Density and Conditional Distribution Based Specification Analysis*

Diep Duong and Norman R. Swanson*
Rutgers University

February 2012

The technique of using densities and conditional distributions to carry out consistent specification testing and model selection amongst multiple diffusion processes have received considerable attention from both financial theoreticians and empirical econometricians over the last two decades. One reason for this interest is that correct specification of diffusion models describing dynamics of financial assets is crucial for many areas in finance including equity and option pricing, term structure modeling, and risk management, for example. In this paper, we discuss advances to this literature introduced by Corradi and Swanson (2005), who compare the cumulative distribution (marginal or joint) implied by a hypothesized null model with corresponding empirical distributions of observed data. We also outline and expand upon further testing results from Bhardwaj, Corradi and Swanson (BCS: 2008) and Corradi and Swanson (2011). In particular, parametric specification tests in the spirit of the conditional Kolmogorov test of Andrews (1997) that rely on block bootstrap resampling methods in order to construct test critical values are first discussed. Thereafter, extensions due to BCS (2008) for cases where the functional form of the conditional density is unknown are introduced, and related continuous time simulation methods are introduced. Finally, we broaden our discussion from single process specification testing to multiple process model selection by discussing how to construct predictive densities and how to compare the accuracy of predictive densities derived from alternative (possibly misspecified) diffusion models. In particular, we generalize simulation Steps outlined in Cai and Swanson (2011) to multifactor models where the number of latent variables is larger than three. These final tests can be thought of as continuous time generalizations of the discrete time "reality check" test statistics of White (2000), which are widely used in empirical finance (see e.g. Sullivan, Timmermann and White (1999, 2001)). We finish the chapter with an empirical illustration of model selection amongst alternative short term interest rate models.

Keywords: multi-factor diffusion process, specification test, out-of-sample forecasts, conditional distribution, model selection,block bootstrap, jump process.

---

* Diep Duong, Department of Economics, Rutgers University, 75 Hamilton Street, New Brunswick, NJ 08901, USA, dduong@econ.rutgers.edu. Norman R. Swanson, Department of Economics, Rutgers University, 75 Hamilton Street, New Brunswick, NJ 08901, USA, nswanson@econ.rutgers.edu. The authors thanks the editor, Cheng-Few Lee for many useful suggestions given during the writing of this paper. Duong and Swanson would like to thank the Research Council at Rutgers University for research support.

1 Introduction

The last three decades have provided a unique opportunity to observe numerous interesting developments in finance, financial econometrics and statistics. For example, although starting as a narrow sub-field, financial econometrics has recently transformed itself into an important discipline, equipping financial economic researchers and industry practitioners with immensely helpful tools for estimation, testing and forecasting. One of these developments has involved the development of "state of the art" consistent specification tests for continuous time models, including not only the geometric Brownian motion process used to describe the dynamics of asset returns (Merton (1973)), but also a myriad of other diffusion models used in finance, such as the Ornstein-Uhlenbeck process introduced by Vacisek (1977), the constant elastic volatility process applied by Beckers (1980), the square root process due to Cox, Ingersoll and Ross (CIR: 1985), the so called CKLS model by Chan, Karolyi, Longstaff and Sanders (CKLS: 1992), various three factor models proposed Chen (1996), stochastic volatility processes such as generalized CIR of Andersen and Lund (1997), and the generic class of affine jump diffusion processes discussed in Duffie, Pan and Singleton (2000).[1]

The plethora of available diffusion models allow decision makers to be flexible when choosing a specification to be subsequently used in contexts ranging from equity and option pricing, to term structure modeling and risk management. Moreover, the use of high frequency data when estimating such models, in continuous time contexts, allows investors to continuously update their dynamic trading strategies in real-time.[2] However, for statisticians and econometricians, the vast number of available models has important implications for formalizing model selection and specification testing methods. This has led to several key papers that have recently been published in the area of parametric and non-parametric specification testing. Most of the papers focus on the ongoing "search" for correct Markov and stationary models that "fit" historical data and associated dynamics. In this literature, it is important to note that correct specification of a joint distribution is not the same as that of a conditional distribution, and hence the recent focus on conditional distributions, given that most models have an interpretation as conditional models. In summary, the key issue in the construction of model selection and specification tests of conditional distributions is the fact that knowledge of the transition density (or conditional distribution) in general cannot be inferred from knowledge of the drift and variance terms of a diffusion model. If the functional form of the density is available parametrically, though, one can 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 and Li (2005) and Hong, Li, and Zhao (2007); the martingalization-type Kolmogorov test of Bai (2003); or via the normality

[1] For complete details, see Section 2.2.

[2] For further discussion, see Duong and Swanson (2010, 2011).

transformation approaches of Bontemps and Meddahi (2005) and Duan (2003). Furthermore, if the transition density is unknown, one can construct a non-parametric test by comparing a kernel density estimator of the actual and simulated data, for example, as in Altissimo and Mele (2009) and Thompson (2008); or by comparing the conditional distribution of the simulated and the historical data, as in Bhardwaj, Corradi, and Swanson (BCS: 2008). One can also use the methods of Aït-Sahalia (2002) and Aït-Sahalia, Fan, and Peng (2009), in which they compare closed form approximations of conditional densities under the null, using data-driven kernel density estimates.

For clarity and ease of presentation, we categorize the above literature into two areas. The first area, initiated by the seminal work of Aït-Sahalia (1996) and later followed by Pritsker (1998) and Jiang (1998), breaks new ground in the continuous time specification testing literature by comparing marginal densities implied by hypothesized null models with nonparametric estimates thereof. These sorts of tests examine one factor specifications. The second area of testing, as initiated in Corradi and Swanson (CS: 2005) does not look at densities. Instead, they compare cumulative distributions (marginal, joint, or conditional) implied by a hypothesized null model with corresponding empirical distributions. A natural extension of these sorts of tests involves model selection amongst alternative predictive densities associated with competing models. While CS (2005) focus on cases where the functional form of the conditional density is known, BCS (2008) use simulation methods to examine testing in cases where the functional form of the conditional density is unknown. Corradi and Swanson (CS: 2011) and Cai and Swanson (2011) take the analysis of BCS (2008) on Step further, and focus on the comparison of out of sample predictive accuracy of possibly misspecified diffusion models, when the conditional distribution is not known in closed form (i.e., they "choose" amongst competing models based on predictive density model performance). The "best" model is selected by constructing tests that compare both predictive densities and/or predictive conditional confidence intervals associated with alternative models

In this paper, we primarily focus our attention on the second area of the model selection and testing literature.[3] One feature of all of the tests that we shall discuss is that, given that they are based on the comparison of CDFs, they obtain parametric rates. Moreover, the tests can be used to evaluate single and multiple factor and dimensional models, regardless of whether or not the functional form of the conditional distribution is known.

In addition to discussing simple diffusion process specification tests of CS (2005), we discuss tests discussed in BCS (2008) and CS (2011), and provide some generalizations and additional results. In particular, parametric specification tests in the spirit of the conditional Kolmogorov test of Andrews (1997) that rely on block bootstrap resampling methods in order to construct test critical values are first discussed. Thereafter, extensions due to BCS (2008) for cases where the

[3] For a recent survey on results in the first area of this literature, see Aït-Sahalia (2007).

functional form of the conditional density is unknown are introduced, and related continuous time simulation methods are introduced. Finally, we broaden our discussion from single dimensional specification testing to multiple dimensional selection by discussing how to construct predictive densities and how to compare the accuracy of predictive densities derived from alternative (possibly misspecified) diffusion models as in CS (2011). In addition, we generalize simulation and testing procedures introduced in Cai and Swanson (2011) to more complicated multi-factor and multi-dimensional models where the number of latent variables larger than three. These final tests can be thought of as continuous time generalizations of the discrete time "reality check" test statistics of White (2000), which are widely used in empirical finance (see e.g. Sullivan, Timmermann and White (1999, 2001)). We finish the chapter with an empirical illustration of model selection amongst alternative short term interest rate models, drawing on BCS (2008), CS (2011) and Cai and Swanson (2011).

Of the final note is that the test statistics discussed here are implemented via use of simple bootstrap methods for critical value simulation. We use the bootstrap because the covariance kernels of the (Gaussian) asymptotic limiting distributions of the test statistics are shown to contain terms deriving from both the contribution of recursive parameter estimation error (PEE) and the time dependence of data. Asymptotic critical value thus cannot be tabulated in a usual way. Several methods can easily be implemented in this context. First one can use block bootstrapping procedures, as discussed below. Second one can use the conditional p-value approach of Corradi and Swanson (2002) which extends the work of Hansen (1996) and Inoue (2001) to the case of non vanishing parameter estimation error. Third is the subsampling method of Politis, Romano and Wolf (1999), which has clear efficiency "costs", but is easy implement. Use of the latter two methods yields simulated (or subsample based) critical values that diverge at rate equivalent to the blocksize length under the alternative. This is the main drawback to their use in our context. We therefore focus on use of a block bootstrap that mimics the contribution of parameter estimation error in a recursive setting and in the context of time series data. In general, use of the block bootstrap approach is made feasible 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 present our setup, and discuss various diffusion models used in finance and financial econometrics. Section 3 outlines the specification testing hypotheses, presents the cumulative distribution based test statistics for one factor and multiple factor models, discusses relevant procedures for simulation and estimation, and outlines bootstrap techniques that can be used for critical value tabulation. In Section 4, we present a small

empirical illustration. Section 5 summarizes and concludes.

2 Setup

2.1 Diffusion Models in Finance and Financial Econometrics

For the past two decades, continuous time models have taken center stage in the field of financial econometrics, particularly in the context of structural modeling, option pricing, risk management, and volatility forecasting. One key advantage of continuous time models is that they allow financial econometricians to use the full information set that is available. With the availability of high frequency data and current computation capability, one can update information, model estimates, and predictions in milliseconds. In this Section we will summarize some of the standard models that have been used in asset pricing as well as term structure modelling. Generally, assume that financial asset returns follow Ito-semimartingale processes with jumps, which are the solution to the following stochastic differential equation system.

$$X(t_-) = \int_0^t b(X(s_-), \theta_0)ds - \lambda_0 t \int_Y y\phi(y)dy + \int_0^t \sigma(X(s_-), \theta_0)dW(s) + \sum_{j=1}^{J_t} y_j, \tag{1}$$

where $X(t_-)$ is a cadlag process (right continuous with left limit) for $t \in \Re^+$, and is an N dimensional vector of variables, $W(t)$ is an $N-$dimensional Brownian motion, $b(\cdot)$ is $N-$dimensional function of $X(t_-)$, and $\sigma(\cdot)$ is an NxN matrix-valued function of $X(t_-)$, where θ_0 is an unknown true parameter. J_t is a Poisson process with intensity parameter λ_0, λ_0 finite, and the $N-$dimensional jump size, y_j, is iid with marginal distribution given by ϕ. Both J_t and y_j are assumed to be independent of the driving Brownian motion, $W(t)$.[4]Also, note that $\int_Y y\phi(y)dy$ denotes the mean jump size, hereafter denoted by μ_0. Over a unit time interval, there are on average λ_0 jumps; so that over the time span $[0, t]$, there are on average $\lambda_0 t$ jumps. The dynamics of $X(t_-)$ is then given by:

$$dX(t) = \left(b(X(t_-), \theta_0) - \lambda_0\mu_{y,0}\right)dt + \sigma(X(t_-), \theta_0)dW(t) + \int_Y yp(dy, dt), \tag{2}$$

where $p(dy, dt)$ is a random Poisson measure giving point mass at y if a jump occurs in the interval dt, and $b(\cdot), \sigma(\cdot)$ are the "drift" and "volatility" functions defining the parametric specification of the model. Hereafter, the same (or similar) notation is used throughout when models are specified.

Though not an exhaustive list, we review some popular models. Models are presented with the "true" parameters.

Diffusion Models Without Jumps:

[4]Hereafter, $X(t_-)$ denotes the cadlag, while X_t denotes discrete skeleton for $t = 1, 2, \ldots$.

Geometric Brownian Motion (log normal model). In this set-up, $b(X(t_-), \theta_0) = b_0 X(t)$ and $\sigma(X(t_-), \theta_0) = \sigma_0 X(t)$

$$dX(t) = b_0 X(t)dt + \sigma_0 X(t)dW(t),$$

where b_0 and σ_0 are constants and and $W(t)$ is a one dimensional standard Brownian motion. (Below, other constants such as α_0 , β_0, λ_0, γ_0, δ_0, η_0, κ_0, and Ω_0 are also used in model specifications.)

This model is popular in the asset pricing literature. For example, one can model equity prices according to this process, especially in the Black-Scholes option set-up or in structured corporate finance.[5] The main drawback of this model is that the return process (log(price)) has constant volatility, and is not time varying. However, it is widely used as a convenient "first" econometric model.

Vasicek (1977) and Ornstein-Uhlenbeck process. The process is used to model asset prices, specifically in term structure modelling, and the specification is:

$$dX(t) = (\alpha_0 + \beta_0 X(t))dt + \sigma_0 dW(t)$$

where $W(t)$ is a standard Brownian motion, and α_0, β_0 and σ_0 are constants. β_0 is negative to ensure the mean reversion of $X(t)$.

Cox, Ingersoll and Ross (1995) use the following square root process to model the term structure of interest rates:

$$dX(t) = \kappa(\alpha_0 - X(t))dt + \sigma_0 \sqrt{X(t)} dW(t)$$

where $W(t)$ is a standard Brownian motion, α_0 is the long-run mean of $X(t)$, κ measures the speed of mean-reversion, and σ_0 is a standard deviation parameter and is assumed to be fixed. Also, non-negativity of the process is imposed, as $2\kappa\beta_0 > \sigma_0^2$.

Wong (1964) points out that in the CIR model, $X(t)$ with the dynamics evolving according to:

$$dX(t) = ((\alpha_0 - \lambda_0) - X(t))dt + \sqrt{\alpha_0 X(t)} dW(t), \ \alpha_0 > 0 \text{ and } \alpha_0 - \lambda_0 > 0 \tag{3}$$

belongs to the linear exponential (or Pearson) family with a closed form cumulative distribution. α_0 and λ_0 are fixed parameters of the model.

The Constant Elasticity of Variance, or CEV model is specified as follows:

$$dX(t) = \alpha_0 X(t)dt + \sigma_0 X(t)^{\beta_0/2} dW(t)$$

where $W(t)$ is a standard Brownian motion and α_0, σ_0 and β_0 are fixed constants.

Of note is that the interpretation of this model depends on β_0, i.e. in the case of stock prices, if $\beta_0 = 2$, then the price process $X(t)$ follows a lognormal diffusion; if $\beta_0 < 2$, then the model captures exactly the leverage effect as price and volatility are inversely correlated.

[5] See Black and Scholes (1973) for details.

Among other authors, *Beckers (1980)* uses this CEV model for stocks, *Marsha and Rosenfeld (1983)* apply a CEV parametrization to interest rates and *Emanuel and Macbeth (1982)* utilize this set-up for option pricing.

The *Generalized constant elasticity of variance model is defined as follows:*

$$dX(t) = (\alpha_0 X(t)^{-(1-\beta_0)} + \lambda_0 X(t))dt + \sigma_0 X(t)^{\beta_0/2} dW(t)$$

where the notation follows the CEV case. λ_0 is another parameter of the model. This process nests log diffusion when $\beta_0 = 2$, and nests square root diffusion when $\beta_0 = 1$.

Brennan and Schwartz (1979) and Courtadon (1982) analyze the model:

$$dX(t) = (\alpha_0 + \beta_0 X(t))dt + \sigma_0 X(t)^2 dW(t)$$

where $\alpha_0, \beta_0, \sigma_0$ are fixed constants and $W(t)$ is a standard Brownian motion.

Duffie and Kan (1993) study the specification:

$$dX(t) = (\alpha_0 - X(t))dt + \sqrt{\beta_0 + \gamma_0 X(t)} dW(t)$$

where $W(t)$ is a standard Brownian motion and α_0, β_0 and γ_0 are fixed parameters.

Aït-Sahalia (1996) looks at a general case with general drift and CEV diffusion:

$$dX(t) = (\alpha_0 + \beta_0 X(t) + \gamma_0 X(t)^2 + \eta_0 / X(t))dt + \sigma_0 X(t)^{\beta_0/2} dW(t)$$

In the above expression, $\alpha_0, \beta_0, \gamma_0, \eta_0, \sigma_0$ and β_0 are fixed constants and $W(t)$ is again a standard Brownian motion.

Diffusion Models with Jumps*:*

For term structure modeling in empirical finance, the most widely studied class of models is the family of affine processes, including diffusion processes that incorporate jumps.

Affine Jump Diffusion Model: $X(t_-)$ is defined to follow an affine jump diffusion if

$$dX(t) = \kappa_0(\alpha_0 - X(t))dt + \Omega_0 \sqrt{D(t)} dW(t) + dJ(t)$$

where $X(t_-)$ is an $N-$dimensional vector of variables of interest and is a cadlag process, $W(t)$ is an $N-$dimensional independent standard Brownian motion, κ_0 and Ω_0 are square $N \times N$ matrices, α_0 is a fixed long-run mean, $D(t)$ is a diagonal matrix with ith diagonal element given by

$$d_{ii}(t) = \theta_{0i} + \delta'_{0i} X(t)$$

In the above expressions, θ_{0i} and δ'_{0i} are constants. The jump intensity is assumed to be a positive, affine function of $X(t)$ and the jump size distribution is assumed to be determined by it's conditional characteristic function. The attractive feature of this class of affine jump diffusions is

that, as shown in Duffie, Pan and Singleton (2000), it has an exponential affine structure that can be derived in closed form, i.e.

$$\Phi(X(t)) = \exp(a(t) + b(t)'X(t))$$

where the functions $a(t)$ and $b(t)$ can be derived from Riccati equations.[6] Given a known characteristic function, one can use either GMM to estimate the parameters of this jump diffusion, or one can use quasi-maximum likelihood (QML), once the first two moments are obtained. In the univariate case without jumps, as a special case, this corresponds to the above general CIR model with jumps.

Multifactor and Stochastic Volatility Model: Multifactor models have been widely used in the literature; particularly in option pricing, term structure, and asset pricing. One general set-up has $(X(t), V(t))' = \left(X(t), V^1(t), ..., V^d(t)\right)'$ where only the first element, the diffusion process X_t, is observed while $V(t) = (V^1(t), ..., V^d(t))'_{dx1}$ is latent. In addition, $X(t)$ can be dependent on $V(t)$. For instance, in empirical finance, the most well-known class of the multifactor models is the stochastic volatility model expressed as:

$$\begin{pmatrix} dX(t) \\ dV(t) \end{pmatrix} = \begin{pmatrix} b_1(X(t), \theta_0) \\ b_2(V(t), \theta_0) \end{pmatrix} dt + \begin{pmatrix} \sigma_{11}(V(t), \theta_0) \\ 0 \end{pmatrix} dW_1(t) + \begin{pmatrix} \sigma_{12}(V(t), \theta_0) \\ \sigma_{22}(V(t), \theta_0) \end{pmatrix} dW_2(t), \quad (4)$$

where $W_1(t)_{1x1}$ and $W_2(t)_{1x1}$ are independent standard Brownian motions and $V(t)$ is latent volatility process. $b_1(\cdot)$ is a function of $X(t)$ and $b_2(\cdot), \sigma_{11}(\cdot), \sigma_{22}(\cdot)$ and $\sigma_{22}(\cdot)$ are general functions of $V(t)$, such that system of equations (4) is well-defined. Popular specifications are the square-root model of Heston (1993), the GARCH diffusion model of Nelson (1990), lognormal model of Hull and White (1987) and the eigenfunction models of Meddahi (2001). Note that in this stochastic volatility case, the dimension of volatility is $d = 1$. More general set-up can involve d driving Brownian motions in $V(t)$ equation.

As an example, *Andersen and Lund (1997)* study the generalized CIR model with stochastic volatility, specifically

$$dX(t) = \kappa_{x0}(\overline{x}_0 - X(t))dt + \sqrt{V(t)}dW_1(t)$$

$$dV(t) = \kappa_{v0}(\overline{v}_0 - V(t))dt + \sigma_{v0}\sqrt{V(t)}dW_2(t)$$

where $X(t)$ and $V(t)$ are price and volatility processes, respectively, $\kappa_{x0}, \kappa_{v0} > 0$ to ensure stationarity, $\overline{x}_0$ is the long-run mean of (log) price process, and $\overline{v}_0$ and σ_{v0} are constants. $W_1(t)$ and $W_2(t)$ are scalar Brownian motions. However, $W_1(t)$ and $W_2(t)$ are correlated such that $dW_1(t)dW_2(t) = \rho dt$ where the correlation ρ is some constant $\rho \in [-1, 1]$. Finally, note that volatility is a square-root diffusion process, which requires that $\kappa_{v0}\overline{v}_0 > \sigma^2_{v0}$.

[6] For details, see Singleton (2006), page 102.

Stochastic Volatility Model with Jumps (SVJ): A standard specification is:

$$\begin{aligned} dX(t) &= \kappa_{x0}\left(\overline{x}_0 - X(t)\right)dt + \sqrt{V(t)}dW_1(t) + J_u dq_u - J_d dq_d, \\ dV(t) &= \kappa_{v0}\left(\overline{v}_0 - V(t)\right)dt + \sigma_{v0}\sqrt{V(t)}dW_2(t), \end{aligned}$$

where q_u and q_d are Poisson processes with jump intensity parameters λ_u and λ_d respectively, and are independent of the Brownian motions $W_1(t)$ and $W_2(t)$. In particular, λ_u is the probability of a jump up, $\Pr\left(dq_u(t)=1\right)=\lambda_u$ and λ_d is the probability of a jump down, $\Pr\left(dq_d(t)=1\right)=\lambda_d$. J_u and J_d are jump up and jump down sizes and have exponential distributions: $f(J_u)=\frac{1}{\zeta_u}\exp\left(-\frac{J_u}{\zeta_u}\right)$ and $f(J_d)=\frac{1}{\zeta_d}\exp\left(-\frac{J_d}{\zeta_d}\right)$, where ζ_u, $\zeta_d>0$ are the jump magnitudes, which are the means of the jumps, J_u and J_d.

Three Factor Model (CHEN): The three factor model combines various features of the above models, by considering a version of the oft examined 3-factor model due to Chan, Karolyi, Longstaff and Sanders (1992), which is discussed in detail in Dai and Singleton (2000). In particular,

$$\begin{aligned} dX(t) &= \kappa_{x0}\left(\theta(t) - X(t)\right)dt + \sqrt{V(t)}dW_1(t), \\ dV(t) &= \kappa_{v0}\left(\overline{v} - V(t)\right)dt + \sigma_{v0}\sqrt{V(t)}dW_2(t), \\ d\theta(t) &= \kappa_{\theta 0}\left(\overline{\theta}_0 - \theta(t)\right)dt + \sigma_{\theta 0}\sqrt{\theta(t)}dW_3(t), \end{aligned} \tag{5}$$

where $W_1(t)$, $W_2(t)$ and $W_3(t)$ are independent Brownian motions, and V and θ are the stochastic volatility and stochastic mean of $X(t)$, respectively. $\kappa_{x0}, \kappa_{v0}, \kappa_{\theta 0}, \overline{v_0}, \overline{\theta_0}, \sigma_{v0}, \sigma_{\theta 0}$ are constants. As discussed above, non-negativity for $V(t)$ and $\theta(t)$ requires that $2\kappa_{v0}\overline{v}_0 > \sigma_{v0}^2$ and $2\kappa_{\theta 0}\overline{\theta_0} > \sigma_{\theta 0}^2$.

Three Factor Jump Diffusion Model (CHENJ): Andersen, Benzoni and Lund (2004) extend the three factor Chen (1996) model by incorporating jumps in the short rate process, hence improving the ability of the model to capture the effect of outliers, and to address the finding by Piazzesi (2004, 2005) that violent discontinuous movements in underlying measures may arise from monetary policy regime changes. The model is defined as follows:

$$dX(t) = \kappa_{x0}\left(\theta(t) - X(t)\right)dt + \sqrt{V(t)}dW_1(t) + J_u dq_u - J_d dq_d, \tag{6}$$

$$\begin{aligned} dV(t) &= \kappa_{v0}\left(\overline{v_0} - V(t)\right)dt + \sigma_{v0}\sqrt{V(t)}dW_2(t), \\ d\theta(t) &= \kappa_{\theta 0}\left(\overline{\theta_0} - \theta(t)\right)dt + \sigma_{\theta 0}\sqrt{\theta(t)}dW_3(t) \end{aligned} \tag{7}$$

where all parameters are similar as in (5), $W_1(t)$, $W_2(t)$ and $W_3(t)$ are independent Brownian motions, q_u and q_d are Poisson processes with jump intensities λ_{u0} and λ_{d0}, respectively, and are independent of the Brownian motions $W_r(t)$, $W_v(t)$ and $W_\theta(t)$. In particular, λ_{u0} is the probability of a jump up, $\Pr\left(dq_u(t)=1\right)=\lambda_{u0}$ and λ_{d0} is the probability of a jump down, $\Pr\left(dq_d(t)=1\right)=\lambda_{d0}$. J_u and J_d are jump up and jump down sizes and have exponential distributions $f(J_u)=\frac{1}{\zeta_{u0}}\exp\left(-\frac{J_u}{\zeta_{u0}}\right)$ and $f(J_d)=\frac{1}{\zeta_{d0}}\exp\left(-\frac{J_d}{\zeta_{d0}}\right)$, where ζ_{u0}, $\zeta_{d0}>0$ are the jump magnitudes, which are the means of the jumps J_u and J_d.

2.2 Overview on Specification Tests and Model Selection

The focus in this paper is specification testing and model selection. The "tools" used in this literature have been long established. Several key classical contributions include the Kolmogorov-Smirnov test (see e.g. Kolmogorov (1933) and Smirnov (1939)), various results on empirical processes (see e.g. Andrews (1993) and the discussion in chapter 19 of van der Vaart (1998) on the contributions of Glivenko, Cantelli, Doob, Donsker and others), the probability integral transform (see e.g. Rosenblatt (1952)), and the Kullback-Leibler Information Criterion (see e.g. White (1982) and Vuong (1989)). For illustration, the empirical distribution mentioned above is crucial in our discussion of predictive densities because it is useful in estimation, testing, and model evaluation. Let Y_t is a variable of interest with distribution F and parameter θ_0. The theory of empirical distributions provides a result that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(1\{Y_t \leq u\} - F(u|\theta_0))$$

satisfies a central limit theorem (with a parametric rate) if T is large (i.e., asymptotically). In the above expression, $1\{Y_t \leq u\}$ is the indicator function which takes value 1 if $Y_t \leq u$ and 0 otherwise. In the case where there is parameter estimation error, we can use more general results in chapter 19 of van der Vaart (1998). Define

$$P_T(f) = \frac{1}{T}\sum_{i=1}^{T} f(Y_i) \text{ and } P(f) = \int f dP$$

where P is a probability measure associated with F. Here, $P_n(f)$ converges to $P(f)$ almost surely for all the measurable functions f for which $P(f)$ is defined . Suppose one wants to test the null hypothesis that P belongs to a certain family $\{P_{\theta_0} : \theta_0 \in \Theta\}$, where θ_0 is unknown; it is natural to use a measure of the discrepancy between P_n and $P_{\widehat{\theta}}$ for a reasonable estimator $\widehat{\theta}_t$ of θ_0. In particular, if $\widehat{\theta}_t$ converges to θ_0 at a root-T rate, $\frac{1}{\sqrt{T}}(P_T - P_{\widehat{\theta}_t})$ has been shown to satisfy a central limit theorem.[7]

With regard to dynamic misspecification and parameter estimation error, the approach discussed for the class of tests in this paper allows for the construction of statistics that admit for dynamic misspecification under both hypotheses. This differs from other classes of tests such as the framework used by Diebold, Gunther and Tay (DGT: 1998), Hong (2001), and Bai (2003) in which correction dynamic specification under the null hypothesis is assumed. In particular, DGT use the probability integral transform to show that $F_t(Y_t|\Im_{t-1}, \theta_0) = \int_{-\infty}^{Y_t} f_t(y|\Im_{t-1}, \theta_0)dy$ is identically and independently distributed as a uniform random variable on $[0; 1]$, where $F_t(\cdot)$ and $f_t(\cdot)$ are a parametric distribution and density with underlying parameter θ_0, Y_t is again our random variable

[7] See Theorem 19.23 in van der Vaart (1998) for details.

of interest, and $\Im_t$ is the information set containing all "relevant" past information. They thus suggest using the difference between the empirical distribution of $F_t(Y_t|\Im_{t-1}, \widehat{\theta}_t)$ and the 45° - degree line as a measure of "goodness of fit", where $\widehat{\theta}_t$ is some estimator of θ_0. This approach has been shown to be very useful for financial risk management (see e.g. Diebold, Hahnand, Tay (1999)), as well as for macroeconomic forecasting (see e.g. Diebold, Tay and Wallis (1998) and Clements and Smith (2000,2002)). Similarly, Bai (2003) develops a Kolmogorov type test of $F_t(Y_t|\Im_{t-1}, \theta_0)$ on the basis of the discrepancy between $F_t(Y_t|\Im_{t-1}, \widehat{\theta}_t)$ and the CDF of a uniform on $[0; 1]$. As the test involves estimator $\widehat{\theta}_t$, the limiting distribution reflects the contribution of parameter estimation error and is not nuisance parameter free. To overcome this problem, Bai (2003) proposes a novel approach based on a martingalization argument to construct a modified Kolmogorov test which has a nuisance parameter free limiting distribution. This test has power against violations of uniformity but not against violations of independence. Hong (2001) proposes another related interesting test, based on the generalized spectrum, which has power against both uniformity and independence violations, for the case in which the contribution of parameter estimation error vanishes asymptotically. If the null is rejected, Hong (2001) also proposes a test for uniformity robust to non independence, which is based on the comparison between a kernel density estimator and the uniform density. Two features differentiate the tests surveyed in this paper from the tests outlined in the other papers mentioned above. First, the tests discussed here assume strict stationarity. Second, they allow for dynamic misspecification under the null hypothesis. The second feature allows us to obtain asymptotically valid critical values even when the conditioning information set does not contain all of the relevant past history. More precisely, assume that we are interested in testing for correct specification, given a particular information set which may or may not contain all of the relevant past information. This is important when a Kolmogorov test is constructed, as one is generally faced with the problem of defining $\Im_{t-1}$. If enough history is not included, then there may be dynamic misspecification. Additionally, finding out how much information (e.g. how many lags) to include may involve pre-testing, hence leading to a form of sequential test bias. By allowing for dynamic misspecification, such pre-testing is not required. Also note that critical values derived under correct specification given $\Im_{t-1}$ are not in general valid in the case of correct specification given a subset of $\Im_{t-1}$. Consider the following example. Assume that we are interested in testing whether the conditional distribution of $Y_t|Y_{t-1}$ follows normal distribution $N(\alpha_1 Y_{t-1}, \sigma_1)$. Suppose also that in actual fact the "relevant" information set has $\Im_{t-1}$ including both $Y_{t-1}$and Y_{t-2}, so that the true conditional model is $Y_t|\Im_{t-1} = Y_t|Y_{t-1}, Y_{t-2} = N(\alpha_1 Y_{t-1} + \alpha_2 Y_{t-2}, \sigma_2)$. In this case, correct specification holds with respect to the information contained in X_{t-1}; but there is dynamic misspecification with respect to $Y_{t-1}$and Y_{t-2}. Even without taking account of parameter estimation error, the critical values obtained assuming correct dynamic specification are invalid,

thus leading to invalid inference. Stated differently, tests that are designed to have power against both uniformity and independence violations (i.e. tests that assume correct dynamic specification under the null) will reject; an inference which is incorrect, at least in the sense that the "normality" assumption is not false. In summary, if one is interested in the particular problem of testing for correct specification for a given information set, then the approach of tests in this paper is appropriate

3 Consistent Distribution-Based Specification Tests and Predictive Density Type Model Selection for Diffusion Processes

3.1 One Factor Models

In this Section we outline the set-up for the general class of one factor jump diffusion specifications. All analysis carry through to the more complicated case of multi-factor stochastic volatility models which we will elaborate upon in the next Subsection. In the presentation of the tests, we follow a view that all candidate models, either single or multiple dimensional ones, are approximations of reality, and can thus be misspecified. The issue of correct specification (or misspecification) of a single model and the model selection test for choosing amongst multiple competing models allow for this feature.

To begin, fix the time interval $[0, T]$, consider a given single one factor candidate model the same as (1), with the true parameters $\theta_0, \lambda_0, \mu_0$ to be replaced by it's the pseudo true analogs $\theta^\dagger, \lambda, \mu$, respectively and $0 \leq t \leq T$:

$$X(t_-) = \int_0^t b(X(s_-), \theta^\dagger)ds - \lambda t \int_Y y\phi(y)dy + \int_0^t \sigma(X(s_-), \theta^\dagger)dW(s) + \sum_{j=1}^{J_t} y_j,$$

or

$$dX(t-) = \Big(b(X(t-), \theta^\dagger) - \lambda\mu\Big)\, dt + \sigma(X(t-), \theta^\dagger)dW(t) + \int_Y yp(dy, dt), \tag{8}$$

where variables are defined the same as in (1) and (2). Note that as the above model is the one factor version of (1) and (2) where the dimension of $X(t_-)$ is 1x1, $W(t)$ is a one-dimensional standard Brownian motion and jump size, and y_j is one dimensional variable for all j. Also note that both J_t and y_j are assumed to be independent of the driving Brownian motion.

If the single model is correctly specified, then $b(X(t-), \theta^\dagger) = b_0(X(t-), \theta_0)$, $\sigma(X(t-), \theta^\dagger) = \sigma_0(X(t-), \theta_0)$, $\lambda = \lambda_0$, $\mu = \mu_0$ and $\phi = \phi_0$ where $b_0(X(t-), \theta_0), \sigma_0(X(t-), \theta_0), \lambda_0, \mu_0\, , \phi_0$ are unknown and belong to the true specification.

Now consider a different case (not a single model) where m candidate models are involved. For model k with $1 \leq k \leq m$, denote it's corresponding specification to be $(b_k(X(t_-), \theta_k^\dagger), \sigma_k(X(t_-), \theta_k^\dagger),$

λ_k, μ_k, ϕ_k). Two scenarios immediate arise. Firstly, if the model k is correctly specified, then $b_k(X(t_-), \theta_k^\dagger) = b_0(X(t_-), \theta_0)$, $\sigma_k(X(t_-), \theta_k^\dagger) = \sigma_0(X(t_-), \theta_0)$, $\lambda_k = \lambda_0$, $\mu_k = \mu_0$ and $\phi_k = \phi_0$ which are similar to the case of a single model. In the second scenario, all the models are likely to be misspecified and modelers are faced with the choice of selecting the "best" one. This type of problem is well-fitted into the class of accuracy assessment tests initiated earlier by Diebold and Mariano (1995) or White (2000).

The tests discussed hereafter are Kolomogorov type tests based on the construction of cumulative distribution functions (CDFs). In a few cases, the CDF is known in closed form. For instance, for the simplified version of the CIR model as in (3), $X(t)$ belongs to the linear exponential (or Pearson) family with the gamma CDF of the form:[8]

$$F(u, \alpha, \lambda) = \frac{\int_0^u (\frac{\lambda}{2})^{-2(1-\alpha/\lambda)-1} \exp(-x/(\frac{\lambda}{2}))dx}{\Gamma(2(1-\alpha/\lambda))}, \text{ where } \Gamma(x) = \int_0^\infty t^x \exp(-t)dt, \tag{9}$$

and α, λ are constants.

Furthermore, if we look at the pure diffusion process without jumps:

$$dX(t) = b(X(t), \theta^\dagger)dt + \sigma(X(t), \theta^\dagger)dW(t), \tag{10}$$

where $b(\cdot)$ and $\sigma = \sigma(\cdot)$ are drift and volatility functions, it is known that the stationary density, say $f(x, \theta^\dagger)$, associated with the invariant probability measure can be expressed explicitly as:[9]

$$f(x, \theta^\dagger) = \frac{c(\theta^\dagger)}{\sigma^2(x, \theta^\dagger)} \exp\left(\int^x \frac{2b(u, \theta^\dagger)}{\sigma^2(u, \theta^\dagger)} du\right)$$

where $c(\theta^\dagger)$ is a constant ensuring that f integrates to one. The CDF, say $F(u, \theta^\dagger) = \int^u f(x, \theta^\dagger)dx$, can then be obtained using available numerical integration procedures.

However, in most cases, it is impossible to derive the CDFs in closed form. To obtain a CDF in such cases, a more general approach is to use simulation. Instead of estimating the CDF directly, simulation techniques estimates the CDF indirectly utilizing it's generated sample paths and the theory of empirical distributions. The specification of a specific diffusion process will dictate the sample paths and thereby corresponding test outcomes.

Note that in the historical context, many early papers in this literature are probability density-based. For example, in a seminal paper, Ait-Sahalia (1996) compares the marginal densities implied by hypothesized null models with nonparametric estimates thereof. Following the same framework of correct specification tests, CS(2005) and BCS (2008), however, do not look at densities. Instead, they compare the cumulative distribution (marginal or joint) implied by a hypothesized null model with the corresponding empirical distribution. While CS (2005) focus on the known unconditional

[8] See Wong (1964) for details.

[9] See Karlin and Taylor (1981) for details.

distribution, BCS (2008) look at the conditional simulated distributions. CS (2011) make extensions to multiple models in the context of out of sample accuracy assessment tests. This approach is somewhat novel to this continuous time model testing literature.

Now suppose we observe a discrete sample path $X_1, X_2, ..., X_T$ (also referred as skeletons).[10] The corresponding hypotheses can be set up as follows:

Hypothesis 1: Unconditional Distribution Specification Test of a Single Model

$H_0 : F(u, \theta^\dagger) = F_0(u, \theta_0)$, for all u, a.s.

$H_A : \Pr\left(F(u, \theta^\dagger) - F_0(u, \theta_0) \neq 0\right) > 0$, for some $u \in U$, with non-zero Lebesgue measure.

where $F_0(u, \theta_0)$ is the true cumulative distribution implied by the above density, i.e. $F_0(u, \theta_0) = \Pr(X_t \leq u)$. $F(u, \theta^\dagger) = \Pr\left(X_t^{\theta^\dagger} \leq u\right)$ is the cumulative distribution of the proposed model. $X_t^{\theta^\dagger}$ is a skeleton implied by model (8).

Hypothesis 2: Conditional Distribution Specification Test of A Single Model

$H_0 : F_\tau(u|X_t, \theta^\dagger) = F_{0,\tau}(u|X_t, \theta_0)$, for all u, a.s.

$H_A : \Pr\left(F_\tau(u|X_t, \theta^\dagger) - F_{0,\tau}(u|X_t, \theta_0) \neq 0\right) > 0$, for some $u \in U$, with non-zero Lebesgue measure.

where $F_\tau(u|X_t, \theta^\dagger) = \Pr\left(X_{t+\tau}^{\theta^\dagger} \leq u | X_t^{\theta^\dagger} = X_t\right)$ is τ-Step ahead conditional distributions and $t = 1, ..., T - \tau$. $F_{0,\tau}(u|X_t, \theta_0)$ is τ-Step ahead true conditional distributions .

Hypothesis 3: Predictive Density Test for Choosing Amongst Multiple Competing Models

The null hypothesis is that no model can outperform model 1 which is the benchmark model.[11]

$$H_0 : \max_{k=2,...,m} E_X \left(\left(F_{X_{1,t+\tau}^{\theta_1^\dagger}(X_t)}(u_2) - F_{X_{1,t+\tau}^{\theta_1^\dagger}(X_t)}(u_1) \right) - (F_0(u_2|X_t) - F_0(u_1|X_t)) \right)^2$$
$$- E_X \left(\left(F_{X_{k,t+\tau}^{\theta_k^\dagger}(X_t)}(u_2) - F_{X_{k,t+\tau}^{\theta_k^\dagger}(X_t)}(u_1) \right) - (F_0(u_2|X_t) - F_0(u_1|X_t)) \right)^2$$

H_A : negation of H_0'

where $F_{X_{k,t+\tau}^{\theta_k^\dagger}(X_t)}(u) = F_k^\tau(u|X_t, \theta_k^\dagger) = P_{\theta_k^\dagger}^\tau\left(X_{t+\tau}^{\theta_k^\dagger} \leq u | X_t^{\theta_k^\dagger} = X_t\right)$, which is the conditional distribution of $X_{t+\tau}$, given X_t, and evaluated at u under the probability law generated by model k. $X_{k,t+\tau}^{\theta_k^\dagger}(X_t)$ with $1 \leq \tau \leq T - t$ is the skeleton implied by model k, parameter $\theta_k^\dagger$ and initial value X_t. Analogously, define $F_0^\tau(u|X_t, \theta_0) = P_{\theta_0}^\tau(X_{t+\tau} \leq u|X_t)$ to be the "true" conditional distribution.

Note that the three hypotheses expressed above apply exactly the same to the case of multifactor

[10] As mentioned earlier, we follow CS (2005) by using notation $X(\cdot)$ when defining continuous time processes and X_t for a skeleton.

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

diffusions. Now, before moving to the statistics description Section, we briefly explain the intuitions in facilitating construction of the tests:

In the first case (Hypothesis 1), CS (2005) construct a Kolomogorov type test based on comparison of the empirical distribution and the unconditional CDF implied by the specification of the drift, variance and jumps. Specifically, one can look at the scaled difference between

$$F(u,\theta^{\dagger}) = \Pr\left(X_t^{\theta^{\dagger}} \leq u\right) = \int^{u} f(x,\theta^{\dagger})dx$$

and estimator of the true $F_0(u|X_t,\theta_0)$, the empirical distribution of X_t defined as:

$$\frac{1}{T}\sum_{t=1}^{T} 1\left\{X_t \leq u\right\}$$

where $1\left\{Y_t \leq u\right\}$ is indicator function which takes value 1 if $Y_t \leq u$ and 0 otherwise.

Similarly for the second case of conditional distribution (Hypothesis 2), the test statistic V_T can be a measure of the distance between the $\tau - Step$ ahead conditional distribution of $X_{t+\tau}^{\theta^{\dagger}}$, given $X_t^{\theta^{\dagger}} = X_t$, as:

$$F_{\tau}(u|X_t,\theta^{\dagger}) = \Pr\left(X_{t+\tau}^{\theta^{\dagger}} \leq u|X_t^{\theta^{\dagger}} = X_t\right),$$

to an estimator of the true $F_{0,\tau}(u|X_t,\theta_0)$, the conditional empirical distribution of $X_{t+\tau}$ conditional on the initial value X_t defined as:

$$\frac{1}{T-\tau}\sum_{t=1}^{T-\tau} 1\left\{X_{t+\tau} \leq u\right\},$$

In the third case (Hypothesis 3), model accuracy is measured in terms of a distributional analog of mean square error. As is commonplace in the out-of-sample evaluation literature, the sample of T observations is divided into two subsamples, such that $T = R + P$, where only the last P observations are used for predictive evaluation. A $\tau-$Step ahead prediction error under model k is $1\{u_1 \leq X_{t+\tau} \leq u_2\} - \left(F_k^{\tau}(u_2|X_t,\theta_k^{\dagger}) - F_k^{\tau}(u_1|X_t,\theta_k^{\dagger})\right)$ where $2 \leq k \leq m$ and similarly for model 1 by replacing index k with index 1. Suppose we can simulate $P-\tau$ paths of $\tau-$Step ahead skeleton[12] using X_t as starting values where $t = R, ..., R+P-\tau$, from which we can construct a sample of $P-\tau$ prediction errors. Then, these prediction errors can be used to construct a test statistic for model comparison. In particular, model 1 is defined to be more accurate than model k if:

$$\begin{aligned} & E\left(\left((F_1^{\tau}(u_2|X_t,\theta_1^{\dagger}) - F_1^{\tau}(u_1|X_t,\theta_1^{\dagger})) - (F_0^{\tau}(u_2|X_t,\theta_0) - F_0^{\tau}(u_1|X_t,\theta_0))\right)^2\right) \\ < \quad & E\left(\left((F_k^{\tau}(u_2^{\tau}|X_t,\theta_k^{\dagger}) - F_k^{\tau}(u_1^{\tau}|X_t,\theta_k^{\dagger})) - (F_0^{\tau}(u_2|X_t,\theta_0) - F_0^{\tau}(u_1|X_t,\theta_0))\right)^2\right). \end{aligned}$$

where $E(\cdot)$ is an expectation operator and $E\left(1\{u_1 \leq X_{t+\tau} \leq u_2\}|X_t\right) = F_0^{\tau}(u_2|X_t,\theta_0) - F_0^{\tau}(u_1|X_t,\theta_0)$. Concretely, model k is worse than model 1 if on average $\tau-$Step ahead prediction errors under model k is larger than that of model 1.

[12] See Section 3.3.1 for model simulation details.

Finally, it is important to point out some main features characterized by all the three test statistics. Processes $X(t)$ hereafter is required to satisfy the regular conditions, i.e. assumptions A1-A8 in CS (2011). Regarding model estimation (in Section 3.3), $\theta^\dagger$ and $\theta_k^\dagger$ are unobserved and need to be estimated. While CS (2005), BCS (2008) utilize (recursive) Simulated General Method of Moments (SGMM), CS (2011) make extension to (recursive) Nonparametric Simulated Quasi Maximum Likelihood (NPSQML). For the unknown distribution and conditional distribution, it will be pointed out in Section 3.3.2 that $F(u,\theta^\dagger)$, $F_\tau(u|X_t,\theta^\dagger)$ and $F_{X_{k,t+\tau}^{\theta_k^\dagger}(X_t)}(u)$ can be replaced by their simulated counterparts using the (recursive) SGMM and NPSQML parameter estimators. In addition, test statistics converge to functional of Gaussian processes with covariance kernels that reflect time dependence of the data and the contribution of parameter estimation error (PEE). Limiting distributions are not nuisance parameter free and critical values thereby cannot be tabulated by the standard approach. All the tests discussed in this paper rely on the bootstrap procedures for obtaining the asymptotically valid critical values, which we will describe in Section 3.4.

3.1.1 Unconditional Distribution Tests

For one-factor diffusions, we outline the construction of unconditional test statistics in the context where CDF is known in closed form. In order to test the **Hypothesis 1**, consider the following statistic:

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

where

$$V_{T,N,h} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big(1\{X_t \le u\} - F(u,\widehat{\theta}_{T,N,h})\Big)$$

In the above expression, U is a compact interval and $\int_U \pi(u)du = 1$, $1\{X_t \le u\}$ is again the indicator function which returns value 1 if $X_t \le u$ and 0 otherwise. Further, as defined in Section 3.3, $\widehat{\theta}_{T,N,h}$ hereafter is a simulated estimator where T is sample size and h is the discretization interval used in simulation. In addition, with the abuse of notation, N is a generic notation throughout this paper, i.e. $N = L$, the length of each simulation path for (recursive) SGMM and $N = M$, the number of random draws (simulated paths) for (recursive) NPQML estimator.[13] Also note in our notation that as the above test is in sample specification test, the estimator and the statistics are constructed using the entire sample, i.e. $\widehat{\theta}_{T,N,h}$.

It has been shown in CS (2005) that under regular conditions and if the estimator is estimated by SGMM, the above statistics converges to a functional of Gaussian process.[14] In particular, pick

[13] M is often chosen to coincide with S, the number of simulated paths used when simulating distributions.

[14] For details and the proof, see Theorem 1 in CS (2005).

the choice $T, N \to \infty, h \to 0, T/N \to 0$ and $Th^2 \to 0$

Under the null,

$$V^2_{T,N,h} \to \int_U Z^2(u)\pi(u)$$

where Z is a Gaussian process with covariance kernel. Hence, the limiting distribution of $V^2_{T,N,h}$ is a functional of a Gaussian process with a covariance kernel that reflects both PEE and the time series nature of the data. As $\widehat{\theta}_{T,N,h}$ is root-T consistent, PEE does not disappear in the asymptotic covariance kernel.

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

$$\lim_{T\to\infty} \Pr(\frac{1}{T} V^2_{T,N,h} > \varepsilon) = 1$$

For the asymptotic critical value tabulation, we use the bootstrap procedure. In order to establish validity of the block bootstrap under SGMM with the presence of PEE, the simulated sample size should be chosen to grow at a faster rate than the historical sample, i.e. $T/N \to 0$.

Thus, we can follow Steps in appropriate bootstrap procedure in Section 3.4. For instance, if the SGMM estimator is used, the bootstrap statistic is

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

where

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

In the above expression, $\widehat{\theta}^*_{T,N,h}$ is the bootstrap analog of $\widehat{\theta}_{T,N,h}$ and is estimated by the bootstrap sample $X^*_1, ..., X^*_T$ (see Section 3.4). With appropriate conditions, CS (2005) show that under the null, $V^{2*}_{T,N,h}$ has a well defined limiting distribution which coincides with that of $V^2_{T,N,h}$. We then can straightforwardly derive the bootstrap critical value by following Step 1-5 Section 3.4. In particular, in Step 5, the idea is to perform B bootstrap replications (B large) and compute the percentiles of the empirical distribution of the B bootstrap statistics. Reject H_0 if $V^2_{T,N,h}$ is greater than the $(1-\alpha)th-$percentile of this empirical distribution. Otherwise, do not reject H_0.

3.1.2 Conditional Distribution Tests

Hypothesis 2 tests correct specification of the conditional distribution, implied by a proposed diffusion model. In practice, the difficulty arises from the fact that the functional form of neither τ-Step ahead conditional distributions $F_\tau(u|X_t, \theta^\dagger)$ nor $F_{0,\tau}(u|X_t, \theta_0)$ is unknown in most cases. Therefore, BCS (2008) develop bootstrap specification test on the basis of simulated distribution

using the SGMM estimator.[15] With the important inputs leading to the test such as simulated estimator, distribution simulation and bootstrap procedures to be presented in the next Section[16], the test statistic is defined as:

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

where

$$Z_T(u,v) = \frac{1}{\sqrt{T-\tau}} \sum_{t=1}^{T-\tau} \left(\frac{1}{S} \sum_{s=1}^{S} 1\left\{ X_{s,t+\tau}^{\widehat{\theta}_{T,N,h}} \leq u \right\} - 1\{X_{t+\tau} \leq u\} \right) 1\{X_t \leq v\},$$

with U and V compact sets on the real line. $\widehat{\theta}_{T,N,h}$ is the simulated estimator using entire sample $X_{1,},...,X_T$ and S is the number of simulated replications used in the estimation of conditional distributions as described in Section 3.3. If SGMM estimator is used (similar to unconditional distribution case and the same as in BCS (2008)), then $N = L$, where L is the simulation length used in parameter estimation.

The above statistic is a simulation-based version of the conditional Kolmogorov test of Andrews (1997), which compare the joint empirical distribution

$$\frac{1}{T-\tau} \sum_{t=1}^{T-\tau} 1\{X_{t+\tau} \leq u\} 1\{X_t \leq v\}$$

with its semi-empirical/semi-parametric analog given by the product of

$$\frac{1}{T-\tau} \sum_{t=1}^{T-\tau} F_{0,\tau}(u|X_t, \theta_0) 1\{X_t \leq v\}.$$

Intuitively, if the null is not rejected, the metric distance between the two should asymptotically disappear. In the simulation context with parameter estimation error, the asymptotic limit of Z_T however is a nontrivial one. BCS (2008) show that with the proper choice of T, N, S, h, i.e. $T, N, S, T^2/S \to \infty$ and $h, T/N, T/S, Nh, h^2T \to 0$, then

$$Z_T \xrightarrow{d} \sup_{u \times v \in U \times V} |Z(u,v)|,$$

where $Z(u,v)$ is a Gaussian process with a covariance kernel that characterizes: 1) long-run variance we would have if we knew $F_{0,\tau}(u|X_1, \theta_0)$; 2) the contribution of parameter estimation error; 3) The correlation between the first two.

Furthermore, under H_A, there exists some $\varepsilon > 0$ such that:

$$\lim_{P \to \infty} \Pr\left(\frac{1}{\sqrt{T}} Z_T > \varepsilon \right) = 1.$$

[15] In this paper, we assume that $X(\cdot)$ satisfies the regularity conditions stated in CS (2011), i.e. assuptions A1-A8. Those conditions also reflect requirements A1-A2 in BCS (2008). Note that, the SGMM estimator used in BCS (2008) satisfies the root-N consistency condition that CS (2011) impose on their parameter estimator (See Assumption 4).

[16] See Sections 3.3 and 3.4 for further details.

As $T/S \to 0$, the contribution of simulation error is asymptotically negligible. The limiting distribution is not nuisance parameter free and hence critical values cannot be tabulated directly from it. The appropriate bootstrap statistic in this context is:

$$Z_T^* = \sup_{u \times v \in U \times V} |Z_T^*(u,v)|,$$

where

$$\begin{aligned} Z_T^*(u,v) &= \frac{1}{\sqrt{T-\tau}} \sum_{t=1}^{T-\tau} \left(\frac{1}{S} \sum_{s=1}^{S} 1\left\{ X_{s,t+\tau}^{\widehat{\theta}_{T,N,h}^*} \leq u \right\} - 1\{X_{t+\tau}^* \leq u\} \right) 1\{X_t^* \leq v\} \\ &\quad - \frac{1}{\sqrt{T-\tau}} \sum_{t=1}^{T-\tau} \left(\frac{1}{S} \sum_{s=1}^{S} 1\left\{ X_{s,t+\tau}^{\widehat{\theta}_{T,N,h}} \leq u \right\} - 1\{X_{t+\tau} \leq u\} \right) 1\{X_t \leq v\} \end{aligned}$$

In the above expression, $\widehat{\theta}_{T,N,h}^*$ is the bootstrap parameter estimated using the resampled data X_t^* for $t = 1, ..., T-\tau$. $X_{s,t+\tau}^{\widehat{\theta}_{T,N,h}^*}$, $s = 1, ..., S$ and $t = 1, ..., T-\tau$ is the simulated data under $\widehat{\theta}_{T,N,h}^*$ and X_t^*, $t = 1, ..., T-\tau$ is a resampled series constructed using standard block-bootstrap methods as described in 3.4. Note that in the original paper, BCS (2008) propose bootstrap SGMM estimator for conditional distribution of diffusion processes. CS (2011) extend the test to the case of simulated recursive NPSQML estimator. Regarding the generation of the empirical distribution of Z_T^* (asthmatically the same as Z_T), follow Step 1-5 in the bootstrap procedure in Section 3.4. This yields B bootstrap replications (B large) of Z_T^*. One can then compare Z_T with the percentiles of the empirical distribution of Z_T^*, and reject H_0 if Z_T is greater than the $(1-\alpha)th$-percentile. Otherwise, do not reject H_0. Tests carried out in this manner are correctly asymptotically sized, and have unit asymptotic power.

3.1.3 Predictive Density Tests for Multiple Competing Models

In many circumstances, one might want to compare one (benchmark) model (model 1) against multiple competing models (models k, $2 \leq k \leq m$). In this case, recall in the null in **Hypothesis 3** is that no model can outperform the benchmark model. In testing the null, we first choose a particular interval i.e., $(u_1, u_2) \in U \text{x} U$ where U is a compact set so that the objective is evaluation of predictive densities for a given range of values. In addition, in the recursive setting (not full sample is used to estimate parameters), if we use the recursive NPSQML estimator, say $\widehat{\theta}_{1,t,N,h}$ and $\widehat{\theta}_{k,t,N,h}$, for models 1 and k, respectively, then the test statistic is defined as

$$D_{k,P,S}^{Max}(u_1,u_2) = \max_{k=2,...,m} D_{k,P,S}(u_1,u_2).$$

where

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

$$- \left[\frac{1}{S} \sum_{i=1}^{S} 1 \left\{ u_1 \leq X_{k,i,t+\tau}^{\widehat{\theta}_{k,t,N,h}}(X_t) \leq u_2 \right\} - 1\{u_1 \leq X_{t+\tau} \leq u_2\} \right]^2 \Bigg).$$

All notation is consistent with previous Sections where S is the number of simulated replications used in the estimation of conditional distributions. $X_{1,i,t+\tau}^{\widehat{\theta}_{1,t,N,h}}(X_t)$ and $X_{k,i,t+\tau}^{\widehat{\theta}_{k,t,N,h}}$, $i = 1, ..., S$, $t = 1, ..., T-\tau$, are the ith simulated path under $\widehat{\theta}_{1,t,N,h}$ and $\widehat{\theta}_{k,t,N,h}$. If models 1 and k are nonnested for at least one $k = 2, ..., m$. Under regular conditions and if P, R, S, h are chosen such as $P, R, N \to \infty$ and $h, P/N, h^2 P \to 0$, $P/R \to \pi$ where π is finite then

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

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

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

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 the covariance kernels that involves error in parameter estimation. Bootstrap statistics are therefore required to reflect this parameter estimation error issue.[17]

In the implementation, we can obtain the asymptotic critical value using a recursive version of the block bootstrap. The idea is that when forming block bootstrap samples in the recursive setting, observations at the beginning of the sample are used more frequently than observations at the end of the sample. We can replicate the Step 1-5 in bootstrap procedure in Section 3.4. It should be stressed the re-sampling in the Step 1 is the recursive one. Specifically, 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 bootstrap estimator $\widehat{\theta}_{k,t,N,h}^*$. Recall that N is chosen in CS (2011) as the number of simulated series used to estimate the parameters ($N = M = S$) and such as $N/R, N/P \to \infty$. Under this condition, simulation error vanishes and there is no need to resample the simulated series.

CS (2011) show that

$$\frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left(\widehat{\theta}_{k,t,N,h}^* - \widehat{\theta}_{k,t,N,h} \right)$$

has the same limiting distribution as

$$\frac{1}{\sqrt{P}} \sum_{t=R}^{T} \left(\widehat{\theta}_{k,t,N,h} - \theta_k^\dagger \right),$$

[17] See CS (2011) for further discussion.

conditional on all samples except a set with probability measure approaching zero. Given this, the appropriate bootstrap statistic is:

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

As the bootstrap statistic is calculated from the last P resampled observations, it is necessary to have each bootstrap term recentered around the (full) sample mean. This is true even in the case there is no need to mimic PEE, i.e. the choice of P, R is such that $P/R \to 0$. In such a case, above statistic can be formed using $\widehat{\theta}_{k,t,N,h}$ rather than $\widehat{\theta}^*_{k,t,N,h}$.

For any bootstrap replication, repeat B times (B large)) bootstrap replications which yield B bootstrap statistics $D^*_{k,P,S}$. Reject H_0 if $D_{k,P,S}$ is greater than the $(1-\alpha)th$-percentile of the bootstrap empirical distribution. For numerical implementation, it is of importance to note that in the case where $P/R \to 0, P, T, R \to \infty$, there is no need to re-estimate $\widehat{\theta}^*_{1,t,N,h}$ ($\widehat{\theta}^*_{k,t,N,h}$). Namely, $\widehat{\theta}_{1,t,N,h}(\widehat{\theta}_{k,t,N,h})$ can be used in all bootstrap experiments.

Of course, the above framework can also be applied using entire simulated distributions rather than predictive densities, by simply estimating parameters once, using the entire sample, as opposed to using recursive estimation techniques, say, as when forming predictions and associated predictive densities.

3.2 Multifactor Models

Now, let us turn our attention to multifactor diffusion models of the form $(X(t), V(t))' = \left(X(t), V^1(t), ..., V^d(t)\right)'$, where only the first element, the diffusion process X_t, is observed while $V(t) = (V^1(t), ..., V^d(t))'$ is latent. The most popular class of the multifactor models is stochastic volatility model expressed as below:

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

where $W_1(t)_{1x1}$ and $W_2(t)_{1x1}$ are independent Brownian Motions.[18] For instance, many term structure models require the multifactor specification of the above form (see Dai and Singleton (2000)). In a more complicated case, the drift function can also be specified to be a stochastic process which poses even more challenges to testing. As mentioned earlier, the hypotheses (**Hypothesis 1,2,3**) and the test construction strategy for multifactor models are the same as for one factor model. All theory essentially applies immediately to multifactor cases. In implementation, the key difference is in the simulated approximation scheme facilitating parameter and CDF estimation. $X(t)$ cannot simply be expressed as a function of $d+1$ driving Brownian motions but instead involves a function of $(W_{jt}, \int_0^t W_{js} dW_{is})$, $i, j = 1, ..., d+1$ (see e.g. Pardoux and Talay (1985) p.30-32 and CS(2005)).

For illustration, we hereafter focus on the analysis of a stochastic volatility model (11) where drift and diffusion coefficients can be written as

$$b = \begin{pmatrix} b_1(X(t), \theta^\dagger)) \\ b_2(V(t), \theta^\dagger)) \end{pmatrix}, \ \sigma = \begin{pmatrix} \sigma_{11}(V(t), \theta^\dagger) & \sigma_{12}(V(t), \theta^\dagger) \\ 0 & \sigma_{22}(V(t), \theta^\dagger) \end{pmatrix}$$

We also examine a three factor model (i.e., the Chen Model as in (5)) and a three factor model with jumps, (i.e., CHENJ as in (6)). By presenting two and three factor models as an extension of our above discussion, we make it clear that specification tests of multiple factor diffusions with $d \geq 3$ can be easily constructed in similar manner.

In distribution estimation, the important challenge for multifactor models lies in the missing variable issue. In particular, for simulation of X_t, one needs initial values of the latent processes $V_1, ..., V_d$, which are unobserved. To overcome this problem, it suffices to simulate the process using different random initial values for the volatility process, then construct the simulated distribution using those initial values and average them out. This allows one to integrate out the effect of a particular choice of volatility initial value.

For clarity of exposition, we sketch out a simulation strategy for a general model of d latent variables in Section 3.3. This generalizes the simulation scheme of three factor models in the Cai and Swanson (2011). As a final remark before moving to the statistic presentation, note that the class of multifactor diffusion processes considered in this paper is required to match the regular conditions as in previous Section (assumption from A1-A8 in CS (2011) with A4 being replaced by A4').

[18] Note that the dimension of $X(\cdot)$ can be higher and we can add jumps to the above specification such that it satisfies the regularity conditions outlined in the one factor case. In addition, CS (2005), provide a detailed discussion of approximation schemes in the context of stochastic volatility models.

3.2.1 Unconditional Distribution Tests

Following the above discussion on test construction, we specialize to the case of two-factor stochastic volatility models. Extension to general multidimensional and multifactor models follows similarly. As the CDF is rarely known in closed form for stochastic volatility models, we rely on simulation technique. With the simulation scheme, estimators, simulated distributed and bootstrap procedures to be presented in the next sections (see Section 3.3 and 3.4), the test statistics for **Hypothesis 1** turns out to be:

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

In the above expression, recall that S is the number of simulation paths used in distribution simulation, $\widehat{\theta}_{T,N,L,h}$ is a simulated estimator (see Section 3.3). N is a generic notation throughout this paper, i.e. $N = L$, the length of each simulation path for SGMM and $N = M$, the number of random draws (simulated paths) for NPQML estimator. h is the discretization interval used in simulation. Note that $\widehat{\theta}_{T,N,L,h}$ is chosen in CS (2005) to be SGMM estimator using full sample and therefore $N = L = S$.[19] To put it simply, one can write $\widehat{\theta}_{T,S,h} = \widehat{\theta}_{T,N,L,h}$.

Under the null, choose T, S to satisfy $T, S \to \infty, Sh \to 0, T/S \to 0$ then:

$$SV_{T,S,h}^2 \to \int_U SV^2(u)\pi(u)$$

where Z is a Gaussian process with covariance kernel that reflects both PEE and the time dependent nature of the data. The relevant bootstrap statistic is:

$$SV_{T,S,h}^{2*} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left((1\{X_t^* \leq u\} - 1\{X_t \leq u\}) - \frac{1}{S} (\sum_{t=1}^{S} 1(X_{t,h}^{\widehat{\theta}_{T,N,L,h}^*} \leq u) - 1(X_{t,h}^{\widehat{\theta}_{T,N,L,h}} \leq u)) \right)$$

where $\widehat{\theta}_{T,S,h}^*$ is the bootstrap analogue of $\widehat{\theta}_{T,S,h}$. Repeat the Step 1-5 in the bootstrap procedure in Section 3.4 to obtain critical value which are the percentiles of the empirical distribution of Z_T^*. Compare $SV_{T,S,h}$ with the percentiles of the empirical distribution of the bootstrap statistic and reject H_0 if $SV_{T,S,h}$ is greater than the $(1-\alpha)th$-percentile thereof. Otherwise, do not reject H_0.

3.2.2 Conditional Distribution Tests

To test **Hypothesis 2** for the multifactor models, first we present the test statistic for the case of the stochastic volatility model (X_t, V_t) in (11), (i.e., for two factor diffusion), and then we discuss testing with the three factor model (X_t, V_t^1, V_t^2) as in (5). Other multiple factor models can be tested analogously. Note that for illustration, we again assume use of the SGMM estimator $\widehat{\theta}_{T,N,L,h}$, as in

[19] As seen in assumption A4' in CS (2011) and Section 3.3 of this paper, $\widehat{\theta}_{T,N,L,h}$ can be other estimators such as the NPSQML estimator. Importantly, $\widehat{\theta}_{T,N,L,h}$ satisfies condition A4' in CS (2011).

the original work of BCS (2008) (namely, $\widehat{\theta}_{T,N,L,h}$ is the simulated estimator described in Section 3.3). Specifically, N is chosen as the length of sample path L used in parameter estimation. The associated test statistic is:

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

$$SZ_T(u,v) = \frac{1}{\sqrt{T-\tau}}\sum_{t=1}^{T-\tau}\left(\frac{1}{NS}\sum_{j=1}^{N}\sum_{i=1}^{S} 1\left\{X_{j,i,t+\tau}^{\widehat{\theta}_{T,N,,L,h}} \le u\right\} - 1\{X_{t+\tau}\le u\}\right) 1\{X_t \le v\}.$$

where $X_{j,i,t+\tau}^{\widehat{\theta}_{T,N,,L,h}}$ is is τ - Step ahead simulated skeleton obtained by simulation procedure for multi-factor model in Subsection 3.4.1.

In a similar manner, the bootstrap statistic analogous to SZ_T is

$$SZ_T^* = \sup_{u\times v\in U\times V} |SZ_T^*(u,v)|,$$

$$\begin{aligned} SZ_T^*(u,v) &= \frac{1}{\sqrt{T-\tau}}\sum_{t=1}^{T-\tau}\left(\frac{1}{NS}\sum_{j=1}^{N}\sum_{i=1}^{S} 1\left\{X_{j,i,t+\tau}^{\widehat{\theta}^*_{T,N,L,h}} \le u\right\} - 1\{X^*_{t+\tau}\le u\}\right) 1\{X^*_t \le v\} \\ &\quad -\frac{1}{\sqrt{T-\tau}}\sum_{t=1}^{T-\tau}\left(\frac{1}{NS}\sum_{j=1}^{N}\sum_{i=1}^{S} 1\left\{X_{j,i,t+\tau}^{\widehat{\theta}_{T,N,L,h}} \le u\right\} - 1\{X_{t+\tau}\le u\}\right) 1\{X_t \le v\}. \end{aligned}$$

where $\widehat{\theta}^*_{T,N,L,h}$ is the bootstrap estimator described in Section 3.4. For the three factor model, the test statistic is defined as

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

$$MZ_T(u,v) = \frac{1}{\sqrt{T-\tau}}\sum_{t=1}^{T-\tau}\left(\frac{1}{L^2S}\sum_{j=1}^{L}\sum_{k=1}^{L}\sum_{i=1}^{S} 1\left\{X_{s,t+\tau}^{\widehat{\theta}_{T,N,L,h}} \le u\right\} - 1\{X_{t+\tau}\le u\}\right) 1\{X_t \le v\}$$

and bootstrap statistics is:

$$\begin{aligned} MZ_T^*(u,v) &= \frac{1}{\sqrt{T-\tau}}\sum_{t=1}^{T-\tau}\left(\frac{1}{L^2S}\sum_{j=1}^{L}\sum_{k=1}^{L}\sum_{i=1}^{S} 1\left\{X_{s,t+\tau}^{\widehat{\theta}^*_{t,N,L,h}} \le u\right\} - 1\{X^*_{t+\tau}\le u\}\right) 1\{X^*_t \le v\} \\ &\quad -\frac{1}{\sqrt{T-\tau}}\sum_{t=1}^{T-\tau}\left(\frac{1}{L^2S}\sum_{j=1}^{L}\sum_{k=1}^{L}\sum_{i=1}^{S} 1\left\{X_{s,t+\tau}^{\widehat{\theta}_{t,N,L,h}} \le u\right\} - 1\{X_{t+\tau}\le u\}\right) 1\{X_t \le v\}. \end{aligned}$$

where $X_{s,t+\tau}^{\widehat{\theta}_{T,N,L,h}} = X_{s,t+\tau}^{\widehat{\theta}_{T,N,L,h}}(X_t, V_j^{1,\widehat{\theta}_{T,N,L,h}}, V_k^{2,\widehat{\theta}_{T,N,L,h}})$ and $X_{s,t+\tau}^{\widehat{\theta}^*_{t,N,L,h}} = X_{s,t+\tau}^{\widehat{\theta}^*_{t,N,L,h}}(X_t, V_j^{1,\widehat{\theta}^*_{t,N,L,h}}, V_k^{2,\widehat{\theta}^*_{t,N,L,h}})$.

The first order asymptotic validity of inference carried out using bootstrap statistics formed as outlined above follows immediately from BCS (2008). For testing decisions, one compares the test statistics $SZ_{T,S,h}$ and $MZ_{T,S,h}$ with the percentiles or the empirical distributions of SZ_T^* and $MZ_{T,S,h}^*$, respectively. Then, reject H_0 if the actual statistic is greater than the $(1-\alpha)th$-percentile of the empirical distribution of the bootstrap statistic, as in Section 3.4. Otherwise, do not reject H_0.

3.2.3 Predictive Density Tests for Multiple Competing Models

For illustration, we present the test for the stochastic volatility model (two factor model). Again, note that extension to other multi-factor models follows immediately. In particular, all steps in the construction of the test in the one factor model case carry through immediately to the stochastic volatility case with the statistic defined as:

$$DV_{P,L,S} = \max_{k=2,...,m} DV_{k,P,L,S}(u_1, u_2)$$

where

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

Critical values for these tests can be obtained using a recursive version of the block bootstrap. The corresponding bootstrap test statistic is:

$$DV^*_{P,L,S} = \max_{k=2,...,m} DV^*_{k,P,L,S}(u_1, u_2)$$

where

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

Of note is that we follow CS (2011) by adopting the recursive NPSQML estimator $\widehat{\theta}_{1,t,N,L,h}$ and $\widehat{\theta}_{k,t,N,L,h}$ for model 1 and k, respectively, as introduced in Section 3.3.4 with the choice $N = M = S$. $\widehat{\theta}^*_{1,t,N,L,h}$ and $\widehat{\theta}^*_{k,t,N,L,h}$ are bootstrap analogs of $\widehat{\theta}_{1,t,N,L,h}$ and $\widehat{\theta}_{k,t,N,L,h}$ respectively (see Section 3.4). In addition, we do not need to resample the volatility process, although volatility is simulated under both $\widehat{\theta}_{k,t,N,L,h}$ and $\widehat{\theta}^*_{k,t,N,L,h}$, $k = 1, ..., m$.

Repeat Steps 1-5 in the bootstrap procedure in Section 3.4 to obtain critical values. Compare $DV_{P,L,S}$ with the percentiles of the empirical distribution of $DV^*_{P,L,S}$, and reject H_0 if $DV_{P,L,S}$ is greater than the $(1-\alpha)th$-percentile. Otherwise, do not reject H_0. Again, in implementation, there is no need to re-estimate $\widehat{\theta}^*_{k,t,N,L,h}$ for each bootstrap replications if $P/R \to 0, P, T, R \to \infty$, ass parameter estimation error vanishes asymptotically in this case.

3.3 Model Simulation and Estimation

3.3.1 Simulating Data

Approximation schemes are used to obtain simulated distributions and simulated parameter estimators, which are needed in order to construct the tests statistics outlined in previous sections. We therefore devote the first part of this section to a discussion of the Milstein approximation schemes that have been used in CS (2005), BCS (2008) and CS (2011). Let L be the length of each simulation path and h be the discretization interval, $L = Qh$ and θ be a generic parameter in simulation expression. We consider three cases:

The pure diffusion process as in (10)*:*

$$\begin{aligned} X^\theta_{qh} - X^\theta_{(q-1)h} &= b(X^\theta_{(q-1)h}, \theta)h + \sigma(X^\theta_{(q-1)h}, \theta)\epsilon_{qh} \\ &\quad -\frac{1}{2}\sigma(X^\theta_{(q-1)h}, \theta)'\sigma(X^\theta_{(q-1)h}, \theta)h \\ &\quad +\frac{1}{2}\sigma(X^\theta_{(q-1)h}, \theta)'\sigma(X^\theta_{(q-1)h}, \theta)\epsilon^2_{qh}, \end{aligned}$$

where

$$\left(W_{qh} - W_{(q-1)h}\right) = \epsilon_{qh} \overset{iid}{\sim} N(0, h),$$

$q = 1, \ldots, Q$, with $\epsilon_{qh} \overset{iid}{\sim} N(0, h)$; and where σ' is the derivative of $\sigma(\cdot)$ with respect to its first argument. Hereafter, X^θ_{qh} denotes the values of the diffusion at time qh, simulated under generic θ, and with a discrete interval equal to h, and so is a fine grain analog of $X^\theta_{t,h}$.

The pure jump diffusion process without stochastic volatility as in (**??**)*:*

$$\begin{aligned} & X^\theta_{(q+1)h} - X^\theta_{qh} \\ = \; & b(X^\theta_{qh}, \theta)h + \sigma(X^\theta_{qh}, \theta)\epsilon_{(q+1)h} - \frac{1}{2}\sigma(X^\theta_{qh}, \theta)'\sigma(X^\theta_{qh}, \theta)h \end{aligned}$$

$$+\frac{1}{2}\sigma(X^\theta_{qh}, \theta)'\sigma(X^\theta_{qh}, \theta)\epsilon^2_{(q+1)h} - \lambda\mu_y h + \sum_{j=1}^{\mathcal{J}} y_j 1\left\{qh \le \mathcal{U}_j \le (q+1)h\right\}, \tag{12}$$

The only difference between this approximation and that used for the pure diffusion is the jump part. Note that the last term on the right-hand-side (RHS) of (12) is nonzero whenever we have one (or more) jump realization(s) in the interval $[(q-1)h, qh]$. Moreover, as neither the intensity nor the jump size is state dependent, the jump component can be simulated without any discretization error,

as follows. Begin by making a draw from a Poisson distribution with intensity parameter $\widehat{\lambda}\tau$, say $\mathcal{J}$. This gives a realization for the number of jumps over the simulation time span. Then, draw $\mathcal{J}$ uniform random variables over $[0, L]$, and sort them in ascending order so that $\mathcal{U}_1 \leq \mathcal{U}_2 \leq ... \leq \mathcal{U}_{\mathcal{J}}$. These provide realizations for the $\mathcal{J}$ jump times. Then, make $\mathcal{J}$ independent draws from ϕ, say $y_1, ..., y_{\mathcal{J}}$.

SV models without jumps as in (4) *(*using a generalized Milstein scheme):

$$\begin{aligned} X^{\theta}_{(q+1)h} &= X^{\theta}_{qh} + \widetilde{b}_1(X^{\theta}_{qh}, \theta)h + \sigma_{11}(V^{\theta}_{qh}, \theta)\epsilon_{1,(q+1)h} \\ &\quad + \sigma_{12}(V^{\theta}_{qh}, \theta)\epsilon_{2,(q+1)h} + \frac{1}{2}\sigma_{22}(V^{\theta}_{qh}, \theta)\frac{\partial \sigma_{12,k}(V^{\theta}_{qh}, \theta)}{\partial V}\epsilon^2_{2,(q+1)h} \\ &\quad + \sigma_{22}(V^{\theta}_{qh}, \theta)\frac{\partial \sigma_{11}(V^{\theta}_{qh}, \theta)}{\partial V}\int_{qh}^{(q+1)h}\left(\int_{qh}^{s} dW_{1,\tau}\right) dW_{2,s} \end{aligned} \tag{13}$$

$$\begin{aligned} V^{\theta}_{(q+1)h} &= V^{\theta}_{qh} + \widetilde{b}_2(V^{\theta}_{qh}, \theta)h + \sigma_{22}(V^{\theta}_{qh}, \theta)\epsilon_{2,(q+1)h} \\ &\quad + \frac{1}{2}\sigma_{22}(V^{\theta}_{qh}, \theta)\frac{\partial \sigma_{22}(V^{\theta}_{qh}, \theta)}{\partial V}\epsilon^2_{2,(q+1)h} \end{aligned} \tag{14}$$

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

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

The last terms on the RHS of (13) involve stochastic integrals and cannot be explicitly computed. However, they can be approximated, up to an error of order $o(h)$ by (see, for example, equation (3.7), pp. 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 $h \to 0$, $p \to \infty$.

Stochastic Volatility with Jumps

Simulation of sample paths of diffusion processes with stochastic volatility and jumps follows straightforwardly from the previous two cases. Whenever both intensity and jump size are not state dependent, a jump component can be simulated and added to either $X(t)$ and/or the $V(t)$ in the same manner as above. Extension to general multidimensional and multifactor models both with and without jumps also follows directly.

3.3.2 Simulating Distributions

In this section we sketch out methods used to construct $\tau-$step ahead simulated conditional distributions using simulated data. In applications, simulation techniques are needed when the functional form conditional distribution is unknown. We first illustrate the technique for one factor models and then discuss multifactor models.

One factor models:

Consider the one factor model as in (8). To estimate the simulated CDFs,

Step 1: Obtain $\widehat{\theta}_{T,N,h}$ (using the entire sample) or $\widehat{\theta}_{t,N,h}$ (recursive estimator) where $\widehat{\theta}_{T,N,h}$ and $\widehat{\theta}_{t,N,h}$ are estimators as discussed in Section 3.3.3 and 3.3.4.

Step 2: Under $\widehat{\theta}_{T,N,h}$ or $\widehat{\theta}_{t,N,h}$[20], simulate S paths of length τ, all having the same starting value, X_t. In particular, for each path $i = 1, ...S$ of length τ, generate $X_{i,t+\tau}^{\widehat{\theta}_{T,N,h}}(X_t)$ according to a Milstein schemes detailed in previous section, with $\theta = \widehat{\theta}_{T,N,h}$ or $\widehat{\theta}_{t,N,h}$. The errors used in simulation are $\epsilon_{qh} \overset{iid}{\sim} N(0,h)$, and $Qh = \tau$. ϵ_{qh} is assumed to be independent across simulations, so that $E(\epsilon_{i,qh}\epsilon_{j,qh}) = 0$,for all $i \neq j$ and $E(\epsilon_{i,qh}\epsilon_{i,qh}) = h$, for any i, j. In addition, as the simulated diffusion is ergodic, the effect of the starting value approaches zero at an exponential rate, as $\tau \to \infty$.

Step 3: If $\widehat{\theta}_{T,N,h}$ ($\widehat{\theta}_{t,N,h}$) is used, an estimate for the distribution, at time $t+\tau$, conditional on X_t, with estimator $\widehat{\theta}_{T,N,h}(\widehat{\theta}_{t,N,h})$, is defined as:

$$\widehat{F}_\tau(u|X_t, \widehat{\theta}_{T,N,h}) = \frac{1}{S}\sum_{i=1}^{S} 1\left\{X_{i,t+\tau}^{\widehat{\theta}_{T,N,h}}(X_t) \leq u\right\}$$

BCS (2008) show that if the model is correctly specified, then $\frac{1}{S}\sum_{i=1}^{S} 1\left\{X_{i,t+\tau}^{\widehat{\theta}_{T,N,h}}(X_t) \leq u\right\}$ provides a consistent estimate of the conditional distribution $F_\tau(u|X_t, \theta^\dagger) = \Pr\left(X_{t+\tau}^{\theta^\dagger} \leq u | X_t^{\theta^\dagger} = X_t\right)$.

Specifically, assume that $T, N, S \to \infty$. Then, for the case of SGMM estimator, if $h \to 0$, $T/N \to 0$, and $h^2 T \to 0$, $T^2/S \to \infty$, the following result holds for any X_t, $t \geq 1$, uniformly in u

$$\widehat{F}_\tau(u|X_t, \widehat{\theta}_{T,N,h}) - F_\tau(u|X_t, \theta^\dagger) \xrightarrow{pr} 0,$$

In addition, if the model is correctly specified (i.e. if $\mu(\cdot,\cdot) = \mu_0(\cdot,\cdot)$ and $\sigma(\cdot,\cdot) = \sigma_0(\cdot,\cdot)$) then:

$$\widehat{F}_\tau(u|X_t, \widehat{\theta}_{T,N,h}) - F_{0,\tau}(u|X_t, \theta_0) \xrightarrow{pr} 0,$$

Step 4: Repeat Steps 1-3 for $t = 1, ..., T-\tau$. This yields $T-\tau$ conditional distributions that are $\tau-$Steps ahead which will be used in the construction of the specification tests.

The CDF simulation in the case selection test of multiple models with recursive estimator is similar. For model k, let $\widehat{\theta}_{k,t,N,h}$ be the recursive estimator of "pseudo true" $\theta_k^\dagger$ computed using all

[20] Note that $N = L$ for the SGMM estimator while $N = M = S$ for NSQML estimator.

observations up to varying time t. Then, $X_{k,i,t+\tau}^{\widehat{\theta}_{k,t,N,h}}(X_t)$ is generated according to a Milstein schemes as in Section 3.3.1, with $\theta = \widehat{\theta}_{k,t,N,h}$ and the initial value X_t, $Qh = \tau$. The corresponding empirical distribution of the simulated series $X_{k,i,t+\tau}^{\widehat{\theta}_{k,t,N,h}}(X_t)$ can then be constructed. Under some regularity conditions,

$$\frac{1}{S}\sum_{i=1}^{S} 1\left\{u_1 \leq X_{k,i,t+\tau}^{\widehat{\theta}_{k,t,N,h}}(X_t) \leq u_2\right\} \xrightarrow{pr} F_{X_{k,t+\tau}^{\theta_k^\dagger}(X_t)}(u_2) - F_{X_{k,t+\tau}^{\theta_k^\dagger}(X_t)}(u_1),\ t = R, ..., T-\tau,$$

where $F_{X_{k,t+\tau}^{\theta_k^\dagger}(X_t)}(u)$ is the marginal distribution of $X_{t+\tau}^{\theta_k^\dagger}(X_t)$ implied by k model (i.e., by the model used to simulate the series), conditional on the (simulation) starting value X_t. Furthermore, the marginal distribution of $X_{t+\tau}^{\theta^\dagger}(X_t)$ is the distribution of $X_{t+\tau}$ conditional on the values observed at time t. Thus, $F_{X_{k,t+\tau}^{\theta_k^\dagger}(X_t)}(u) = F_k^\tau(u|X_t, \theta_k^\dagger)$.

Of important note is that in the simulation of $X_{k,i,t+\tau}^{\widehat{\theta}_{k,t,N,h}}(X_t)$, $i = 1, ..., S$, for each t, $t = R, ..., T-\tau$, we must use the same set of randomly drawn errors and similarly the same draws for numbers of jumps, jump times and jump sizes. Thus, we only allow for the starting value to change. In particular, for each $i = 1, ..., S$, we generate $X_{k,i,R+\tau}^{\widehat{\theta}_{k,R,N,h}}(X_R), ..., X_{k,i,T}^{\widehat{\theta}_{k,T-\tau,N,h}}(X_{T-\tau})$. This yields an SxP matrix of simulated values, where $P = T - R - \tau + 1$ refers to the length of the out-of-sample period. $X_{k,i,R+j+\tau}^{\widehat{\theta}_{k,R+j,N,h}}(X_{R+j})$ (at time $R + j + \tau$) can be seen as τ periods ahead value "predicted" by model k using all available information up to time $R + j_{R+j}$, $j = 1, ..., P$ (the initial value X_{R+j} and $\widehat{\theta}_{k,R+j,N,h}$ estimated using $X_1, ..., X_{R+j}$). The key feature of this setup is that it enables us to compare "predicted " τ periods ahead values (i.e. $X_{k,i,R+j+\tau}^{\widehat{\theta}_{k,R+j,N,h}}(X_{R+j})$) with actual values that are τ periods ahead (i.e., $X_{R+j+\tau}$), for $j = 1, ..., P$. In this manner, simulation based tests under *ex-ante* predictive density comparison framework can be constructed.

Multifactor model:

Consider the multi-factor model with a skeleton $\left(X_t, V_t^1, ..., V_t^d\right)'$ (e.g. stochastic mean, stochastic volatility models, stochastic volatility of volatility, etc.) where only the first element X_t is observed. For simulation of the CDF, the difficulty arises as we do not know the initial values of latent variables $(V_t^1, ..., V_t^d)'$ at each point in time. We generalize the simulation plan of BCS (2008) and Cai and Swanson (2011) to the case of d dimensions. Specifically, to overcome the initial value difficulty, a natural strategy is to simulate a long path of length L for each latent variable $V_t^1, ..., V_t^d$, use them to construct $X_{t+\tau}$ and the corresponding simulated CDF of $X_{t+\tau}$; and finally, we average out the volatility values. Note that there are L^d combinations of the initial values $V_t^1, ..., V_t^d$. For illustration, consider the case of stochastic volatility ($d = 1$) and the Chen three factor model as in (5) ($d = 2$), using recursive estimators.

For the case of stochastic volatility ($d = 1$), i.e. $(X_t, V_t)'$, the steps are as follows:

Step 1: Estimate $\widehat{\theta}_{t,N,L,h}$ using recursive SGMM or NSQML estimation methods.

Step 2. Using the scheme in (14) with $\theta = \widehat{\theta}_{t,N,L,h}$, generate the path $V_{qh}^{\widehat{\theta}_{t,N,L,h}}$ for $q = 1/h, ..., Qh$ with $Qh = L$ and hence obtain $V_{j}^{\widehat{\theta}_{t,N,L,h}}$ $j = 1, ...L$.

Step 3: Using schemes in (13), (14), simulate LxS paths of length τ, setting the initial value for the observable state variable to be X_t. For the initial values of unobserved volatility, use $V_{j,qh}^{\widehat{\theta}_{t,N,L,h}}$, $j = 1, ...L$ as retrieved in Step 2. Also, keep the simulated random innovations (i.e.$\epsilon_{1,qh}$,.$\epsilon_{1,qh}$, $\int_{qh}^{(q+1)h} \left(\int_{qh}^{s} dW_{1,\tau} \right) dW_{2,s}$) to be constant across each j and t. Hence, for each replication i,using initial values X_t and $V_{j,qh}^{\widehat{\theta}_{T,N,h}}$, we obtain $X_{j,i,t+\tau}^{\widehat{\theta}_{t,N,L,h}}(X_t)$ which is a τ - step ahead simulated value.

Step 4: Now the estimator of $F_\tau(u|X_t, \theta^\dagger)$ is defined as:

$$\widehat{F}_\tau(u|X_t, \widehat{\theta}_{t,N,L,h}) = \frac{1}{LS} \sum_{j=1}^{L} \sum_{i=1}^{S} 1 \left\{ X_{j,i,t+\tau}^{\widehat{\theta}_{t,N,h}}(X_t) \leq u \right\}$$

Note that, by averaging over the initial value of the volatility process, we have integrated out it's effect. In other words, $\frac{1}{S} \sum_{i=1}^{S} 1 \left\{ X_{j,i,t+\tau}^{\widehat{\theta}_{t,N,h}}(X_t) \leq u \right\}$ is an estimate of $F_\tau(u|X_t, V_{j,h}^{\widehat{\theta}_{t,N,h}}, \theta^\dagger)$.

Step 5: Repeat the Steps 1-4 for $t = 1, ..., T - \tau$. This yields $T - \tau$ conditional distributions that are $\tau-$steps ahead which will be used in the construction of the specification tests.

For three factor model ($d = 2$), i.e., (X_t, V_t^1, V_t^2), consider model (5), where $W_t = \left(W_t^1, W_t^2, W_t^3\right)$ are mutually independent standard Brownian motions.

Step 1: Estimate $\widehat{\theta}_{t,N,L,h}$ using SGMM or NSQML estimation methods.

Step 2*:* Given the estimated parameter $\widehat{\theta}_{t,N,L,h}$, generate the path $V_{qh}^{1,\widehat{\theta}_{t,N,L,h}}$ and $V_{ph}^{2,\widehat{\theta}_{t,N,L,h}}$ for $q, p = 1/h, ..., Qh$ with $Qh = L$ and hence obtain $V_{j}^{1,\widehat{\theta}_{T,N,L,h}}$,$V_{k}^{2,\widehat{\theta}_{T,N,L,h}}$ $j, k = 1, ..., L$.

Step 3: Given the observable X_t and the $L \times L$ simulated latent paths ($V_{j}^{1,\widehat{\theta}_{t,N,L,h}}$ and $V_{k}^{2,\widehat{\theta}_{t,N,L,h}}$ $j, k = 1, ..., L$) as the start values , we simulate τ-Step ahead $X_{t+\tau}^{\widehat{\theta}_{t,N,L,h}}(X_t, V_{j}^{1,\widehat{\theta}_{t,N,L,h}}, V_{k}^{2,\widehat{\theta}_{t,N,L,h}})$. Since the start values for the two latent variables are $L \times L$ length, so for each X_t we have N^2 path. Now to integrate out the initial effect of latent variables, form the estimate of conditional distribution as

$$\widehat{F}_{\tau,s}(u|X_t, \widehat{\theta}) = \frac{1}{L^2} \sum_{j=1}^{L} \sum_{k=1}^{L} 1 \left\{ X_{s,t+\tau}^{\widehat{\theta}_{t,N,L,h}}(X_t, V_{j}^{1,\widehat{\theta}_{t,N,L,h}}, V_{k}^{2,\widehat{\theta}_{t,N,L,h}}) \leq u \right\},$$

where s denotes the sth simulation.

Step 4*:* Simulate $X_{s,t+\tau}^{\widehat{\theta}_{t,N,L,h}}$ S times, that is, repeat Step 3 S times i.e. $s = 1, ..., S$. The estimate of $F_\tau(u|X_t, \theta^\dagger)$ is

$$\widehat{F}_\tau(u|X_t, \widehat{\theta}) = \frac{1}{S} \sum_{i=1}^{S} \widehat{F}_{\tau,s}(u|X_t, \widehat{\theta}_{T,N,h})$$

Step 5: Repeat the Steps 1-4 for $t = 1, ..., T - \tau$. This yields $T - \tau$ conditional distributions that are $\tau-$steps ahead which will be used in the construction of the specification tests.

As a final remark, for the case of multiple competing models, we can proceed similarly. In addition, in the next two subsections, we present the exactly identified simulated (recursive) general method of moments and recursive nonparametric simulated quasi-maximum likelihood estimators that can be used in simulating distributions as well as constructing test statistics described in Section 3.2. The bootstrap analogs of those estimators will be discussed in Section 3.4.

3.3.3 Estimation: (Recursive) Simulated General Method of Moments (SGMM) Estimators

Suppose that we observe a discrete sample (skeleton) of T observations, say $(X_1, X_2, ..., X_T)'$, from the underlying diffusion in (8). The (recursive) SGMM estimator $\widehat{\theta}_{t,L,h}$ with $1 \leq t \leq T$ is specified as:

$$\widehat{\theta}_{t,L,h} = \arg\min_{\theta\in\Theta} \left(\frac{1}{t}\sum_{j=1}^{t} g(X_j) - \frac{1}{L}\sum_{j=1}^{L} g(X_{j,h}^{\theta})\right)' W_t^{-1} \left(\frac{1}{t}\sum_{j=1}^{t} g(X_j) - \frac{1}{L}\sum_{j=1}^{L} g(X_{j,h}^{\theta})\right) \tag{15}$$

$$= \arg\min_{\theta\in\Theta} G_{t,L,h}(\theta)' W_t G_{t,L,h}(\theta), \tag{16}$$

where g is a vector of p moment conditions, $\Theta \subset \Re^p$ (so that we have as many moment conditions as parameters), and $X_{j,h}^{\theta} = X_{[Qjh/L]}^{\theta}$, with $L = Qh$ is the simulated path under generic parameter θ and with discrete interval h. $X_{j,h}^{\theta}$ is simulated using the Milstein schemes.

Note that in the above expression, in the context of the specification test $\widehat{\theta}_{t,L,h}$ is estimated using the whole sample, i.e. $t = T$. In the out of sample context, the recursive SGMM estimator $\widehat{\theta}_{t,L,h}$ is estimated recursively using the using sample from 1 up to t.

Typically, the p moment conditions are based on the difference between sample moments of historical and simulated data or, between sample moments and model implied moments, whenever the latter are known in closed form. Finally, W_t is the heteroskedasticity and autocorrelation (HAC) robust covariance matrix estimator, defined as

$$W_t^{-1} = \frac{1}{t}\sum_{\nu=-l_t}^{l_t} w_\nu \sum_{j=\nu+1+l_t}^{t-l_t} \left(g(X_j) - \frac{1}{t}\sum_{j=1}^{t} g(X_j)\right)\left(g(X_{j-\nu}) - \frac{1}{t}\sum_{j=1}^{t} g(X_j)\right)', \tag{17}$$

where $w_v = 1 - v/(l_T + 1)$. Further, the pseudo true value, $\theta^{\dagger}$, is defined to be:

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

where

$$G_{\infty}(\theta)' W_0 G_{\infty}(\theta) = p \lim_{L,T\to\infty, h\to 0} G_{T,L,h}(\theta)' W_T G_{T,L,h}(\theta);$$

and where $\theta^{\dagger} = \theta_0$, if the model is correctly specified.

In the above set up, the exactly identified case is considered rather than the overidentified (S)GMM. This choice guarantees that $G_\infty(\theta^\dagger) = 0$ even under misspecification, in the sense that the model differs from the underlying DGP. As pointed out by Hall and Inoue (2003), the root-N consistency does not hold for overidentified (S)GMM estimators of misspecified models. In addition,

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

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

Also note that other available estimation methods using moments include the efficient method of moments (EMM) estimator as proposed by Gallant and Tauchen (1996, 1997), which calculates moment functions by simulating the expected value of the score implied by an auxiliary model. In their setup, parameters are then computed by minimizing a chi-square criterion function.

3.3.4 Estimation: Recursive Nonparametric Simulated Quasi Maximum Likelihood Estimators

In this section we outline a recursive version of the NPSQML estimator of Fermanian and Salani´e (2004), proposed by CS (2011). The bootstrap counterpart of the recursive NPSQML estimator will be presented in the next section.

One factor models:

Hereafter, let $f\left(X_t|X_{t-1}, \theta^\dagger\right)$ be the conditional density associated with the above jump diffusion. If f is known in closed form, we can just estimate $\theta^\dagger$ recursively, using standard QML as:[21]

$$\widehat{\theta}_t = \arg\max_{\theta\in\Theta} \frac{1}{t} \sum_{j=2}^{t} \ln f\left(X_j|X_{j-1}, \theta\right), \; t = R, ..., R+P-1.. \tag{18}$$

Note that, similarly to the case of SGMM, the pseudo true value $\theta^\dagger$ is optimal in the sense:

$$\theta^\dagger = \arg\max_{\theta\in\Theta} E\left(\ln f\left(X_t|X_{t-1}, \theta\right)\right). \tag{19}$$

for the case f is not known in closed form, we can follow Kristensen and Shin (2008) and CS (2011) to construct the simulated analog $\widehat{f}$ of f and then use it to estimate $\theta^\dagger$. $\widehat{f}$ is estimated as function of the simulated sample paths $X^\theta_{t,i}(X_{t-1})$, for $t = 2, ..., T-1$, $i = 1, ..., M$. First, generate $T-1$ paths of length one for each simulation replication, using X_{t-1} with $t = 1, ...T$ as starting values. Hence, at time t and simulation replication i we obtain skeletons $X^\theta_{t,i}(X_{t-1})$, for $t = 2, ..., T-1$, $i = 1, ..., M$ where M is the number of simulation paths (number of random draws or $X^\theta_{t,j}(X_{t-1})$

[21] 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.

and $X_{t,l}^{\theta}(X_{t-1})$ are i.i.d.) for each simulation replication. M is fixed across all initial values. Then the recursive NPSQML estimator is defined as follows:

$$\widehat{\theta}_{t,M,h} = \arg\max_{\theta\in\Theta} \frac{1}{t}\sum_{i=2}^{t} \ln \widehat{f}_{M,h}\left(X_i|X_{i-1},\theta\right)\tau_M\left(\widehat{f}_{M,h}\left(X_i|X_{i-1},\theta\right)\right),\ t\geq R,$$

where

$$\widehat{f}_{M,h}\left(X_t|X_{t-1},\theta\right) = \frac{1}{M\xi_M}\sum_{i=1}^{M} K\left(\frac{X_{t,i,h}^{\theta}(X_{t-1}) - X_t}{\xi_M}\right).$$

Note that with abuse of notation, we define $\widehat{\theta}_{t,L,h}$ for SGMM and $\widehat{\theta}_{t,M,h}$ for NPSQML estimators where L and M have different interpretations (L is the length of each simulation path and M is number of random draws).

The function $\tau_M\left(\widehat{f}_{M,h}\left(X_t|X_{t-1},\theta\right)\right)$ is a trimming function. It has some characteristics such as positive and increasing, $\tau_M\left(\widehat{f}_{M,h}\left(X_t,X_{t-1},\theta\right)\right)=0$, if $\widehat{f}_{M,h}\left(X_t,X_{t-1},\theta\right)<\xi_M^{\delta}$, and $\tau_M\left(\widehat{f}_{M,h}\left(X_t,X_{t-1},\theta\right)\right)=1$, if $\widehat{f}_{M,h}\left(X_t,X_{t-1},\theta\right)>2\xi_M^{\delta}$, for some $\delta>0$.[22] Note that when the log density is close to zero, the derivative tends to infinity and thus even very tiny simulation errors can have a large impact on the likelihood. The introduction of the trimming parameter into the optimization function ensures the impact of this case to be minimal asymptotically.

Multifactor Models:

Since volatility is not observable, we cannot proceed as in the single factor case when estimating the SV model using NPSQML estimator. Instead, let V_j^{θ} be generated according to (14), setting $qh=j$, and $j=1,...,L$. The idea is to simulate L different starting values for unobservable volatility, construct the simulated likelihood functions accordingly and then average them out. For each simulation replication at time t, we simulate L different values of X_t (X_{t-1},V_j^{θ}) by generating L paths of length one, using fixed observable X_{t-1} and unobservable V_j^{θ}, $j=1,...,L$ as starting values. Repeat this procedure for any $t=1,...,T-1$, and for any set j, $j=1,...,L$ of random errors $\epsilon_{1,t+(q+1)h,j}$ and $\epsilon_{2,t+(q+1)h,j}$, $q=1,...,1/h$. Note that it is important to use the same set of random errors $\epsilon_{1,t+(q+1)h,j}$ and $\epsilon_{2,t+(q+1)h,j}$ across different initial values for volatility. Denote the simulated value at time t, simulation replication i, under generic parameter θ, using X_{t-1},V_j^{θ} as starting values as $X_{t,i,h}^{\theta}(X_{t-1},V_j^{\theta})$. Then:

$$\widehat{f}_{M,L,h}\left(X_t|X_{t-1},\theta\right) = \frac{1}{L}\sum_{j=1}^{L}\frac{1}{M\xi_M}\sum_{i=1}^{M} K\left(\frac{X_{t,i,h}^{\theta}(X_{t-1},V_j^{\theta}) - X_t}{\xi_M}\right),$$

[22] Fermanian and Salanie (2004) suggest using the following triming function:

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

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

and note that by averaging over the initial values for the unobservable volatility, its effect is integrated out. Finally, define:[23]

$$\widehat{\theta}_{t,M,L,h} = \arg\min_{\theta\in\Theta} \frac{1}{t}\sum_{s=2}^{t} \ln \widehat{f}_{M,L,h}\left(X_s|X_{s-1},\theta\right)\tau_M\left(\widehat{f}_{M,L,h}\left(X_s|X_{s-1},\theta\right)\right),\ t \geq R.$$

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 of true data generating process, the joint likelihood cannot be expressed as the product of the conditional and marginal distributions. Thus, $\widehat{\theta}_{t,M,L,h}$ is necessarily a QML estimator. Furthermore, note that $\nabla_\theta f(X_t|X_{t-1},\theta^\dagger)$ is no longer a martingale difference sequence; therefore, we need to use HAC robust covariance matrix estimators, regardless of whether the model is the "correct" model or not.

3.4 Bootstrap Critical Value Procedures

The test statistics presented in Section 3.1 and 3.2 are implemented using critical values constructed via the bootstrap. As mentioned earlier, motivation for using the bootstrap is clear. The covariance kernel of the statistics limiting distributions contain both parameter estimation error and the data related time dependence components. Asymptotic critical value cannot thus be tabulated in a usual way. Several methods have been proposed to tackle this issue. One is the block bootstrap procedures which we discuss. Others have been mentioned above.

With regarding to the validity of the bootstrap, note that, in the case of dependent observations without PEE, we can tabulate valid critical value using a simple empirical version of the Künsch (1989) block bootstrap. Now, the difficulty in our context lies in accounting for parameter estimation error. Goncalves and White (2002) establish the first order validity of the block bootstrap for QMLE (or m-estimator) for dependent and heterogeneous data. This is an important result for the class of SGMM and NSQML estimators surveyed in this paper, and allows Corradi and Swanson in CS (2011) and elsewhere to develop asymptotically valid version of the bootstrap that can be applied under generic model misspecification, as assumed throughout this paper.

For the SGMM estimator, as shown in CS (2005) the first order validity of the block bootstrap is valid in the exact identification case, and when $T/S \to 0$. In this case, SGMM is asymptotically equivalent to GMM, and consequently there is no need to bootstrap the simulated series. In addition, in the exact identification case, GMM estimators can be treated the same way that QMLE estimators are treated. For the NSQML estimator, CS (2011) point out that the NPSQML estimator is asymptotically equivalent to the QML estimator. Thus, we do not need to resample the simulated observations as the negligible contribution of simulation errors.

[23] For discussion of asymptotic properties of $\widehat{\theta}_{k,t,M,L,h}$, as well as of regularity conditions, see CS(2011).

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

Now the details of bootstrap procedure for critical value tabulation can be outlined in 5 steps as follows:

Step 1: Let $T = bl$, where b denotes the number of blocks and l denotes the length of each block. We first draw a discrete uniform random variable, I_1, that can take values $0, 1, ..., T-l$ with probability $1/(T-l+1)$. The first block is given by $X_{I_1+1}, ..., X_{I_1+l}$. We then draw another discrete uniform random variable, say I_2, and a second block of length l is formed, say $X_{I_2+1}, ..., X_{I_2+l}$. Continue in the same manner, until you draw the last discrete uniform say I_b, and so the last block is $X_{I_b+1}, ..., X_{I_b+l}$. Let's call the X_t^* the resampled series, and note that $X_1^*, X_2^*, ..., X_T^*$ corresponds to $X_{I_1+1}, X_{I_1+2}, ..., X_{I_b+l}$. Thus, conditional on the sample, the only random element is the beginning of each block. In particular

$$X_1^*, ..., X_l^*, X_{l+1}^*, ..., X_{2l}^*, X_{T-l+1}^*, ..., X_T^*,$$

conditional on the sample, can be treated as b *iid* blocks of discrete uniform random variables. For a simple illustration the link between the bootstrap sample and the original sample. Note that it can be shown that except a set of probability measure approaching zero,

$$E^*\left(\frac{1}{T}\sum_{t=1}^{T} X_t^*\right) = \frac{1}{T}\sum_{t=1}^{T} X_t + O_P^*(l/T) \tag{20}$$

$$\begin{aligned} Var^*\left(\frac{1}{T^{1/2}}\sum_{t=1}^{T} X_t^*\right) &= \frac{1}{T}\sum_{t=l}^{T-l}\sum_{i=-l}^{l}(X_t - \frac{1}{T}\sum_{t=1}^{T} X_t)(X_{t+i} - \frac{1}{T}\sum_{t=1}^{T} X_t) \\ &\quad + O_{P^*}(l^2/T), \end{aligned} \tag{21}$$

where E^* and Var^* denotes the expectation and the variance operators with respect to P^* (the probability law governing the resampled series or the probability law governing the *iid* uniform random variables, conditional on the sample), and where $O_{P^*}(l/T)$ ($O_{P^*}(l^2/T)$) denotes a term converging in probability P^* to zero, as $l/T \to 0$ ($l^2/T \to 0$).

In the case of recursive estimators, we proceed the bootstrap similarly as follows. 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 $Z^{t,*} = (X_t^*, X_{t+1}^*, ..., X_{t+\tau}^*)$, $t = 1, ..., T-\tau$.

Step 2: Re-estimate $\widehat{\theta}^*_{t,N,h}(\widehat{\theta}^*_{T,N,L,h})$ using the bootstrap sample $Z^{t,*} = (X_t^*, X_{t+1}^*, ..., X_{t+\tau}^*)$, $t = 1, ..., T-\tau$ (or full sample $X_1^*, X_2^*, ..., X_T^*$). Recall that if we use the entire sample for the estimation, as the specification test in CS(2005) and BCS(2008), then $\widehat{\theta}^*_{t,N,h}$ is denoted as $\widehat{\theta}^*_{T,N,h}$. The bootstrap estimators for SGMM and NPSQML are presented below:

Bootstrap (recursive) SGMM Estimators

If the full sample is used in the specification test as in CS (2005) and BCS(2008), the bootstrap estimator is constructed straightforward as

$$\widehat{\theta}^*_{T,L,h} = \arg\min_{\theta\in\Theta} \left(\frac{1}{T}\sum_{j=1}^{T} g(X_j^*) - \frac{1}{L}\sum_{i=1}^{L} g(X_{j,h}^{\theta})\right)' W_T^{*-1} \left(\frac{1}{T}\sum_{j=1}^{T} g(X_j^*) - \frac{1}{L}\sum_{i=1}^{L} g(X_{j,h}^{\theta})\right),$$

where W_T^{-1} and $g(.)$ are defined in (17) and L is the length of each simulation path.

Note that it is convenient not to resample the simulated series as the simulation error vanishes asymptotically. In implementation, we do not have mimic its contribution to the covariate kernel.

In the case of predictive density type model selection where recursive estimators are needed, define the bootstrap analog as

$$\begin{aligned}
\widehat{\theta}^*_{t,L,h} &= \arg\min_{\theta\in\Theta} \left(\frac{1}{t}\sum_{j=1}^{t}\left(\left(g(X_j^*) - \frac{1}{T}\sum_{j'=1}^{T} g(X_{j'})\right) - \left(\frac{1}{L}\sum_{i=1}^{L} g(X_{j,h}^{\theta}) - \frac{1}{L}\sum_{i=1}^{L} g(X_{j,h}^{\widehat{\theta}_{t,L,h}})\right)\right)\right)' \\
&\quad \Omega_t^{*-1} \left(\frac{1}{t}\sum_{j=1}^{t}\left(\left(g(X_j^*) - \frac{1}{T}\sum_{j'=1}^{T} g(X_{j'})\right) - \left(\frac{1}{L}\sum_{i=1}^{L} g(X_{j,h}^{\theta}) - \frac{1}{L}\sum_{i=1}^{L} g(X_{j,h}^{\widehat{\theta}_{t,L,h}})\right)\right)\right) \\
&= \arg\min_{\theta\in\Theta} G^*_{t,L,h}(\theta)'\Omega_t^{*-1} G^*_{t,L,h}(\theta),
\end{aligned}$$

where

$$\Omega_t^{*-1} = \frac{1}{t}\sum_{\nu=-l_t}^{l_t} w_{\nu,t} \sum_{j=\nu+1+l_t}^{t-l_t} \left(g(X_j^*) - \frac{1}{T}\sum_{j'=1}^{T} g(X_{j'})\right)\left(g(X_{j-\nu}^*) - \frac{1}{T}\sum_{j'=1}^{T} g(X_{j'})\right)$$

Note that each bootstrap term is recentered around the (full) sample mean. The intuition behind the particular recentering in bootstrap recursive SGMM estimator is that it ensures that the mean of the bootstrap moment conditions, evaluated at $\widehat{\theta}_{t,L,h}$ is zero, up to a negligible term. Specifically, we have

$$E^*\left(\frac{1}{t}\sum_{j=1}^{t}\left(g(X_j^*)-\frac{1}{T}\sum_{j'=1}^{T}g(X_{j'})\right)-\left(\frac{1}{L}\sum_{i=1}^{L}g(X_{j,h}^{\widehat{\theta}^*_{t,L,h}})-\frac{1}{L}\sum_{i=1}^{L}g(X_{j,h}^{\widehat{\theta}_{t,L,h}})\right)\right)$$

$$= E^*(g(X_j^*))-\frac{1}{T}\sum_{j'=1}^{T}g(X_{j'})=O(l/T),\ \text{with } l=o(T^{1/2}),$$

where the $O(l/T)$ term is due to the end block effect (see Corradi and Swanson (2007b) for further discussion).

Bootstrap Recursive NPSQML Estimators

Let $Z^{t,*}=(X_t^*,X_{t+1}^*,...,X_{t+\tau}^*)$, $t=1,...,T-\tau$. For each simulation replication, generate $T-1$ paths of length one, using $X_1^*,...,X_{T-1}^*$ as starting values, and so obtaining $X_{t,j}^{\theta}(X_{t-1}^*)$, for $t=2,...,T-1$, $i=1,...,M$. Further, let:

$$\widehat{f}^*_{M,h}\left(X_t^*|X_{t-1}^*,\theta\right)=\frac{1}{M\xi_M}\sum_{i=1}^{M}K\left(\frac{X_{t,i,h}^{\theta}(X_{t-1}^*)-X_t^*}{\xi_M}\right),$$

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

$$\begin{aligned}\widehat{\theta}^*_{t,M,h} &= \arg\max_{\theta\in\Theta}\frac{1}{t}\sum_{l=2}^{t}\left(\ln\widehat{f}_{M,h}\left(X_l^*|X_{l-1}^*,\theta\right)\tau_M\left(\widehat{f}_{M,h}\left(X_l^*|X_{l-1}^*,\theta\right)\right)\right.\\
&-\theta'\left(\frac{1}{T}\sum_{l'=2}^{T}\frac{\nabla_\theta\widehat{f}_{M,h}\left(X_{l'}|X_{l'-1},\theta\right)}{\widehat{f}_{M,h}\left(X_{l'}^*|X_{t-l'}^*,\theta\right)}\Big|_{\theta=\widehat{\theta}_{t,M,h}}\tau_M\left(\widehat{f}_{M,h}\left(X_{l'}|X_{l'-1},\widehat{\theta}_{t,M,h}\right)\right)\right.\\
&\left.\left.+\tau'_M\left(\widehat{f}_{M,h}\left(X_{l'}|X_{l'-1},\widehat{\theta}_{t,M,h}\right)\right)\nabla_\theta\widehat{f}_{M,h}\left(X_{l'}|X_{l'-1},\theta\right)\Big|_{\widehat{\theta}_{t,M,h}}\ln\widehat{f}_{M,h}\left(X_{l'}|X_{l'-1},\widehat{\theta}_{t,M,h}\right)\right)\right),\end{aligned}$$

where $\tau'_M(\cdot)$ denotes the derivative of $\tau_M(\cdot)$ 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 $\widehat{\theta}_{t,M,h}$. This ensures that the bootstrap score has mean zero, conditional on the sample. The recentering term requires computation of $\nabla_\theta\widehat{f}_{M,h}\left(X_{l'}|X_{l'-1},\widehat{\theta}_{t,M,h}\right)$, which is not known in closed form. Nevertheless, it can be computed numerically, by simply taking the numerical derivative of the simulated likelihood.

Bootstrap Estimators for Multifactor Model

The SGMM and the bootstrap SGMM estimators in the case of multifactor model are similar as in one factor model. The difference is that the simulation scheme (13) and (14) are used instead of (12).

For recursive NPSQML estimators, to construct the bootstrap counterpart $\widehat{\theta}^*_{t,M,L,h}$ of $\widehat{\theta}_{t,M,L,h}$, since $M/T\to\infty$ and $L/T\to\infty$, 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:

$$\widehat{f}_{M,L,h}\left(X_t^*|X_{t-1}^*,\theta\right)=\frac{1}{L}\sum_{j=1}^{L}\frac{1}{M\xi_M}\sum_{i=1}^{M}K\left(\frac{X_{t,i,h}^{\theta}(X_{t-1}^*,V_j^{\theta})-X_t^*}{\xi_M}\right).$$

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

$$
\begin{aligned}
&\widehat{\theta}^{*}_{t,M,L,h} \\
= \; & \arg\max_{\theta\in\Theta} \frac{1}{t}\sum_{l=2}^{t}\left(\log \widehat{f}_{t,M,L,h}\left(X^{*}_{l}|X^{*}_{l-1},\theta\right)\tau_{M}\left(\widehat{f}_{t,M,L,h}\left(X^{*}_{l}|X^{*}_{l-1},\theta\right)\right)\right. \\
& -\theta'\left(\frac{1}{T}\sum_{l'=2}^{T}\frac{\nabla_{\theta}\widehat{f}_{t,M,L,h}\left(X_{l'}|X_{l'-1},\theta\right)}{\widehat{f}_{t,M,L,h}\left(X^{*}_{l'}|X^{*}_{t-l'},\theta\right)}\Big|_{\theta_{t,M,L,h}}\tau_{M}\left(\widehat{f}_{t,M,L,h}\left(X_{l'}|X_{l'-1},\widehat{\theta}_{t,M,L,h}\right)\right)\right. \\
& \left.\left.+\tau'_{M}\left(\widehat{f}_{t,M,L,h}\left(X_{l'}|X_{l'-1},\widehat{\theta}_{t,M,L,h}\right)\right)\nabla_{\theta}\widehat{f}_{t,M,L,h}\left(X_{l'}|X_{l'-1},\theta\right)\Big|_{\widehat{\theta}_{t,M,L,h}}\ln\widehat{f}_{t,M,L,h}\left(X_{l'}|X_{l'-1},\widehat{\theta}_{t,M,L,h}\right)\right)\right),
\end{aligned}
$$

where $\tau'_{M}(\cdot)$ denotes the derivative with respect to its argument.

Of note is 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. In addition, in the case $P/R \to 0$, so that there is no need to mimic parameter estimation error, the bootstrap statistics can be constructed using $\widehat{\theta}_{t,M,L,h}$ instead of $\widehat{\theta}^{*}_{t,M,L,h}$.

Step 3: Using the same set of random variables used in the construction of the actual statistics, construct $X^{\widehat{\theta}^{*}_{t,N,h}}_{i,t+\tau,*}$ or $X^{\widehat{\theta}^{*}_{t,N,h}}_{k,i,t+\tau,*}$, $i = 1, ..., S$ and $t = 1, ..., T - \tau$. Note that we do not need resample the simulated series (as $L/T \to \infty$,simulation error is asymptotically negligible). Instead, simulate the series using bootstrap estimators and using bootstrapped values as starting values.

Step 4: Corresponding bootstrap statistics $V^{2*}_{T,N,h}$ (or $Z^{*}_{T,N,h}$, $D^{*}_{k,P,S}$, $SV^{2*}_{T,N,h}$, $SZ^{*}_{T,N,h}$, $SD^{*}_{k,P,S}$ depending on the types of tests) which are built on $\widehat{\theta}^{*}_{t,N,h}$ ($\widehat{\theta}^{*}_{t,N,L,h}$) then are followed correspondingly. For the numerical implementation, again, of importance note is that in the case where we pick the choice $P/R \to 0, P, T, R \to \infty$, there is no need to re-estimate $\widehat{\theta}^{*}_{t,N,h}(\widehat{\theta}^{*}_{t,N,L,h})$. $\widehat{\theta}_{t,N,h}$ ($\widehat{\theta}^{*}_{t,N,L,h}$) can be used in all the bootstrap replications.

Step 5: Repeat the bootstrap Steps 1-4 B times, and generate the empirical distribution of the B bootstrap statistics.

4 Summary of Empirical Applications of the Tests

In this section, we briefly review some empirical applications of the methods discussed above. We start with unconditional distribution test, as in CS (2005), then give a specific empirical example using the conditional distribution test from BCS (2008). Finally, we briefly discuss on conditional distribution specification test applied to multiple competing models. The list of the diffusion models considered are provided in Table 1.

Table 1: Specification Test Hypotheses of Continuous Time Spot Rate Process[24]

[24]Note that the 3rd column, "Reference & Data" provides the referenced papers and data used in empirical

Model	Specification	Reference & Data	Hypothesis
Wong[25]	$dr(t) = (\alpha - \lambda - r(t))dt + \sqrt{\alpha r(t)}dW_r(t)$	CS (2005) Simulated Data	H1
CIR	$dr(t) = k_r(\overline{r} - r(t))\,dt + \sqrt{V(t)}dW_r(t),$	BCS (2008) Eurodollar Rate (1971-2005)	H2
		Cai & Swanson (2011) Eurodollar Rate (1971-2008)	H2,H3
CEV	$dr(t) = k_r(\overline{r} - r(t))\,dt + \sigma_r r(t)^{\rho} dW_r(t)$	Cai & Swanson (2011) Eurodollar Rate (1971-2008)	H2,H3
SV[26]	$dr(t) = k_r(\overline{r} - r(t))\,dt + \sqrt{V(t)}dW_r(t),$ $dV(t) = k_v(\overline{v} - V(t))\,dt + \sigma_v\sqrt{V(t)}dW_v(t),$	BCS (2008) Cai & Swanson (2011)	H2 H2,H3
SVJ[27]	$dr(t) = k_r(\overline{r} - r(t))\,dt + \sqrt{V(t)}dW_r(t) + J_u dq_u - J_d dq_d,$ $dV(t) = k_v(\overline{v} - V(t))\,dt + \sigma_v\sqrt{V(t)}dW_v(t),$	BCS (2008) Cai & Swanson (2011)	H2 H2,H3
CHEN	$dr(t) = \kappa_r(\theta(t) - r(t))\,dt + \sqrt{V(t)}dW_r,$ $dV(t) = \kappa_v(\overline{v} - V(t))\,dt + \sigma_v\sqrt{V(t)}dW_v(t),$ $d\theta(t) = \kappa_\theta(\overline{\theta} - \theta(t))\,dt + \sigma_\theta\sqrt{\theta(t)}dW_\theta(t),$	Cai & Swanson (2011) Euro Dollar Rate (1971-2008)	H2, H3
CHENJ	$dr(t) = \kappa_r(\theta(t) - r(t))\,dt + \sqrt{V(t)}dW_r(t) + J_u dq_u - J_d dq_d,$ $dV(t) = \kappa_v(\overline{v} - V(t))\,dt + \sigma_v\sqrt{V(t)}dW_v(t),$ $d\theta(t) = \kappa_\theta(\overline{\theta} - \theta(t))\,dt + \sigma_\theta\sqrt{\theta(t)}dW_\theta(t),$	Cai & Swanson (2011 Euro Dollar Rate (1971-2008)	H2, H3

Note that specification testing of the first model - a simplified version of the CIR model (we refer to this model as Wong) is carried out using the unconditional distribution test. With the cumulative distribution function known in closed form as in (9), the test statistic can be straightforwardly calculated. It is also convenient to use GMM estimation in this case as the first two moments are known in closed form, i.e. $\alpha - \lambda$ and $\alpha/2(\alpha - \beta)$, respectively. CS (2005) examine Hypothesis 1 using simulated data. Their Monte Carlo experiments suggest that the test is useful, even for samples as small as 400 observations.

Hypothesis 2 is tested in BCS (2008) and Cai and Swanson (2011). For illustration, we focus on the results in BCS (2008) where CIR, SV and SVJ models are empirically tested using the one-month Eurodollar deposit rate (as a proxy for short rate) for the sample period January 6, 1971 - September 30, 2005, which yields 1,813 weekly observations. Note that one might apply these tests to other datasets including the monthly federal funds rate, the weekly 3-month T-bill rate, the weekly US dollar swap rate, the monthly yield on zero-coupon bonds with different maturities,

applications. In the 4th column, H1, H2 and H3 denote Hypothesis 1, Hypothesis 2 and Hypothesis 3, respectively. The hypotheses are presented corresponding to the references in the third column. For example, for CIR model, H2 corresponds to BCS (2008) and H2, H3 correspond to Cai and Swanson (2011).

[25] This model is a simplified version of the CIR model. For convenience, we refer to this model as Wong.

[26] Data used for Stochatic Volatility (SV) model is the same as in CIR Model.

[27] Data used for Stochatic Volatility and Jump (SVJ) model is the same as in CIR Model.

and the 6-month LIBOR. Some of these variables have been examined elsewhere, for example in Ait-Sahalia (1999), Andersen, Benzoni and Lund (2004), Dai and Singleton (2000), Diebold and Li (2006, 2007), and Piazzesi (2001).

The statistic needed to apply the test discussed in Section 3.1.2 is:

$$Z_T = \sup_{v \in V} |Z_T(v)|,$$

where

$$Z_T(v) = \frac{1}{\sqrt{T-\tau}} \sum_{t=1}^{T-\tau} \left(\frac{1}{S} \sum_{s=1}^{S} 1\left\{ \underline{u} \leq X_{s,t+\tau}^{\widehat{\theta}_{T,N,h}} \leq \overline{u} \right\} - 1\{\underline{u} \leq X_{t+\tau} \leq \overline{u}\} \right) 1\{X_t \leq v\};$$

and

$$Z_T^* = \sup_{v \in V} |Z_T^*(v)|,$$

where

$$\begin{aligned} Z_T^*(v) &= \frac{1}{\sqrt{T-\tau}} \sum_{t=1}^{T-\tau} \left(\frac{1}{S} \sum_{s=1}^{S} 1\left\{ \underline{u} \leq X_{s,t+\tau,*}^{\widehat{\theta}^*_{T,N,h}} \leq \overline{u} \right\} - 1\{\underline{u} \leq X_{t+\tau}^* \leq \overline{u}\} \right) 1\{X_t^* \leq v\} \\ &\quad - \frac{1}{\sqrt{T-\tau}} \sum_{t=1}^{T-\tau} \left(\frac{1}{S} \sum_{s=1}^{S} 1\left\{ \underline{u} \leq X_{s,t+\tau}^{\widehat{\theta}_{T,N,h}} \leq \overline{u} \right\} - 1\{\underline{u} \leq X_{t+\tau} \leq \overline{u}\} \right) 1\{X_t \leq v\}. \end{aligned}$$

For the case of stochastic volatility models, similarly we have:

$$SZ_T = \sup_{v \in V} |SZ_T(v)|,$$

where

$$SZ_T(v) = \frac{1}{\sqrt{T-\tau}} \sum_{t=1}^{T-\tau} \left(\frac{1}{LS} \sum_{j=1}^{L} \sum_{s=1}^{S} 1\left\{ \underline{u} \leq X_{j,s,t+\tau}^{\widehat{\theta}_{T,N,h}} \leq \overline{u} \right\} - 1\{\underline{u} \leq X_{t+\tau} \leq \overline{u}\} \right) 1\{X_t \leq v\};$$

and its bootstrap analog

$$SZ_T^* = \sup_{v \in V} |SZ_T^*(v)|,$$

where

$$\begin{aligned} SZ_T^*(v) &= \frac{1}{\sqrt{T-\tau}} \sum_{t=1}^{T-\tau} \left(\frac{1}{LS} \sum_{j=1}^{L} \sum_{s=1}^{S} 1\left\{ \underline{u} \leq X_{j,s,t+\tau,*}^{\widehat{\theta}^*_{i,T,N,h}} \leq \overline{u} \right\} - 1\{\underline{u} \leq X_{t+\tau}^* \leq \overline{u}\} \right) 1\{X_t^* \leq v\} \\ &\quad - \frac{1}{\sqrt{T-\tau}} \sum_{t=1}^{T-\tau} \left(\frac{1}{LS} \sum_{j=1}^{L} \sum_{s=1}^{S} 1\left\{ \underline{u} \leq X_{j,s,t+\tau}^{\widehat{\theta}_{i,T,N,h}} \leq \overline{u} \right\} - 1\{\underline{u} \leq X_{t+\tau} \leq \overline{u}\} \right) 1\{X_t \leq v\}. \end{aligned}$$

BCS (2008) carry out these tests using $\tau - Step$ ahead confidence intervals. They set $\tau = \{1, 2, 4, 12\}$ which corresponds to one week, two week, one month, and one quarter ahead intervals and set $(\underline{u}, \overline{u}) = (\overline{X} \pm 0.5\sigma_X, \overline{X} \pm \sigma_X)$, covering 46.3% and 72.4% coverage, respectively. $\overline{X}$ and σ_X are the mean and variance of an initial sample of data. In addition, $S = \{10T, 20T\}$ and $l = \{5, 10, 20, 50\}$.

For illustrative purposes, we report one case from BCS (2008). The test is implemented by setting $S = 10T$ and $l = 25$ for the calculation of both Z_T and SZ_T. In the Table 2, single, double, and triple starred entries represent rejection using $20\%, 10\%$, and 5% size tests, respectively. Not surprisingly, the findings are consistent with some other papers in the specification test literature such as such as Aït-Sahalia (1996) and Bandi (2002). Namely, the CIR model is rejected using 5% size tests in almost all cases. When considering SV and SVJ models, smaller confidence intervals appear to lead to more model rejections. Moreover, results are somewhat mixed when evaluating the SVJ model, with a slightly higher frequency of rejection than in the case of SV models.

Table 2: Empirical Illustration of Specification Testing - CIR, SV, SVJ Models

	$(\underline{u}, \overline{u})$	CIR			SV			SVJ		
		Z_T	5% CV	10% CV	SZ_T	5% CV	10% CV	SZ_T	5% CV	10% CV
					$l = 25$					
1	$\overline{X} \pm 0.5\sigma_X$	0.5274***	0.2906	0.3545	0.9841***	0.8729	0.9031	1.1319	1.8468	2.1957
	$\overline{X} \pm \sigma_X$	0.4289***	0.2658	0.3178	0.6870	0.6954	0.7254	1.2272*	1.1203	1.3031
2	$\overline{X} \pm 0.5\sigma_X$	0.6824***	0.4291	0.4911	0.4113	1.3751	1.4900	0.9615*	0.8146	1.1334
	$\overline{X} \pm \sigma_X$	0.4897*	0.4264	0.5182	0.3682	1.1933	1.2243	1.2571	1.3316	1.4096
4	$\overline{X} \pm 0.5\sigma_X$	0.8662**	0.7111	0.8491	1.2840	2.3297	2.6109	1.5012*	1.1188	1.6856
	$\overline{X} \pm \sigma_X$	0.8539*	0.7521	0.9389	1.0472	2.2549	2.2745	0.9901*	0.9793	1.0507
12	$\overline{X} \pm 0.5\sigma_X$	1.1631*	1.0087	1.3009	1.7687	4.9298	5.2832	2.4237*	2.0818	3.0640
	$\overline{X} \pm \sigma_X$	1.0429	1.4767	2.0222	1.7017	5.2601	5.6522	1.4522	1.7400	2.1684

(*) Notes: Tabulated entries are test statistics and 5%, 10% and 20% level critical values. Test intervals are given in the second column of the table, for τ =1, 2, 4, 12. All tests are carried out using historical one-month Eurodollar deposit rate data for the period January 1971 - September 2005, measured at a weekly frequency. Single, double, and triple starred entries denote rejection at the 20%,10%, and 5% levels, respectively. Additionally, $\overline{X}$ and σ_X are the mean and standard deviation of the historical data. See above for complete details.

Finally, turning to **Hypothesis 3,** Cai and Swanson (2011) use an extended version of the above dataset, i.e. the one-month Eurodollar deposit rate from January 1971 to April 2008 (1,996 weekly observations). Specifically, they examine whether the CHEN model is the "best" model amongst multiple alternative models including those outlined in Table 1. The answer is "yes". In this example, the test was implemented using $D_{k,p,N}(u1, u2)$, as described in Sections 3.1 and 3.2, where $P = T/2$ and predictions are constructed using recursively estimated models and the simulation sample length used to address latent variable initial values is set at $L = 10T$. The choice of other inputs to the test such as τ and interval $(\underline{u}, \overline{u})$ are the same as in BCS (2008). The number of replications S, the block length l and number of bootstrap replications are $S = 10T$, $l = 20$ and $B = 100$.

Cai and Swanson (2011) also compare the Chen model with the so called Smooth Transition Autoregression Model (STAR) model defined as follows:

$$r_t = (\theta_1 + \beta_1 r_{t-1})G(\gamma, z_t, c) + (\theta_1 + \beta_2 r_{t-1})(1 - G(\gamma, z_t, c)) + u_t$$

where u_t is a disturbance term, θ_1, β_1, γ , β_2, and c are constants, $G(\cdot)$ is the logistic CDF (i.e. $G(\gamma, z_t, c) = \frac{1}{1+\exp(\gamma(z_t - c))}$), and the number of lags, p is selected via the use of Schwarz information criterion. Test statistics and predictive density type "mean square forecast errors" (MSFEs) values are again calculated as in Section 3.1 and 3.2.[28] Their results indicate that at a 90% level of confidence, one cannot reject the null hypothesis that the CHEN model generates predictive densities at least as accurate as the STAR model, regardless of forecast horizon and confidence interval width. Moreover, in almost all cases, the CHEN model has lower MSFE, and the magnitude of the MSFE differential between the CHEN model and STAR model rises as the forecast horizon increases. This confirms their in-sample findings that the CHEN model also wins when carrying out in-sample tests.

5 Conclusion

This paper reviews a class of specification and model selection type tests developed by CS (2005), BCS (2008) and CS (2011) for continuous time models. We begin with outlining the setup used to specify the types of diffusion models considered in this paper. Thereafter, diffusion models in finance are discussed, and testing procedures are outlined. Related testing procedures are also discussed, both in contexts where models are assumed to be either correctly specified under the null hypothesis or generically misspecified under both the null and alternative test hypotheses. In addition to discussing tests of correct specification and test for selecting amongst alternative competing models, using both in-sample methods and via comparison of predictive accuracy, methodology is outlined allowing for parameter estimation, model and data simulation, and bootstrap critical value construction.

Several extensions that are left to future research are as follows. First, it remains to construct specification tests that do not integrate out the effects of latent factors. Additionally, it remains to examine the finite sample properties of the estimators and bootstrap methods discussed in this paper.

[28] See Table 6 in Cai and Swanson (2011) for complete details.

6 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., (2007), Estimating Continuous-Time Models Using Discretely Sampled Data, Econometric Society World Congress Invited Lecture, in Advances in Economics and Econometrics, Theory and Applications, Ninth World Congress, edited by Richard Blundell, Persson Torsten and Whitney K. Newey, Econometric Society Monographs, Cambridge University Press.

Aït-Sahalia, Y., J. Fan and H. Peng, (2009), Nonparametric Transition-Based Tests for Diffusions, *Journal of the American Statistical Association*, 104, 1102-1116.

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

Andersen, T.G., and J. Lund, (1997), Estimating Continuous Time Stochastic Volatility Models of Short Term Interest Rates, *Journal of Econometrics*, 72, 343-377.

Andersen, T.G., L. Benzoni and J. Lund, (2004), Stochastic Volatility, Mean Drift, and Jumps in the Short-Term Interest Rate, Working Paper, Northwestern University.

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

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

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

Bandi, F., (2002), Short-Term Interest Rate Dynamics: A Spatial Approach, *Journal of Financial Economics,* 65, 73-110.

Beckers, S. (1980), The Constant Elasticity of Variance Model and US Implications for Option Pricing. *Journal of Finance,* 35, 661-673.

Bhardwaj, G., V. Corradi and N.R. Swanson, (2008), A Simulation Based Specification Test for Diffusion Processes, *Journal of Business and Economic Statistics*, 26, 176-93.

Black, F. and M. Scholes, (1973), The pricing of options and corporate liabilities, *Journal of Political Economy*, 81, 637–654.

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

Brennan, M.J., and E.S. Schwartz, (1979), A Continuous-Time Approach to the Pricing Bonds. *Journal of Banking and Finance,* 3, 133-155.

Cai, L. and N.R. Swanson, (2011), An Empirical Assessment of Spot Rate Model Stability. *Journal of Empirical Finance*, 18, 743–764.

Chan, K.C., G. A. Karolyi, F. A. Longstaff and A. B. Sanders, (1992), An Empirical Comparison of Alternative Models of the Short-Term Interest Rate, *Journal of Finance,* 47, 1209-1227.

Chen, L, (1996), Stochastic Mean and Stochastic Volatility - A Three-Factor Model of Term Structure of Interest Rates and its Application to the Pricing of Interest Rate Derivatives, Blackwell Publishers, Oxford, UK.

Chen, B. and Y. Hong, (2005), Diagnostics of Multivariate Continuous-Time Models with Application to Affine Term Structure Models, Working Paper, Cornell University.

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

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

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

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

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

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

Corradi, V. and N.R. Swanson, (2011), Predictive Density Construction and Accuracy Testing with Multiple Possibly Misspecified Diffusion Models, *Journal of Econometrics,* 161,304 - 324..

Courtadon, G., (1982), The Pricing of Options on Default-Free Bonds. *Journal of Financial and Quantitative Analysis*, 17, 75-100.

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

Dai, Q. and Kenneth J. Singleton, (2000), Specification Analysis of Affine Term Structure Models, *Journal of Finance,* 55, 1943-1978.

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

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

Diebold, F.X. and C. Li, (2006), Forecasting the Term Structure of Government Bond Yields, *Journal of Econometrics*, 130, 337-364.

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

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

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

Duffie, D. and R. Kan, (1996), A Yield-Factor Model of Interest Rates. *Mathematical Finance,* 6:379-406.

Duffie, D., J. Pan, and K. Singleton, (2000), Transform Analysis and Asset Pricing for Affine Junp-Diffusions. *Econometrica*, 68, 1343–1376.

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

Duong, D., and N.R. Swanson, (2010), Empirical Evidence on Jumps and Large Fluctuations in Individual Stocks. Working Paper, Rutgers University.

Duong, D., and N.R. Swanson, (2011), Volatility in Discrete and Continuous Time Models: A Survey with New Evidence on Large and Small Jumps, *Missing Data Methods: Advances in Econometrics,* 27B, 179-233.

Emanuel, D. and J. Mcbeth, (1982), Further Results on the Constant Elasticity of Variance Call Option Pricing Model, *Journal of Financial Quantitative Analysis,* 17, 533-554.

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.

Gallant, A.R., and G. Tauchen (1997), Estimation of Continuous Time Models for Stock Returns and Interest Rates, *Macroeconomic Dynamics,* 1, 135-168.

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

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

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

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

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

Hong, Y.M., and H.T. 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.T. Li, and F. Zhao, (2007), Can the random walk model be beaten in out-of-sample density forecasts? Evidence from intraday foreign exchange rates, *Journal of Econometrics*, 141, 736-776.

Hull, J., and A. White, (1987), The Pricing of Options on Assets with Stochastic Volatility, *Journal of Finance*, 42, 281-300.

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

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.

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

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

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

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.

Marsh, T. and E. Rosenfeld, (1983), Stochastic Processes for Interest Rates and Equilibrium Bond Prices. *Journal of Finance,* 38, 635-646.

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

Merton, C.R., (1973), Theory of Rational Option Pricing, *The Bell Journal of Economics and Management Science,* 4, 141-183

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

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

Piazzesi, M., (2001), Macroeconomic jump effects and the yield curve, working paper, UCLA.

Piazzesi, M., (2004), Affine Term Structure Models, Manuscript prepared for the Handbook of Financial Econometrics, University of California at Los Angeles.

Piazzesi, M., (2005), Bond Yields and the Federal Reserve, *Journal of Political Economy,* 113, 311-344.

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

Pritsker, M., (1998), Nonparametric Density Estimators and Tests of Continuous Time Interest Rate Models, *Review of Financial Studies*, 11, 449-487.

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

Singleton, K.J. (2006), **Empirical Dynamic Asset Pricing - Model Specification and Econometric Assessment**, Princeton University Press: Princeton and Oxford.

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

Sullivan R., A. Timmermann, and H. White, (1999), Data-Snooping, Technical Trading Rule Performance, and the Bootstrap, *The Journal of Finance*, vol. 54, issue 5, 1647-1691.

Sullivan, R., A. Timmermann, and H. White, (2001), Dangers of data-driven inference: The case of calendar effects in stock returns, *Journal of Econometrics*, 249– 286.

Thompson, S.B., (2008), Identifying Term Structure Volatility from the LIBOR-swap Curve, *Review of Financial Studies*, 21, 819-854.

Vasicek, O. A., (1977), An Equilibrium Characterization of the Term Structure, *Journal of Financial Economics,* 5, 177-188.

Vuong, Q., (1989), Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses, *Econometrica*, 57, 307-333

White, H., (1982), Maximum Likelihood Estimation of Misspecified Models, *Econometrica*, 50, 1-25

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

Wong, E., (1960), The Construction of a Class of Stationary Markov Processes, in Sixtenn Symposia in Applied Mathematics: Stochastic Processes in Mathematical Physics and Engineering, ed. by R. Bellman, American Mathematical Society, Providence, R.I.