Predictive Density Construction and Accuracy Testing with Multiple Possibly Misspecified Diffusion Models*

Valentina Corradi[1], Norman R. Swanson[2]

[1]University of Warwick and [2]Rutgers University

April 2009

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. Then, we construct accuracy assessment 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.

JEL classification: C22, C51.
Keywords: block bootstrap, diffusion processes, jumps, nonparametric simulated quasi maximum likelihood, parameter estimation error, recursive estimation, stochastic volatility.

----

*We would like to thank 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 Leuen econometrics workshop, the 2007 Summer Meeting of the Econometric Society, the Marseille conference in honor of Russell Davidson, and at the following universities: University of Montreal, LSE, Michigan State University, New York University, University of Chicago GBS, University of Maryland, for useful many comments on earlier versions of this paper.

Valentina Corradi, Department of Economics, University of Warwick, Coventry CV4 7AL, UK, v.corradi@warwick.ac.uk. Norman R. Swanson, Department of Economics, Rutgers University, 75 Hamilton Street, New Brunswick, NJ 08901, USA, nswanson@econ.rutgers.edu. 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.

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), Durham (2003)). 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 multiple factor 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 of this problem. Nevertheless, correct specification of the joint distribution is not equivalent to 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, Gunther and Tay (1998); the cross spectrum approach of Hong (2001), Hong, Li and Zhao (2004), 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, Corradi and Swanson (2008); or by using the approaches of Aït-Sahalia (2002) and Aït-Sahalia, Fan and Peng (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 $E\left(\left(F_k^\tau(u|X_t, \theta_k^\dagger) - F_0^\tau(u|X_t, \theta_0)\right)^2\right) <$ $E\left(\left(F_1^\tau(u|X_t, \theta_1^\dagger) - F_0^\tau(u|X_t, \theta_0)\right)^2\right)$. 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 not known 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 $P - \tau$ $\tau-$step ahead paths, using $X_R, ..., X_{R+P-\tau}$ 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 (functional 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 both simulated generalized method of moments (SGMM) and 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.

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 nonparametric simulated (quasi) maximum likelihood (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, assume that $O_P^*(1)$ $\Pr -P$ denotes 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 $k = 1, ..., m$ consider[1]:

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

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 $\int_Y y\phi_k(y)dy$ 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 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:

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

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:

$$\begin{aligned} & E\left(\left((F_1^\tau(u_2|X_t, \vartheta_1^\dagger) - F_1^\tau(u_1|X_t, \vartheta_1^\dagger)) - (F_0^\tau(u_2|X_t, \vartheta_0) - F_0^\tau(u_1|X_t, \vartheta_0))\right)^2\right) \\ < \quad & E\left(\left((F_k^\tau(u_2^\tau|X_t, \vartheta_k^\dagger) - F_k^\tau(u_1^\tau|X_t, \vartheta_k^\dagger)) - (F_0^\tau(u_2|X_t, \vartheta_0) - F_0^\tau(u_1|X_t, \vartheta_0))\right)^2\right). \end{aligned}$$

This measure defines a norm and implies a standard goodness of fit measure (see e.g. Corradi and Swanson (2005b). Recalling that $E\left(1\{u_1 \leq X_{t+\tau} \leq u_2\}|X_t\right) = F_0^\tau(u_2|X_t, \vartheta_0) - F_0^\tau(u_1|X_t, \vartheta_0)$, it is straightforward to construct a sequence of $P - \tau$ $\tau-$step ahead prediction errors under model k as $1\{u_1 \leq X_{t+\tau} \leq u_2\} - \left(F_k^\tau(u_2|X_t, \widehat{\vartheta}_{k,t,N,h}) - F_k^\tau(u_1|X_t, \widehat{\vartheta}_{k,t,N,h})\right)$, for $t = R, ..., R + P - \tau$, 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 $X_{R+1}, ..., X_{R+P-\tau}$ as starting values, and using the recursively estimated parameters, $\widehat{\vartheta}_{k,t,N,h}$, $t = R, ..., R + P - \tau$. 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, Corradi and Swanson (2008):

$$\frac{1}{N}\sum_{i=1}^N 1\left\{u_1 \leq X_{k,t+\tau,i}^{\widehat{\vartheta}_{k,t,N,h}}(X_t) \leq u_2\right\} \stackrel{pr}{\to} F_{X_{k,t+\tau}^{\vartheta_k^\dagger}(X_t)}(u_2) - F_{X_{k,t+\tau}^{\vartheta_k^\dagger}(X_t)}(u_1),\ t = R, ..., T - \tau, \tag{2}$$

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

where $F_{X^{\vartheta_k^\dagger}_{k,t+\tau}(X_t)}(u)$ is the marginal distribution of $X^{\vartheta_k^\dagger}_{t+\tau}(X_t)$ 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 $X^{\vartheta^\dagger}_{t+\tau}(X_t)$ is the distribution of $X_{t+\tau}$ conditional on the values observed at time t. Thus, $F_{X^{\vartheta_k^\dagger}_{k,t+\tau}(X_t)}(u) = F_k^\tau(u|X_t, \vartheta_k^\dagger)$. In the above expression, $X^{\widehat{\vartheta}_{k,t,N,h}}_{k,t+\tau,i}(X_t)$ is generated according to a Milstein scheme, where

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

with $\epsilon_{qh} \overset{iid}{\sim} N(0,h)$, $q = 1, \dots, Q$; and where σ' is the derivative of $\sigma(\cdot)$ with respect to its first argument. Additionally, the argument X_t in $X^{\widehat{\vartheta}_{k,t,N,h}}_{k,t+\tau,i}(X_t)$ denotes that the starting value for the simulation is X_t. Note, that the last term on the 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 are 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 $\mathcal{U}_1 \le \mathcal{U}_2 \le ... \le \mathcal{U}_{\mathcal{J}_k}$. These provide realizations for the $\mathcal{J}_k$ jump times. Then, make $\mathcal{J}_k$ independent draws from ϕ_k, say $y_{k,1}, ..., y_{k,\mathcal{J}_k}$. An important feature of this simulation procedure is that to generate $X^{\widehat{\vartheta}_{k,t,N,h}}_{k,t+\tau,i}(X_t)$, $i = 1, ..., N$, for $t = R, ..., T-\tau$ 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 change. More precisely, the errors used in simulation are defined to be $\epsilon_{qh,i} \overset{iid}{\sim} N(0,h)$, with $Qh = \tau$, $i = 1, ..., N$.

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

Turning now to the case of stochastic volatility 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 stochastic volatility 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, $k = 1, ..., m$, defined as follows:

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

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

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

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

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 eq. (3.7), p.347 in Kloeden and Platen (1999)):

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

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 stochastic volatility models, proceed as follows:
Step 1: Using the schemes in (5) and (6), simulate $(P-\tau)\times S\times N$ paths of length τ, setting the initial values for the observable state variable equal to the initial value X_t, $t = R+1, ..., R+P-\tau$; and for each X_t, using the S different starting values for volatility (i.e. $V_j^{\widehat{\theta}_{k,t,N,S,h}}$, $j = 1, ..., S$). Thus, there are S paths rather than one, for each starting value of X_t. For any initial value X_t and $V_j^{\widehat{\theta}_{k,t,N,S,h}}$, $t = R+1, ..., R+P-\tau$ and $j = 1, ..., S$, generate N independent paths of length τ. Also, keep the simulated randomness (i.e. $\epsilon_{1,qh}, \epsilon_{2,qh}$, $\int_{qh}^{(q+1)h}\left(\int_{qh}^{s} dW_{1,\tau}\right) dW_{2,s}$) 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 $\tau-$step ahead value for the return series simulated (under model k), at replication i, $i = 1, ..., N$, 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 $\frac{1}{NS}\sum_{j=1}^{S}\sum_{i=1}^{N} 1\left\{u_1 \leq X_{k,t+\tau,i,j}^{\widehat{\theta}_{k,t,N,S,h}}(X_t, V_{k,t,j}^{\widehat{\theta}_{k,t,N,S,h}}) \leq u_2\right\}$, where $V_{k,t,j}^{\widehat{\theta}_{k,t,N,S,h}}$ denote 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 $\beta-$mixing diffusion; and (ii) $\int_Y y^p \phi_k(y)dy < \infty$ for some $p > 2$.

Assumption A2: For $k = 1, ..., m$, $b_k(\cdot, \theta^{\dagger})$ 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, 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': Let $b_k(\cdot)$ and $\sigma_k(\cdot)$ (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, with 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 $k = 1, ..., m$.

Assumption A3: For $k = 1, ..., m$: (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 strictly stationary; (ii) $X_{k,t+\tau,i}^{\vartheta_k}$ is continuously differentiable in the interior of Θ_k, for $i = 1, ..., N$; and (iii) $\nabla_{\theta_k} X_{k,t+\tau,i}^{\vartheta_k}$ is $r-$dominated in Θ_k, uniformly in i for $r > 4$.

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

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

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

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

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

Assumption A4': For each model $k = 1, ..., m$ the parameters $\widehat{\vartheta}_{k,t,N,S,h}$ admit the following expansion:

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

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

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

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

Assumption A1(i) requires the diffusion, $X(t)$, to be geometrically ergodic and $\beta-$mixing. In the case of no jumps, conditions for (geometric) $\beta-$mixing for (multivariate) diffusions which can be relatively easily verified are provided by Meyn and Tweedie (1993). Such conditions suffice also for the case of jump diffusions, when both the intensity parameters and the jump sizes are independent of the state of the system. Recently, Makuda (2004) has extended the conditions for $\beta-$mixing to the case of jump diffusions in which the intensity parameter is constant, but the size of the jumps is state dependent.

Assumption A4 and A4' requires that the contribution of (recursive) parameter estimation error is $\sqrt{P}-$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 $E\left(\psi_{k,t,N,h}\left(\theta_k^{\dagger}\right)\right)$ and $E\left(\psi_{k,t,N,S,h}\left(\theta_k^{\dagger}\right)\right)$ are $o\left(P^{-1/2}\right)$, regardless of whether or not the model is misspecified. We shall show that NPSQMLE and exactly identified SGMM satisfy this requirement. Needless to say, in some cases the transition density is known in closed form, and can be used to obtain quasi maximum likelihood 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 use of the conditional empirical characteristic function (see e.g. Singleton (2001)).

3 Test Statistics

3.1 One Factor Models

First, consider comparing the predictive accuracy of 2 possibly misspecified diffusion models. The hypotheses of interest are:

$$H_0 : E_X\left(\left(F_{X_{1,t+\tau}^{\vartheta_1^\dagger}(X_t)}(u_2) - F_{X_{1,t+\tau}^{\vartheta_1^\dagger}(X_t)}(u_1)\right) - \left(F_0^\tau(u_2|X_t) - F_0^\tau(u_1|X_t)\right)\right)^2$$

$$-E_X\left(\left(F_{X_{k,t+\tau}^{\vartheta_k^\dagger}(X_t)}(u_2) - F_{X_{k,t+\tau}^{\vartheta_k^\dagger}(X_t)}(u_1)\right) - \left(F_0^\tau(u_2|X_t) - F_0^\tau(u_1|X_t)\right)\right)^2 = 0$$

$$H_A: \text{ negation of } H_0$$

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

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

Theorem 1: *Let Assumptions A1-A4 hold. Also, assume that models* 1 *and* k *are nonnested. If as* $P, R, N \to \infty$, $h \to 0$, $P/N \to 0$, $h^2 P \to 0$, *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, $\Pr\left(\frac{1}{\sqrt{P}} |D_{k,P,N}(u_1,u_2)| > \varepsilon\right) \to 0$.

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

$$\frac{1}{N}\sum_{i=1}^{N} 1\left\{X_{k,t+\tau,i}^{\widehat{\vartheta}_{k,t,N,h}}(X_t) \leq u\right\} = \frac{1}{N}\sum_{i=1}^{N} 1\left\{X_{k,t+\tau,i}^{\vartheta_k^\dagger}(X_t) \leq u\right\}$$

$$+E\left(f_{X_{k,t+\tau,i}^{\theta_k^\dagger}(X_t)}(u)\nabla_{\theta_k} X_{k,t+\tau,i}^{\vartheta_k^\dagger}(X_t)\right)\frac{1}{\sqrt{P}}\sum_{t=R}^{T}\left(\widehat{\vartheta}_{k,t,N,h} - \vartheta^\dagger\right) + o_P(1)$$

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

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.[2] More specifically, the hypotheses of interest are:

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

$$\left. - E_X \left(\left(F_{X^{\vartheta_k^\dagger}_{k,t+\tau}(X_t)}(u_2) - F_{X^{\vartheta_k^\dagger}_{k,t+\tau}(X_t)}(u_1) \right) - \left(F_0(u_2|X_t) - F_0(u_1|X_t) \right) \right)^2 \right) \leq 0$$

$$H_A' : \text{ negation of } H_0'$$

The statistic for testing these hypotheses is:

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

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

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

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*

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

for $j = 1, ..., m$, *and where* $(Z_1(u_1, u_2), ..., Z_m(u_1, u_2))$ *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

[2] See White (2000) for a discussion of a discrete time series analog to this case, where point rather than density based loss is considered; Corradi and Swanson (2007a) for an extension of White (2000) that allows for parameter estimation error; and Corradi and Swanson (2006a) for an extention of Corradi and Swanson (2007a) to the case of the comparison of conditional distributions and densities in a discrete time series context.

based $m-$estimators constructed using a recursive estimation scheme is outlined in Corradi and Swanson (2007a). Here we extend the results of Corradi and Swanson (2007a) by establishing asymptotic results for cases where 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 $X_t, X_{t+1}, ..., X_{t+\tau}$. More precisely, let $Z^{t,\tau} = (X_t, X_{t+1}, ..., X_{t+\tau})$, $t = 1, ..., T-\tau$. Now, resample b overlapping blocks of length l from $Z^{t,\tau}$. This yields $Z^{t,*} = (X_t^*, X_{t+1}^*, ..., X_{t+\tau}^*)$, $t = 1, ..., T-\tau$. 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 shall assume $N/R, N/P \to \infty$, simulation error vanishes and there is no need to resample the simulated series. Proceed by assuming that first order asymptotic validity of the bootstrap estimator can be established, as outlined in the following assumption (in Section 4 we shall provide primitive conditions under which NPSQMLE and SGMM satisfy this assumption).

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

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

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

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

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 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* $P, R, N \to \infty$, $h \to 0$, $P/N \to 0$, $h^2 P \to 0$, $l \to \infty$, $l/T^{1/4} \to 0$, *and* $P/R \to \pi$, *where* $0 < \pi < \infty$, *then:*

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

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

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

$$\to \quad 0.$$

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 lesser than the $\alpha/2th$ 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 inference drawn on them is conservative.

3.2 Stochastic Volatility Models

The test statistic for comparing two models is:

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

and the bootstrap test statistic is:

$$\begin{aligned}
& DV_{k,P,S,N}^*(u_1,u_2) \\
= & \frac{1}{\sqrt{P}} \sum_{t=R}^{T-\tau} \left\{ \left(\left[\frac{1}{NS} \sum_{j=1}^{S} \sum_{i=1}^{N} 1\left\{ u_1 \leq X_{1,t+\tau,i,j}^{\widehat{\theta}^*_{1,t,N,S,h}}(X_t^*, V_{1,j}^{\widehat{\theta}^*_{1,t,N,S,h}}) \leq u_2 \right\} - 1\{u_1 \leq X_{t+\tau}^* \leq u_2\} \right]^2 \right. \right. \\
& \left. - \left(\frac{1}{T} \sum_{l=1}^{T} \left[\frac{1}{NS} \sum_{j=1}^{S} \sum_{i=1}^{N} 1\left\{ u_1 \leq X_{1,t+\tau,i,j}^{\widehat{\theta}_{1,t,N,S,h}}(X_l, V_{1,j}^{\widehat{\theta}_{1,t,N,S,h}}) \leq u_2 \right\} - 1\{u_1 \leq X_{l+\tau} \leq u_2\} \right]^2 \right) \right) \\
& - \left(\left[\frac{1}{NS} \sum_{j=1}^{S} \sum_{i=1}^{N} 1\left\{ u_1 \leq X_{k,t+\tau,i,j}^{\widehat{\theta}^*_{k,t,N,S,h}}(X_t^*, V_{k,j}^{\widehat{\theta}^*_{k,t,N,S,h}}) \leq u_2 \right\} - 1\{u_1 \leq X_{t+\tau}^* \leq u_2\} \right]^2 \right. \\
& \left. \left. - \left(\frac{1}{T} \sum_{l=1}^{T} \left[\frac{1}{NS} \sum_{j=1}^{S} \sum_{i=1}^{N} 1\left\{ u_1 \leq X_{k,t+\tau,i,j}^{\widehat{\theta}_{k,t,N,S,h}}(X_l, V_{k,j}^{\widehat{\theta}_{k,t,N,S,h}}) \leq u_2 \right\} - 1\{u_1 \leq X_{l+\tau} \leq u_2\} \right]^2 \right) \right) \right\}.
\end{aligned}$$

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

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

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

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

Theorem 3: *Let Assumptions A1, A2', A3 and A4' hold. Also, assume that models* 1 *and* k *are nonnested. If as* $P, R, S, N \to \infty$, $h \to 0$, $P/N \to 0$, $P/S \to 0$, $h^2 P \to 0$, *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,*

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

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

Theorem 4: *Let Assumptions A1, A2', A3, and A4'-A5' hold. Also, assume that models* 1 *and* k *are nonnested. If as* $P, R, S, N \to \infty$, $h \to 0$, $P/N \to 0$, $P/S \to 0$, $h^2 P \to 0$, $l \to \infty$, $l/T^{1/4} \to 0$, *and* $P/R \to \pi$, *where* $0 < \pi < \infty$, *then:*

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

and

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

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

4 Recursive Nonparametric Simulated Quasi Maximum Likelihood Estimators

In this chapter 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 NPSQML and the corresponding recursive QMLE 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).[3]

[3]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, Monfort and Renault 1993, Dridi, Guay and Renault 2007), efficient method of moment (Gallant and Tauchen 1996) and simulated GMM with a continuum of moment conditions (Carrasco, Chernov, Florens and Ghysels 2007). We leave this to future research.

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 first order validity thereof.

Hereafter, let $f_k\left(X_t|X_{t-1},\vartheta_k^{\dagger}\right)$ 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 QMLE as:[4]

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

Now, define:

$$\vartheta_k^{\dagger}=\arg\max_{\vartheta_k\in\Theta_k}E\left(\ln f_k\left(X_t|X_{t-1},\vartheta_k\right)\right). \tag{7}$$

Following Kristensen and Shin (2008), generate $T-1$ paths of length one for each simulation replication, using $X_1,...,X_{T-1}$ as starting values; and hence construct $X_{k,t,j}^{\vartheta}(X_{t-1})$, for $t=2,...,T-1$, $j=1,...,N$. Note that we keep the N random draws fixed across different initial values. Then, define the following estimator of the conditional density:

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

Further, define the recursive NPSQML estimator as follows:

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

where the trimming function $\tau_N\left(\widehat{f}_{k,N,h}\left(X_t|X_{t-1},\vartheta_k\right)\right)$ is a positive and increasing function, such that $\tau_N\left(\widehat{f}_{k,N,h}\left(X_t,X_{t-1},\vartheta_k\right)\right)=0$, if $\widehat{f}_{k,N,h}\left(X_t,X_{t-1},\vartheta_k\right)<\xi_N^{\delta}$, and $\tau_N\left(\widehat{f}_{k,N,h}\left(X_t,X_{t-1},\vartheta_k\right)\right)=1$, if $\widehat{f}_{k,N,h}\left(X_t,X_{t-1},\vartheta_k\right)>2\xi_N^{\delta}$, for some $\delta>0$.[5] 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.

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

[5]As an example of a trimming function, Fermanian and Salanie (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 $a_N\leq x\leq 2a_N$.

Assumption A3': For $k = 1, ..., m$: (i) $X_i^{\vartheta_k}(x)$ and $X_{i,h}^{\vartheta_k}(x)$ are geometrically ergodic and strictly stationary, (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 6: Let $\mathcal{N}_{\vartheta_k^\dagger}$ be a neighborhood of $\vartheta_k^\dagger$, $E\left(\sup_{\vartheta_k \in \mathcal{N}_{\vartheta_k^\dagger}} \left\| \frac{\partial \ln f_k(X_t|X_{t-1},\vartheta_k)}{\partial \vartheta_k} \right\|^r\right) < \infty$, $E\left(\sup_{\vartheta_k \in \mathcal{N}_{\vartheta_k^\dagger}} \left\| \frac{\partial f_k(X_t|X_{t-1},\vartheta_k)}{\partial \vartheta_k} \right\|^r\right) < \infty$, $E\left(\sup_{\vartheta_k \in \mathcal{N}_{\vartheta_k^\dagger}} \left\| \frac{\partial X_i^{\vartheta_k}(X_{t-1})}{\partial \vartheta_k} \right\|^r\right) < \infty$, $E\left(\sup_{\vartheta_k \in \mathcal{N}_{\vartheta_k^\dagger}} \left\| \frac{\partial X_{i,h}^{\vartheta_k}(X_{t-1})}{\partial \vartheta_k} \right\|^r\right) < \infty$, for $k = 1, ..., m$ and for $r > 4$.
Assumption 7: For $k = 1, ..., m$: (i) $\vartheta_k^\dagger$ is uniquely identified (i.e. $E(\ln f_k(X_t|X_{t-1}, \vartheta_k)) < E(\ln f_k\left(X_t|X_{t-1}, \vartheta_k^\dagger\right))$ 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 $\nabla_{\vartheta_k} f_k(X_t|X_{t-1}, \vartheta_k)$, $\nabla_{\vartheta_k}^2 f_k(X_t|X_{t-1}, \vartheta_k)$, $\nabla_{\vartheta_k} \ln f_k(X_t|X_{t-1}, \vartheta_k)$ and $\nabla_{\vartheta_k} \ln f_k(X_t|X_{t-1}, \vartheta_k)$ are $r-$dominated on Θ_k, with $r > 4$; and (iv) $E\left(-\nabla_\theta^2 \ln f_k(\vartheta_k)\right)$ is positive definite, uniformly on Θ_k.
Assumption 8: 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: $\int K(u)\mathrm{d}u = 1$, $\int u^{s-1}(u)\mathrm{d}u = 0$, $s \geq 2$. 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 $N = T^a$ $a > 1$. *If as* $T, P, N \to \infty$, *(a)* $T^{\frac{r}{2(r-1)}} \xi_N^\delta \left|\ln \xi_N\right|^{\frac{r+1}{2r-1}} \to 0$, (b) $T^{1/2}\xi_N^{s-\delta}\left|\ln \xi_N\right| \to 0$, (c) $T^{(1-a)}\xi_N^{-4-2\delta}(\ln \xi_N^2) \ln T^a \to 0$, (d) $T^{1/2}\xi_N^{-(\delta+3)}h\left|\ln \xi_N^\delta\right| \to 0$. Then, for $k = 1, ..., m$: (i) $\sup_{t \geq R}\left(\widehat{\vartheta}_{k,t,N,h} - \vartheta_k^\dagger\right) \xrightarrow{p} 0$ and (ii) $\frac{1}{\sqrt{P}}\sum_{t=R}^T \left(\widehat{\vartheta}_{k,t,N,h} - \vartheta_k^\dagger\right) \xrightarrow{d} N(0, 2\Pi A_k^\dagger V_k^\dagger A_k^\dagger)$, *where* $A_k^\dagger = E\left(-\nabla_{\theta_k} f_k\left(X_t|X_{t-1}, \vartheta_k^\dagger\right)\right)$, $V_k^\dagger = \sum_{i=-\infty}^\infty E\left(\nabla_{\theta_k} f_k\left(X_2|X_1, \vartheta_k^\dagger\right)\nabla_{\theta_k} f_k\left(X_{2+i}|X_{1+i}, \vartheta_k^\dagger\right)'\right)$ *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 $T^{1/2}\xi_N^{-(d+3)}h^{1/2}\left|\ln \xi_N\right| \to 0$.

From Theorem 5 is immediate to see that NPSQMLE satisfies Assumption 4; and is asymptotically equivalent to unfeasible QMLE, which is constructed by maximizing the likelihood of model k. An interesting alternative nonparametric simulated maximum likelihood estimator has recently been suggested by Altissimo and Mele (2005). 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 MLE.

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 $Z^{t,*} = (X_t^*, X_{t+1}^*, ..., X_{t+\tau}^*)$, $t = 1, ..., T - \tau$ be as outlined in Section 3. For each simulation replication, generate $T - 1$ paths of length one, using as starting values $X_1^*, ..., X_{T-1}^*$; and so obtaining $X_{k,t,j}^{\vartheta_k}(X_{t-1}^*)$, for $t = 2, ..., T - 1$, $j = 1, ..., N$. Further, let:

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

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

$$
\begin{aligned}
\widehat{\vartheta}^{*}_{k,t,N,h} &= \arg\max_{\vartheta_k\in\Theta_k}\frac{1}{t}\sum_{l=2}^{t}\left(\ln\widehat{f}_{k,N,h}\left(X^{*}_{l}|X^{*}_{l-1},\vartheta_k\right)\tau_N\left(\widehat{f}_{k,N,h}\left(X^{*}_{l}|X^{*}_{l-1},\vartheta_k\right)\right)\right.\\
&\quad -\vartheta_k'\left(\frac{1}{T}\sum_{l'=2}^{T}\frac{\nabla_{\vartheta_k}\widehat{f}_{k,N,h}\left(X_{l'}|X_{l'-1},\vartheta_k\right)}{\widehat{f}_{k,N,h}\left(X_{s'}|X_{t-s'},\vartheta_k\right)}\Big|_{\vartheta_k=\widetilde{\vartheta}_{k,t,N,h}}\tau_N\left(\widehat{f}_{k,N,h}\left(X_{l'}|X_{l'-1},\widehat{\vartheta}_{k,t,N,h}\right)\right)\right.\\
&\quad \left.\left.+\tau_N'\left(\widehat{f}_{k,N,h}\left(X_{l'}|X_{l'-1},\widehat{\vartheta}_{k,t,N,h}\right)\right)\nabla_{\vartheta_k}\widehat{f}_{k,N,h}\left(X_{l'}|X_{l'-1},\vartheta_k\right)\Big|_{\widehat{\vartheta}_{k,t,N,h}}\ln\widehat{f}_{k,N,h}\left(X_{l'}|X_{l'-1},\widehat{\vartheta}_{k,t,N,h}\right)\right)\right),
\end{aligned}
$$

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 $\nabla_{\theta_k}\widehat{f}_{k,N,h}\left(X_{l'}|X_{l'-1},\widehat{\vartheta}_{k,t,N,h}\right)$, 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 $N=T^a$ $a>1$. *If as $T,N,l\to\infty$, $l/T^{1/4}\to 0$, (a) $T^{\frac{r}{2(r-1)}}\xi_N^{\delta}\left|\ln\xi_N\right|^{\frac{r+1}{2r-1}}\to 0$,* (b) $T^{1/2}\xi_N^{s-\delta}\left|\ln\xi_N\right|\to 0$, (c) $T^{(1-a)}\xi_N^{-4-2\delta}(\ln\xi_N^2)\ln T^a\to 0$, (d) $T^{1/2}\xi_N^{-(\delta+3)}h\left|\ln\xi_N^{\delta}\right|\to 0$. Then, for $k=1,...,m$:

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

where P^{*}_{T} denotes the probability law of the resampled series, conditional on the (entire) sample.

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

4.2 Stochastic Volatility Models

Since volatility is not observable, we cannot proceed as in the single factor case when estimating the SV model using NPSQMLE. Instead, let $V_s^{\theta_k}$ be generated according to (4), setting $qh=s$, $q=1,...,1/h$, and $s=1,...,S$. For each model $k=1,..m$, and at each simulation replication, $i=1,...,N$, 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. $V_s^{\theta_k}$, $s=1,...,S$). Thus, for any $t=1,...,T-1$, and for any set i, $i=1,...,N$ of random errors $\epsilon_{1,t+(q+1)h,i}$ and $\epsilon_{2,t+(q+1)h,i}$, $q=1,...,1/h$, 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 $t=2,...,T$, $i=1,...,N$ and $s=1,...,S$. Now, define:

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

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

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

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

Assumption 9: Let $\mathcal{N}_{\vartheta_k^\dagger}$ be a neighborhood of $\vartheta_k^\dagger$, $E\left(\sup_{\vartheta_k \in \mathcal{N}_{\vartheta_k^\dagger}} \left\| \frac{\partial X_i^{\vartheta_k}(X_{t-1}, V_j^{\vartheta_k})}{\partial \vartheta_k} \right\|^r\right) < \infty$, $E\left(\sup_{\vartheta_k \in \mathcal{N}_{\vartheta_k^\dagger}} \left\| \frac{\partial X_{i,h}^{\vartheta_k}(X_{t-1}, V_j^{\vartheta_k})}{\partial \vartheta_k} \right\|^r\right) < \infty$, for $k = 1, ..., m$ and for $r > 4$.

Theorem 7: *Let Assumptions A1,A2'-A3', and A6-A9 hold. Let $T = R + P$, $P/R \to \pi$, where $0 < \pi < \infty$. Let $N = P^a$, for $\varepsilon > 0$ arbitrarily small, $h^{1-\varepsilon}N \to 0$, $a > 1$. If as $T, N \to \infty$, (a) $T^{\frac{r}{2(r-1)}} \xi_N^{\delta} \left|\ln \xi_N\right|^{\frac{r+1}{2r-1}} \to 0$, (b) $T^{1/2}\xi_N^{s-\delta}\left|\ln \xi_N\right| \to 0$, (c) $T^{(1-a)}\xi_N^{-4-2\delta}(\ln \xi_N^2)\ln T^a \to 0$, (d) $T^{1/2(1-a)}\xi_N^{-(\delta+3)}h\left|\ln \xi_N^{\delta}\right| \to 0$, (e) $T^{1/2}S^{-1/2}\xi_N^{-(1+3\delta)} \to 0$. Then for $k = 1, ..., m$:*
(i) $\sup_{t\geq R}\left(\widehat{\theta}_{k,t,N,S,h} - \theta_k^\dagger\right) \xrightarrow{p} 0$ and (ii) $\frac{1}{\sqrt{P}}\sum_{t=R}^T\left(\widehat{\theta}_{k,t,N,S,h} - \theta_k^\dagger\right) \xrightarrow{d} N(0, 2\Pi A_k^\dagger V_k^\dagger A_k^\dagger)$, *where* $A_k^\dagger = E\left(-\nabla_{\theta_k} f_k\left(X_t|X_{t-1}, \theta_k^\dagger\right)\right)$, $V_k^\dagger = \sum_{i=-\infty}^{\infty} E\left(\nabla_{\theta_k} f_k\left(X_2|X_1, \theta_k^\dagger\right)\nabla_{\theta_k} f_k\left(X_{2+i}|X_{1+i}, \theta_k^\dagger\right)'\right)$, *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 where 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 QMLE estimator. Furthermore, note that $\nabla_{\theta_k} f(X_t|X_{t-1}, \theta_k^\dagger)$ is no longer a martingale difference sequence, and 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:

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

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

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

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 $N = T^a$ $a > 1$. *If as $T, N, l \to \infty$, $l/T^{1/4} \to 0$, (a) $T^{\frac{r}{2(r-1)}}\xi_N^{\delta}\left|\ln \xi_N\right|^{\frac{r+1}{2r-1}} \to 0$,* (b) $T^{1/2}\xi_N^{s-\delta}\left|\ln \xi_N\right| \to 0$, (c) $T^{(1-a)}\xi_N^{-4-2\delta}(\ln \xi_N^2)\ln T^a \to 0$, (d) $T^{1/2}\xi_N^{-(\delta+3)}h\left|\ln \xi_N^{\delta}\right| \to 0$, *(e)* $T^{1/2}S^{-1/2}\xi_N^{-(1+3\delta)} \to 0$. Then, for $k = 1, ..., m$:

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

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

5 Empirical Illustration: Choosing Between CIR, SV, and SVJ Models

In this section, we choose between 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 tests 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 sub-sample 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, Johannes, and Polson (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, Corradi and Swanson (BCS: 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: $dX(t) = \kappa_1(\alpha_1 - X(t))\,dt + \gamma_1\sqrt{X(t)}dW_1(t)$, where $\kappa_1 > 0$, $\gamma_1 > 0$ and $2\kappa_1\alpha_1 \geq \gamma_1^2$,

SV: $dX(t) = \kappa_2(\alpha_2 - X(t))\,dt + \sqrt{V(t)}dW_1(t)$, and $dV(t) = \kappa_3(\alpha_3 - V(t))\,dt + \gamma_2\sqrt{V(t)}dW_2(t)$, where $W_1(t)$ and $W_2(t)$ 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: $dX(t) = \kappa_4(\alpha_4 - X(t))\,dt + \sqrt{V(t)}dW_1(t) + J_u dq_u - J_d dq_d$, and $dV(t) = \kappa_5(\alpha_5 - V(t))\,dt + \gamma_3\sqrt{V(t)}dW_2(t)$, where $W_r(t)$ and $W_v(t)$ 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 exponential distributions, with densities: $f(J_u) = \frac{1}{\zeta_u}\exp\left(-\frac{J_u}{\zeta_u}\right)$ and $f(J_d) = \frac{1}{\zeta_d}\exp\left(-\frac{J_d}{\zeta_d}\right)$. 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.

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

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

and

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

depending upon whether we are predicting using one factor or stochastic volatility models. We define the CIR model to be our "benchmark", against which the other models are compared. 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 95^{th}, 90^{th}, 85^{th}, and 80^{th} 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 $\tau - step$ 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 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 2 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 95^{th} 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, the absolute magnitude of the SV $PDMSFEs$ are actually substantively lower than those for the CIR model, in Tables 3 and 4, 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 importantly, 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.[6]

In Figures 1 and 2, predictive densities are plotted for various evaluation points given a particular set of recursively estimated 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. Figure 1 reports densities for our first sample period and Figure 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 lesser dispersion that its competitor model. Interestingly, in Figure 1 it is quite apparent that the SVJ model is preferred, although none of the models are particularly well centered for evaluation points not equal to the mean of 0.055. In Figure 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 Figures 1a and 1c). 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.

[6] One might be tempted to think that the SVJ model should always be selected, as it "nests" the other models. However, as we are performing true ex ante prediction experiments using predictive densities, this is clearly not the case; more parsimonious models should be expected to perform better, particularly if they are "better approximations" of the true underlying DGP.

6 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:

$$
\begin{aligned}
& \frac{1}{\sqrt{P}} \sum_{t=R}^{T-\tau} \left(\frac{1}{N} \sum_{i=1}^{N} 1 \left\{ X_{1,t+\tau,i}^{\widehat{\vartheta}_{1,t,N,h}}(X_t) \leq u \right\} - 1\{X_{t+\tau} \leq u\} \right)^2 \\
= & \frac{1}{\sqrt{P}} \sum_{t=R}^{T-\tau} \left(\frac{1}{N} \sum_{i=1}^{N} 1 \left\{ X_{1,t+\tau,i}^{\vartheta_1^\dagger}(X_t) \leq u \right\} - 1\{X_{t+\tau} \leq u\} \right)^2 \\
& + \frac{1}{\sqrt{P}} \sum_{t=R}^{T-\tau} \left(\frac{1}{N} \sum_{i=1}^{N} \left(1 \left\{ X_{1,t+\tau,i}^{\widehat{\vartheta}_{1,t,N,h}}(X_t) \leq u \right\} - 1 \left\{ X_{1,t+\tau,i}^{\vartheta_1^\dagger}(X_t) \leq u \right\} \right) \right)^2 \\
& + \frac{2}{\sqrt{P}} \sum_{t=R}^{T-\tau} \left[\left(\frac{1}{N} \sum_{i=1}^{N} 1 \left\{ X_{1,t+\tau,i}^{\vartheta_1^\dagger}(X_t) \leq u \right\} - 1\{X_{t+\tau} \leq u\} \right) \right. \\
& \left. \times \left(\frac{1}{N} \sum_{i=1}^{N} \left(1 \left\{ X_{1,t+\tau,i}^{\widehat{\vartheta}_{1,t,N,h}}(X_t) \leq u \right\} - 1 \left\{ X_{1,t+\tau,i}^{\vartheta_1^\dagger}(X_t) \leq u \right\} \right) \right) \right] \\
= & I_{P,N,h} + II_{P,N,h} + III_{P,N,h} \qquad (8)
\end{aligned}
$$

Now,

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

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

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

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

$$
\begin{aligned}
III_{P,N,h} = & \frac{1}{\sqrt{P}} \sum_{t=R}^{T-\tau} \left[\left(F_{X_{1,t+\tau}^{\vartheta_1^\dagger}}(u|X_t) - 1\{X_{t+\tau} \leq u\} \right) \right. \\
& \left. \times \left(\frac{1}{N} \sum_{i=1}^{N} \left(1 \left\{ X_{1,t+\tau,i}^{\widehat{\vartheta}_{1,t,N,h}}(X_t) \leq u \right\} - 1 \left\{ X_{1,t+\tau,i}^{\vartheta_1^\dagger}(X_t) \leq u \right\} \right) \right) \right] + o_P(1) \\
= & \frac{\mu_{F_1}}{\sqrt{P}} \sum_{t=R}^{T-\tau} \left(\frac{1}{N} \sum_{i=1}^{N} \left(1 \left\{ X_{1,t+\tau,i}^{\widehat{\vartheta}_{1,t,N,h}}(X_t) \leq u \right\} - 1 \left\{ X_{1,t+\tau,i}^{\vartheta_1^\dagger}(X_t) \leq u \right\} \right) \right) + o_P(1)
\end{aligned}
$$

$$
\begin{aligned}
= \quad & \frac{\mu_{F_1}}{\sqrt{P}} \sum_{t=R}^{T-\tau} \left(\left(\frac{1}{N} \sum_{i=1}^{N} 1 \left\{ X_{1,t+\tau,i}^{\vartheta_1^\dagger}(X_t) \leq u - \left(X_{1,t+\tau,i}^{\widehat{\vartheta}_{1,t,N,h}}(X_t) - X_{1,t+\tau,i}^{\vartheta_1^\dagger}(X_t) \right) \right\} \right. \right. \\
& -F_{X_{1,t+\tau}^{\vartheta_1^\dagger}} \left(\left(u - \left(X_{1,t+\tau,i}^{\widehat{\vartheta}_{1,t,N,h}}(X_t) - X_{1,t+\tau,i}^{\vartheta_1^\dagger}(X_t) \right) \right) |X_t \right) \Bigg) - \left(\frac{1}{N} \sum_{i=1}^{N} 1 \left\{ X_{1,t+\tau,i}^{\vartheta_1^\dagger}(X_t) \leq u \right\} - F_{X_{1,t+\tau}^{\vartheta_1^\dagger}}(u|X_t) \right) \Bigg) \\
& + \frac{\mu_{F_1}}{\sqrt{P}} \sum_{t=R}^{T-\tau} \frac{1}{N} \sum_{i=1}^{N} \left(F_{X_{1,t+\tau}^{\vartheta_1^\dagger}} \left(\left(u - \left(X_{1,t+\tau,i}^{\widehat{\vartheta}_{1,t,N,h}}(X_t) - X_{1,t+\tau,i}^{\vartheta_1^\dagger}(X_t) \right) \right) |X_t \right) - F_{X_{1,t+\tau}^{\vartheta_1^\dagger}}(u|X_t) \right) + o_p(1). \qquad (9)
\end{aligned}
$$

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 (9) 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 (9) can be written as:

$$
\begin{aligned}
& \frac{1}{\sqrt{P}} \sum_{t=R}^{T-\tau} \left(\frac{1}{N} \sum_{i=1}^{N} f_1 \left(\left(u - \left(X_{1,t+\tau,i}^{\overline{\vartheta}_{1,t,N,h}}(X_t) - X_{1,t+\tau,i}^{\vartheta_1^\dagger}(X_t) \right) \right) |X_t \right) \nabla_{\theta_1} X_{1,t+\tau,i}^{\overline{\vartheta}_{1,t,N,h}}(X_t) \right) \\
& \times \left(\widehat{\vartheta}_{1,t,N,h} - \vartheta_1^\dagger \right), \qquad (10)
\end{aligned}
$$

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 (8). By treating model k in the same manner as model 1, we have that,

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

Now, let $\mu_{f_k,\theta_k^\dagger}(u)' = E_X \left(f_{X_{1,t+\tau}^{\vartheta_k^\dagger}}(u|X_t)\, E_N \left(\nabla_{\theta_k} X_{k,t+\tau,i}^{\overline{\vartheta}_{k,t,N,h}}(X_t) \right)' \right)$, 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:

$$
\begin{aligned}
& D_{k,P,N}(u) \\
= \; & \frac{1}{\sqrt{P}} \sum_{t=R}^{T-\tau} \left(\left(F_{X_{1,t+\tau}^{\vartheta_1^\dagger}}(u|X_t) - F_0(u|X_t) \right)^2 - \left(F_{X_{k,t+\tau}^{\vartheta_k^\dagger}}(u|X_t) - F_0(u|X_t) \right)^2 \right) \\
& + \frac{1}{\sqrt{P}} \sum_{t=R}^{T-\tau} \left((F_0(u|X_t) - 1\{X_{t+\tau} \leq u\}) \left(F_{X_{1,t+\tau}^{\vartheta_1^\dagger}}(u|X_t) - F_{X_{k,t+\tau}^{\vartheta_k^\dagger}}(u|X_t) \right) \right) \\
& + \mu_{F_1} \mu_{f_1,\vartheta_1^\dagger}(u) A_1^\dagger \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \psi_{1,t,N,h} \left(\vartheta_1^\dagger \right) - \mu_{F_k} \mu_{f_k,\vartheta_k^\dagger}(u) A_k^\dagger \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} \psi_{k,t,N,h} \left(\vartheta_k^\dagger \right) + o_P(1).
\end{aligned}
$$

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

$$
\begin{aligned}
W_k(u) \;=\; & C(u) + V(u) + CV(u) + P_{11}(u) \\
& + P_{kk}(u) - P_{1k}(u) + P_1C(u)P_kC(u) + P_1V(u) - P_kV(u),
\end{aligned}
$$

and where, recalling A4,

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

$$
\begin{aligned}
V(u) = & \sum_{j=0}^{\infty} E \left(\left((F_0^\tau(u|X_1) - 1\{X_{1+\tau} \leq u\}) \left(F_{X_{1,1+\tau}^{\vartheta_1^\dagger}}(u|X_1) - F_{X_{k,1+\tau}^{\vartheta_k^\dagger}}(u|X_1) \right) \right) \right. \\
& \left. \left((F_0^\tau(u|X_{1+j}) - 1\{X_{1+j+\tau} \leq u\}) \left(F_{X_{1,1+j+\tau}^{\vartheta_1^\dagger}}(u|X_{1+j}) - F_{X_{k,1+j+\tau}^{\vartheta_k^\dagger}}(u|X_{1+j}) \right) \right) \right)
\end{aligned}
$$

$$
\begin{aligned}
CV(u) \;=\; & \sum_{j=0}^{\infty} E \left(\left(\left(F_{X_{1,1+\tau}^{\theta_1^\dagger}}(u|X_1) - F_0(u|X_1) \right)^2 - \left(F_{X_{k,1+\tau}^{\theta_k^\dagger}}(u|X_1) - F_0(u|X_1) \right)^2 \right) \right. \\
& \left. \left((F_0^\tau(u|X_{1+j}) - 1\{X_{1+j+\tau} \leq u\}) \left(F_{X_{1,1+j+\tau}^{\theta_1^\dagger}}(u|X_{1+j}) - F_{X_{k,1+j+\tau}^{\theta_k^\dagger}}(u|X_{1+j}) \right) \right) \right)
\end{aligned}
$$

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

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

$$P_1C(u) = 4\Pi\mu_{F_1}(u)\mu_{f_1,\vartheta_1^\dagger}(u)A_1^\dagger\sum_{j=0}^{\infty}E\left(\psi_{1,1}\left(\theta_1^\dagger\right)\right.$$

$$\left.\left(\left(F_{X^{\vartheta_1^\dagger}_{1,1+j+\tau}}(u|X_{1+j})-F_0(u|X_{1+j})\right)^2-\left(F_{X^{\vartheta_k^\dagger}_{k,1+j+\tau}}(u|X_{1+j})-F_0(u|X_{1+j})\right)^2\right)\right),$$

and

$$P_1V_1(u) = 4\Pi\mu_{F_1}(u)\mu_{f_1,\theta_1^\dagger}(u)'A_1^\dagger\sum_{j=0}^{\infty}E\left(\psi_{1,1}\left(\theta_1^\dagger\right)\right.$$

$$\left.\left(\left(F_0^\tau(u|X_{1+j})-1\{X_{1+j+\tau}\leq u\}\right)\left(F_{X^{\theta_1^\dagger}_{1,1+j+\tau}}(u|X_{1+j})-F_{X^{\theta_k^\dagger}_{k,1+j+\tau}}(u|X_{1+j})\right)\right)\right),$$

(ii) Under H_A, $\frac{1}{\sqrt{P}}\sum_{t=R}^{T-\tau}\left(\left(F_{X^{\vartheta_1^\dagger}_{1,t+\tau}}(u|X_t)-F_0(u|X_t)\right)^2-\left(F_{X^{\vartheta_k^\dagger}_{k,t+\tau}}(u|X_t)-F_0(u|X_t)\right)^2\right)$ 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,

$$\left(\left(D_{2,P,N}(u_1,u_2)-\mu_2(u_1,u_2)\right),...,\left(D_{m,P,N}(u_1,u_2)-\mu_m(u_1,u_2)\right)\right)$$

converges to a $m-$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.

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:

$$\begin{aligned}
= \; & \frac{1}{\sqrt{P}}\sum_{t=R}^{T-\tau}\left(\left[\frac{1}{N}\sum_{i=1}^{N}1\left\{X^{\widehat{\vartheta}_{1,t,N,h}}_{1,t+\tau,i}(X_t^*)\leq u\right\}-1\{X^*_{t+\tau}\leq u\}\right]^2\right.\\
& \left.-\left(\frac{1}{T}\sum_{j=1}^{T}\left[\frac{1}{N}\sum_{i=1}^{N}1\left\{X^{\widehat{\vartheta}_{1,t,N,h}}_{1,t+\tau,i}(X_j)\leq u\right\}-1\{X_{j+\tau}\leq u\}\right]^2\right)\right)\\
& +\frac{1}{\sqrt{P}}\sum_{t=R}^{T-\tau}\left[\frac{1}{N}\sum_{i=1}^{N}\left(1\left\{X^{\widehat{\vartheta}^*_{1,t,N,h}}_{1,t+\tau,i}(X_t^*)\leq u\right\}-1\left\{X^{\widehat{\vartheta}_{1,t,N,h}}_{1,t+\tau,i}(X_t^*)\leq u\right\}\right)\right]^2\\
& +2\frac{1}{\sqrt{P}}\sum_{t=R}^{T-\tau}\left[\left(\frac{1}{N}\sum_{i=1}^{N}1\left\{X^{\widehat{\vartheta}_{1,t,N,h}}_{1,t+\tau,i}(X_t^*)\leq u\right\}-1\{X^*_{t+\tau}\leq u\}\right)\right.\\
& \left.\times\left(\frac{1}{N}\sum_{i=1}^{N}\left(1\left\{X^{\widehat{\vartheta}^*_{1,t,N,h}}_{1,t+\tau,i}(X_t^*)\leq u\right\}-1\left\{X^{\widehat{\vartheta}_{1,t,N,h}}_{1,t+\tau,i}(X_t^*)\leq u\right\}\right)\right)\right] \qquad (11)
\end{aligned}$$

First, note that:

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

Also, by the same arguments as those used in the proof of Theorem 4 in Bhardwaj, Corradi and Swanson (2008),

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

Thus, from Theorem 3.5 in Künsch (1989), it follows that the first term on the RHS of the last equality in (11) has the same limiting distribution as:

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

Now, $\frac{1}{N}\sum_{i=1}^{N} 1\left\{X_{1,t+\tau,i}^{\vartheta_1^{\dagger}}(X_t)\leq u\right\} - F_{X_{1,t+\tau}^{\vartheta_1^{\dagger}}}(u|X_t) = O_N(N^{-1/2})$, and as $N/P\to\infty$, the third term on the RHS of (11) can be written as:

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

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

$$
\begin{aligned}
& \frac{1}{\sqrt{P}}\sum_{t=R}^{T-\tau}\left(\frac{1}{N}\sum_{i=1}^{N}\left(1\left\{X_{1,t+\tau,i}^{\widehat{\vartheta}^{*}_{1,t,N,h}}(X_t^{*})\leq u\right\} - 1\left\{X_{1,t+\tau,i}^{\widehat{\vartheta}_{1,t,N,h}}(X_t^{*})\leq u\right\}\right)\right) \\
&= \frac{1}{\sqrt{P}}\sum_{t=R}^{T-\tau}\left(\left(\frac{1}{N}\sum_{i=1}^{N} 1\left\{X_{1,t+\tau,i}^{\widehat{\vartheta}_{1,t,N,h}}(X_t)\leq u - \left(X_{1,t+\tau,i}^{\widehat{\vartheta}^{*}_{1,t,N,h}}(X_t) - X_{1,t+\tau,i}^{\widehat{\vartheta}_{1,t,N,h}}(X_t)\right)\right\}\right.\right. \\
&\quad - F_{X_{1,t+\tau}^{\widehat{\vartheta}_{1,t,N,h}}}\left(\left(u - \left(X_{1,t+\tau,i}^{\widehat{\vartheta}^{*}_{1,t,N,h}}(X_t) - X_{1,t+\tau,i}^{\widehat{\vartheta}_{1,t,N,h}}(X_t)\right)\right)|X_t\right)\Bigg) - \\
&\quad \left(\frac{1}{N}\sum_{i=1}^{N} 1\left\{X_{1,t+\tau,i}^{\widehat{\vartheta}_{1,t,N,h}}(X_t)\leq u\right\} - F_{X_{1,t+\tau}^{\widehat{\vartheta}_{1,t,N,h}}}(u|X_t)\right)\Bigg) \\
&\quad + \frac{1}{\sqrt{P}}\sum_{t=R}^{T-\tau}\frac{1}{N}\sum_{i=1}^{N}\left(F_{X_{1,t+\tau}^{\widehat{\vartheta}_{1,t,N,h}}}\left(\left(u - \left(X_{1,t+\tau,i}^{\widehat{\vartheta}^{*}_{1,t,N,h}}(X_t) - X_{1,t+\tau,i}^{\widehat{\vartheta}_{1,t,N,h}}(X_t)\right)\right)|X_t\right) - F_{X_{1,t+\tau}^{\widehat{\vartheta}_{1,t,N,h}}}(u|X_t)\right). \qquad (12)
\end{aligned}
$$

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

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

Finally, the last term on the RHS of (12), conditional on the sample, and for all samples except a set with probability measure approaching zero, has the same limiting distribution as:

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

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

Proof of Corollary 2: Given Theorem 2, the result follows directly upon application of the Cramer-Wold device and the continuous mapping theorem.

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:

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

The statement follows by the same argument as that used in Theorem 1, as by Proposition 5 in Bhardwaj, Corradi and Swanson (2008),

$$
E\left(\frac{1}{NS}\sum_{j=1}^{S}\sum_{i=1}^{N}1\left\{X_{1,t+\tau,i,j}^{\theta_1^\dagger}(X_t)\leq u\right\}-F_{X_{1,t+\tau}^{\vartheta_1^\dagger}}(u|X_t,V_{j,h}^{\theta_1^\dagger})\right)=0,
$$

and for $S/P\to\infty$ and $N/P\to\infty$,

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

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,

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

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:

$$L_{t,h}^N \left(\vartheta \right) = \frac{1}{t} \sum_{j=1}^{t} \ln \widehat{f}_{k,N,h} \left(X_j | X_{j-1}, \vartheta \right) \tau_N \left(\widehat{f}_{k,N,h} \left(X_j | X_{j-1}, \vartheta \right) \right), \tag{13}$$

$$L_t^N \left(\vartheta \right) = \frac{1}{t} \sum_{j=1}^{t} \ln \widehat{f}_{k,N} \left(X_j | X_{j-1}, \vartheta \right) \tau_N \left(\widehat{f}_{k,N} \left(X_j | X_{j-1}, \vartheta \right) \right), \tag{14}$$

and

$$L_t \left(\vartheta \right) = \frac{1}{t} \sum_{j=1}^{t} \ln f \left(X_j | X_{j-1}, \vartheta \right), \tag{15}$$

where $L_t\left(\vartheta\right)$ 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:

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

Step 2:

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

Step 3:

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

Step 4:

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

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, FS), 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 $\beta-$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, A8 implies K0, in FS. A1(ii)-(iii) and A6-A7 imply L1 and L2, with $\beta = r$, and L3, with $\gamma = \gamma' = r > 4$ in FS. A3' implies M1 with $s_0 = 0$, and M2 with $r_0 = s_1 = 0$ and $p_0 = \zeta = r > 4$, in FS. It remains to check that the rate conditions T1, R1, T2, R2 and R3 in FS 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), $\Pr\left(\sup_t |X_t| > \varepsilon T^a\right) \leq \sum_{t=1}^{T} \Pr\left(|X_t| > \varepsilon T^a\right) \leq \frac{1}{\varepsilon^r} T^{1-ar} E(|X_t|^r)$, and as $a > 1$ and $r > 4$, (c) in the statement of the theorem implies T2 (and hence T1) in FS for $v = 1$ and $\gamma = \gamma' = \zeta = r > 4$. As $\Pr\left(\inf_{\vartheta \in \mathcal{N}_{\vartheta_k^\dagger}} f\left(X_j|X_{j-1},\vartheta\right) < \xi_N^\delta\right) = O\left(\xi_N^\delta\right)$, it follows that (a) is equivalent to R3 in FS, for $\gamma = r$. Finally, (c) and (b) are equivalent to R2 in FS, for $m = 1$ and $r_0 = 0$.

As FS's proof is based on the rate at which $1\{\|X_t, X_{t-1}\| < N\} \sup_{\vartheta\in\Theta} \left|\ln \widehat{f}_N\left(X_t|X_{t-1},\vartheta\right) - \ln f\left(X_j|X_{j-1},\vartheta\right)\right|$ and $1\{\|X_t, X_{t-1}\| > N\} \sup_{\vartheta\in\Theta} \left|\ln \widehat{f}_N\left(X_t|X_{t-1},\vartheta\right)\right|$ 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 FS. However, given the geometric $\beta-$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. Moreover, Step 2 follows by the same argument. Hence, it remains to prove Step 4.

Proof of Step 4:

$$
\begin{aligned}
&\sup_{\vartheta\in\mathcal{N}_{\vartheta_k^\dagger}} \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left|\nabla_\vartheta L_{t,h}^N(\vartheta) - \nabla_\vartheta L_t^N(\vartheta)\right| \leq \sup_{\vartheta\in\mathcal{N}_{\vartheta_k^\dagger}} \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left(\left| \frac{1}{t}\sum_{j=1}^{t} \tau_N\left(\widehat{f}_N\left(X_j|X_{j-1},\vartheta\right)\right) \frac{1}{\widehat{f}_N\left(X_j|X_{j-1},\vartheta\right)} \right.\right. \\
&\times \left.\left(\frac{\partial \widehat{f}_{N,h}\left(X_j|X_{j-1},\vartheta\right)}{\partial\vartheta} - \frac{\partial \widehat{f}_N\left(X_j|X_{j-1},\vartheta\right)}{\partial\vartheta}\right)\right| \\
&+ \left| \frac{1}{t}\sum_{j=1}^{t} \left(\frac{\tau_N\left(\widehat{f}_{N,h}\left(X_j|X_{j-1},\vartheta\right)\right)}{\widehat{f}_{N,h}\left(X_j|X_{j-1},\vartheta\right)} - \frac{\tau_N\left(\widehat{f}_N\left(X_j|X_{j-1},\vartheta\right)\right)}{\widehat{f}_N\left(X_j|X_{j-1},\vartheta\right)}\right) \frac{\partial \widehat{f}_{N,h}\left(X_j|X_{j-1},\vartheta\right)}{\partial\vartheta}\right| \\
&+ \left| \frac{1}{t}\sum_{j=1}^{t} \tau_N'\left(\widehat{f}_{N,h}\left(X_j|X_{j-1},\vartheta\right)\right) \frac{\partial \widehat{f}_{N,h}\left(X_j|X_{j-1},\vartheta\right)}{\partial\vartheta} \ln \widehat{f}_{N,h}\left(X_j|X_{j-1},\vartheta\right)\right| \\
&+ \left.\left| \frac{1}{t}\sum_{j=1}^{t} \tau_N'\left(\widehat{f}_N\left(X_j|X_{j-1},\vartheta\right)\right) \frac{\partial \widehat{f}_N\left(X_j|X_{j-1},\vartheta\right)}{\partial\vartheta} \ln \widehat{f}_N\left(X_j|X_{j-1},\vartheta\right)\right|\right) \\
= \;& A_{1,T,N,h} + A_{2,T,N,h} + A_{3,T,N,h} + A_{4,T,N,h}.
\end{aligned}
$$

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

$$
\begin{aligned}
& A_{1,T,N,h} \\
\leq\; & \xi_N^{-\delta}\sqrt{P} \sup_{t\geq R} \sup_{\vartheta\in\mathcal{N}_{\vartheta_k^\dagger}} \left| \frac{1}{t}\sum_{j=1}^{t} \frac{1}{N\xi_N}\sum_{i=1}^{N} \left(\nabla_\vartheta K\left(\frac{X_{j,i,h}^{\vartheta}(X_{j-1}) - X_j}{\xi_N}\right) - \nabla_\vartheta K\left(\frac{X_{t,i}^{\vartheta}(X_{j-1}) - X_j}{\xi_N}\right)\right)\right| \\
\leq\; & \xi_N^{-(\delta+3)}\sqrt{P} \sup_{t\geq R} \sup_{\vartheta\in\mathcal{N}_{\vartheta_k^\dagger}} \left| \frac{1}{t}\sum_{j=1}^{t} \frac{1}{N}\sum_{i=1}^{N} \nabla_\vartheta^2 K\left(\frac{X_{j,i,h}^{\vartheta}(X_{j-1}) - X_j}{\xi_N}\right)\right|_{\overline{X}_{j,i,h}^{\vartheta}(X_{j-1})} \\
& \left(X_{j,i,h}^{\vartheta}(X_{j-1}) - X_{j,i}^{\vartheta}(X_{j-1})\right)\Big| = O_p\left(\sqrt{P}\xi_N^{-(\delta+3)}h\right), \qquad (16)
\end{aligned}
$$

and given A6,

$$
\begin{aligned}
& A_{2,T,N,h} \\
\leq \quad & \sqrt{P} \sup_{t \geq R} \sup_{\vartheta \in \mathcal{N}_{\vartheta_k^\dagger}} \left| \frac{1}{t} \sum_{j=1}^{t} \frac{\tau_N\left(\widehat{f}_N\left(X_j|X_{j-1},\vartheta\right)\right)\left(\widehat{f}_{N,h}\left(X_j|X_{j-1},\vartheta\right)-\widehat{f}_N\left(X_j|X_{j-1},\vartheta\right)\right)}{\widehat{f}_{N,h}\left(X_j|X_{j-1},\vartheta\right)} \right. \\
& \left. \frac{\partial \ln \widehat{f}_N\left(X_j|X_{j-1},\vartheta\right)}{\partial \vartheta} \right| \\
& + \sqrt{P} \sup_{t \geq R} \sup_{\vartheta \in \mathcal{N}_{\vartheta_k^\dagger}} \left| \frac{1}{t} \sum_{j=1}^{t} \left(\frac{\tau_N\left(\widehat{f}_N\left(X_j|X_{j-1},\vartheta\right)\right)-\tau_N\left(\widehat{f}_{N,h}\left(X_j|X_{j-1},\vartheta\right)\right)}{\widehat{f}_{N,h}\left(X_j|X_{j-1},\vartheta\right)} \right) \right. \\
& \left. \times \widehat{f}_N\left(X_j|X_{j-1},\vartheta\right) \frac{\partial \ln \widehat{f}_N\left(X_j|X_{j-1},\vartheta\right)}{\partial \vartheta} \right| = O_p\left(\sqrt{P}\xi_N^{-(\delta+3)} h \left|\ln \xi_N^\delta\right|\right). \qquad (17)
\end{aligned}
$$

Given the rate conditions in (a),(b), and (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 FS.

Proof of Theorem 6: Define,

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

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

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

Step 1:

$$
\sup_{\vartheta\in\Theta}\sup_{t\geq R}\left|L_{t,h}^{*N}(\vartheta) - L_t^*(\vartheta)\right| = o_{p^*}(1).
$$

Step 2:

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

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

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

$$
\begin{aligned}
& \arg\max_{\vartheta\in\Theta_k}\sup_{t\geq R}\left|\frac{1}{t}\sum_{l=2}^{t}\left(\ln \widehat{f}_{N,h}\left(X_l^*|X_{l-1}^*,\vartheta\right)\tau_N\left(\widehat{f}_{N,h}\left(X_l^*|X_{l-1}^*,\vartheta\right)\right) - \ln f\left(X_l^*|X_{l-1}^*,\vartheta\right)\right)\right| \\
= \quad & o_{p^*}(1) \qquad (18)
\end{aligned}
$$

and

$$
\sup_{t\geq R}\frac{1}{T}\sum_{i=1}^{T}\left(\nabla_\vartheta L_{i,h}^{*N}\left(\widehat{\vartheta}_{t,N,h}\right) - \nabla_\vartheta L_i^*\left(\widehat{\vartheta}_t\right)\right) = o_{p^*}(1). \qquad (19)
$$

Now, (18) 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 (19) is majorized by:

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

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

$$\sup_{t\geq R}\left|\frac{1}{T}\sum_{i=1}^{T}\left(\nabla_{\vartheta}^{2}L_{i}^{*}\left(\overline{\vartheta}_{t,N,h}\right)\left(\widehat{\vartheta}_{t,N,h}-\widehat{\vartheta}_{t}\right)\right)\right|\leq\sup_{t\geq R}\frac{1}{T}\sum_{i=1}^{T}\left|\nabla_{\vartheta}^{2}L_{i}^{*}\left(\overline{\vartheta}_{t,N,h}\right)\right|\sup_{t\geq R}\left(\widehat{\vartheta}_{t,N,h}-\widehat{\vartheta}_{t}\right)=O_{p^*}(1)o_{p}(1).$$

Proof of Step 2: Follows directly from (19) and from Steps 2 and 4 in the proof Theorem 5

Proof of Theorem 7: Let:

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

We show that:

$$\sup_{\theta\in\Theta}\sup_{t\geq R}\left|L_{t,h}^{N,S}(\theta)-L_{t,h}^{N}(\theta)\right|=o_{p}(1)\tag{20}$$

and

$$\sup_{\theta\in\mathcal{N}_{\theta_k^{\dagger}}}\frac{1}{\sqrt{P}}\sum_{t=R}^{T}\left|\nabla_{\theta}L_{t,h}^{N,S}(\theta)-\nabla_{\theta}L_{t,h}^{N}(\theta)\right|=o_{p}(1).\tag{21}$$

The desired outcome then follows from Theorem 5. Note first that (20) can be written as:

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

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

$$\begin{aligned}
K\left(\frac{X_{l,i,h}^{\theta}(X_{l-1})-X_{l}}{\xi_{N}}\right)&=\int_{V}K\left(\frac{X_{l,i,h}^{\theta}(X_{l-1},v^{\theta})-X_{l}}{\xi_{N}}\right)f_{\vartheta}(v)dv\\
&=E_{S}\left(K\left(\frac{X_{l,i,h}^{\theta}(X_{l-1},v^{\theta})-X_{l}}{\xi_{N}}\right)\right),
\end{aligned}$$

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

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

Also,

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

Given the rate condition in (e), this proves (20). Turning now to (21), note that after few simple manipulations:

$$
\sup_{\theta \in \mathcal{N}_{\theta^\dagger}} I_{T,N,S,h} \leq
$$

$$
\begin{aligned}
& \sup_{\theta \in \mathcal{N}_{\theta^\dagger}} \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left(\left| \frac{1}{t} \sum_{l=2}^{t} \frac{\tau_{N,S}\left(\widehat{f}_{N,S,h}\left(X_l|X_{l-1},\theta\right) \right)}{\widehat{f}_{N,S,h}\left(X_l|X_{l-1},\theta\right)} \left(\frac{\partial \widehat{f}_{N,S,h}\left(X_l|X_{l-1},\theta\right)}{\partial \theta} - \frac{\partial \widehat{f}_{N,h}\left(X_l|X_{l-1},\theta\right)}{\partial \theta} \right) \right| \right. \\
& + \left| \frac{1}{t} \sum_{l=2}^{t} \tau_{N,S}\left(\widehat{f}_{N,S,h}\left(X_l|X_{l-1},\theta\right) \right) \frac{\partial \widehat{f}_{N,h}\left(X_l|X_{l-1},\theta\right)}{\partial \theta} \frac{\left(\widehat{f}_{N,h}\left(X_l|X_{l-1},\theta\right) - \widehat{f}_{N,S,h}\left(X_l|X_{l-1},\theta\right) \right)}{\widehat{f}_{N,S,h}\left(X_l|X_{l-1},\theta\right) \widehat{f}_{N,h}\left(X_l|X_{l-1},\theta\right)} \right| \\
& + \left| \frac{1}{t} \sum_{l=2}^{t} \left(\tau_{N,S}\left(\widehat{f}_{N,S,h}\left(X_l|X_{l-1},\theta\right) \right) - \tau_{N}\left(\widehat{f}_{N,h}\left(X_l|X_{l-1},\theta\right) \right) \right) \frac{\partial \ln \widehat{f}_{N,h}\left(X_l|X_{l-1},\theta\right)}{\partial \theta} \right| \\
& + \left| \frac{1}{t} \sum_{l=2}^{t} \tau_{N,S}\left(\widehat{f}_{N,S,h}\left(X_l|X_{l-1},\theta\right) \right) \left(\ln \widehat{f}_{N,S,h}\left(X_l|X_{l-1},\theta\right) - \ln \widehat{f}_{N,h}\left(X_l|X_{l-1},\theta\right) \right) \right| \\
& + \left| \frac{1}{t} \sum_{l=2}^{t} \tau_{N,S}'\left(\widehat{f}_{N,S,h}\left(X_l|X_{l-1},\theta\right) \right) \left(\frac{\partial \widehat{f}_{N,S,h}\left(X_l|X_{l-1},\theta\right)}{\partial \theta} - \frac{\widehat{f}_{N,h}\left(X_l|X_{l-1},\theta\right)}{\partial \theta} \right) \ln \widehat{f}_{N,h}\left(X_l|X_{l-1},\theta\right) \right| \\
& + \left| \frac{1}{t} \sum_{l=2}^{t} \left(\tau_{N,S}'\left(\widehat{f}_{N,S,h}\left(X_l|X_{l-1},\theta\right) \right) - \tau_{N}'\left(\widehat{f}_{N,h}\left(X_l|X_{l-1},\theta\right) \right) \right) \right. \\
& \left. \left. \times \frac{\widehat{f}_{N,h}\left(X_l|X_{l-1},\theta\right)}{\partial \theta} \ln \widehat{f}_{N,h}\left(X_l|X_{l-1},\theta\right) \right| \right) \\
= \quad & \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 A9,

$$\begin{aligned}
& \sup_{\theta\in\mathcal{N}_{\theta^\dagger}} V_{1,T,N,S,h}(\theta) \\
\leq\ & \sup_{\theta\in\mathcal{N}_{\theta^\dagger}} \xi_N^{-(\delta+2)} \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left| \frac{1}{t} \sum_{l=2}^{t} \frac{1}{N} \sum_{i=1}^{N} \left(\frac{1}{S} \sum_{s=1}^{S} \left(\frac{\partial X_{l,i,h}^{\theta}(X_{j-1}, V_s^{\theta})}{\partial\theta} K' \left(\frac{X_{l,i,h}^{\theta}(X_{j-1}, V_s^{\theta}) - X_j}{\xi_N} \right) \right.\right.\right. \\
& \left.\left.\left. - \frac{\partial X_{l,i,h}^{\theta}(X_{j-1})}{\partial\theta} K' \left(\frac{X_{l,i,h}^{\theta}(X_{j-1}) - X_j}{\xi_N} \right) \right) \right) \right| \\
=\ & O_p\left(P^{1/2} \xi_N^{-(\delta+2)} S^{-1/2} \right)
\end{aligned}$$

$$\begin{aligned}
& \sup_{\theta\in\mathcal{N}_{\theta^\dagger}} V_{2,T,N,S,h}(\theta) \\
\leq\ & \sup_{\theta\in\mathcal{N}_{\theta^\dagger}} \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left| \frac{1}{t} \sum_{l=2}^{t} \frac{\tau_{N,S}\left(\widehat{f}_{N,S,h}(X_l|X_{l-1},\theta) \right)}{\widehat{f}_{N,S,h}(X_l|X_{l-1},\theta)\, \widehat{f}_{N,h}(X_l|X_{l-1},\theta)} \frac{\partial \overline{f}_{N,S,h}(X_l|X_{l-1},\theta)}{\partial\theta} \right. \\
& \left. \left(\left(\widehat{f}_{N,h}(X_l|X_{l-1},\theta) - \widehat{f}_{N,S,h}(X_l|X_{l-1},\theta) \right) \right) \right| \\
& + \sup_{\theta\in\mathcal{N}_{\theta^\dagger}} \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left| \frac{1}{t} \sum_{l=2}^{t} \frac{\tau_{N,S}\left(\widehat{f}_{N,S,h}(X_l|X_{l-1},\theta) \right)}{\widehat{f}_{N,S,h}(X_l|X_{l-1},\theta)\, \widehat{f}_{N,h}(X_l|X_{l-1},\theta)} \left(\frac{\partial \widehat{f}_{N,h}(X_l|X_{l-1},\theta)}{\partial\theta} \right.\right. \\
& \left.\left. - \frac{\partial f(X_l|X_{l-1},\theta)}{\partial\theta} \right) \left(\widehat{f}_{N,h}(X_l|X_{l-1},\theta) - \widehat{f}_{N,S,h}(X_l|X_{l-1},\theta) \right) \right) \right| \qquad (22)
\end{aligned}$$

Given Steps 2 and 4 in the proof of Proposition 2, it is immediate to see that that the second term on the RHS of (22) is of smaller order than the first. Now, the first term on the RHS of (22) is majorized by:

$$\begin{aligned}
& \xi_N^{-2\delta} \sup_{\theta\in\mathcal{N}_{\theta^\dagger}} \sup_{t\geq R} \left| \frac{1}{t} \sum_{l=2}^{t} \left(\frac{\partial \widehat{f}_{N,h}(X_l|X_{l-1},\theta)}{\partial\theta} \right)^r \right|^{1/r} \times \\
& \sup_{\theta\in\mathcal{N}_{\theta^\dagger}} \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left| \left(\frac{1}{t} \sum_{l=2}^{t} \left(\widehat{f}_{N,h}(X_l|X_{l-1},\theta) - \widehat{f}_{N,S,h}(X_l|X_{l-1},\theta) \right)^{(r-1)/r} \right)^{1-1/r} \right| \\
=\ & O_P\left(\sqrt{P} S^{-1/2} \xi_N^{-(1+2\delta)} \right),
\end{aligned}$$

and so

$$\begin{aligned}
& \sup_{\theta\in\mathcal{N}_{\theta^\dagger}} V_{3,T,N,S,h}(\theta) \\
\leq\ & \sup_{\theta\in\mathcal{N}_{\theta^\dagger}} \frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left| \frac{1}{t} \sum_{l=2}^{t} \left(\tau_{N,S}\left(\widehat{f}_{N,S,h}(X_l|X_{l-1},\theta) \right) - \tau_N\left(\widehat{f}_{N,h}(X_l|X_{l-1},\theta) \right) \right) \right. \\
& \left. \frac{\partial \ln f(X_l|X_{l-1},\theta)}{\partial\theta} \right| + \text{term of smaller order} \\
=\ & O_p\left(\sqrt{P} S^{-1/2} \xi_N^{-(1+\delta)} \right).
\end{aligned}$$

By a similar argument as that used in the proof of (20), $\sup_{\theta\in\mathcal{N}_{\theta^\dagger}} V_{4,T,N,S,h}(\theta) = O_p\left(\xi_N^{-(\delta+1)} \sqrt{P} S^{-1/2} \right)$; $V_{5,T,N,S,h}(\theta)$, (other than a log term), can be treated as $V_{1,T,N,S,h}(\theta)$, and so $\sup_{\theta\in\mathcal{N}_{\theta^\dagger}} V_{5,T,N,S,h}(\theta) =$

$O_p\left(P^{1/2}\xi_N^{-(\delta+2)}S^{-1/2}\left|\ln\xi_N^{-\delta}\right|\right)$. Finally, by a similar argument as that used to examine $V_{3,T,N,S,h}(\theta)$:

$$
\begin{aligned}
& \sup_{\theta\in\mathcal{N}_{\theta^\dagger}} V_{5,T,N,S,h}(\theta) \\
\leq & \sup_{\theta\in\mathcal{N}_{\theta^\dagger}} \frac{1}{\sqrt{P}}\sum_{t=R}^{T}\left|\frac{1}{t}\sum_{l=2}^{t}\left(\tau'_{N,S}\left(\widehat{f}_{N,S,h}(X_l|X_{l-1},\theta)\right)-\tau'_N\left(\widehat{f}_{N,h}(X_l|X_{l-1},\theta)\right)\right)\right. \\
& \left.\times\frac{f(X_l|X_{l-1},\theta)}{\partial\theta}\ln\widehat{f}_{N,h}(X_l|X_{l-1},\theta)\right|\Bigg) + \text{terms of smaller order} \\
= & O_p\left(\sqrt{P}S^{-1/2}\xi_N^{-(1+2\delta)}\right).
\end{aligned}
$$

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

7 References

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., J. Fan and H. Peng, (2009), Nonparametric Transition-Based Tests for Diffusions, forthcoming, *Journal of the American Statistical Association*.

Altissimo, F., and A. Mele, (2009), Simulated Nonparametric Estimation of Dynamic Models with Application in Finance, *Review of Economic Studies*, 76, 413-450.

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

Bandi, F. M. and R. Reno, (2008), Nonparametric Stochastic Volatility, Working Paper, University of Chicago.

Bhardwaj, G., V. Corradi and N.R. Swanson, (2008), A Simulation Based Specification Test for Diffusion Processes, *Journal of Business and Economic Statistics,* 26, 176-193.

Bontemps, C., and N. Meddahi, (2005), Testing Normality: a GMM approach, *Journal of Econometrics,* 124, 149-186.

Carrasco, M., M. Chernov, J.P. Florens and E. Ghysels ,(2007), Efficient Estimation of General Dynamic Models with a Continuum of Moment Conditions, *Journal of Econometrics,* 140, 529-543.

Chacko, G., and L.M. Viceira, (2003), Spectral GMM estimation of Continuous-Time Processes, *Journal of Econometrics,* 116, 259-292.

Cont, R., and P. Tankov, (2004), *Financial Modelling with Jump Processes,* Chapman and Hall, London

Corradi, V., and N.R. Swanson (2005a), Bootstrap Tests for Diffusion Processes, *Journal of Econometrics,* 124, 117-148, 2005.

Corradi, V., and N.R. Swanson (2005b), A Test for Comparing Multiple Misspecified Conditional Intervals, *Econometric Theory,* 21, 991-1016.

Corradi, V., and N.R. Swanson (2006a), Predictive Density and Conditional Confidence Interval Accuracy Tests", *Journal of Econometrics*, 135, 187-228.

Corradi, V., and N.R. Swanson (2006b), Predictive Density Evaluation, in: *Handbook of Economic Forecasting*, eds. Clive W.J. Granger, Graham Elliot and Allan Timmermann, Elsevier, Amsterdam, pp. 197-284.

Corradi, V. and N.R. Swanson, (2007a), Nonparametric Bootstrap Procedures for Predictive Inference Based on Recursive Estimation Schemes", *International Economic Review,* 48, 67-109.

Corradi, V., and N.R. Swanson (2007b), Evaluation of Dynamic Stochastic General Equilibrium Models Based on Distributional Comparison of Simulated and Historical Data, *Journal of Econometrics,* 136, 699-723.

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

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

Doukhan, P. (1995), *Mixing Properties and Examples,* Springer and Verlag, New York.

Duan, J.C., (2003), A Specification Test for Time Series Models by a Normality Transformation, Working Paper, University of Toronto.

Dridi, R., A. Guay and E. Renault, (2007), Indirect Inference and Calibration of Dynamic Stochastic General Equilibrium Models, *Journal of Econometrics,* 136, 397-439.

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

Eraker, B., M. Johannes, and N. Polson, (2003), The Impact of Jumps in Volatility and Returns, *Journal of Finance, 58,* 1269-1300.

Durham, G.B., (2003), Likelihood-Based Specification Analysis of Continuous-Time Models of the Short-Term Interest Rate, *Journal of Financial Economics*, 70, 463-487.

Fermanian, J.-D. and B. Salanié, (2004), A Nonparametric Simulated Maximum Likelihood Estimation Method, *Econometric Theory, 20*, 701-734.

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

Gourieroux, C., A. Monfort and E. Renault, (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., and H. Li, (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., H. Li, and F. Zhao, (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., and J.L. Knight, (2002), Estimation of Continuous-Time Processes via the Empirical Characteristic Function, *Journal of Business and Economic Statistics,* 20, 198-212.

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

Kristensen, D. and Y. Shin, (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., (2004), Ergodicity and Exponential $\beta-$Mixing for a Strong Solution of Levy-Driven Stochastic Differential Equations, Working Paper, Kyushu University, MHF-2004-19.

Pardoux, E. and D. Talay, (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.

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	$D^{Max}_{k,P,S,N}(u_1, u_2)$	$PDMSFE_{CIR}$	$PDMSFE_{SV}$	$PDMSFE_{SVJ}$	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, predicitve 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 95^{th}, 90^{th}, 85^{th}, and 80^{th} 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	$D^{Max}_{k,P,S,N}(u_1, u_2)$	$PDMSFE_{CIR}$	$PDMSFE_{SV}$	$PDMSFE_{SVJ}$	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

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	$D^{Max}_{k,P,S,N}(u_1, u_2)$	$PDMSFE_{CIR}$	$PDMSFE_{SV}$	$PDMSFE_{SVJ}$	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	$D^{Max}_{k,P,S,N}(u_1, u_2)$	$PDMSFE_{CIR}$	$PDMSFE_{SV}$	$PDMSFE_{SVJ}$	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

Figure 1: Predictive Densities for CIR, SV and SVJ Models - 01:1989-12:1998

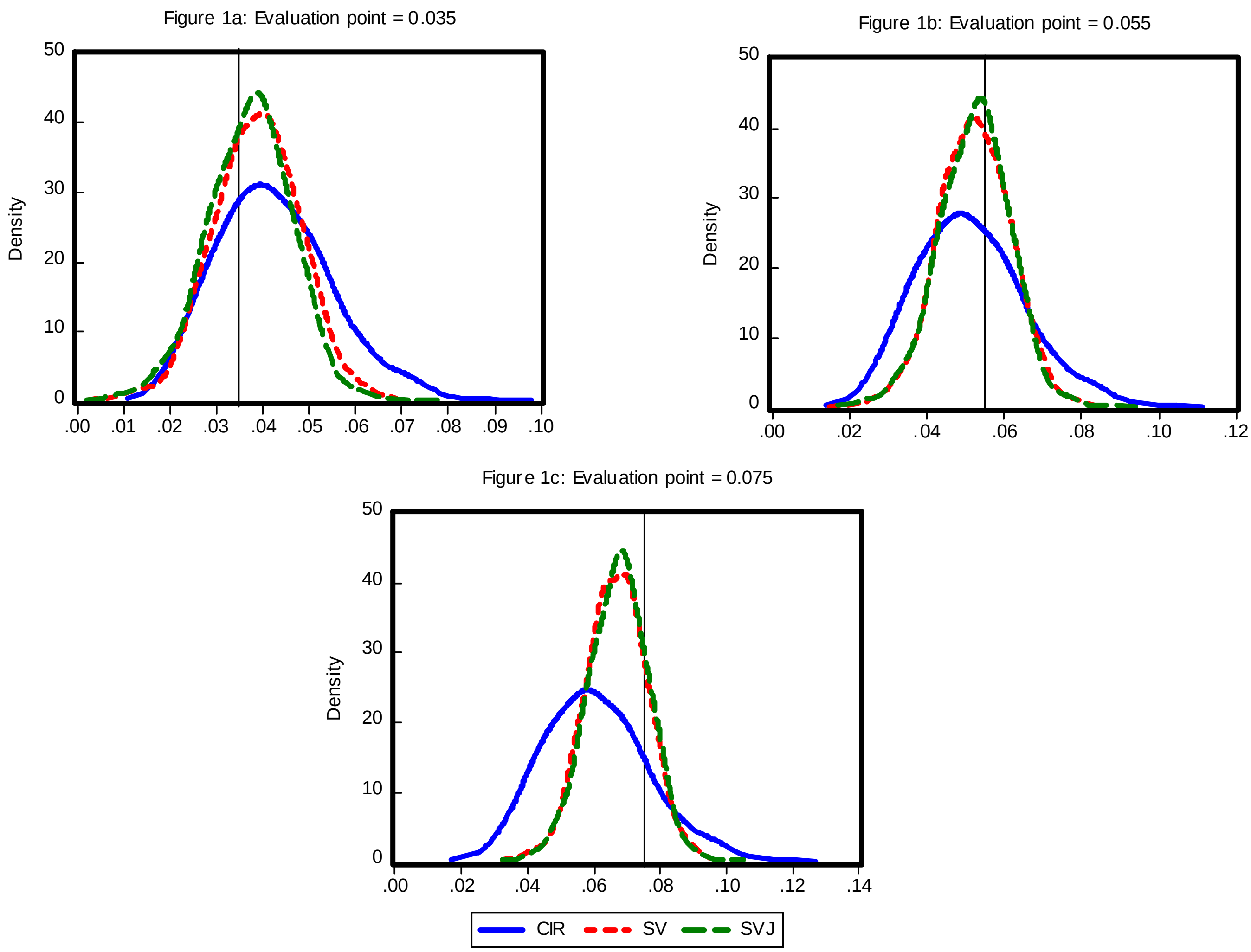

Figure 2: Predictive Densities for CIR, SV and SVJ Models - 01:1999-04:2008