sup_{t\ge R}\left|\frac{1}{T}\sum_{i=1}^{T}(\nabla_\vartheta L_i^*(\widehat{\vartheta}_{t,N,h})-\nabla_\vartheta L_i^*(\widehat{\vartheta}_t))\right|.$$

The first term above is $o_{p^*}(1)$ as a direct consequence of Steps 3 and 4 in the proof of Theorem 5. The second term is majorized by

$$\sup_{t\ge R}\left|\frac{1}{T}\sum_{i=1}^{T}(\nabla^2_\vartheta L_i^*(\overline{\vartheta}_{t,N,h})(\widehat{\vartheta}_{t,N,h}-\widehat{\vartheta}_t))\right|$$
$$\le \sup_{t\ge R}\frac{1}{T}\sum_{i=1}^{T}|\nabla^2_\vartheta L_i^*(\overline{\vartheta}_{t,N,h})|\sup_{t\ge R}(\widehat{\vartheta}_{t,N,h}-\widehat{\vartheta}_t) = O_{p^*}(1)o_p(1). \quad \square$$

Proof of Step 2. Follows directly from Steps 2 and 4 in the proof Theorem 5.

Proof of Theorem 7. Let:

$$L_{t,h}^{N,S}(\theta) = \frac{1}{t}\sum_{l=2}^{t} \ln \widehat{f}_{N,S,h}(X_l|X_{l-1},\theta)\tau_{N,S} \times (\widehat{f}_{N,S,h}(X_l|X_{l-1},\theta)).$$

We show that:

$$\sup_{\theta\in\Theta}\sup_{t\geq R} |L_{t,h}^{N,S}(\theta) - L_{t,h}^{N}(\theta)| = o_p(1) \quad (25)$$

and

$$\sup_{\theta\in\mathcal{N}_{\theta_k^\dagger}} \frac{1}{\sqrt{P}}\sum_{t=R}^{T} |\nabla_\theta L_{t,h}^{N,S}(\theta) - \nabla_\theta L_{t,h}^{N}(\theta)| = o_p(1). \quad (26)$$

The desired outcome then follows from Theorem 5. Note first that (25) can be written as:

$$\begin{aligned}
&\sup_{\theta\in\Theta}\sup_{t\geq R} |L_{t,h}^{N,S}(\theta) - L_{t,h}^{N}(\theta)| \\
&= \sup_{\theta\in\Theta}\sup_{t\geq R}\left|\frac{1}{t}\sum_{l=2}^{t}(\ln \widehat{f}_{N,S,h}(X_l|X_{l-1},\theta)\tau_{N,S}(\widehat{f}_{N,S,h}(X_l|X_{l-1},\theta)) - \ln \widehat{f}_{N,h}(X_l|X_{l-1},\theta)\tau_{N}(\widehat{f}_{N,h}(X_l|X_{l-1},\theta)))\right| \\
&\leq \sup_{\theta\in\Theta}\sup_{t\geq R}\left|\frac{1}{t}\sum_{l=2}^{t}\tau_{N,S}(\widehat{f}_{N,S,h}(X_l|X_{l-1},\theta))(\ln \widehat{f}_{N,S,h}(X_l|X_{l-1},\theta) - \ln \widehat{f}_{N,h}(X_l|X_{l-1},\theta))\right| \\
&\quad + \sup_{\theta\in\Theta}\sup_{t\geq R}\left|\frac{1}{t}\sum_{l=2}^{t}(\tau_{N,S}(\widehat{f}_{N,S,h}(X_l|X_{l-1},\theta)) - \tau_{N}(\widehat{f}_{N,h}(X_l|X_{l-1},\theta)))\ln \widehat{f}_{N,h}(X_l|X_{l-1},\theta)\right| \\
&= \sup_{\theta\in\Theta}\sup_{t\geq R}(I_{t,N,S,h} + II_{t,N,S,h}).
\end{aligned}$$

Let $\bar{f}_{N,S,h}(X_l|X_{l-1},\theta) \in (\widehat{f}_{N,S,h}(X_l|X_{l-1},\theta), \widehat{f}_{N,h}(X_l|X_{l-1},\theta))$, and note that for all i, j

$$K\left(\frac{X_{l,i,h}^{\theta}(X_{l-1}) - X_l}{\xi_N}\right) = \int_V K\left(\frac{X_{l,i,h}^{\theta}(X_{l-1},v^\theta) - X_l}{\xi_N}\right) f_\theta(v)\mathrm{d}v = E_S\left(K\left(\frac{X_{l,i,h}^{\theta}(X_{l-1},v^\theta) - X_l}{\xi_N}\right)\right),$$

where E_S denotes the expectation with respect to the simulated initial values of volatility. By a mean value expansion,

$$\begin{aligned}
I_{t,N,S,h} &\leq \xi_N^{-\delta}\left|\frac{1}{t}\sum_{l=2}^{t}(\widehat{f}_{N,S,h}(X_l|X_{l-1},\theta) - \widehat{f}_{N,h}(X_l|X_{l-1},\theta))\right| \\
&= \xi_N^{-(\delta+1)}\left|\frac{1}{t}\sum_{l=2}^{t}\frac{1}{N}\sum_{i=1}^{N} \times \left(\frac{1}{S}\sum_{s=1}^{S} K\left(\frac{X_{l,i,h}^{\theta}(X_{j-1},V_s^\theta) - X_j}{\xi_N}\right) - K\left(\frac{X_{l,i,h}^{\theta}(X_{j-1}) - X_j}{\xi_N}\right)\right)\right| \\
&= O_p(\xi_N^{-(\delta+1)}S^{-1/2}), \quad \text{uniformly in } t \text{ and } \theta.
\end{aligned}$$

Also,

$$\begin{aligned}
II_{t,N,S,h} &\leq \left|\frac{1}{t}\sum_{l=2}^{t}\tau_{N,S}'(\bar{f}_{N,S,h}(X_l|X_{l-1},\theta))\ln \widetilde{f}_{N,h}(X_l|X_{l-1},\theta) \times (\widehat{f}_{N,S,h}(X_l|X_{l-1},\theta) - \widehat{f}_{N,h}(X_l|X_{l-1},\theta))\right| \\
&= O_p(\xi_N^{-(\delta+1)}S^{-1/2}|\ln \xi_N^{-\delta}|), \quad \text{uniformly in } t \text{ and } \theta.
\end{aligned}$$

Given the rate condition in (e), this proves (25). Turning now to (26), note that after few simple manipulations:

$$\begin{aligned}
&\sup_{\theta\in\mathcal{N}_{\theta_k^\dagger}} \frac{1}{\sqrt{P}}\sum_{t=R}^{T} |\nabla_\theta L_{t,h}^{N,S}(\theta) - \nabla_\theta L_{t,h}^{N}(\theta)| \\
&\leq \sup_{\theta\in\mathcal{N}_{\theta^\dagger}} \frac{1}{\sqrt{P}}\sum_{t=R}^{T}\Bigg(\left|\frac{1}{t}\sum_{l=2}^{t}\frac{\tau_{N,S}(\widehat{f}_{N,S,h}(X_l|X_{l-1},\theta))}{\widehat{f}_{N,S,h}(X_l|X_{l-1},\theta)} \times \left(\frac{\partial \widehat{f}_{N,S,h}(X_l|X_{l-1},\theta)}{\partial\theta} - \frac{\partial \widehat{f}_{N,h}(X_l|X_{l-1},\theta)}{\partial\theta}\right)\right| \\
&\quad + \left|\frac{1}{t}\sum_{l=2}^{t}\tau_{N,S}(\widehat{f}_{N,S,h}(X_l|X_{l-1},\theta))\frac{\partial \widehat{f}_{N,h}(X_l|X_{l-1},\theta)}{\partial\theta} \times \frac{(\widehat{f}_{N,h}(X_l|X_{l-1},\theta) - \widehat{f}_{N,S,h}(X_l|X_{l-1},\theta))}{\widehat{f}_{N,S,h}(X_l|X_{l-1},\theta)\widehat{f}_{N,h}(X_l|X_{l-1},\theta)}\right| \\
&\quad + \left|\frac{1}{t}\sum_{l=2}^{t}(\tau_{N,S}(\widehat{f}_{N,S,h}(X_l|X_{l-1},\theta)) - \tau_{N}(\widehat{f}_{N,h}(X_l|X_{l-1},\theta))) \times \frac{\partial \ln \widehat{f}_{N,h}(X_l|X_{l-1},\theta)}{\partial\theta}\right| \\
&\quad + \left|\frac{1}{t}\sum_{l=2}^{t}\tau_{N,S}(\widehat{f}_{N,S,h}(X_l|X_{l-1},\theta))(\ln \widehat{f}_{N,S,h}(X_l|X_{l-1},\theta) - \ln \widehat{f}_{N,h}(X_l|X_{l-1},\theta))\right| \\
&\quad + \left|\frac{1}{t}\sum_{l=2}^{t}\tau_{N,S}'(\widehat{f}_{N,S,h}(X_l|X_{l-1},\theta)) \times \left(\frac{\partial \widehat{f}_{N,S,h}(X_l|X_{l-1},\theta)}{\partial\theta} - \frac{\widehat{f}_{N,h}(X_l|X_{l-1},\theta)}{\partial\theta}\right) \times \ln \widehat{f}_{N,h}(X_l|X_{l-1},\theta)\right| \\
&\quad + \left|\frac{1}{t}\sum_{l=2}^{t}(\tau_{N,S}'(\widehat{f}_{N,S,h}(X_l|X_{l-1},\theta)) - \tau_{N}'(\widehat{f}_{N,h}(X_l|X_{l-1},\theta))) \times \frac{\widehat{f}_{N,h}(X_l|X_{l-1},\theta)}{\partial\theta}\ln \widehat{f}_{N,h}(X_l|X_{l-1},\theta)\right|\Bigg) \\
&= \sup_{\theta\in\mathcal{N}_{\theta^\dagger}}\left(V_{1,T,N,S,h}(\theta) + V_{2,T,N,S,h}(\theta) + V_{3,T,N,S,h}(\theta) + V_{4,T,N,S,h}(\theta) + V_{5,T,N,S,h}(\theta) + V_{6,T,N,S,h}(\theta)\right).
\end{aligned}$$

Now, recalling Assumption A9,

$$\sup_{\theta\in\mathcal{N}_{\theta^\dagger}} V_{1,T,N,S,h}(\theta)$$

$$\leq \sup_{\theta\in\mathcal{N}_{\theta^\dagger}} \xi_N^{-(\delta+2)} \frac{1}{\sqrt{P}} \sum_{t=R}^{T}$$

$$\times \left| \frac{1}{t}\sum_{l=2}^{t} \frac{1}{N}\sum_{i=1}^{N} \left(\frac{1}{S}\sum_{s=1}^{S} \left(\frac{\partial X_{l,i,h}^{\theta}(X_{j-1}, V_s^{\theta})}{\partial\theta} \right.\right.\right.$$

$$\times K'\left(\frac{X_{l,i,h}^{\theta}(X_{j-1}, V_s^{\theta}) - X_j}{\xi_N} \right)$$

$$\left.\left.\left. - \quad \frac{\partial X_{l,i,h}^{\theta}(X_{j-1})}{\partial\theta} K'\left(\frac{X_{l,i,h}^{\theta}(X_{j-1}) - X_j}{\xi_N} \right) \right)\right)\right|$$

$$= O_p(P^{1/2}\xi_N^{-(\delta+2)}S^{-1/2}) = o_p(1) \quad \text{because of (e).}$$

Then,

$$\sup_{\theta\in\mathcal{N}_{\theta^\dagger}} V_{2,T,N,S,h}(\theta)$$

$$\leq \xi_N^{-2\delta} \sup_{\theta\in\mathcal{N}_{\theta^\dagger}} \sup_{t\geq R} \left| \frac{1}{t}\sum_{l=2}^{t} \left(\frac{\partial \widehat{f}_{N,h}(X_l|X_{l-1},\theta)}{\partial\theta} \right)^r \right|^{1/r}$$

$$\times \sup_{\theta\in\mathcal{N}_{\theta^\dagger}} \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left| \left(\frac{1}{t}\sum_{l=2}^{t} (\widehat{f}_{N,h}(X_l|X_{l-1},\theta) \right.\right.$$

$$\left.\left. -\widehat{f}_{N,S,h}(X_l|X_{l-1},\theta))^{r/(r-1)} \right) \right|^{(r-1)/r}$$

$$= O_P(\sqrt{P}S^{-1/2}\xi_N^{-(1+2\delta)}),$$

and so

$$\sup_{\theta\in\mathcal{N}_{\theta^\dagger}} V_{3,T,N,S,h}(\theta)$$

$$\leq \sup_{\theta\in\mathcal{N}_{\theta^\dagger}} \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left| \frac{1}{t}\sum_{l=2}^{t} (\tau_{N,S}(\widehat{f}_{N,S,h}(X_l|X_{l-1},\theta)) \right.$$

$$\left. - \quad \tau_N(\widehat{f}_{N,h}(X_l|X_{l-1},\theta))) \frac{\partial \ln f(X_l|X_{l-1},\theta)}{\partial\theta} \right| (1+o_p(1))$$

$$= O_p(\sqrt{P}S^{-1/2}\xi_N^{-(1+\delta)}).$$

By a similar argument as that used in the proof of (25), $\sup_{\theta\in\mathcal{N}_{\theta^\dagger}} V_{4,T,N,S,h}(\theta) = O_p(\xi_N^{-(\delta+1)}\sqrt{P}S^{-1/2})$; $V_{5,T,N,S,h}(\theta)$, (other than a log term), can be treated as $V_{1,T,N,S,h}(\theta)$, and so $\sup_{\theta\in\mathcal{N}_{\theta^\dagger}} V_{5,T,N,S,h}(\theta) = O_p(P^{1/2}\xi_N^{-(\delta+2)}S^{-1/2}|\ln \xi_N^{-\delta}|)$. Finally, by a similar argument as that used to examine $V_{3,T,N,S,h}(\theta)$:

$$\sup_{\theta\in\mathcal{N}_{\theta^\dagger}} V_{5,T,N,S,h}(\theta)$$

$$\leq \sup_{\theta\in\mathcal{N}_{\theta^\dagger}} \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left| \frac{1}{t}\sum_{l=2}^{t} (\tau'_{N,S}(\widehat{f}_{N,S,h}(X_l|X_{l-1},\theta)) \right.$$

$$- \tau'_N(\widehat{f}_{N,h}(X_l|X_{l-1},\theta)))$$

$$\left. \times \frac{f(X_l|X_{l-1},\theta)}{\partial\theta} \ln \widehat{f}_{N,h}(X_l|X_{l-1},\theta) \right| (1+o_p(1))$$

$$= O_p(\sqrt{P}S^{-1/2}\xi_N^{-(1+2\delta)}). \quad \square$$

Proof of Theorem 8. Follows immediately, given Theorem 7, and by the same arguments as those used in the proof of Theorem 6. □

References

Altissimo, F., Mele, A., 2009. Simulated nonparametric estimation of dynamic models with application in finance. Review of Economic Studies 76, 413–450.
Aït-Sahalia, Y., 1996. Testing continuous time models of the spot interest rate. Review of Financial Studies 9, 385–426.
Aït-Sahalia, Y., 2002. Maximum likelihood estimation of discretely sampled diffusions: a closed form approximation approach. Econometrica 70, 223–262.
Aït-Sahalia, Y., Fan, J., Peng, H., 2009. Nonparametric transition-based tests for diffusions. Journal of the American Statistical Association 104, 1102–1116.
Bai, J., 2003. Testing parametric conditional distributions of dynamic models. Review of Economics and Statistics 85, 531–549.
Bandi, F.M., Reno, R., 2008. Nonparametric Stochastic Volatility. Working Paper, University of Chicago.
Bhardwaj, G., Corradi, V., Swanson, N.R., 2008. A simulation based specification test for diffusion processes. Journal of Business and Economic Statistics 26, 176–193.
Bontemps, C., Meddahi, N., 2005. Testing normality: a GMM approach. Journal of Econometrics 124, 149–186.
Carrasco, M., Chernov, M., Florens, J.P., Ghysels, E., 2007. Efficient estimation of general dynamic models with a continuum of moment conditions. Journal of Econometrics 140, 529–543.
Chacko, G., Viceira, L.M., 2003. Spectral GMM estimation of continuous-time Processes. Journal of Econometrics 116, 259–292.
Corradi, V., Swanson, N.R., 2005a. Bootstrap tests for diffusion processes. Journal of Econometrics 124, 117–148.
Corradi, V., Swanson, N.R., 2005b. A test for comparing multiple misspecified conditional intervals. Econometric Theory 21, 991–1016.
Corradi, V., Swanson, N.R., 2006a. Predictive density and conditional confidence interval accuracy tests. Journal of Econometrics 135, 187–228.
Corradi, V., Swanson, N.R., 2006b. Predictive density evaluation. In: Granger, C.W.J., Elliot, G., Timmermann, A. (Eds.), Handbook of Economic Forecasting. Elsevier, Amsterdam, pp. 197–284.
Corradi, V., Swanson, N.R., 2007. Nonparametric bootstrap procedures for predictive inference based on recursive estimation schemes. International Economic Review 48, 67–109.
Diebold, F.X., Gunther, T., Tay, A.S., 1998. Evaluating density forecasts with applications to finance and management. International Economic Review 39, 863–883.
Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13, 253–263.
Doukhan, P., 1995. Mixing Properties and Examples. Springer and Verlag, New York.
Dridi, R., Guay, A., Renault, E., 2007. Indirect inference and calibration of dynamic stochastic general equilibrium models. Journal of Econometrics 136, 397–439.
Duan, J.C., 2003. A Specification Test for Time Series Models by a Normality Transformation. Working Paper, University of Toronto.
Duffie, D., Singleton, K., 1993. Simulated moment estimation of Markov models of asset prices. Econometrica 61, 929–952.
Eraker, B., Johannes, M., Polson, N., 2003. The impact of jumps in volatility and returns. Journal of Finance 58, 1269–1300.
Fermanian, J.-D., Salanié, B., 2004. A nonparametric simulated maximum likelihood estimation method. Econometric Theory 20, 701–734.
Gallant, A.R., Tauchen, G., 1996. Which moments to match. Econometric Theory 12, 657–681.
Gourieroux, C., Monfort, A., Renault, E., 1993. Indirect inference. Journal of Applied Econometrics 8, 203–227.
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., Li, H., 2005. Nonparametric specification testing for continuous time models with applications to term structure interest rates. Review of Financial Studies 18, 37–84.
Hong, Y.M., Li, H., Zhao, F., 2002. Out of sample performance of spot interest rate models. Journal of Business Economics and Statistics 22, 457–473.
Jiang, G.J., 1998. Nonparametric modelling of US term structure of interest rates and implications on the prices of derivative securities. Journal of Financial and Quantitative Analysis 33, 465–497.
Jiang, G.J., Knight, J.L., 2002. Estimation of continuous-time processes via the empirical characteristic function. Journal of Business and Economic Statistics 20, 198–212.
Kloeden, P.E., Platen, E., 1999. Numerical Solution of Stochastic Differential Equations. Springer and Verlag, New York.
Kristensen, D., Shin, Y., 2008. Estimation of Dynamic Models with Nonparametric Simulated Maximum Likelihood. CREATES Research Paper 2008-58, University of Aarhus and Columbia University.
Künsch, H.R., 1989. The jackknife and the bootstrap for general stationary observations. Annals of Statistics 17, 1217–1241.
Masuda, H., 2007. Ergodicity and exponential β-mixing bounds for multivariate diffusions with jumps. Stochastic Processes and Their Applications 117, 35–56.
Meyn, S.P., Tweedie, R.L., 1993. Markov Chains and Stochastic Stability. Spinger and Verlag, New York.
Pardoux, E., Talay, D., 1985. Discretization and simulation of stochastic differential equations. Acta Applicandae Mathematicae 3, 23–47.
Pritsker, M., 1998. Nonparametric density estimators and tests of continuous time interest rate models. Review of Financial Studies 11, 449–487.
Singleton, K.J., 2001. Estimation of affine asset pricing models using empirical characteristic function. Journal of Econometrics 102, 111–141.
Thompson, S.B., 2008. Identifying term structure volatility from the LIBOR-swap curve. Review of Financial Studies 21, 819–854.
White, H., 2000. A reality check for data snooping. Econometrica 68, 1097–1126.