In- and Out-of-Sample Specification Analysis of Spot Rate Models: Further Evidence for the Period 1982-2008*

Lili Cai[1] and Norman R. Swanson[2]

[1]Shanghai Jiao Tong University and [2]Rutgers University

May 2011

Abstract

We review and construct consistent in-sample specification and out-of-sample model selection tests on conditional distributions and predictive densities associated with continuous multifactor (possibly with jumps) and (non)linear discrete models of the short term interest rate. The results of our empirical analysis are used to carry out a "horse-race" comparing discrete and continuous models across multiple sample periods, forecast horizons, and evaluation intervals. Our evaluation involves comparing models during two distinct historical periods, as well as across our entire weekly sample of Eurodollar deposit rates from 1982-2008. Interestingly, when our entire sample of data is used to estimate competing models, the "best" performer in terms of distributional "fit" as well as predictive density accuracy, both in-sample and out-of-sample, is the three factor Chen (CHEN: 1996) model examined by Andersen, Benzoni and Lund (2004). Just as interestingly, a logistic type discrete smooth transition autoregression (STAR) model is preferred to the "best" continuous model (i.e. the one factor Cox, Ingersoll, and Ross (CIR: 1985) model) when comparing predictive accuracy for the "Stable 1990s" period that we examine. Moreover, an analogous result holds for the "Post 1990s" period that we examine, where the STAR model is preferred to a two factor stochastic mean model. Thus, when the STAR model is parameterized using only data corresponding to a particular sub-sample, it outperforms the "best" continuous alternative during that period. However, when models are estimated using the entire dataset, the continuous CHEN model is preferred, regardless of the variety of model specification (selection) test that is carried out. Given that it is very difficult to ascertain the particular future regime that will ensue when constructing ex ante predictions, thus, the CHEN model is our overall "winning" model, regardless of sample period.

Keywords: interest rate, multi-factor diffusion process, specification test, out-of-sample forecasts, conditional distribution, model selection,block bootstrap, jump process
JEL Classification: C1, C5, G0.

---

* Lili Cai, Department of Finance, Shanghai Jiao Tong University, 535 Fahuazhen Road, Shanghai, 200052, China, lilycc@gmail.com. Norman R. Swanson, Department of Economics, Rutgers University, 75 Hamilton Street, New Brunswick, NJ 08901, USA, nswanson@econ.rutgers.edu. The authors owe many thanks to the editor, Franz Palm, and to three anonymous referees, all of whom provided numerous useful comments and suggestions on earlier versions of this paper. The authors are also grateful to Torben Andersen, Valentina Corradi, Frank Diebold, Walter Distaso, Elena Goldman, Roger Klein, John Landon Lane, and George Tauchen for helpful comments on earlier version of the paper. The authors also own many thanks to seminar participants at the University of Chicago and Rutgers University for numerous useful suggestions. Swanson gratefully acknowledges financial support from a Rutgers University Research Council grant.

1 Introduction

Diffusion processes are used in virtually all aspects of continuous time finance from yield curve to exchange rate modeling, and for the purposes of prediction, simulation and pricing. This has led to many papers recently being published in the field, numerous of which are a part of an ongoing effort to specify models that adequately capture the dynamics of financial variables across reasonable spans of time, rather than across specific historical episodes. In this paper we first review recent methodological advances in the area of specification and predictive accuracy testing, and subsequently undertake a specification search of alternative short rate models, thereby adding to the rich literature begun by the key research of Chan, Karolyi, Longstaff and Sanders (CKLS: 1992). Our search focuses on a variety of multi factor continuous models both with and without jumps as well as simple and nonlinear discrete models.

One characteristic of continuous time models that is crucial to the application of such models is that only a few of those currently in use by practitioners have closed form solutions (see e.g. Vasicek model (1977), Cox, Ingersoll, and Ross (CIR: 1985), Black and Scholes (1973), and Hull and White (1990)). Indeed, many do not have closed form solutions, particularly those involving one or multiple latent variables (see e.g. the stochastic mean model of Balduzzi et al. (1998), the stochastic volatility model of Heston (1993), the three-factor model of Chen (1996), and the three-factor model with jumps discussed in the noteworthy paper by Andersen, Benzoni and Lund (ABL: 2004)). This issue has implications not only for pricing formulae derived from these models, but also for estimation. In recent years, many new methods have been developed for the estimation of continuous time models and the (often unknown in closed form) conditional densities associated with them. For example, Aït-Sahalia (1999, 2002, 2008) provides closed form approximations of (unknown) conditional densities using Hermite polynomials, for one-factor, stochastic volatility, and multi-factor models, respectively. These and other approximations (as well as general work on conditional Kolmogorov testing - see e.g. Andrews (1997) and Corradi and Swanson (2005a)) have led to the development of numerous consistent specification tests for evaluating individual models. Some of the earliest key papers on "goodness of fit" testing of continuous time models include those by Stanton (1997), Conley, Hansen, Luttmer, and Scheinkman (1997), Jiang (1998), and Jones(2003). Many specification tests for continuous models fall within one of two different categories. One category focuses on nonparametric tests. For example, tests characterized by comparing model implied transition densities with their nonparametric estimated (e.g. using kernels) counterparts (see e.g. Aït-Sahalia(1996, 2002), Aït-Sahalia, Fan and Peng (2009)); and tests involving the examination of generalized cross spectra (see e.g. Hong and Li (2005) and Chen and Hong (2008)). Another category that includes papers by Gallant and Tauchen (1997), Andersen and Lund (1997), Dai and Singleton (2000), Ahn, Dittmar and Gallant (2002), ABL (2004), Thompson (2008), Aït-Sahalia and Kimmel (2007), and Corradi and Swanson (2005a), to name but a few, who use parametric methods to examine the "goodness of fit" of models. The testing approaches reviewed and used in this paper falls within this category. Namely, we review, extend and implement the simulation based test for the correct specification of a diffusion process due to Bhardwaj, Corradi and Swanson (BCS: 2008). This test is in the spirit of the conditional Kolmogorov test of Andrews (1997). In addition, we discuss a simple extension to the test of Corradi and Swanson (CS: 2011) for comparing the accuracy of predictive densities derived from (possibly misspecified) diffusion models. These tests are continuous time generalizations of the discrete time, point mean square forecast error, model selection test statistics of White (2000) which are widely used in empirical finance (see e.g. Sullivan, Timmermann and White (1999, 2001)).

It should be noted that the tests used in this paper are also closely related to the interesting nonparametric specification tests of Hong (2002), Hong, Li and Zhao (2004, 2007), and Chen and Hong (2008), some of which are based upon use of the conditional characteristic function (ccf) in conjunction with the generalized cross spectrum. Our in-sample specification test is in the same spirit as these tests. Both, for example, are motivated by the classical Kolmogorov-Smirnov test, and our test along with many of their tests do not require a closed form solution for the transition density. However, our tests converge at a parametric rate while theirs converge at nonparametric rates. Moreover, our out-of-sample predictive density type model selection tests have the added feature that estimation is recursive, parameter estimation error does not vanish asymptotically and is explicitly accounted for, and multiple (possibly misspecified) models are jointly compared.

Of further note is that the difference between our approaches to in-sample (and out-of-sample) specification testing (and predictive density type model selection) and that taken elsewhere can be easily motivated within the framework used by Diebold, Gunther and Tay (DGT: 1998), Bai (2003), Hong (2002) and Hong, Li and Zhao (2004). In their paper, DGT use the probability integral transform (see e.g. Rosenblatt (1952)) to show that $F_t(y_t|\Im_{t-1},\theta_0)$, is identically and independently distributed as a uniform random variable on $[0,1]$, where $F_t(\cdot|\Im_{t-1},\theta_0)$ is a parametric distribution with underlying parameter θ_0, y_t is the random variable of interest, and $\Im_{t-1}$ is the information set containing all "relevant" past information (see below for further discussion). 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, Hahn and Tay (1998)), as well as for macroeconomic forecasting (see e.g. Diebold, Tay and Wallis (1998) and Clements and Smith (2000, 2002)). Likewise, Bai (2003) proposes a Kolmogorov type test of $F_t(u|\Im_{t-1}, \theta_0)$ based on the comparison of $F_t(y_t|\Im_{t-1}, \widehat{\theta}_T)$ with the CDF of a uniform on $[0, 1]$. As a consequence of using estimated parameters, the limiting distribution of his test reflects the contribution of parameter estimation error and is not nuisance parameter free. To overcome this problem, Bai (2003) uses 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. Two features differentiate our approach from that taken in the above papers. First, we assume strict stationarity, while they do not. Second, we allow for dynamic misspecification under the null hypothesis, while they do not. While our approach is clearly less general because of the first feature, 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, 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 relevant 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, we do not require such pre-testing. Another key feature of our approach concerns the fact that the limiting distribution of Kolmogorov type tests is affected by dynamic misspecification. 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 $y_t|y_{t-1}$ is $N(\alpha_1^\dagger y_{t-1}, \sigma_1)$. 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 $y_t|\Im_{t-1} = y_t|y_{t-1}, y_{t-2} = N(\alpha_1 y_{t-1} + \alpha_2 y_{t-2}, \sigma_2)$, where $\alpha_1^\dagger$ differs from α_1. In this case, we have correct specification with respect to the information contained in y_{t-1}; but we have dynamic misspecification with respect to y_{t-1}, 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 H_0) 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 our approach is appropriate.

One feature of the current literature is that the application of different specification tests and different data sets have led to a variety of different conclusions. For example, Aït-Sahalia (1996) test fails to reject the CKLS (1992) model and the nonlinear drift model of Aït-Sahalia (1996). On the other hand, Hong and Li (2005) strongly reject all univariate affine models of the Euro dollar rate, and suggest that even very sophisticated models (including GARCH, regime switching, and jumps) do not adequately capture interest rate dynamics. BCS (2008) reject the CIR (1985) model, and conclude that stochastic volatility models are superior to the CIR model. To some extent, one might argue that the mixed evidence in the extant literature can be attributed to the fact that numerous analyses have been carried out using (relatively) small numbers of models and varying data samples, many of which can be tied to particular historical "episodes". In light of this, we examine multiple time periods and a relatively rich set of models in this paper. We then compare our findings based on multiple sample periods with those based on analysis of our entire sample period of Eurodollar deposit rates from 1982-2008.

Interestingly, when our entire sample of data is used to estimate competing models, the "best" performer in terms of distributional "fit" as well as predictive density accuracy, both in-sample and out-of-sample, is the three factor Chen (CHEN: 1996) model examined by ABL (2004). This model is selected from a group of continuous time models including one, two, and three factor variants (with and without jumps), a simple discrete AR(p) benchmark model with lags (p) selected recursively using the Schwarz information criterion, and a nonlinear (logistic type) smooth transition autoregression (STAR) model. Just as interestingly, the STAR model is preferred to the "best" continuous model (i.e. the one factor CIR (1985) model) when comparing predictive accuracy for the "Stable 1990s" period that we examine. Moreover, an analogous result holds for the "Post 1990s" period that we examine, where the STAR model is preferred to a two factor stochastic mean model. Thus, when the STAR model is parameterized using only data corresponding to a particular sub-sample, it outperforms the "best" continuous alternative during that period. However, when models are estimated using the entire dataset, the continuous CHEN model is preferred, regardless of the variety of model specification (selection) test that is carried out. Given that it is very difficult to ascertain the particular future regime that will ensue when constructing ex ante predictions, thus, the CHEN model is our overall "winning" model, regardless of sample period. These findings are largely in agreement with the caveat pointed out by Hong, Li and Zhao (2004) that a model which fits historical data well is not guaranteed to have better out-of-sample performance. Indeed, this feature of empirical models is one of the main reasons why we carry out both in-sample and out-of-sample tests in our analysis.

In summary, our empirical analysis suggests that continuous time models do actually "hold their own" against discrete time models, even in forecasting competitions involving comparison of predictive density accuracy. Thus, the oft asked question of whether continuous time models that impose certain critical restrictions (such as no arbitrage) actually compare favorably with discrete time models is answered in the affirmative in our experiments. That said, it should be stressed that further research comparing our continuous time models with even more dynamically complex discrete models remains to be undertaken.

The rest of the paper is organized as follows. In Section 2, we review recent in-sample and out-of-sample specification tests that are used in the sequel. Section 3 summarizes the short term interest rate models that we examine. Finally, Section 4 discusses the data used in our experiments and empirical results are gathered in Section 5. Concluding remarks are given in Section 6.

2 Specification Tests and Model Selection

Our testing methodology follows that of BSC (2008) and CS (2011). In particular, we first implement in-sample consistent specification tests that due to BCS (2008). These tests, including a generalization to the case of three factor models, are Kolmogorov type tests utilizing a simple simulation based approach to construct conditional distributions when the functional form of the conditional density is unknown. The distributions are in turn used to form predictive confidence intervals for time period $t + \tau$, given information up to period t. The specification test measures the difference between simulated conditional confidence distributions generated using fitted parametric models and empirical conditional distributions implied by historical data. This approach is closely related to the specification testing approach discussed in Corradi and Swanson (2005a), Aït-Sahalia (1996, 2006), Hong and Li (2005) and Hong, Li and Zhao (2007). As opposed to the above approach of testing each individual model to ascertain whether it is correctly specified, we also implement out-of-sample model selection type tests due to CS (2011) that jointly compare all models based on predictive accuracy. These tests are in the spirit of the predictive accuracy tests due to Diebold and Mariano (1995) and White (2000), and involve using a simulation based approach closely related to the approach used in BCS (2008). Of note is that all models in both types of tests carried out in this paper are (possibly) misspecified. For a complete discussion of the importance of model misspecification in the case of specification testing and model selection, please refer to Corradi and Swanson (2006a). In the following sub-sections, we briefly discuss the in-sample and out-of-sample tests discussed above.

2.1 A Consistent Specification Test

The BCS (2008) test is based on the comparison of the empirical cumulative distribution function (CDF) and the cumulative distribution function implied by the specification of the drift and the variance under a given null model. To illustrate the idea, consider a parametric diffusion process:

$$dX_t = b(X_t, \theta^{\dagger})dt + \sigma(X_t, \theta^{\dagger})dW_t, \tag{1}$$

where W_t is a Brownian motion, and the true parameter vector is $\theta_0 = (b_0', \sigma_0')' \in \Theta$, Θ is a compact subset of $\Re^K$. Thus, we are assuming that X_t is a one-dimensional diffusion process solution to the above stochastic differential equation, where the true parameters, θ_0, have been replaced by their pseudo true analogs, $\theta^{\dagger}$. The discrete analog to the above process is a model in which X_t is a scalar. Multi-dimensional diffusion processes are discussed later.

Under correct specification of the diffusion process, we have that $b(\cdot,\cdot) = b_0(\cdot,\cdot)$ and $\sigma(\cdot,\cdot) = \sigma_0(\cdot,\cdot)$, that is $\theta^{\dagger} = \theta_0$. Note that the stationary density, $f(x, \theta^{\dagger})$, and its associated invariant probability measure are uniquely determined by $b(\cdot)$ and $\sigma^2(\cdot)$ (the drift and variance terms in the model). The alternative hypothesis is that the parameters in the above diffusion process do not coincide with the true parameters. Instead of comparing transition densities directly (see Aït-Sahalia et al. (2009)), we compare the cumulative distribution function. The null and alternative hypotheses are:

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

To construct a specification test, we follow BCS (2008) by defining the $\tau - step$ ahead conditional distribution of $X_{t+\tau}^{\theta^{\dagger}}$, given $X_t^{\theta^{\dagger}} = X_t$, as:

$$F_{\tau}(u|X_t, \theta^{\dagger}) = \Pr\left(X_{t+\tau}^{\theta^{\dagger}} \leq u | X_t^{\theta^{\dagger}} = X_t\right), \tag{2}$$

where $t = 1, 2, 3, ..., T - \tau$. Instead of comparing $F_{\tau}(u|X_t, \theta^{\dagger})$ and $F_{0,\tau}(u|X_t, \theta_0)$, we need to replace $F_{\tau}(u|X_t, \theta^{\dagger})$ with its simulated counterpart. Namely:

$$\widehat{F}_{\tau}(u|X_t, \widehat{\theta}_{T,N,h}) = \frac{1}{S}\sum_{s=1}^{S} 1\left\{X_{s,t+\tau}^{\theta_{T,N,h}} \leq u\right\}, \tag{3}$$

where $\widehat{\theta}_{T,N,h}$ is estimated by using the whole sample of T observations. Here, $\widehat{\theta}_{T,N,h}$ converges to $\theta^{\dagger}$, and S is the number of simulation paths. BCS (2008) show that $\frac{1}{S}\sum_{s=1}^{S} 1\left\{X_{s,t+\tau}^{\theta_{T,N,h}} \leq u\right\}$

is a consistent estimate of $F_\tau(u|X_t,\theta^\dagger)$. Moreover, under the null hypothesis of correct specification, $\frac{1}{S}\sum_{s=1}^{S}1\left\{X_{s,t+\tau}^{\theta_{T,N,h}}\leq u\right\}$ is also a consistent estimator of $F_{0,\tau}(u|X_t,\theta_0)$. Thus, we can implement a simulation-based version of the conditional Kolmogorov tests of Andrews (1997), in which we compare the joint empirical distribution $\frac{1}{T-\tau}\sum_{t=1}^{T-\tau}1\{X_{t+\tau}\leq u\}1\left\{X_t\leq v\right\}$ with its semi-empirical/semi-parametric analog given by the product of $\frac{1}{T-\tau}\sum_{t=1}^{T-\tau}F_{0,\tau}(u|X_t,\theta_0)1\left\{X_t\leq v\right\}$. Intuitively, if the null model used for simulating the data is correct, then the difference between the two approaches zero, and has a well-defined limiting distribution when properly scaled.

Namely, in order to test the above hypotheses, we measure the departure from the null hypothesis by defining the test statistic $Z_T=\sup_{u\times v\in U\times V}\left|Z_T(u,v)\right|$, where

$$Z_T(u,v)=\frac{1}{\sqrt{T-\tau}}\sum_{t=1}^{T-\tau}\left(\frac{1}{S}\sum_{s=1}^{S}1\left\{X_{s,t+\tau}^{\theta_{T,N,h}}\leq u\right\}-1\{X_{t+\tau}\leq u\}\right)1\left\{X_t\leq v\right\}, \tag{4}$$

and U and V are compact sets on the real line. BCS outline block-bootstrap methods for constructing critical values for this test. Specifically, the bootstrap statistic is $Z_T^*=\sup_{u\times v\in U\times V}\left|Z_T^*(u,v)\right|$, where

$$\begin{aligned}Z_T^*(u,v)=&\frac{1}{\sqrt{T-\tau}}\sum_{t=1}^{T-\tau}\left(\frac{1}{S}\sum_{s=1}^{S}1\left\{X_{s,t+\tau}^{\theta^*_{T,N,h}}\leq u\right\}-1\{X^*_{t+\tau}\leq u\}\right)1\left\{X^*_t\leq v\right\}\\&-\frac{1}{\sqrt{T-\tau}}\sum_{t=1}^{T-\tau}\left(\frac{1}{S}\sum_{s=1}^{S}1\left\{X_{s,t+\tau}^{\theta_{T,N,h}}\leq u\right\}-1\{X_{t+\tau}\leq u\}\right)1\left\{X_t\leq v\right\},\end{aligned} \tag{5}$$

and X_t^* is a resampled series constructed using standard block-bootstrap methods, $\widehat{\theta}^*_{T,N,h}$ is estimated parameter using the resampled data, X_t^*, and is in turn used to construct simulated paths, $X_{s,t+\tau}^{\theta^*_{T,N,h}}$, $s=1,...,S$ and $t=1,...,T-\tau$. In order to generate the empirical distribution of Z_T^*, one performs B bootstrap replications (B is large). Then, one compares Z_T with the percentiles of the empirical distribution of Z_T^*, and rejects H_0 if Z_T is greater than the $(1-\alpha)th$-percentile. Otherwise, one fails to reject. BCS (2008) has proved that the test carried out in this manner is correctly asymptotically sized, and has unit asymptotic power.

For two-factor models (e.g. stochastic mean and stochastic volatility models, where $X_t=\left(X_t^1,X_t^2\right)'$), the difficulty lies in dealing with the initial value for the simulation process, given that the latent variable in X_t is unobservable. BCS (2008) integrate out this effect by first simulating a long path of length N observations for latent variable X_t^2. Second, they take the simulated values in the first step as starting values for the latent variable and simulate $S\times N$ paths in order to form, $\widehat{F}_\tau(u|X_t,\widehat{\theta}_{T,N,h})$. The associated test statistic in Eq(4) is:

$$Z_T(u,v) = \frac{1}{\sqrt{T-\tau}} \sum_{t=1}^{T-\tau} \left(\frac{1}{NS} \sum_{j=1}^{N} \sum_{s=1}^{S} 1\left\{ X_{j,s,t+\tau}^{1,\theta_{T,N,h}} \le u \right\} - 1\{X_{t+\tau}^1 \le u\} \right) 1\left\{ X_t^1 \le v \right\}. \quad (6)$$

Similarly, the bootstrap statistic analogous to that given in Eq(5) is $Z_T^* = \sup_{u\times v \in U\times V} |V_T^*(u,v)|$, where

$$Z_T^*(u,v) = \frac{1}{\sqrt{T-\tau}} \sum_{t=1}^{T-\tau} \left(\frac{1}{NS} \sum_{j=1}^{N} \sum_{s=1}^{S} 1\left\{ X_{j,s,t+\tau}^{1,\theta_{T,N,h}^*} \le u \right\} - 1\{X_{t+\tau}^{1*} \le u\} \right) 1\left\{ X_t^{1*} \le v \right\}$$
$$- \frac{1}{\sqrt{T-\tau}} \sum_{t=1}^{T-\tau} \left(\frac{1}{NS} \sum_{j=1}^{N} \sum_{s=1}^{S} 1\left\{ X_{j,s,t+\tau}^{1,\theta_{T,N,h}} \le u \right\} - 1\{X_{t+\tau}^1 \le u\} \right) 1\left\{ X_t^1 \le v \right\}. \quad (7)$$

Of note that we use $X_{t+\tau}^1$ and $X_{j,s,t+\tau}^{1,\theta_{T,N,h}}$ to construct the conditional interval because only X_t^1 is observable in X_t.

Now, consider a three-factor model (see e.g. the "CHEN" and "CHENJ" models discussed below), where $X_t = \left(X_t^1, X_t^2, X_t^3\right)'$, and $W_t = \left(W_t^1, W_t^2, W_t^3\right)$ are mutually independent standard Brownian motions in Eq(1). The key issue concerns how to construct the conditional distribution $F_\tau(u|X_t, \theta^\dagger) = \Pr\left(X_{t+\tau}^{\theta^\dagger} \le u | X_t^{\theta^\dagger} = X_t\right)$ without knowing the starting value of X_t^2 and X_t^3. To deal with this issue, we propose a simple simulation based method that is an immediate consequence of the approach discussed in BCS for one and two factor models:

Step 1: Given the estimated parameter $\widehat{\theta}_{T,N,h}$, generate a path of length N (a large number) for $X_t^{\theta_{T,N,h}}$. The key here is to use the mean of stochastic volatility and the mean of stochastic mean in $\widehat{\theta}_{T,N,h}$ as the initial start values for these two latent variables. Retrieve $X_t^{2,\theta_{T,N,h}}$ and $X_t^{3,\theta_{T,N,h}}$, $t = 1, 2, ..., N$ from the path.

Step 2: Given the observable X_t^1 and the $N \times N$ simulated latent paths ($X_j^{2,\theta_{T,N,h}} \times X_m^{3,\theta_{T,N,h}}$ $j, m = 1, ..., N$) as the start values , we simulate τ-step ahead $X_{t+\tau}^{1,\theta_{T,N,h}}$. Since the start values for the two latent variables are $N \times N$ length, so for each X_t^1 we have N^2 path, that is $F_{\tau,i}(u|X_t, \widehat{\theta}) = \frac{1}{N^2}\sum_{m=1}^{N}\sum_{j=1}^{N} 1\left\{X_{j,m,t+\tau}^{1,\theta_{T,N,h}} \le u\right\}$, where i denotes the ith simulation.

Step 3: Simulate $X_{t+\tau}^{1,\theta_{T,N,h}}$ S times, that is to repeat step 2 S times. $\frac{1}{S}\sum_{i=1}^{S} F_{\tau,i}(u|X_t, \widehat{\theta}_{T,N,h})$ is the estimator of $F_\tau(u|X_t, \theta^\dagger)$.

Step 4: Construct the statistic for the null of correct specification of the conditional distribution:

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

where

$$Z_T(u,v) = \frac{1}{\sqrt{T-\tau}} \sum_{t=1}^{T-\tau} \left(\frac{1}{N^2 S} \sum_{m=1}^{N} \sum_{j=1}^{N} \sum_{s=1}^{S} 1\left\{ X_{j,m,s,t+\tau}^{1,\theta_{T,N,h}} \le u \right\} - 1\{X_{t+\tau}^1 \le u\} \right) 1\left\{ X_t^1 \le v \right\}. \quad (8)$$

It follows immediately from BCS (2008) that $T, N, S \to \infty$. Then, if $h \to 0$, $T/N \to 0$, $T/S \to 0$, $h^2 T \to 0$, and the model is correctly specified, the following result holds for any X_t^1, $t \geq 1$:

$$\frac{1}{N^2 S} \sum_{m=1}^{N} \sum_{j=1}^{N} \sum_{s=1}^{S} 1 \left\{ X_{j,m,s,t+\tau}^{1,\theta_{T,N,h}} \leq u \right\} - F_0(u|X_t, \theta_0) \xrightarrow{p} 0, \text{uniformly in } u.$$

Moreover, we can implement the same bootstrap method as that outlined in BCS (2008) in order to form the resampled series, X_t^{1*} and construct bootstrap statistics. Namely:

Step 1: Resample X_t^1. In particular, we draw b blocks (with replacement) of length l, where $bl = T$. Thus, each block is equal to $X_{i+1}^1, ..., X_{i+l}^1$, for some $i = 0, ..., T-l$, with probability $1/(T-l)$. More formally, let I_k, $k = 1, ..., b$ be iid discrete uniform random variables on $[0, 1, ..., T-l]$. Then, the resampled series, X_t^{1*} is such that $\{X_1^{1*}, X_2^{1*}, ..., X_l^{1*}, X_{l+1}^{1*}, ..., X_T^{1*}\} = \{X_{I_1+1}^1, X_{I_1+2}^1, ..., X_{I_1+l}^1, X_{I_2}^1, ..., X_{I_b+l}^1\}$, and so a resampled series consists of b blocks that are discrete iid uniform random variables, conditional on the sample. Use these data to construct $\widehat{\theta}_{T,N,h}^*$.

Step 2: Repeat Steps 1-3 in constructing $Z_T(u, v)$, but we use X_t^{1*} and $\widehat{\theta}_{T,N,h}^*$ to replace X_t^1 and $\widehat{\theta}_{T,N,h}$ respectively, to construct the conditional distribution for $X_{t+\tau}^{1*}$. *Particularly, $X_{j,s,m,t+\tau}^{1,\theta_{T,N,h}^*}$ is the simulated value at simulation s, constructed using $\widehat{\theta}_{T,N,h}^*$, and $X_t^{1,\theta_{T,N,h}^*}$, $X_{j,h}^{2,\theta_{T,N,h}^*}$, $X_{m,h}^{3,,\theta_{T,N,h}^*}$ as initial value.* Of note that we use the same set of random errors used in $X_{j,m,s,t+\tau}^{1,\theta_{T,N,h}}$ to construct $X_{j,m,s,t+\tau}^{1,\theta_{T,N,h}^*}$.

Step 3: Construct the bootstrap statistic, which is the bootstrap counterpart of Z_T:

$$Z_T^* = \sup_{u \times v \in U \times V} |Z_T^*(u, v)|, \tag{9}$$

where

$$\begin{aligned} Z_T^*(u, v) = & \frac{1}{\sqrt{T-\tau}} \sum_{t=1}^{T-\tau} \left(\frac{1}{N^2 S} \sum_{m=1}^{N} \sum_{j=1}^{N} \sum_{s=1}^{S} 1 \left\{ X_{j,m,s,t+\tau}^{1,\theta_{T,N,h}^*} \leq u \right\} - 1\{X_{t+\tau}^{1*} \leq u\} \right) 1 \left\{ X_t^{1*} \leq v \right\} \\ & - \frac{1}{\sqrt{T-\tau}} \sum_{t=1}^{T-\tau} \left(\frac{1}{N^2 S} \sum_{m=1}^{N} \sum_{j=1}^{N} \sum_{s=1}^{S} 1 \left\{ X_{j,m,s,t+\tau}^{1,\theta_{T,N,h}} \leq u \right\} - 1\{X_{t+\tau}^{1} \leq u\} \right) 1 \left\{ X_t^{1} \leq v \right\}, \end{aligned}$$

Step 4: Repeat step 1-3 B times to generate the empirical distribution of the B bootstrap statistics.

2.2 An Out-of-Sample Model Selection Test

The approach of CS (2011) is to measure predictive accuracy using a distributional generalization of mean square error, as defined in Corradi and Swanson (2005b). Namely, let $F_k^\tau(u|X_t, \vartheta_k^\dagger)$ be the distribution of $X_{t+\tau}$ given X_t, evaluated at u, implied by diffusion model k, where $\theta_k^\dagger$ is a parameter vector,and let $F_0^\tau(u|X_t,\vartheta_0)$ be the distribution associated with the underlying and unknown "true" model. Now, choose model k over model 1, say, if $E\left(\left(F_k^\tau(u|X_t,\vartheta_k^\dagger)-F_0^\tau(u|X_t,\vartheta_0)\right)^2\right)<E\left(\left(F_1^\tau(u|X_t,\vartheta_1^\dagger)-F_0^\tau(u|X_t,\vartheta_0)\right)^2\right)$. The test can be viewed as distributional generalizations of both Diebold and Mariano (1995) and White (2000). Note that in the case of continuous time models, if we knew $F_k^\tau(u|X_t,\vartheta_k^\dagger)$ in closed form, then we could proceed as in Corradi and Swanson (2006a,b). However, the functional form of the model implied conditional distribution is unknown in closed form, in general, and hence we rely on a simulation-based approach to facilitate testing. As is customary in the out-of-sample evaluation literature, the sample of T observations is split into two subsamples, such that $T=R+P$, where only the last P observations are used for predictive evaluation. We first simulate S paths of length $P-\tau$ (each path is $\tau-$steps ahead) using $X_R,...,X_{R+P-\tau}$ as starting values. Then, a scaled difference between the conditional distribution, estimated with historical as well as simulated data, is used to construct our test statistic.

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

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

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

The test statistic is:

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

where the test statistic is constructed over an interval $(u_1,u_2)\in U\text{x}U$ and

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

with $X_{k,t+\tau,i}^{\vartheta_{k,t,h}}$ the simulated τ-step ahead value given X_t and the recursively estimated parameter $\widehat{\vartheta}_{k,t,h}$. Of note is that $\widehat{\vartheta}_{k,t,h}$ is the parameter implied by model k and it is estimated from the first t observations instead of total sample T as in BCS (2008) test, and more precisely, it is updated prior to each new forecast construction. A version of the test that applies for evaluation of the entire predictive density is discussed in CS (2011). The appropriate bootstrap statistic for critical value construction is:

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

where $X^{*}_{t+\tau}$ is from a bootstrap sample constructed using the standard block bootstrap and $\widehat{\vartheta}^{*}_{1,t,N,h}$ are parameter estimates based on the bootstrap sample (see CS (2011) for complete details). 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. Thus, even if $P/R\to 0$, so that there is no need to mimic parameter estimation error (and hence the above statistic can be constructed using $\widehat{\vartheta}_{k,t,N,h}$ instead of $\widehat{\vartheta}^{*}_{k,t,N,h}$), it remains the case that recentering of all bootstrap terms around the (full) sample mean is necessary.

3 The Models

In our empirical analysis we consider a variety of continuous and discrete models, both with and without jumps. The jump processes that we model are driven by two separate Poisson draws (up jumps and down jumps) with different intensities, and different jump magnitudes, as discussed in Chacko and Das (2002). Overall, our models are most closely related to those examined in ABL (2004), and Hong and Li (2005), although we consider both continuous time and discrete models. In total, we consider six affine models, one nonaffine model, and two discrete models of the short rate, all of which are outlined briefly below.

The Cox-Ingersoll-Ross (CIR) Model: We follow CIR (1985) and posit that:

$$dr(t)=\kappa_r(\theta-r(t))\,dt+\sigma_r\sqrt{r(t)}dW_r(t)\,, \tag{10}$$

where $W_r(t)$ is standard Brownian motion, θ is the long-run mean of the interest rate, κ_r measures the speed of mean-reversion, and σ_r is a standard deviation parameter that is assumed to be fixed. Also, non-negativity is imposed, as $2\kappa_r\theta > \sigma_r^2$.

Stochastic Mean Model (SM): As above, but $\theta(t)$ is a mean reverting process that converges to its unconditional mean:

$$
\begin{aligned}
dr(t) &= \kappa_r(\theta(t) - r(t))\,dt + \sigma_r dW_r(t), \\
d\theta(t) &= \kappa_\theta\left(\overline{\theta} - \theta(t)\right)dt + \sigma_\theta\sqrt{\theta(t)}dW_\theta(t),
\end{aligned}
\tag{11}
$$

where $W_r(t)$ and $W_\theta(t)$ are independent Brownian motions, and $\overline{\theta}$ and σ_θ are the mean and standard deviation of $\theta(t)$, respectively. Of note is that the mean process, $\theta(t)$, cannot take negative values provided that $2\kappa_\theta\overline{\theta} > \sigma_\theta^2$. Stationarity requires that κ_r and κ_θ (which control mean reversion speeds) are greater than zero.

Stochastic Volatility Model (SV): We estimate the Heston (1993) model (see also Chen (1996), Andersen, and Lund (1997), and Aït-Sahalia and Kimmel (2007)). Namely:

$$
\begin{aligned}
dr(t) &= \kappa_r(\overline{r} - r(t))\,dt + \sqrt{V(t)}dW_r(t), \\
dV(t) &= \kappa_v(\overline{v} - V(t))\,dt + \sigma_v\sqrt{V(t)}dW_v(t),
\end{aligned}
\tag{12}
$$

where κ_r,$\kappa_v > 0$ in order to ensure stationarity, and $\overline{v}$ and σ_v are the mean and standard deviation of $V(t)$. As above, $W_r(t)$ and $W_v(t)$ are scalar Brownian motions in some probability measure. However, we now assume that $W_r(t)$ and $W_v(t)$ are correlated, such that $dW_r(t)\,dW_v(t) = \rho dt$, where the correlation, ρ, is some constant in [-1,1]. Finally, note that volatility is a square-root diffusion process, which requires that $2\kappa_v\overline{v} > \sigma_v^2$.

Stochastic Volatility Model with Jumps (SVJ): We add Poisson-exponential jumps to the previous model. Namely:

$$
\begin{aligned}
dr(t) &= \kappa_r(\overline{r} - r(t))\,dt + \sqrt{V(t)}dW_r(t) + J_u dq_u - J_d dq_d, \\
dV(t) &= \kappa_v(\overline{v} - V(t))\,dt + \sigma_v\sqrt{V(t)}dW_v(t),
\end{aligned}
\tag{13}
$$

where q_u and q_d are Poisson processes with jump intensity λ_u and λ_d respectively, and are independent of the Brownian motions $W_r(t)$ and $W_v(t)$. In particular, λ_u is the probability of a jump up, $\Pr(dq_u(t) = 1) = \lambda_u$ and λ_d is the probability of a jump down, $\Pr(dq_d(t) = 1) = \lambda_d$. 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): We combine various features of the above models, by considering a version of the oft examined 3-factor model due to CKLS (1992), which is discussed in detail in Dai and Singleton (2000). In particular, we consider the Chen (1996) 3-factor model:

$$\begin{aligned} dr(t) &= \kappa_r(\theta(t) - r(t))\,dt + \sqrt{V(t)}dW_r(t), \\ dV(t) &= \kappa_v(\overline{v} - V(t))\,dt + \sigma_v\sqrt{V(t)}dW_v(t), \\ d\theta(t) &= \kappa_\theta\left(\overline{\theta} - \theta(t)\right)dt + \sigma_\theta\sqrt{\theta(t)}dW_\theta(t), \end{aligned} \tag{14}$$

where $W_r(t)$, $W_v(t)$ and $W_\theta(t)$ are independent Brownian motions, and V and θ are the stochastic volatility and stochastic mean of short rate r, respectively. As discussed above, non-negativity for $V(t)$ and $\theta(t)$ requires that $2\kappa_v\overline{v} > \sigma_v^2$ and $2\kappa_\theta\overline{\theta} > \sigma_\theta^2$.

Three Factor Jump Diffusion Model (CHENJ): ABL (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 addressing the finding of Piazzesi (2004, 2005) that violent discontinuous movements may arise from monetary policy regime changes. The model is defined as follows:

$$\begin{aligned} dr(t) &= \kappa_r(\theta(t) - r(t))\,dt + \sqrt{V(t)}dW_r(t) + J_u dq_u - J_d dq_d, \\ dV(t) &= \kappa_v(\overline{v} - V(t))\,dt + \sigma_v\sqrt{V(t)}dW_v(t), \\ d\theta(t) &= \kappa_\theta\left(\overline{\theta} - \theta(t)\right)dt + \sigma_\theta\sqrt{\theta(t)}dW_\theta(t), \end{aligned} \tag{15}$$

where parameters are as defined for the above SVJ and CHEN models.

General Single-Factor Diffusion Model (CEV):

The best known model that incorporates the dependence of volatility on the level of the interest level is probably the nonaffine constant elasticity of volatility (CEV) model discussed in CKLS (1992), which has been studied by numerous authors (see e.g. Aït-Sahalia(1996), Stanton (1997), Durham (2003), Hong and Li (2005), and Bali and Wu (2006)). In light of this, we also consider following CEV model

$$dr(t) = \kappa_r(\theta - r(t))\,dt + \sigma_r r(t)^\rho dW_r(t), \tag{16}$$

where parameters are defined as in the CIR model. We also assume that $\rho > 1$ (see e.g. Aït-Sahalia(1996)). The CEV model is probably the best known model that incorporates the dependence of volatility on the level of the interest level.

Linear Autoregression Model (AR(p)):

As a discrete benchmark model, we estimate a standard AR(p) model:

$$r_t = \theta + \beta r_{t-1} + u_t, \tag{17}$$

where u_t is a disturbance term, and the number of lags, p, is selected via the use of the Schwarz information criterion, and is re-estimated prior to each new prediction when the model is used to construct predictions.

Smooth Transition Autoregression Model (STAR):

Finally, we estimate a standard version of the STAR model developed by Chan and Tong (1986) (see also Teräsvirta and Anderson (1992), Teräsvirta(1994, 1998), and the references therein):

$$r_t = (\theta_1 + \beta_1 r_{t-1})G(\gamma, z_t, c) + (\theta_2 + \beta_2 r_{t-1})(1 - G(\gamma, z_t, c)) + u_t, \tag{18}$$

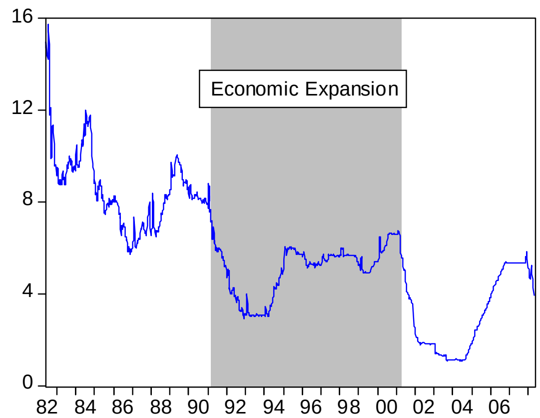
In our analysis, the transition function, $G(\cdot)$ is defined to be the logstic CDF (i.e. $G(\gamma, z_t, c) = \frac{1}{1+e^{\gamma(z_t - c)}}$) and z_t, is an exogenous transition variable, defined in the same manner as the moving average process used in Franses and van Dijk (2000).

Of note is that there are many estimation techniques that can be used to estimate the above continuous time models. For example, there are many simulation based methods, such as efficient method of moments (see Gallant and Tauchen (1996,1997), Chernov and Ghysels (2000), and the references cited therein) and nonparametric simulated maximum likelihood (see e.g. Fermanian and Salanié (2004) and CS (2011)). For further discussion, see Aït-Sahalia (2007) which provides a survey on estimating continuous models using discrete observations. In the current paper, we follow the simulated generalized method of moments estimation approach discussed in Duffie and Singleton (1993). Of course, it should also be pointed out that it is possible in the case of the affine models considered here to compute in closed form the conditional characteristic function, and hence follow Singleton (2001), Jiang and Knight (1997, 2002), Duffie, Pan and Singleton(2000), or Chacko and Viceira (2003). Finally, note that our discrete models are estimated using maximum likelihood, and as discussed in Franses and van Dijk (2000), and van Dijk, Teräsvirta, and Franses (2002). Given that we estimate our parameters recursively, the values are not reported here, for the sake of brevity (complete results are available upon request from the authors).

4 Data

In order to facilitate the comparison of our empirical findings with those of BCS (2008), we use the same dataset as they do. Namely, we use data collected on the one-month Eurodollar deposit rate as our proxy of the short rate. Our data ranges from June 1982 to April 2008 (1,396 weekly observations). Other yields that are often considered in the literature include the monthly federal funds rate (Aït-Sahalia (1999)), monthly yields on zero-coupon bonds with different maturities (see Duffie (2002) and Diebold and Li (2006, 2007)), and the weekly 3-month T-bill (see ABL (2004) and Durham (2003)).

Figure 1: Eurodollar rate 06/04/1982 to 04/25/2008



Notes: This figure plots empirical weekly data on the Eurodollar rate for the period 06/04/1982 to 04/24/2008. The shadowed is the "Stable 1990s" period, which covers period 03/1991- 05/2001.

Bai and Perron (1998, 2003) extend the single unknown break point testing approach (see Andrews (1993), Andrews and Ploberger (1994)) to testing for no structural breaks against an unknown number of breaks, (i.e. 0 against l breaks and l against $l+1$ breaks, where l is arbitrary but fixed number). They apply a dynamic programming algorithm to estimate breakpoints sequentially by adding one breakpoint at a time and minimizing the resulting sum of squared residuals. We implement the tests outlined in Bai and Perron (1998, 2003) to test and date structural changes in our dataset. In particular, we test whether the mean of the spot interest rate changes over time. We find two breakpoints, corresponding to February 1991 and May 2001. These results lead to our consideration of three 3 different sample periods (see Figure 1). The first sub-sample corresponds to the “Stable 1990s” period (i.e. March 1991 to May 2001). The second sub-sample is the so-called “Post 1990s” period from June 2001 to April 2008. Finally, we consider our entire sample from June 1982 to April 2008 (denoted as “Whole Sample” in the sequel). Of note is that the “Stable 1990s” is the longest economic expansion that the United States has ever experienced. In early 1990s, the Federal Reserve returned to targeting the federal funds rate. In particular, the Federal Reserve fixed the federal funds rate at 3% from late 1992 until February 1994, and increased the rate to 6% by early 1995. Only in January 2001 did the Federal Reserve begin to cut rates again. Of further note is that our 2 sub-samples are consistent with break dates found using the multiple break point test developed by Bai (1997) and Bai and Perron (1998, 2003). Details of these calculations are available upon request from the authors. For an excellent and exhaustive survey of the extant literature on testing for structural breaks in financial time series, the reader is referred to Andreou and Ghysels (2009).

5 Empirical Results

In this section, we first present various summary measures for our dataset, then we summarize our main findings based on specification tests and model selection tests.

5.1 Summary Statistics

Table 1: Summary Statistics for Different Subsamples

Sample period	Date	Mean	Median	Std. Dev.	Skewness	Kurtosis	Jarque Berra
Whole Sample	06/1982-04/2008	0.0568	0.0563	0.0259	0.2890	3.0451	19.0971
Pre 1990s[1]	06/1982-02/1991	0.0851	0.0825	0.0160	1.1003	5.7137	232.6905
The Stable 1990s	03/1991-05/2001	0.0508	0.0540	0.0106	-0.7292	2.4262	55.4637
Post 1990s	06/2001-04/2008	0.0302	0.0283	0.0161	0.3002	1.5218	38.5192

Notes: This table reports summary statistics for our historical data. The“Whole Sample” stands for the period from June 1982 to April 2008. The “Stable 1990” period is from March 1991 to May 2001. The “Pre 1990s” period covers June 1982 to February 1991. The “Post 1990s" period is from June 2001 to April 2008. See section 4 for details.

Table 1 reports various summary statistics for the data, including mean, median, variance, skewness, kurtosis and Jarque-Bera test statistics. The “Whole sample” data has a mean of 0.0568 or 5.68%, a standard deviation of 0.0259, negligible positive skewness of 0.289, and substantial kurtosis relative to that of a normal distribution. The Jarque-Bera test indicates that the sample does not follow a normal distribution. The three sub-samples are also not normally distributed; and they have quite different distributional properties. Although not examined further in this paper, note that the “Pre 1990s” period is quite volatile, and has the highest mean 8.51%. This is due to the high inflation/interest rate regime in the 1980s. In contrast, the “Stable 1990s” period is much more stable, with a mean of 5.08%, standard deviation of 0.0106, and smaller skewness and

[1] Of note is that during the period prior to 1990, the Federal Reserve targeted, at various times, monetary aggregates and interest rates, which resulted to some extent in elevated interest rate volatility relative to that experienced during the 1990s. Moreover, high inflation and recessions in the 1980s also increased volatility over this period and unit root tests for this period suggest that interest rate are nonstationary. As Pritsker (1998) documents, specification tests that assume stationarity can lead to over-rejection of the null hypothesis, in such cases (see also Aït-Sahalia and Park (2009)). For this reason, we do not separately report results for the "Pre 1990s" period.

kurtosis. Of note is that the "Stable 1990s" period is negatively skewed (-0.729), although all other samples have significant positive skewness. As opposed to the other two sub-samples, the "Post 1990s" period demonstrates a bimodal distribution[2]. Of note is that in 2004 the Eurodollar rate reached a low of 1.04%. Moreover, it is not surprising that the Eurodollar rate data that we examine shares the same patterns of increase and decrease as the federal funds rate, which explains the sharp decreases and increases in the Eurodollar rate in the "Post 1990s" period. Compared with other samples, the "Post 1990s" period has the lowest mean (3.02%), but has a relatively high standard deviation. These results suggest that interest rate models that are "regime-dependent" may provide a better good fit to the data and possibly also good predictions. Hence our specification of the STAR model. Of course, a difficulty with modelling individual regimes, in general, is ascertaining whether the period to be simulated can be expected to remain within the regime.

5.2 Consistent Specification Test Results

Tables 2-4 report our in-sample specification test results for the 3 sample periods outlined above. Tests are carried out using $\tau-$step ahead confidence intervals. We set $\tau = \{1, 2, 4, 12\}$, corresponding to one week, two weeks, one month, and one quarter ahead conditional distribution evaluation periods. The confidence intervals that we use in test statistic construction are chosen based on the properties of our historical data. In particular, we set $\underline{u}$ and $\overline{u}$ equal to $\overline{X} \pm \sigma_X$ and $\overline{X} \pm 0.5\sigma_X$, respectively. Additionally, we set our simulation sample length as $S = 10T$, where T is the historical sample length. The simulation sample length for latent variables is set at $N = 10T$. When simulating using our discrete models, errors are drawn randomly from the empirical distribution of the residuals. In our implementation of the bootstrap, we set block length to be 20, and carry out 100 bootstrap replications. In the tables, test statistics (denoted by Z_T) and 5%, 10%, 15% and 20% bootstrap critical values are given. Single starred entries denote rejection at the 10% significant level.

5.2.1 Whole sample

Results are presented in Table 2 for the "Whole Sample" case. Not surprisingly, the CIR model is rejected for all τ, regardless of confidence interval width. Moreover, the SV model performs the best amongst the CIR, SV and SVJ models. These results are in line with those reported in BCS (2008). Interestingly, though, we find that the three-factor model (CHEN) not considered in BCS

[2] It should be noted that the "Post 1990s" period has also been more volatile, as the federal funds rate was first lowered in order to stimulate the economy after the high-tech bubble crash, was later increased to accommodate increasing inflation and concerns about the booming housing market, and has recently again been lowered due to global financial concerns surrounding the recent crash of the U.S. housing market and related problems.

(2008) performs at least as well as the SV model, and indeed fails to reject at higher significance levels, suggesting that the CHEN model may actually be marginally better than the SV model. Overall, it seems that increased model complexity may help capture spot rate dynamics when long samples of data across many historical regimes are used to calibrate models. Of further note is that the CHENJ model performs marginally better than the SVJ model. These results are in accordance with the findings of Hong, Li and Zhao (2004) that more complicated models tend to have better in-sample fit since they have more parameters to catch different aspects of interest rate dynamics. Turning now to our discrete models, note that the AR(p) and STAR models are not rejected for any τ interval combinations, as expected. In the next sub-section, we shall examine whether or not this in-sample dominance of discrete models parlays into superior ex-ante forecasting performance.

To further illustrate the findings of Table 2, we plot kernel densities of our simulated data and actual data in Figure 2, for selected models. Specifically, we choose points that represent the left tail, middle points and right tail of our historical data as evaluation points, and construct kernel density estimates. Figure 2a contains plots of the simulated density at $x = 0.03$. Compared with the CIR and the SV models, note that simulated data from CHEN model are more concentrated around the actual data point. Moreover, the 3-factor CHEN model has higher kurtosis than the SV model. These findings are consistent with the above results suggesting that CHEN is superior to the other candidate models. Similar results are reported in Figures 2b-d for other evaluation points.

5.2.2 The stable 1990s

Table 3 presents results for the "Stable 1990s" period. The most noteworthy result is that the univariate CIR model beats all multifactor models. The CIR model fails to be rejected for all values of τ, and for all confidence intervals, at almost any confidence level. This result is in stark contrast to our findings for the whole sample period. As discussed in Section 5.1, this is perhaps not surprising given the stable monetary regime of the 1990s; and reminds us that our different models perform very differently depending upon the particular regime from whence data are generated. Figures 3a-d display plots of the simulated densities for the CIR, SM and CHENJ models. We choose the SM model instead of the SV model in Panel A (plots a and b) of Figure 3 because the SM model is the "second best" from amongst the continuous candidate models. The CHENJ model is depicted in order to illustrate how the jump process distorts the simulated densities in this stable period. As evidenced upon inspection of Panel A of Figure 3, the CIR model is superior to the SM model in that CIR-simulated data is more concentrated around the actual evaluation point. As expected, the simulated density for the CHENJ model is far from the evaluation point. Similar conclusions emerge upon examination of Panel A (plots c and d) of Figure 3. Finally, note that the CIR model outperforms the discrete models when the simulation period is long ($\tau = 12$) and the confidence interval is wide ($\overline{X} \pm \sigma_X$), which suggests that the CIR model is superior to our dynamically more flexible discrete models for long-run modeling, and points to the need for careful examination and comparison of the ex-ante predictive performance of the continuous and discrete models.

5.2.3 Post-1990s

The bimodality associated with the "Post 1990s" period makes our specification test results quite different from those reported for our other sample periods. The SM model "outperforms" other models. The rejection frequency for the CIR model is low as well. Additionally, we fail to reject the null of correct specification for the complex three-factor with jumps CHENJ model. These results are sensible given the underlying economic conditions prevailing during this period. As is apparent upon inspection of Table 4, the SM model captures the changing mean in this period quite well. However, since the Federal Reserve Bank changed the target rate with more frequently and with greater magnitude during this period, the CHENJ model which combines stochastic mean and jump components also performs well.[3] Moreover, the CHENJ model appears to be able to capture possible outliers in this period via its inclusion of two latent variables. Panel B (plots e and f) of Figure 3 contain plots of empirical densities for the SM, CIR and SV models at the two modal points ($x = 0.02$ and $x = 0.05$). The densities associated with the SM model have higher peak and narrower tails than those for the CIR model, at both evaluation points. As expected, the SV model has densities with very fat tails; but these are centered far from the evaluation point. Of final note is that there are evaluation points for which other models do beat the SM model, but the SM performs best, overall. This is a feature which should be expected, and underscores the importance of using not only portmanteau type specification tests but also inspecting individual densities when evaluating alternative models.

5.2.4 Is there an overall winner?

The above question can be answered, to some extent, by fixing the parameterization estimated for the "best" model in our "Whole Sample" analysis (i.e. the CHEN model) and comparing the performance of this model with that of the "winners" during our two sub-sample periods. Results

[3] The Federal Reserve Board changed interest rate target 11 times in 2001, once in 2002 and 2003, 5 times in 2004, 8 times in 2005, 4 times in 2006 and 2008, and 3 times in 2007.

of consistent specification tests constructed in this manner are contained in Table 5. Both panels of this table report test results for the CHEN model. The results in Panel A should be compared with the CIR results from Table 3, while the results in Panel B should be compared with the SM model results from Table 4. Interestingly, the CHEN model fails to reject for all significance levels for the "Post 1990s" period, when using the "Whole Sample" parameterization. However, perhaps not surprisingly given our above discussion, the CHEN model is rejected during the "Stable 1990s". Overall, thus, the CHEN model appears to be a reasonably adequate model, nesting all but one historical episode at least as well as any other model. In the next section, we shall ascertain whether this finding translates into superior predictive performance when comparing the CHEN to all other models, including our discrete alternatives.

5.3 Prediction Based Model Selection Test Results

Given our above findings, a natural next step is the construction of ex-ante prediction type model selection tests in order to determine whether the CHEN model dominates all other models for the entire sample. This is done by applying the $D_{k,P,N}(u_1, u_2)$ test discussed above. In particular, tests are carried out using $\tau-$step ahead predictive confidence intervals (with the ex-ante period $P = T/2$) constructed using recursively estimated models. For ease of comparison, we focus only on the "best" continuous and the "best" discrete models, according to the finding of the previous section. In addition, we set $\tau = \{1, 2, 3, 4, 5, 6, 12\}$, and the confidence intervals that we use in test statistic construction are chosen based on the properties of our historical data. In particular, we set $\underline{u}$ and $\overline{u}$ equal to $\overline{X} \pm \sigma_X$ and $\overline{X} \pm 0.5\sigma_X$, respectively, as above. Finally, we set our simulation sample length to be $S = 10T$, and the simulation sample length for latent variable integration to be $N = 10T$. In our implementation of the bootstrap, we set block length to be 20, and carry out 100 bootstrap replications. In the tables, test statistics (denoted by $D_{k,P,N}$) and predictive density type "mean square forecast errors" (MSFEs - see footnote to Table 6 for further details) values are reported. Single starred entries denote rejection at the 10% significant level. Turning to our results, note that upon inspection of the entries in Table 6 it is clear that the null hypothesis that the CHEN model generates predictive densities at least as accurate as the STAR model (our "best" discrete model) is not rejected at a 90% level of confidence, regardless of forecast horizons and interval width.[4] Moreover, in almost all cases, the CHEN model has lower MSFE, and the MSFE spread between the CHEN model and STAR model increases in magnitude as the forecast horizon

[4]Both the AR(p) and the STAR models perform well when evaluated using our in-sample specification tests, as the null hypothesis of correct specification generally fails to reject for both models (see above discussion). However, given that the STAR model is more flexible, and given the clear evidence of structural and regime shifts in our dataset, we use the STAR model as our discrete competitor in our prediction experiments.

increases. This confirms our in-sample findings that the CHEN model is our "global winner". To further illustrate this finding, we provide plots of simulated predictive densities for the CHEN and STAR models in Panel A of Figure 4. Note that the simulated predictive densities associated with CHEN model are highly concentrated around the true evaluation points. Moreover, though the STAR model also has quite peaked densities, they are not generally as centered around the evaluation points.

In order to shed further light on this interesting conclusion that continuous models do indeed in certain cases outperform discrete models, even when comparing predictive accuracy, we carry out the same analysis for the two sub-samples examined in the previous section. Namely, we set CIR as the benchmark for the "Stable 1990s" and SM as the benchmark for the "Post 1990s". Results of comparisons of these models with the STAR model are collected in Tables 7 and 8. Interestingly, the STAR model is preferred to the CIR model (note that the null of equal predictive accuracy is rejected in virtually all cases in Table 7) when comparing predictive accuracy for the "Stable 1990s". An analogous result hold for the "Post 1990s", where the STAR model is preferred to the SM model. These findings are confirmed by plots of empirical densities in Panels B and C of Figure 4. The STAR model has more "correctly" centered and highly concentrated densities when compared with the CIR and SM models. Thus, when the STAR model is parameterized using only data corresponding to a particular sub-sample, it outperforms the "best" continuous alternative. However, and as just noted, when models are estimated using the entire dataset, the continuous CHEN model is preferred. Given that it is very difficult to ascertain what regime will ensue when constructing ex ante predictions, thus, the CHEN model is our overall "winner" when it comes to ex-ante prediction, just as it was carrying out in-sample specification analysis. Still, the discrete STAR model is clearly also promising, and further analysis using alternative STAR and related discrete specifications is warranted. This, however, is left to future research.

6 Concluding Remarks

This paper reviews and implements simulation based specification testing methodology in order to study the in-sample and out-of-sample performance of different affine and non-affine multifactor diffusion processes and discrete time models across various historical sample periods, for various (predictive) horizons, and for various density and distributional evaluation intervals. Interestingly, when our entire sample of weekly data from 1982 to 2008 is used to estimate competing models, the "best" performer, both in-sample and out-of-sample is the three factor Chen (1996) model examined by Andersen, Benzoni and Lund (2004). Just as interestingly, a logistic type discrete STAR model is preferred to the CIR model when comparing predictive accuracy for the "Stable 1990s", during which period the CIR model performs best from amongst continuous models. Moreover, an analogous result holds for the "Post 1990s", where the STAR model is preferred to the two factor SM model. Thus, when the STAR model is parameterized using only data corresponding to a particular sub-sample, it outperforms the "best" continuous alternative, regardless of sub-sample. However, when models are estimated using the entire dataset, the continuous CHEN model is preferred. Given that it is very difficult to ascertain what regime will ensue when constructing ex ante predictions, we conclude that the CHEN model is our overall "winner" when it comes to ex-ante prediction (as well as in-sample specification analysis).

Many topics for further research remain. For example, from a theoretical perspective, it remains to construct specification tests that do not integrate out the effects of latent factors. From an empirical perspective, it remains to determine whether it is in some sense "optimal" to fit models to shorter data samples when simulating future scenarios, and if so, exactly how "short" should samples be? Moreover, there are many STAR and related discrete models for which it may be fruitful to examine ex-ante predictive performance, given that our logistic-STAR model performed so well in the different subsamples examined in this paper.

7 References

Ahn, D.-H., R.F. Dittmar, and A.R. Gallant, 2002, Quadratic Term Structure Models: Theory and Evidence, Review of Financial Studies, 15, 243-288.

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

Aït-Sahalia, Y., 1999, Transition Densities for Interest Rate and Others Nonlinear Diffusions, Journal of Finance, 54,1361-1395.

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 Monographs, Cambridge University Press.

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

Aït-Sahalia, Y. and R. Kimmel, 2007, Maximum Likelihood Estimation of Stochastic Volatility Models, Journal of Financial Economics, 83, 413-452.

Aït-Sahalia, Y., 2008, Closed-Form Likelihood Expansions for Multivariate Diffusions, Annals of Statistics, 36, 906-937.

Aït-Sahalia, Y., and J. Park, 2009, Stationarity-Based Specification Tests for Diffusions When the Process is Nonstationary, working paper, Princeton University.

Andersen, T.G., and J. Lund, 1997, Estimating Continuous Time Stochastic Volatility Models of Short Term Interest Rates, Journal of Econometrics, 77, 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.

Andreou, E., and E. Ghysels, 2009, Structural Breaks in Financial Time Series, Handbook of Financial Time Series, Springer, Berlin, 839-870.

Andrews, D.W.K., 1993, Tests for Parameter Instability and Structural Change with Unknown Change Point, Econometrica, 61, 821-856.

Andrews, D.W.K., and W. Ploberger,1994, Optimal Tests when a Nuisance Parameter is Present Only Under the Alternative, Econometrica, 62,1383- 1414.

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

Bai, J., 1997, Estimating multiple breaks one at a time, Econometric Theory, 13, 315-352.

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

Bai, J., and P. Perron, 1998, Estimating and testing linear models with multiple structural changes, Econometrica, 66, 47-78.

Bai, J., and P. Perron, 2003, Computation and Analysis of Multiple Structural Change Models, Journal of Applied Econometrics, 18, 1-22.

Balduzzi, P., S.R. Das, and S. Foresi, 1998, The Central Tendency: A second Factor in bond Yields, Review of Economic and Statistics, 80, 62-72.

Bali, T.G., and L.Wu, 2006, A comprehensive analysis of the short-term interest-rate dynamics, Journal of Banking and Finance, 30, 1269-1290.

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

Black, F., and M. Scholes, 1973, The pricing of options and corporate liabilities, Journal of Political Economy, 81, 637–659.

Chacko, G., and S. Das, 2002, Pricing Interest Rate Derivatives: A General Approach, Review of Financial Studies, 15, 195-241.

Chacko, G., and L. Viceira, 2003, Spectral GMM Estimation of Continuous-Time Processes, Journal of Econometrics, 116, 259-292.

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.

Chan, K.S., and H. Tong, 1986, On estimating thresholds in autoregressive models, Journal of Times Series Analysis, 7, 179-190.

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, 2008, Characteristic Function-Based Testing for Multifactor Continuous-Time Markov Models Via Nonparametric Regression, Working Paper, University of Rochester.

Chernov, M., and E. Ghysels, 2000, A Study Towards a Unified Approach to the Joint Estimation of Objective and Risk Neutral Measures for the Purpose of Options Valuation, Journal of Financial Economics, 56, 407-458.

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.

Conley, T.G., L.P. Hansen, E.G.J. Luttmer, and J.A. Scheinkman, 1997, Short-Term Interest Rates as Subordinated Diffusions, Review of Financial Studies, 10, 525-578.

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

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

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

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

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.

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 K.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 Calibration in 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., C. Li, and V.Z. Yue, 2007, Global Yield Curve Dynamics and Interactions: A Generalized Nelson-Siegel Approach, NBER working paper.

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.

Duffie, G. R., 2002, Term Premia and Interest Rate Forecasts in Affine Models, Journal of Finance, 57, 405-443.

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

Duffie, D., J. Pan, and K.J. Singleton, 2000, Transform Analysis and Asset Pricing for Affine Jump-Diffusions, Econometrica, 68, 1343-1376.

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

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

Franses, P.H., and van Dijk, 2000, Nonlinear Time Series Models in Empirical Finance, Cambridge University Press, Cambridge.

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.

Heston, S., 1993, A closed-form solution for options with stochastic volatility with applications to bonds and currency options, Review of Financial Studies, 6, 327-343.

Hong, Y., 2002, Evaluation of Out of Sample Probability Density Forecasts with Applications to S&P 500 Stock Prices, Mimeo, Cornell University.

Hong, Y., and H. Li, 2005, Nonparametric Specification Testing for Continuous Time Models with Applications to Term Structure Interest Rates, Review of Financial Studies, 18, 37-84.

Hong, Y., H. Li, and F. Zhao, 2004, Out-of-Sample Performance of Discrete-Time Spot Interest Rate Models, Journal of Business and Economic Statistics, 22, 457-473.

Hong, Y., H. 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.

Hull, J., and A. White, 1990, Pricing interest-rate derivative securities, Review of Financial Studies, 3, 573-592.

Jiang, G., 1998, Nonparametric modeling of US interest rate term structure dynamics and implications on the prices of derivative securities, Journal of Financial and Quantitative Analysis, 33, 465–497.

Jiang, G., and J. Knight, 1997, A Nonparametric Approach to the Estimation of Diffusion Processes - with an application to a short-term interest rate model, Econometric Theory, 13, 615-645.

Jiang, G., and J. Knight, 2002, Estimation of Continuous Time Processes via Empirical Characteristic Function, Journal of Business and Economic Statistics, 20, 198–212.

Jones, C.S., 2003, Nonlinear mean reversion in the short-term interest rate. Review of Financial Studies, 16, 793–843.

Piazzesi, M., 2004, Affine Term Structure Models, Handbook of Financial Econometrics, Volume 1, Chapter 12, 691-766.

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

Pritsker, M., 1998, Nonparametric Density Estimation 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., 2001, Estimation of Affine Asset Pricing Models Using the Empirical Characteristic Function, Journal of Econometrics, 102, 111–141.

Stanton, R., 1997, A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk, Journal of Finance, 52, 1973-2002.

Sullivan R., A. Timmermann, and H. White, 1999, Data-Snooping, Technical Trading Rule Performance, and the Bootstrap, Journal of Finance, 54, 1647-1691.

Sullivan, R., A. Timmermann, and H. White, 2001, Dangers of data-driven inference: The case of calendar effects in stock returns, 105, Journal of Econometrics, 249– 286.

Teräsvirta, T., 1994, Specification, estimation, and evaluation of smooth transition autoregressive models, Journal of the American Statistical Association, 89, 208–218.

Teräsvirta, T., 1998, Modelling economic relationships with smooth transition regressions, Handbook of Applied Economic Statistics, 507–552.

Teräsvirta, T., and H.M. Anderson, 1992, Characterizing nonlinearities in business cycles using smooth transition autoregressive models, Journal of Applied Econometrics, 7, 119–136.

Thompson, S., 2008, Identifying Term Structure Volatility from the LIBOR-Swap Curve, Review of Financial Studies, 21, 819-854

van Dijk, D., T. Teräsvirta, and P.H. Franses, 2002, Smooth Transition Autoregressive Models - A Survey of Recent Developments, Econometric Reviews, 21, 1-47.

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

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

Table 2: Consistent Specification Test Results - “Whole Sample” Period

τ	$(\underline{u},\overline{u})$	Z_T	5% CV	10% CV	15% CV	20% CV
		Panel A: CIR model				
1	$\overline{X} \pm 0.5\sigma_X$	3.97*	2.4471	2.0527	1.8945	1.7273
	$\overline{X} \pm \sigma_X$	3.2325*	1.9326	1.693	1.5043	1.4479
2	$\overline{X} \pm 0.5\sigma_X$	5.0864*	3.0005	2.6345	2.4161	2.1491
	$\overline{X} \pm \sigma_X$	4.731*	2.5565	2.346	2.0451	1.945
4	$\overline{X} \pm 0.5\sigma_X$	4.7771*	3.5572	3.2868	2.7203	2.6807
	$\overline{X} \pm \sigma_X$	5.3449*	3.2166	2.9516	2.6501	2.517
12	$\overline{X} \pm 0.5\sigma_X$	4.1669	4.5583	4.2035	3.9597	3.6599
	$\overline{X} \pm \sigma_X$	5.0172*	4.5625	3.9379	3.5979	3.4429
		Panel B: SM model				
1	$\overline{X} \pm 0.5\sigma_X$	4.1029*	2.5726	2.23	1.9945	1.8777
	$\overline{X} \pm \sigma_X$	2.0027*	2.1117	1.7365	1.6419	1.5218
2	$\overline{X} \pm 0.5\sigma_X$	4.6284*	3.2467	2.8227	2.4263	2.2936
	$\overline{X} \pm \sigma_X$	2.6905*	2.6297	2.353	2.1306	1.966
4	$\overline{X} \pm 0.5\sigma_X$	4.4489*	3.724	3.258	2.9065	2.7414
	$\overline{X} \pm \sigma_X$	2.6585	3.1818	2.9199	2.7031	2.5381
12	$\overline{X} \pm 0.5\sigma_X$	4.3481*	4.5506	4.2088	3.954	3.6575
	$\overline{X} \pm \sigma_X$	3.4824	4.5509	3.9313	3.6011	3.445
		Panel C: SV model				
1	$\overline{X} \pm 0.5\sigma_X$	3.5915*	2.4105	2.0225	1.9037	1.7712
	$\overline{X} \pm \sigma_X$	1.6012	2.0766	1.6535	1.5281	1.4746
2	$\overline{X} \pm 0.5\sigma_X$	4.9061*	3.1205	2.9481	2.4408	2.297
	$\overline{X} \pm \sigma_X$	1.6014	2.717	2.3173	2.1534	2.0269
4	$\overline{X} \pm 0.5\sigma_X$	5.4131*	3.7593	3.222	2.933	2.7622
	$\overline{X} \pm \sigma_X$	2.3225	3.303	2.9283	2.764	2.5846
12	$\overline{X} \pm 0.5\sigma_X$	5.7327*	4.55	4.2148	3.9518	3.6573
	$\overline{X} \pm \sigma_X$	2.9625	4.5623	3.9369	3.5926	3.4326
		Panel D: SVJ model				
1	$\overline{X} \pm 0.5\sigma_X$	3.9196*	2.4181	2.182	1.7904	1.7674
	$\overline{X} \pm \sigma_X$	1.8143*	1.9758	1.6048	1.5391	1.4311
2	$\overline{X} \pm 0.5\sigma_X$	4.9829*	2.9419	2.5982	2.3168	2.2438
	$\overline{X} \pm \sigma_X$	2.2453	2.4841	2.2923	2.0251	1.8797
4	$\overline{X} \pm 0.5\sigma_X$	5.267*	3.6363	3.2304	2.783	2.7213
	$\overline{X} \pm \sigma_X$	2.0763	3.0767	2.8884	2.6128	2.4034
12	$\overline{X} \pm 0.5\sigma_X$	5.5122*	4.5414	4.2131	3.9532	3.6432
	$\overline{X} \pm \sigma_X$	2.9531	4.5539	3.9344	3.5836	3.4191
		Panel E: CHEN model				
1	$\overline{X} \pm 0.5\sigma_X$	1.414	1.5987	1.4189	1.2948	1.2445
	$\overline{X} \pm \sigma_X$	0.7071	1.2897	1.2155	1.0209	0.9702
2	$\overline{X} \pm 0.5\sigma_X$	3.8087*	2.319	2.016	1.8504	1.6675
	$\overline{X} \pm \sigma_X$	0.9211	2.0005	1.5503	1.4258	1.3638
4	$\overline{X} \pm 0.5\sigma_X$	4.6069*	3.0663	2.6457	2.5571	2.3363
	$\overline{X} \pm \sigma_X$	1.5877	2.795	2.1432	1.9696	1.9169
12	$\overline{X} \pm 0.5\sigma_X$	5.8624*	4.4227	4.0782	3.8436	3.6186
	$\overline{X} \pm \sigma_X$	3.2534	4.3738	3.719	3.473	3.2859
		Panel F: CHENJ model				
1	$\overline{X} \pm 0.5\sigma_X$	3.9669*	2.3709	2.1614	1.8063	1.7563
	$\overline{X} \pm \sigma_X$	1.868*	1.9173	1.6121	1.565	1.4412
2	$\overline{X} \pm 0.5\sigma_X$	5.3442*	2.9574	2.6332	2.3669	2.2517
	$\overline{X} \pm \sigma_X$	2.0366	2.5034	2.3738	2.1151	1.9418
4	$\overline{X} \pm 0.5\sigma_X$	5.903*	3.6088	3.3246	2.75	2.7323
	$\overline{X} \pm \sigma_X$	2.3689	3.1013	2.9374	2.6094	2.4238
12	$\overline{X} \pm 0.5\sigma_X$	5.8253*	4.5338	4.2128	3.9499	3.6475
	$\overline{X} \pm \sigma_X$	3.9154	4.5467	3.9285	3.5873	3.3963
		Panel G: CEV model				
1	$\overline{X} \pm 0.5\sigma_X$	2.1957*	1.2735	1.1942	1.0523	1.0303
	$\overline{X} \pm \sigma_X$	0.3725	1.1157	0.9739	0.8576	0.83
2	$\overline{X} \pm 0.5\sigma_X$	3.7249*	1.8083	1.6121	1.5055	1.3886
	$\overline{X} \pm \sigma_X$	0.6918	1.3775	1.2876	1.1652	1.091
4	$\overline{X} \pm 0.5\sigma_X$	5.5548*	2.6036	2.1885	2.0861	2.0038
	$\overline{X} \pm \sigma_X$	1.8547*	1.813	1.7217	1.5895	1.4788
12	$\overline{X} \pm 0.5\sigma_X$	7.8409*	4.0685	3.7453	3.5839	3.4703
	$\overline{X} \pm \sigma_X$	4.5488*	3.6259	3.3886	2.938	2.7929

Table 2: Continued

τ	$(\underline{u},\overline{u})$	Z_T	5% CV	10% CV	15% CV	20% CV
		Panel H: AR(p) model				
1	$\overline{X} \pm 0.5\sigma_X$	0.0444	0.4183	0.3752	0.3542	0.3398
	$\overline{X} \pm \sigma_X$	0.4062	0.4488	0.4177	0.4016	0.3819
2	$\overline{X} \pm 0.5\sigma_X$	0.1865	0.7351	0.6871	0.6545	0.6406
	$\overline{X} \pm \sigma_X$	0.6387	0.7876	0.7119	0.6891	0.6486
4	$\overline{X} \pm 0.5\sigma_X$	0.5506	1.4469	1.3152	1.2448	1.1821
	$\overline{X} \pm \sigma_X$	0.8003	1.3486	1.2614	1.164	1.1345
12	$\overline{X} \pm 0.5\sigma_X$	2.551	3.5545	3.4269	3.1836	3.0745
	$\overline{X} \pm \sigma_X$	0.8417	3.0839	2.9831	2.8258	2.7236
		Panel I: STAR model				
1	$\overline{X} \pm 0.5\sigma_X$	0.3183	4.6367	4.2365	3.8501	3.6128
	$\overline{X} \pm \sigma_X$	0.294	8.9534	8.4574	8.163	7.8028
2	$\overline{X} \pm 0.5\sigma_X$	0.6115	5.8465	5.1462	4.8444	4.2511
	$\overline{X} \pm \sigma_X$	0.486	9.273	8.5681	8.2037	7.9961
4	$\overline{X} \pm 0.5\sigma_X$	0.9774	5.2979	4.7667	4.3599	4.1673
	$\overline{X} \pm \sigma_X$	0.679	6.0715	5.6856	5.168	5.0809
12	$\overline{X} \pm 0.5\sigma_X$	1.228	3.9953	3.4217	3.1175	2.9458
	$\overline{X} \pm \sigma_X$	0.7424	4.0021	3.6167	3.3738	3.0446

Notes: Numerical entries in the table are specification test statistics (Z_T) and 5% , 10% ,15% & 20% nominal level critical values, for tests constructed using intervals given in the second column of the table, and for τ =1, 2, 4, 12 (see discussion in Section 5 for complete details). Single starred entries denote rejection at the 10% level. The simulation periods considered is $10T$, where T denotes the number of observations in the sample. The block length is set equal to 20 observations, and empirical bootstrap distributions are constructed using 100 bootstrap replications. See Section 2 for further discussion of the test reported on in this table.

Table 3: Consistent Specification Test Results - "Stable 1990s" Period

τ	$(\underline{u},\overline{u})$	Z_T	5% CV	10% CV	15% CV	20% CV
		Panel A: CIR model				
1	$\overline{X} \pm 0.5\sigma_X$	1.743	2.6772	2.519	2.3908	2.2301
	$\overline{X} \pm \sigma_X$	1.4215	1.7076	1.5674	1.4895	1.3517
2	$\overline{X} \pm 0.5\sigma_X$	1.9155	2.9635	2.6364	2.4659	2.3378
	$\overline{X} \pm \sigma_X$	1.4151	2.3901	2.1124	1.9591	1.7915
4	$\overline{X} \pm 0.5\sigma_X$	1.7242	3.2826	2.9931	2.7374	2.501
	$\overline{X} \pm \sigma_X$	2.3068	3.2457	3.0113	2.4816	2.3736
12	$\overline{X} \pm 0.5\sigma_X$	1.7088	3.47	2.9462	2.7378	2.3888
	$\overline{X} \pm \sigma_X$	2.6992	3.557	3.1466	2.9817	2.8276
		Panel B: SM model				
1	$\overline{X} \pm 0.5\sigma_X$	2.2283*	1.8019	1.5585	1.4458	1.3906
	$\overline{X} \pm \sigma_X$	0.2205	0.4275	0.3689	0.3253	0.2926
2	$\overline{X} \pm 0.5\sigma_X$	3.3549*	2.4955	2.2605	1.8627	1.7101
	$\overline{X} \pm \sigma_X$	0.3074	0.6781	0.5977	0.5494	0.493
4	$\overline{X} \pm 0.5\sigma_X$	4.3109*	3.1	2.5784	2.347	2.0711
	$\overline{X} \pm \sigma_X$	0.4181	1.1967	1.0937	0.9361	0.86
12	$\overline{X} \pm 0.5\sigma_X$	3.1586	3.8378	3.2441	3.0625	2.7503
	$\overline{X} \pm \sigma_X$	0.8719	2.6774	2.4825	2.4048	2.3169
		Panel C: SV model				
1	$\overline{X} \pm 0.5\sigma_X$	2.1092	2.4951	2.1144	1.9801	1.8458
	$\overline{X} \pm \sigma_X$	0.2203	0.4064	0.3616	0.3616	0.3169
2	$\overline{X} \pm 0.5\sigma_X$	4.3226*	3.6056	2.8024	2.6646	2.5099
	$\overline{X} \pm \sigma_X$	0.2205	0.6309	0.5861	0.5412	0.4964
4	$\overline{X} \pm 0.5\sigma_X$	5.6127*	3.7268	3.2304	2.6068	2.3773
	$\overline{X} \pm \sigma_X$	0.221	1.171	1.1218	1.0363	1.032
12	$\overline{X} \pm 0.5\sigma_X$	7.2161*	4.8256	4.3276	4.0473	3.6399
	$\overline{X} \pm \sigma_X$	1.1136	3.0511	2.9087	2.8182	2.5466
		Panel D: SVJ model				
1	$\overline{X} \pm 0.5\sigma_X$	3.9202*	2.3198	1.972	1.8152	1.7045
	$\overline{X} \pm \sigma_X$	1.2275*	1.0654	0.9135	0.8631	0.7727
2	$\overline{X} \pm 0.5\sigma_X$	5.6993*	2.91	2.5822	2.2936	2.1378
	$\overline{X} \pm \sigma_X$	3.4687*	2.2945	1.9649	1.9119	1.6689
4	$\overline{X} \pm 0.5\sigma_X$	8.0184*	3.1449	2.8082	2.6843	2.3019
	$\overline{X} \pm \sigma_X$	4.6234*	3.1318	2.9289	2.3785	2.2673
12	$\overline{X} \pm 0.5\sigma_X$	7.1228*	3.5211	2.9906	2.5983	2.4542
	$\overline{X} \pm \sigma_X$	3.2521	3.6154	3.2698	3.0074	2.8243
		Panel E: CHEN model				
1	$\overline{X} \pm 0.5\sigma_X$	1.1898*	1.2982	1.1381	1.0555	0.9659
	$\overline{X} \pm \sigma_X$	0.0441	0.3134	0.2686	0.2686	0.2686
2	$\overline{X} \pm 0.5\sigma_X$	3.7492*	3.4063	2.779	2.5541	2.3748
	$\overline{X} \pm \sigma_X$	0.1764	0.6259	0.5363	0.4915	0.4915
4	$\overline{X} \pm 0.5\sigma_X$	5.3033*	3.7219	3.1929	2.5194	2.3847
	$\overline{X} \pm \sigma_X$	0.8397	1.3842	1.2046	1.1148	0.9984
12	$\overline{X} \pm 0.5\sigma_X$	9.3987*	3.8648	3.4118	3.197	2.9153
	$\overline{X} \pm \sigma_X$	3.3408*	3.1894	2.9631	2.7976	2.6462
		Panel F: CHENJ model				
1	$\overline{X} \pm 0.5\sigma_X$	1.2177	2.661	2.2292	1.9944	1.788
	$\overline{X} \pm \sigma_X$	2.024*	2.249	1.9447	1.8089	1.5967
2	$\overline{X} \pm 0.5\sigma_X$	1.9958	3.5217	3.0749	2.8393	2.6268
	$\overline{X} \pm \sigma_X$	3.5283*	3.2828	2.8685	2.608	2.3076
4	$\overline{X} \pm 0.5\sigma_X$	3.6745*	3.462	3.0426	2.8255	2.5908
	$\overline{X} \pm \sigma_X$	7.8632*	3.403	3.0572	2.7751	2.5688
12	$\overline{X} \pm 0.5\sigma_X$	4.0284*	3.5635	3.0355	2.7559	2.3334
	$\overline{X} \pm \sigma_X$	9.4419*	3.4723	3.0325	2.9193	2.7301
		Panel G: CEV model				
1	$\overline{X} \pm 0.5\sigma_X$	1.909	2.7254	2.2576	2.0871	1.9806
	$\overline{X} \pm \sigma_X$	2.5445*	0.4025	1.2714	1.1803	1.0555
2	$\overline{X} \pm 0.5\sigma_X$	2.0945	2.8286	2.6007	2.1757	2.1263
	$\overline{X} \pm \sigma_X$	2.8457*	1.8066	1.5382	1.5142	1.4221
4	$\overline{X} \pm 0.5\sigma_X$	1.9559	3.0052	2.7391	2.4398	2.3126
	$\overline{X} \pm \sigma_X$	3.4883*	2.6681	2.4337	2.1319	2.0166
12	$\overline{X} \pm 0.5\sigma_X$	1.1626	3.0454	2.798	2.4371	2.3048
	$\overline{X} \pm \sigma_X$	3.1809	3.4433	3.3017	3.0855	2.9496
		Panel H: AR(p) model				
1	$\overline{X} \pm 0.5\sigma_X$	0.4117	0.4918	0.4532	0.395	0.3444
	$\overline{X} \pm \sigma_X$	0.1058	0.3456	0.3066	0.2778	0.2695
2	$\overline{X} \pm 0.5\sigma_X$	0.6637	0.7995	0.6725	0.6119	0.5697
	$\overline{X} \pm \sigma_X$	0.1763	0.5604	0.5116	0.4703	0.438
4	$\overline{X} \pm 0.5\sigma_X$	0.6973	1.241	1.1054	1.0731	0.9899
	$\overline{X} \pm \sigma_X$	0.8105	0.9317	0.9054	0.8772	0.8564
12	$\overline{X} \pm 0.5\sigma_X$	0.7951	2.4187	2.2064	2.0915	2.0343
	$\overline{X} \pm \sigma_X$	2.779*	2.5623	2.4562	2.3126	2.2323
		Panel I: STAR model				
1	$\overline{X} \pm 0.5\sigma_X$	0.4062	1.2184	1.0819	0.9901	0.9315
	$\overline{X} \pm \sigma_X$	0.1052	1.1126	0.9989	0.8855	0.8115
2	$\overline{X} \pm 0.5\sigma_X$	0.6334	1.3577	1.2446	1.1096	0.9951
	$\overline{X} \pm \sigma_X$	0.2104	1.4041	1.3429	1.2668	1.2481
4	$\overline{X} \pm 0.5\sigma_X$	0.5263	1.6865	1.4475	1.2701	1.0981
	$\overline{X} \pm \sigma_X$	0.9331	1.6739	1.5739	1.4553	1.3774
12	$\overline{X} \pm 0.5\sigma_X$	0.8341	2.4829	2.0842	1.9407	1.8637
	$\overline{X} \pm \sigma_X$	3.2387*	2.8234	2.6331	2.5385	2.4277

Notes: See notes to Table 2.

Table 4: Consistent Specification Test Results - "Post 1990s" Period

τ	$(\underline{u},\overline{u})$	Z_T	5% CV	10% CV	15% CV	20% CV
		Panel A: CIR model				
1	$\overline{X}\pm 0.5\sigma_X$	0.4175	1.1416	0.969	0.896	0.8026
	$\overline{X}\pm \sigma_X$	1.4744*	1.6624	1.3556	1.2694	1.1859
2	$\overline{X}\pm 0.5\sigma_X$	0.9852	1.9374	1.6573	1.5173	1.2902
	$\overline{X}\pm \sigma_X$	1.7948	2.5406	2.272	1.907	1.8487
4	$\overline{X}\pm 0.5\sigma_X$	1.5983	3.0862	2.5308	2.2108	1.8881
	$\overline{X}\pm \sigma_X$	1.904	3.4499	3.166	2.6652	2.4737
12	$\overline{X}\pm 0.5\sigma_X$	1.3776	3.6922	3.0861	2.7607	2.6046
	$\overline{X}\pm \sigma_X$	4.8384*	4.9892	4.3831	4.113	3.8376
		Panel B: SM model				
1	$\overline{X}\pm 0.5\sigma_X$	1.6069	3.5855	3.3937	3.1795	2.8959
	$\overline{X}\pm \sigma_X$	3.0268	4.1986	3.9257	3.4385	3.3061
2	$\overline{X}\pm 0.5\sigma_X$	1.5331	3.6565	3.3421	3.2023	2.8789
	$\overline{X}\pm \sigma_X$	2.0864	4.2483	3.7961	3.3232	3.2102
4	$\overline{X}\pm 0.5\sigma_X$	1.4989	3.4702	3.1957	3.0996	2.8728
	$\overline{X}\pm \sigma_X$	1.786	4.2643	3.764	3.3165	3.1994
12	$\overline{X}\pm 0.5\sigma_X$	1.0684	3.3503	2.9153	2.6545	2.473
	$\overline{X}\pm \sigma_X$	2.7817	4.4871	3.9436	3.752	3.4919
		Panel C: SV model				
1	$\overline{X}\pm 0.5\sigma_X$	0.3163	0.6696	0.5703	0.4852	0.461
	$\overline{X}\pm \sigma_X$	1.0861*	1.252	1.0647	0.9806	0.8843
2	$\overline{X}\pm 0.5\sigma_X$	0.5631	1.1241	0.9481	0.8816	0.8139
	$\overline{X}\pm \sigma_X$	1.8053*	1.7908	1.6109	1.4518	1.3552
4	$\overline{X}\pm 0.5\sigma_X$	1.3555	1.9716	1.6916	1.3984	1.3023
	$\overline{X}\pm \sigma_X$	2.9899*	2.8014	2.6081	2.2876	2.1157
12	$\overline{X}\pm 0.5\sigma_X$	3.7886*	3.6672	3.2848	2.7219	2.2963
	$\overline{X}\pm \sigma_X$	4.2665*	4.7145	4.0067	3.769	3.5775
		Panel D: SVJ model				
1	$\overline{X}\pm 0.5\sigma_X$	2.4083*	1.8806	1.8207	1.6944	1.6047
	$\overline{X}\pm \sigma_X$	6.1295*	3.5185	3.3276	3.1516	2.7994
2	$\overline{X}\pm 0.5\sigma_X$	5.8102*	3.5004	2.9123	2.7156	2.5503
	$\overline{X}\pm \sigma_X$	9.4311*	4.5093	3.8751	3.5117	3.2937
4	$\overline{X}\pm 0.5\sigma_X$	13.7926*	3.7881	3.4008	3.1998	3.0058
	$\overline{X}\pm \sigma_X$	9.4604*	4.416	3.8425	3.6305	3.3264
12	$\overline{X}\pm 0.5\sigma_X$	15.1804*	3.7861	3.5181	3.2501	3.1429
	$\overline{X}\pm \sigma_X$	9.568*	5.2128	4.1002	3.9262	3.7118
		Panel E: CHEN model				
1	$\overline{X}\pm 0.5\sigma_X$	0.6316	1.1613	1.0557	0.9502	0.8974
	$\overline{X}\pm \sigma_X$	4.4211*	3.8638	3.2832	2.9446	2.7554
2	$\overline{X}\pm 0.5\sigma_X$	3.4429	3.9571	3.4538	3.1609	3.0027
	$\overline{X}\pm \sigma_X$	4.4799*	4.1628	3.4456	3.2871	2.8943
4	$\overline{X}\pm 0.5\sigma_X$	8.6148*	6.2591	5.8883	4.8281	4.7224
	$\overline{X}\pm \sigma_X$	4.4924*	4.4947	3.9868	3.4023	3.2448
12	$\overline{X}\pm 0.5\sigma_X$	3.4744*	3.2064	2.9047	2.5646	2.4222
	$\overline{X}\pm \sigma_X$	4.3134*	6.8154	6.5045	6.0347	5.5766
		Panel F: CHENJ model				
1	$\overline{X}\pm 0.5\sigma_X$	1.8064	3.6712	3.4365	3.0852	2.82
	$\overline{X}\pm \sigma_X$	1.9258	4.1339	3.5039	3.3921	3.1739
2	$\overline{X}\pm 0.5\sigma_X$	1.1563	3.5044	3.363	3.105	2.9053
	$\overline{X}\pm \sigma_X$	3.4039	4.2549	3.6242	3.3535	3.2984
4	$\overline{X}\pm 0.5\sigma_X$	1.6572	3.4857	3.1576	3.0951	2.8673
	$\overline{X}\pm \sigma_X$	4.7218*	4.356	3.5848	3.3104	3.1274
12	$\overline{X}\pm 0.5\sigma_X$	1.6528	3.2081	2.9162	2.6076	2.4327
	$\overline{X}\pm \sigma_X$	5.6215	4.5307	4.0599	3.7381	3.4972
		Panel G: CEV model				
1	$\overline{X}\pm 0.5\sigma_X$	0.5804	1.008	0.89	0.7492	0.6945
	$\overline{X}\pm \sigma_X$	1.6882*	1.4622	1.2565	1.0053	0.9334
2	$\overline{X}\pm 0.5\sigma_X$	1.1163	1.6335	1.4195	1.1507	1.0892
	$\overline{X}\pm \sigma_X$	2.3544*	2.0459	1.6594	1.4875	1.2815
4	$\overline{X}\pm 0.5\sigma_X$	1.7105	2.4312	2.0406	1.8266	1.7001
	$\overline{X}\pm \sigma_X$	3.0353*	2.6707	2.3156	2.0381	1.8591
12	$\overline{X}\pm 0.5\sigma_X$	2.5103	3.0373	2.635	2.5062	2.3773
	$\overline{X}\pm \sigma_X$	4.4257*	4.1555	3.5208	3.2686	3.1363
		Panel H: AR(p) model				
1	$\overline{X}\pm 0.5\sigma_X$	0.0698	0.2327	0.1988	0.1856	0.175
	$\overline{X}\pm \sigma_X$	0.1035	0.3220	0.2842	0.2637	0.2387
2	$\overline{X}\pm 0.5\sigma_X$	0.1181	0.4517	0.3979	0.3847	0.3628
	$\overline{X}\pm \sigma_X$	0.2423	0.6653	0.5455	0.4939	0.4595
4	$\overline{X}\pm 0.5\sigma_X$	0.1032	0.7963	0.6644	0.6244	0.6069
	$\overline{X}\pm \sigma_X$	0.4478	1.2500	1.0588	0.9773	0.8505
12	$\overline{X}\pm 0.5\sigma_X$	0.8256	1.9280	1.6573	1.4849	1.4053
	$\overline{X}\pm \sigma_X$	1.0225	2.8022	2.5988	2.3693	2.2606
		Panel I: STAR model				
1	$\overline{X}\pm 0.5\sigma_X$	0.0325	3.2741	2.8362	2.4874	2.3679
	$\overline{X}\pm \sigma_X$	0.1441	2.4418	2.0411	1.7679	1.6145
2	$\overline{X}\pm 0.5\sigma_X$	0.0456	1.8045	1.5337	1.4612	1.2953
	$\overline{X}\pm \sigma_X$	0.2875	2.4282	2.0894	1.7271	1.6094
4	$\overline{X}\pm 0.5\sigma_X$	0.1238	1.8647	1.6553	1.3366	1.069
	$\overline{X}\pm \sigma_X$	0.4477	2.9479	2.6598	2.4275	2.3758
12	$\overline{X}\pm 0.5\sigma_X$	0.9691	2.6775	2.2027	2.1023	1.855
	$\overline{X}\pm \sigma_X$	0.9583	3.5868	3.4019	3.1724	2.9641

Notes: See notes to Table 2.

Table 5: Is There An Overall Winner?

τ	$(\underline{u},\overline{u})$	Z_T	5% CV	10% CV	15% CV	20% CV
	Panel A: Compare CHEN model from the whole sample to CIR model from the Stable 1990s					
1	$\overline{X} \pm 0.5\sigma_X$	2.3388	2.7842	2.6247	2.4437	2.2867
	$\overline{X} \pm \sigma_X$	4.1080*	1.9003	1.7177	1.628	1.4355
2	$\overline{X} \pm 0.5\sigma_X$	3.0640*	2.9183	2.7458	2.6412	2.4222
	$\overline{X} \pm \sigma_X$	5.6435*	2.5118	2.2071	2.0496	1.9132
4	$\overline{X} \pm 0.5\sigma_X$	3.6340*	3.2459	2.7962	2.585	2.4011
	$\overline{X} \pm \sigma_X$	7.4294*	2.9834	2.7666	2.4598	2.3156
12	$\overline{X} \pm 0.5\sigma_X$	3.6097*	3.001	2.7578	2.4197	2.2585
	$\overline{X} \pm \sigma_X$	9.3599*	3.4502	3.3648	3.1227	2.8756
	Panel B: Compare CHEN model from whole sample to SM model from the Post 1990s					
1	$\overline{X} \pm 0.5\sigma_X$	1.9108	2.4683	2.1285	2.0068	1.9351
	$\overline{X} \pm \sigma_X$	1.9867	2.7849	2.58	2.4232	2.1239
2	$\overline{X} \pm 0.5\sigma_X$	2.2265	3.0523	2.7975	2.6313	2.4691
	$\overline{X} \pm \sigma_X$	2.2016	3.4218	2.9702	2.7547	2.5439
4	$\overline{X} \pm 0.5\sigma_X$	2.0204	3.3607	3.0712	2.9132	2.8101
	$\overline{X} \pm \sigma_X$	2.1265	3.9741	3.4823	3.1128	2.9762
12	$\overline{X} \pm 0.5\sigma_X$	1.2257	3.176	3.0793	2.8596	2.6381
	$\overline{X} \pm \sigma_X$	2.7404	4.7821	3.8476	3.7598	3.5084

Notes: See notes to Table 2.

Table 6: Prediction Based Model Selection Test Results (CHEN is the benchmark) – "Whole Sample" Period

τ	$(\underline{u},\overline{u})$	$D_{k,P,N}$	MSFE_{CHEN}	MSFE_{STAR}	5% CV	10% CV	15% CV	20% CV
1	$\overline{X} \pm 0.5\sigma_X$	-1.2533	0.3487	1.602	0.2807	0.1267	0.0622	0.0253
	$\overline{X} \pm \sigma_X$	0.0524	0.1628	0.1103	0.5598	0.3925	0.3337	0.191
2	$\overline{X} \pm 0.5\sigma_X$	-2.1733	0.5868	2.7601	0.5563	0.3636	0.2688	0.2397
	$\overline{X} \pm \sigma_X$	-0.3811	0.0821	0.4632	0.716	0.6226	0.4921	0.3754
3	$\overline{X} \pm 0.5\sigma_X$	-2.1516	0.8527	3.0043	0.8915	0.6571	0.5959	0.5335
	$\overline{X} \pm \sigma_X$	-1.6442	0.166	1.8103	0.8152	0.6854	0.5327	0.3758
4	$\overline{X} \pm 0.5\sigma_X$	-2.0863	1.0789	3.1652	1.1685	0.9784	0.9076	0.8682
	$\overline{X} \pm \sigma_X$	-2.207	0.2523	2.4594	0.5984	0.4198	0.2145	0.1305
5	$\overline{X} \pm 0.5\sigma_X$	-2.0529	1.3319	3.3847	1.6892	1.3829	1.259	1.1313
	$\overline{X} \pm \sigma_X$	-2.5249	0.3417	2.8666	0.7489	0.5714	0.413	0.2415
6	$\overline{X} \pm 0.5\sigma_X$	-1.8553	1.8636	3.7189	2.1268	1.8348	1.7182	1.642
	$\overline{X} \pm \sigma_X$	-2.7208	0.4247	3.1454	0.9263	0.6119	0.4027	0.308
12	$\overline{X} \pm 0.5\sigma_X$	-2.3028	2.9195	5.2223	3.5577	3.386	3.0012	2.8985
	$\overline{X} \pm \sigma_X$	-3.6963	0.6874	4.3836	1.8131	1.5523	1.3649	1.2853

Notes: See also notes to Table 2. Entries in the table are ($D_{k,P,N}$) test statistics, and predictive density type MSFE values constructed using intervals given in the second column of the table, for τ =1, 2, 3, 4, 5, 6, 12 step(s) ahead prediction periods. (Note that MSFE values are simply the (appropriately scaled) individual terms appearing in the $D_{k,P,N}$ test statistic (see Section 2 for further details)). The simulation period considered is $10T$, where T denotes the number of observations in the sample. Additionally, the ex-ante prediction period $P = T/2$, and all model parameters are estimated recursively. For critical value construction, the block length is set equal to 20 observations, and empirical bootstrap distributions are constructed using 100 bootstrap replications. Finally, critical values for various percentiles of the bootstrap distribution are reported. Finally, $\overline{X}$ and σ_X are the mean and variance of an initial sample of data.

Table 7: Prediction Based Model Selection Test Results (CIR is the benchmark) – "Stable 1990s" Period

τ	$(\underline{u},\overline{u})$	$D_{k,P,N}$	MSFE$_{CIR}$	MSFE$_{STAR}$	5% CV	10% CV	15% CV	20% CV
1	$\overline{X} \pm 0.5\sigma_X$	0.7303*	4.2423	3.512	-0.044	-0.2124	-0.3583	-0.5073
	$\overline{X} \pm \sigma_X$	-0.5505	2.2923	2.8428	-0.3547	-0.4337	-0.5596	-0.6139
2	$\overline{X} \pm 0.5\sigma_X$	0.3494*	3.1674	2.818	-0.4476	-0.5722	-0.6403	-0.7515
	$\overline{X} \pm \sigma_X$	-0.5537*	2.0901	2.6438	-0.8879	-1.1865	-1.3721	-1.4046
3	$\overline{X} \pm 0.5\sigma_X$	0.4079*	3.2219	2.814	-0.4925	-0.7128	-0.8527	-0.9482
	$\overline{X} \pm \sigma_X$	0.6978*	2.642	1.9442	-1.3579	-1.5276	-1.5741	-1.7505
4	$\overline{X} \pm 0.5\sigma_X$	0.4366*	3.4308	2.9942	-0.5974	-0.765	-0.909	-0.9631
	$\overline{X} \pm \sigma_X$	0.455*	2.5844	2.1294	-1.0219	-1.2414	-1.3949	-1.4966
5	$\overline{X} \pm 0.5\sigma_X$	0.2775*	3.4751	3.1975	-0.4922	-0.6762	-0.7841	-0.8996
	$\overline{X} \pm \sigma_X$	0.2747*	2.5483	2.2736	-1.2877	-1.5109	-1.6417	-1.8064
6	$\overline{X} \pm 0.5\sigma_X$	0.0689*	3.4393	3.3704	-0.3984	-0.7375	-0.9125	-1.0316
	$\overline{X} \pm \sigma_X$	0.3425*	2.9586	2.6161	-1.6824	-1.8853	-2.0199	-2.0992
12	$\overline{X} \pm 0.5\sigma_X$	-0.3276	3.5943	3.9219	-0.1181	-0.2504	-0.3928	-0.4924
	$\overline{X} \pm \sigma_X$	-1.4209	3.0032	4.424	-1.1303	-1.3781	-1.4943	-1.6525

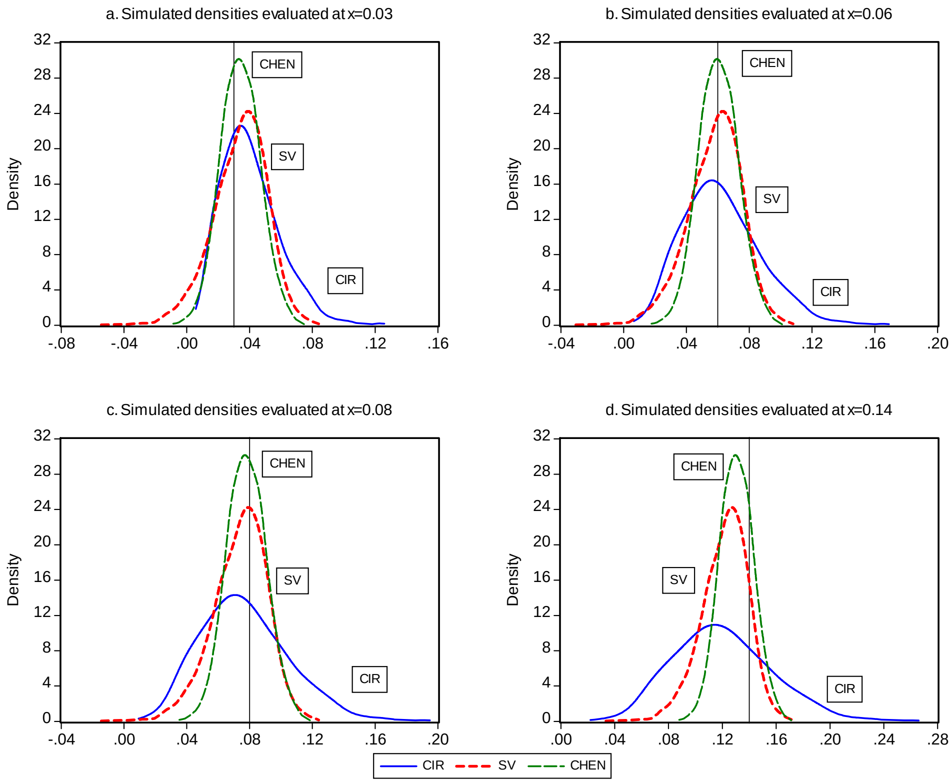
Notes: See notes to Table 6.

Table 8: Prediction Based Model Selection Test Results (SM is the benchmark) – "Post 1990s" Period

τ	$(\underline{u},\overline{u})$	$D_{k,P,N}$	MSFE$_{SM}$	MSFE$_{STAR}$	5% CV	10% CV	15% CV	20% CV
1	$\overline{X} \pm 0.5\sigma_X$	4.7024*	4.9844	0.282	3.6057	3.3603	3.1556	3.0401
	$\overline{X} \pm \sigma_X$	4.8998*	5.0651	0.1653	4.7062	4.4529	4.3239	4.1003
2	$\overline{X} \pm 0.5\sigma_X$	4.4802*	5.0155	0.5353	2.6576	2.4374	2.316	2.1975
	$\overline{X} \pm \sigma_X$	4.9238*	5.2451	0.3212	3.6789	3.5336	3.2743	3.0682
3	$\overline{X} \pm 0.5\sigma_X$	3.7971*	4.6144	0.8173	2.0624	1.9927	1.7629	1.6506
	$\overline{X} \pm \sigma_X$	4.6335*	5.1087	0.4751	2.8992	2.7232	2.4348	2.3148
4	$\overline{X} \pm 0.5\sigma_X$	3.1860*	4.2874	1.1014	1.5053	1.4101	1.3229	1.2321
	$\overline{X} \pm \sigma_X$	4.2125*	4.7276	0.5151	2.4542	2.0808	1.900	1.7100
5	$\overline{X} \pm 0.5\sigma_X$	2.6939*	4.1103	1.4164	1.3738	1.1603	1.0765	0.9844
	$\overline{X} \pm \sigma_X$	3.9075*	4.5096	0.6021	2.2932	1.631	1.5173	1.4187
6	$\overline{X} \pm 0.5\sigma_X$	2.5253*	4.1587	1.6333	1.3632	1.0159	0.901	0.8401
	$\overline{X} \pm \sigma_X$	3.7711*	4.413	0.6419	2.1365	1.4571	1.292	1.0949
12	$\overline{X} \pm 0.5\sigma_X$	0.8495*	3.5725	2.723	0.7741	0.4611	0.3268	0.0644
	$\overline{X} \pm \sigma_X$	2.9897*	4.2874	1.2977	0.786	0.5256	0.349	0.0396

Notes: See notes to Table 6.

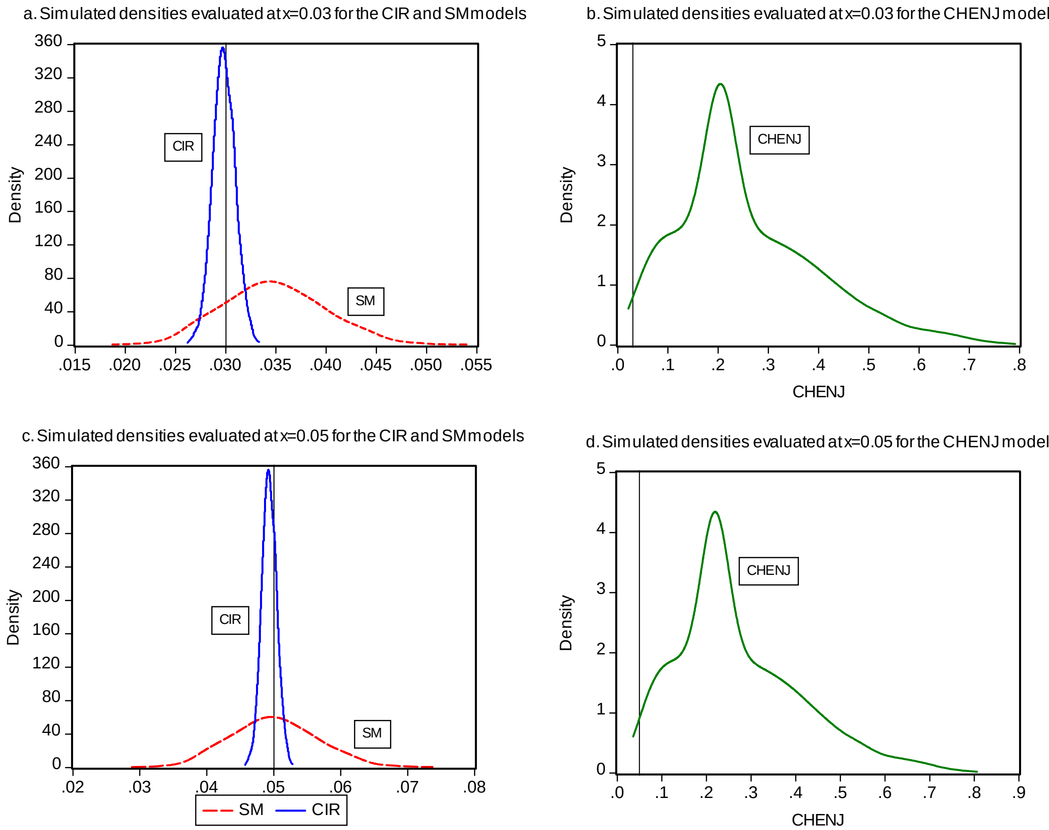
Figure 2: Simulated Densities For The CIR, SV And CHEN Models - "Whole Sample" (06/1982-04/2008)



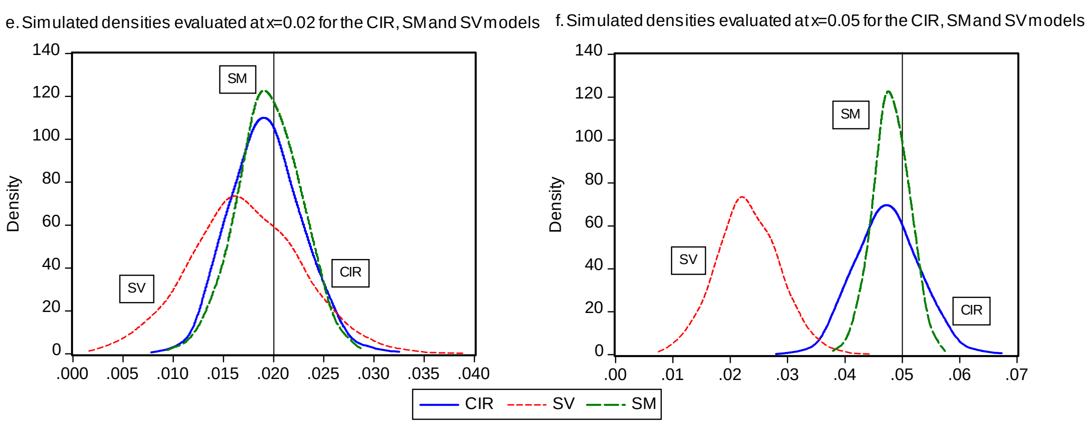
Note: This figure contained kernel density estimates for selected models and selected evaluation points, where evaluation points are taken from the support of the historical data, and correspond roughly to regions of the support sociated with mean or model behavior, as well as tail behavior.

Figure 3: Simulated Densities For ``Stable 1990s'' and ``Post 1990s'' Periods

Panel A: Simulated densities for the CIR, SM and CHENJ models - "Stable 1990s" (03/1991-05/2001)



Panel B: Simulated densities for the CIR, SM and SV models - "Post 1990s" (06/2001-04/2008)

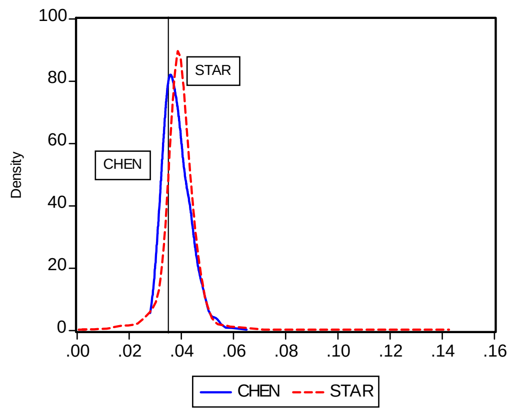


Notes: See Figure 2.

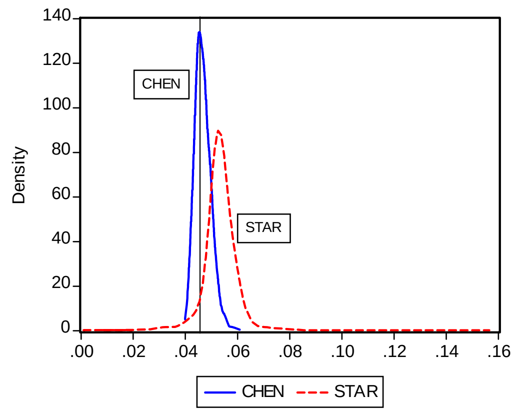
Figure 4: Out-of-sample Simulated Densities Comparison

Panel A: Simulated densities for the CHEN and STAR models - "Whole Sample" (06/1982-04/2008)

a. Simulated densities evaluated at x=0.035 for the CHEN and STAR models

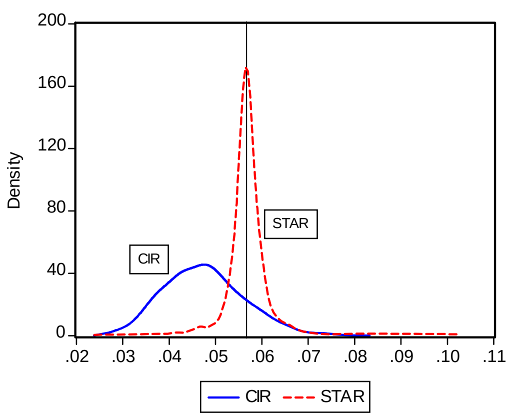


b. Simulated densities evaluated at x=0.0456 for the CHEN and STAR models

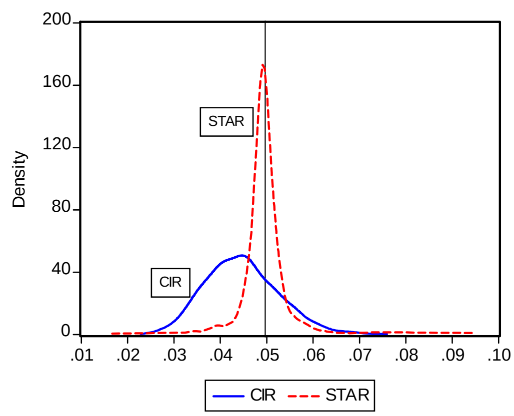


Panel B: Simulated densities for the CIR and STAR models - "Stable 1990s" (03/1991-05/2001)

c. Simulated densities evaluated at x=0.0567 for the CIR and STAR models

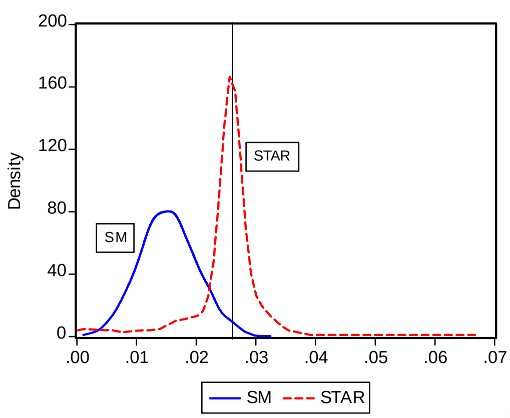


d. Simulated densities evaluated at x=0.0497 for the CIR and STAR models

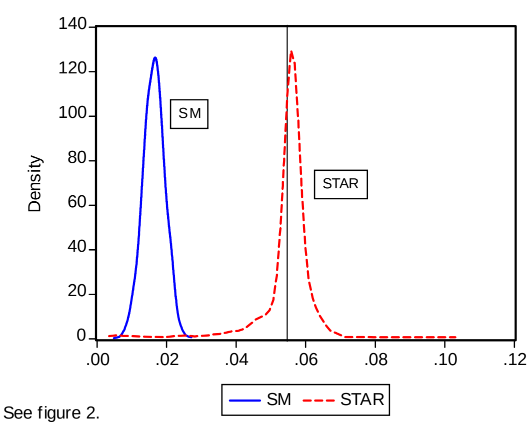


Panel C: Simulated densities for the SM and STAR models - "Post 1990s" (06/2001-04/2008)

e. Simulated densities evaluated at x=0.0261 for the SM and STAR models



f. Simulated densities evaluated at x=0.0547 for the SM and STAR models



Note: See figure 2.