

Nonparametric Specification Testing for Continuous-Time Models with Application to Spot Interest Rates

Yongmiao Hong* and Haitao Li**

March, 2002

*Hong is from Department of Economics and Department of Statistical Science, Cornell University, Ithaca, New York 14853, Phone: (607) 255 5130, Fax: (607) 255 2818, E-mail: yh20@cornell.edu.

**Li is from Johnson Graduate School of Management, Cornell University, Ithaca, NY 14853, Phone: (607) 255 4961, Fax: (607) 254 4590, E-mail: hl70@cornell.edu.

We would like to thank seminar participants at Cornell University and the 2002 American Finance Association Annual Meeting in Atlanta, especially our discussant Andrew Ang for their comments, and Yacine Ait-Sahalia for providing us with the Eurodollar interest rate data. Support from the National Science Foundation under Grant SES-0111769 and computational assistance from the Parker Center at Johnson Graduate School of Management and Doug Elias are greatly appreciated.

Nonparametric Specification Testing for Continuous-Time Models with Application to Spot Interest Rates

Abstract

We propose two nonparametric transition density-based specification tests for continuous-time diffusion models. In contrast to marginal density as used in the literature, transition density can capture the full dynamics of a diffusion process, and in particular, can distinguish processes with the same marginal density but different transition densities. To address the concerns of the finite sample performance of nonparametric methods in the literature, we introduce an appropriate data transformation and correct the boundary bias of kernel estimators. As a result, our tests are robust to persistent dependence in data and provide reliable inferences for sample sizes often encountered in empirical finance. Simulation studies show that our tests have reasonable size and good power against a variety of alternatives in finite samples even for data with highly persistent dependence. Besides the single-factor diffusion models, our tests can be applied to a broad class of dynamic economic models, such as discrete time series models, time-inhomogeneous diffusion models, stochastic volatility models, jump-diffusion models, and multi-factor term structure models. When applied to daily Eurodollar interest rates, our tests overwhelmingly reject some popular spot rate models, including those with nonlinear drifts that some existing tests can not reject after correcting size distortions. We find that models with nonlinear drifts do not significantly improve the goodness-of-fit, and the main source of model inadequacy seems to be the violation of the Markov assumption. We also find that GARCH, regime switching and jump diffusion models perform significantly better than single-factor diffusion models, although they are far from being adequate to fully capture the interest rate dynamics. Our study shows that nonparametric methods are a reliable and powerful tool for analyzing financial data.

Key words: Boundary bias, Continuous-time model, Hellinger metric, Kernel method, Parameter estimation uncertainty, Probability integral transform, Quadratic form, Short-term interest rate, Transition density.

JEL Classifications: C4, E4, G0.

Continuous-time diffusion models have been widely used in finance to capture the dynamics of important economic variables, such as interest rates, exchange rates and stock prices. The well-known option pricing model of Black and Scholes (1973), and the term structure models of Cox, Ingersoll and Ross (1985) and Heath, Jarrow and Morton (1992), for example, all assume that the underlying state variables follow a diffusion process. Economic theories typically do not suggest functional forms for diffusion models, and researchers often consider parametric specifications that are convenient for deriving closed-form solutions for various security prices.

The last decade has seen the development of a large and still growing academic literature on estimation and testing of continuous-time models.¹ One major focus in the literature is to develop rigorous econometric methods to estimate continuous-time models using discretely-sampled data. This is largely motivated by Lo's (1988) finding that estimating the discretized version of a continuous-time model can result in inconsistent parameter estimates. Available estimation procedures include the nonparametric methods of Ait-Sahalia (1996) and Stanton (1997), the simulated method of moments of Duffie and Singleton (1993), the efficient method of moments of Gallant and Tauchen (1996,2001), the generalized method of moments of Hansen and Scheinkman (1995), and the maximum likelihood methods of Lo (1988) and Ait-Sahalia (2001), among many others. Asymptotic properties of these estimators have been well established and inference procedures based on the asymptotic theory have been developed.

The validity of these asymptotic theories and inference procedures, however, crucially depends on the maintained assumption that the underlying model is correctly specified. In the present context, model misspecification generally renders inconsistent estimators of parameters and their variance-covariance matrices, which could lead to misleading conclusions in inference and hypothesis testing. Furthermore, a misspecified model can yield large errors in pricing derivatives and measuring financial risk (e.g., values at risk). It is therefore important to develop reliable specification tests for continuous-time diffusion models.

In contrast to the rapid development of parameter estimation methods, there is relative little effort on specification analysis for continuous-time models (cf. Ait-Sahalia 1996, Gallant and Tauchen 1996, Conley, Hansen, Luttmer and Scheinkman 1997). In a pioneering paper, Ait-Sahalia (1996) develops probably the first nonparametric test for time-homogeneous single-factor diffusion models. Observing the fact that the drift and diffusion functions completely characterize the stationary (or marginal) density of a diffusion model, Ait-Sahalia (1996) compares a model-implied marginal density estimator with a nonparametric kernel density estimator based on a

¹Sundaresan (2001) states that "perhaps the most significant development in the continuous-time field during the last decade has been the innovations in econometric theory and in the estimation techniques for models in continuous time." For other reviews of the recent literature, see, e.g., Melino (1994), Tauchen (1997, 2000), and Campbell, Lo and MacKinlay (1997).

discretely sampled data. The parametric model is rejected if the difference between the two estimators is sufficiently large. The nonparametric test does not make restrictive assumptions on the data generating process and can detect a wide range of alternatives. This appealing power property is not shared by parametric approaches such as conditional moment tests. The latter has optimal power against certain alternatives (depends on the choice of moment functions) but may be completely silent against other alternatives. In an application to daily Eurodollar interest rates, Ait-Sahalia (1996) rejects all existing one-factor spot rate models and find that “the principal source of rejection of existing models is the strong nonlinearity of the drift.” Stanton (1997), using nonparametric kernel regression, also finds evidence of a nonlinear drift. These findings have motivated the development of nonlinear term structure models such as that of Ahn and Gao (1999).

The evidence of a nonlinear drift, however, has been recently challenged by Chapman and Pearson (2000). They find that the nonparametric methods of Ait-Sahalia (1996) and Stanton (1997) produce biased estimates near the boundaries of the observations, which could have produced spurious nonlinear drifts. Pritsker (1998), on the other hand, documents that the size performance of Ait-Sahalia’s (1996) test appears inadequate even for rather large samples: it requires 2,755 years of daily interest rate data generated by an empirically relevant Vasicek (1977) model to attain the accuracy of a kernel density estimator implied by its asymptotic distribution. The main reasons, as pointed out in Pritsker (1998), are the highly persistent dependence of the interest rate data and the slow convergence of nonparametric estimators. The asymptotic distribution of Ait-Sahalia’s (1996) test statistic remains the same whether the sample is independent and identically distributed (i.i.d.) or persistently dependent, but the level of dependent persistence severely affects the finite sample distribution. The findings of Chapman and Pearson (2000) and Pritsker (1998) thus cast serious doubts on the applicability of nonparametric methods in finance, since persistent dependence is a stylized fact for interest rates and many other high frequency financial data.

Gallant and Tauchen (1996) propose a minimum chi-square specification test for diffusion models using the Efficient Method of Moment (hereafter EMM). They examine the simulation-based expectation of a semi-nonparametric density derivative under the model-implied distribution, which takes the value of zero under correct model specification. Among other things, the greatest appeal of Gallant and Tauchen’s EMM method is that it applies to both single factor and multi-factor diffusion models and in addition to the minimum chi-square test, it also provides a spectrum of constructive individual t -statistics that are informative in revealing possible sources of model misspecification. While the EMM test can detect a wide range of model misspecifications, it may still have no power against certain alternatives, because the semi-nonparametric density derivative may still have zero expectation under the distribution of a misspecified model. Thus,

as pointed out by Gallant and Tauchen (1998, p.), one still cannot conclude that a diffusion model is correctly specified even when the minimum chi-square EMM test statistic is insignificant.

In this paper, we propose two new nonparametric transition density-based specification tests for diffusion models that share the appealing features of both Ait-Sahalia (1996) and Gallant and Tauchen (1996). By using an appropriate data transformation and correcting the boundary bias of kernel estimators, our tests are robust to persistent dependence in data and provide reliable inferences for sample sizes often encountered in empirical finance, with superior finite sample performances. Our results indicate that nonparametric methods remain a reliable and powerful tool for studying financial data. Specifically, our tests contribute to the literature in several directions.

First, unlike Ait-Sahalia's (1996) marginal density-based test, our tests are based on transition density, which, under the maintained Markov assumption, captures the full dynamics of a diffusion process. The use of the marginal density is computationally convenient and can detect a wide range of alternatives. However, it may easily miss the alternatives that have the same marginal density as the null model but have different transition densities. In contrast, the transition density can effectively pick them up. Our tests are computationally more expensive than marginal density-based tests, because the transition density usually has no closed-form solution for most diffusion models. One can approximate the transition density using the simulation methods developed by Pedersen (1995), Brandt and Santa-Clara (2001), and Elerian, Chib, and Shephard (2000). The recent important work of Ait-Sahalia (2001) provides a more convenient method that allows us to obtain an accurate closed-form approximation of the transition density.

Second, to achieve robustness to persistent dependence, we transform the data via a dynamic probability integral transform using the model-implied transition density. The dynamic probability integral transform has been well known in statistics (e.g., Rosenblatt 1952) and is more recently used in out-of-sample density forecasts in discrete time series analysis (e.g., Diebold, Gunther and Tay 1998). In the present continuous-time context, the transformed sequence is i.i.d. $U[0, 1]$ under correct model specification, irrespective of the dependence structure of the original data. We propose two tests for the joint hypothesis of i.i.d. and $U[0, 1]$ by comparing a kernel estimator of the joint density of the transformed data with the product of two $U[0, 1]$ densities. As there is no serial dependence in the transformed series under correct model specification, nonparametric density estimators and related tests are expected to perform well in finite samples.

Third, we introduce a boundary-modified kernel that effectively eliminates the important "boundary bias" of nonparametric kernel estimators as documented in Chapman and Pearson (2000). Surprisingly, the boundary bias has important impact on not only the finite sample performance but also the asymptotic behavior of the test statistics. Simulation studies show that, with the help of the probability integral transform and the boundary-modified kernel, our tests

have reasonable size and good power against a variety of alternatives in finite samples even for data with highly persistent dependence.

Fourth, our transition density-based tests are omnibus tests for model misspecification. After rejecting a misspecified model, our methods can provide additional diagnostics about the possible sources for the rejection by using the probability integral transformed series, which constitutes a “generalized residual” of the model. Specifically, our tests contain information of the autocorregrams in all the moments of the generalized residuals. Thus, to check for possible sources of the rejection, we can construct a spectrum of autocorregram-based tests in every moment of the residuals, which are very informative. Thus, our approach shares the appeals of Gallant and Tauchen’s (1996) EMM method in providing constructive diagnostic tests for the possible sources of model misspecification.

Fifth, in addition to a test based on a quadratic form, we also propose a test based on the Hellinger metric to reduce the impact of parameter estimation uncertainty. It is well-known in the literature (e.g., Merton 1980) that drift parameters are much more difficult to estimate than diffusion parameters: the drift parameter estimates may be imprecise even for relatively large samples. These imprecise estimates may have significant impact on the finite sample distribution of the test statistics, though not on their asymptotic distribution. The Hellinger metric is well known in statistics for its robustness to outliers in data, and we expect that it will alleviate the impact of parameter estimation uncertainty on the finite sample distribution of our test statistics.

Finally, the idea of probability integral transform is very general and as a result our tests can be applied to a wide variety of time series models in addition to the strictly stationary, single-factor diffusion models covered by Ait-Sahalia’s (1996) test. For example, our tests are applicable to time-inhomogeneous diffusion models,² discrete time series models such as GARCH and regime-switching models,³ partially observable nonlinear dynamic models, such as stochastic volatility models (both in discrete and continuous time) and multi-factor term structure models.⁴ The integral transformed data via the model-implied transition density in all the above models

²Time-inhomogeneous models for stock prices and interest rates have been proposed in the literature. Examples include Black, Derman and Toy (1990), Derman and Kani (1994), Hull and White (1990, 1993), and Rubinstein (1994), among others. Ait-Sahalia’s (2001) Hermite approximation approach, which has been extended by Egorov, Li and Xu (2001) to time-inhomogeneous diffusions, can be used to obtain an accurate closed-form approximation of the transition density of those models.

³These models have been widely used in the literature to model important economic and financial time series. The probability integral transform is very easy to perform for discrete time series models, as their conditional probability distribution are typically known in closed form.

⁴Stochastic volatility models have been widely used to model interest rate and stock return dynamics (e.g. see Ghysels, Harvey, and Renault 1995). In those models, the dynamics of the economic variables that we are interested in depend on some underlying latent factors. To obtain the model transition density of the observable variables conditional only on its past history, we need to integrate out the latent variables.

should be i.i.d. $U[0, 1]$ if the model is correctly specified. Any deviation from i.i.d. $U[0, 1]$ would indicate model misspecification, and can be tested by our procedures.

As an application of our new tests, we apply them to the daily Eurodollar interest rates to examine the adequacy of a wide variety of spot rate models including the single-factor diffusion models considered in Ait-Sahalia (1996) and some popular discrete time series models. Based on the same data, Ait-Sahalia (1996) rejects all the diffusion models with linear drift using asymptotic critical values. However, using the empirical critical values obtained through simulation in Pritsker (1998), Ait-Sahalia's (1996) test would not reject certain linear drift models. In contrast, all diffusion models are overwhelmingly rejected by our more reliable procedures. Unlike Ait-Sahalia (1996), we find that to include nonlinear drift does not significantly improve the goodness of fit. Moreover, we find that model misspecification is mainly due to the underprediction of the movements of interest rates around the mean and the violation of the Markov assumption.

Besides the continuous-time diffusion models, a large number of discrete time series interest rate models have also been proposed in the literature. For example, GARCH models have been widely used to capture volatility clustering in the data, regime-switching models have been used to model the potential structural breaks in the data generating process, and jumps have been shown to be essential to model excess kurtosis in interest rate data. We show that these more complex models provide improvements over the single-factor diffusion models, especially in terms of modeling the marginal density of the data and the conditional variance, skewness, and kurtosis for the generalized residuals. Despite these improvements, all the discrete time models are still overwhelmingly rejected by our tests. Our results show that we need to consider more complicated interest rate models in the future.

The rest of the paper is organized as follows. In Section I, we state the hypotheses of interest and review the related literature. In Section II, we introduce our tests and discuss their asymptotic properties. In Section III, we discuss the extensions of our methods. In Section IV, we examine the finite sample size and power performance of the proposed tests via empirically relevant Monte Carlo simulations. In Section V, we apply our tests to daily Eurodollar interest rates to evaluate some popular spot rate models. Section VI concludes. In the Appendix, we provide the regularity conditions and mathematical proofs for the asymptotic theory. A GAUSS code for implementing the proposed tests is available from the authors upon request.

I. Hypotheses and Literature Review

As our tests are most closely related to that of Ait-Sahalia (1996), we first follow Ait-Sahalia (1996) and consider single-factor diffusion processes, for ease of comparison. In Section III, we will discuss the applications of our tests to other dynamic economic models. Suppose that a state

variable $\{X_t\}$ follows a diffusion process

$$dX_t = \mu_0(X_t, t)dt + \sigma_0(X_t, t)dW_t, \quad (1)$$

where $\mu_0(X_t, t)$ and $\sigma_0(X_t, t)$ are the (possibly time-inhomogeneous) drift and diffusion functions respectively, and W_t is a standard Brownian motion. (Ait-Sahalia (1996) considers single factor stationary time-homogeneous diffusion processes.) Often it is assumed that $\mu_0(X_t, t)$ and $\sigma_0(X_t, t)$ belong to some parametric families; that is,

$$\begin{aligned} \mu_0 &\in \mathcal{M}_\mu \equiv \{\mu(\cdot, \cdot, \theta), \theta \in \Theta\}, \\ \sigma_0 &\in \mathcal{M}_\sigma \equiv \{\sigma(\cdot, \cdot, \theta), \theta \in \Theta\}, \end{aligned}$$

where Θ is a finite-dimensional parameter space. We say that the models \mathcal{M}_μ and \mathcal{M}_σ are correctly specified for the drift and diffusion $\mu_0(X_t, t)$ and $\sigma_0(X_t, t)$ respectively if

$$H_0 : P[\mu(X_t, t, \theta_0) = \mu_0(X_t, t), \sigma(X_t, t, \theta_0) = \sigma_0(X_t, t)] = 1 \text{ for some } \theta_0 \in \Theta. \quad (2)$$

Various econometric methods have been developed in the literature to estimate the unknown parameter θ_0 , taking (1) as given. However, these methods generally cannot deliver consistent parameter estimates if the model \mathcal{M}_μ or \mathcal{M}_σ is misspecified in the sense that

$$H_A : P[\mu(X_t, t, \theta) = \mu_0(X_t, t), \sigma(X_t, t, \theta) = \sigma_0(X_t, t)] < 1 \text{ for all } \theta \in \Theta. \quad (3)$$

Under H_A , there exists no parameter value θ such that the drift model $\mu(X_t, t, \theta)$ and the diffusion model $\sigma(X_t, t, \theta)$ coincide with the true drift $\mu_0(X_t, t)$ and the true diffusion $\sigma_0(X_t, t)$ respectively. As a result, misleading conclusions in inference and hypothesis testing can be reached based on the biased parameter estimates. In addition, a misspecified model can also produce large errors for pricing derivatives and calculating financial risk. In this paper, we are interested in testing whether a continuous-time diffusion model is correctly specified using a discrete sample of X_t observed over a time span T at interval of Δ , $\{X_{\tau\Delta}\}_{\tau=1}^n$, where $n = T/\Delta$.

In an influential paper, Ait-Sahalia (1996) observes that for a stationary time-homogeneous diffusion process

$$dX_t = \mu_0(X_t)dt + \sigma_0(X_t)dW_t, \quad (4)$$

a pair of drift and diffusion models $\mu(X_t, \theta)$ and $\sigma(X_t, \theta)$ uniquely determines the stationary (or marginal) density of X_t ,

$$\pi(x, \theta) = \frac{\xi(\theta)}{\sigma^2(x, \theta)} \exp \left[\int_{x_0}^x \frac{2\mu(u, \theta)}{\sigma^2(u, \theta)} du \right], \quad (5)$$

where the standardization factor $\xi(\theta)$ ensures that $\pi(x, \theta)$ integrates to 1 for every $\theta \in \Theta$, and x_0 is the lower bound of the support of X_t . Ait-Sahalia (1996) compares a parametric marginal

density estimator $\pi(x, \hat{\theta})$ with a nonparametric density estimator $\hat{\pi}_0(x)$ via a quadratic form

$$M \equiv \int_{x_0}^{x_1} \left[\hat{\pi}_0(x) - \pi(x, \hat{\theta}) \right]^2 \hat{\pi}_0(x) dx, \quad (6)$$

where x_1 is the upper bound for X_t , $\hat{\theta}$ is a minimum distance estimator

$$\hat{\theta} \equiv \arg \min_{\theta \in \Theta} n^{-1} \sum_{\tau=1}^n [\hat{\pi}_0(X_{\tau\Delta}) - \pi(X_{\tau\Delta}, \theta)]^2, \quad (7)$$

and $\hat{\pi}_0(x)$ is a kernel density estimator

$$\hat{\pi}_0(x) \equiv n^{-1} \sum_{\tau=1}^n K_h(x - X_{\tau\Delta}) \quad (8)$$

with $K_h(x - y) \equiv h^{-1}k\left(\frac{x-y}{h}\right)$. The kernel function $k(u)$ is a prespecified symmetric probability density function, and $h = h(n)$ is a bandwidth such that $h \rightarrow 0, nh \rightarrow \infty$ as $n \rightarrow \infty$. Ait-Sahalia (1996) uses a Gaussian kernel.

The marginal density-based test is simple and easy to implement. It also has power against a wide range of misspecifications that render the model-implied marginal density $\pi(x, \theta)$ different from the true marginal density $\pi_0(x)$. However, recent studies have pointed out several limitations of this test that may make its empirical applicability questionable.

First, as already pointed out in Ait-Sahalia (1996), the marginal density cannot capture the full dynamics of the underlying process. In particular, it cannot distinguish two diffusion models that have the same marginal density but different transition densities.⁵

Second, subject to some regularity conditions on the temporal dependence in X_t , the asymptotic distribution of the quadratic form M in (6) remains the same whether the sample $\{X_{\tau\Delta}\}_{\tau=1}^n$ is i.i.d. or highly persistently dependent (Ait-Sahalia 1996). This convenient asymptotic property unfortunately results in a substantial discrepancy between the asymptotic and finite sample distributions, as documented in Pritsker (1998). This discrepancy and the slow convergence of nonparametric kernel estimators are the main reasons identified by Pritsker (1998) for the poor finite sample performance of the test.

Third, as pointed out by Chapman and Pearson (2000), a conventional kernel density estimator produces biased estimates near the boundaries of the observations, due to the asymmetric coverage of the kernel for the data in the boundary regions. This phenomenon has been well documented in econometrics and statistics literature (e.g. see Härdle 1990). Among other things, the boundary bias can generate spurious nonlinear drifts, giving misleading conclusions on the dynamics of X_t .

While the first two features have been well understood in the literature, we are surprised to find that the boundary bias has significant impact on not only the finite sample performance but also

⁵A simple example is the Vasicek model, in which if we vary the speed of mean reversion and the scale of diffusion in the same proportion, the marginal density will remain unchanged, but the transition density will be different.

the asymptotic behavior of the test statistics such as M in (6), although the boundary regions are of size h , which vanishes as $n \rightarrow \infty$. For ease of exposition, suppose that the kernel $k(u)$ is a symmetric bounded probability density with support $[-1, 1]$ so that $\int_{-1}^1 k(u)du = 1$, $\int_{-1}^1 uk(u)du = 0$, and $\int_{-1}^1 u^2k(u)du < \infty$. One example is the quartic kernel:

$$k(u) = \frac{15}{16}(1 - u^2)^2 \mathbf{1}(|u| \leq 1), \quad (9)$$

where $\mathbf{1}(|u| \leq 1)$ is the indicator function, taking value 1 if $|u| \leq 1$ and value 0 otherwise. Further suppose that X_t has bounded support $[0, 1]$.⁶ Then the kernel density estimator $\hat{\pi}_0(x)$ is not asymptotically unbiased in the boundary regions $[0, h)$ and $(1 - h, 1]$ in the sense that $\sup_{x \in [0, h) \cup (1 - h, 1]} |E\hat{\pi}_0(x) - \pi_0(x)|$ does not vanish as $n \rightarrow \infty$. This occurs due to the lack of symmetric coverage of observations for $k(\cdot)$ in the boundary regions. For any interior point $x \in [h, 1 - h]$, by using change of variable and the identities that $\int_{-1}^1 k(u)du = 1$, $\int_{-1}^1 uk(u)du = 0$, and assuming that $\pi_0(\cdot)$ is twice continuously differentiable on $[0, 1]$, we obtain the bias

$$\begin{aligned} E\hat{\pi}_0(x) - \pi_0(x) &= \int_0^1 K_h(x - y)\pi_0(y)dy - \pi_0(x) \\ &= \int_{-1}^1 k(u)\pi_0(x + hu)du - \pi_0(x) \\ &= \frac{1}{2}h^2\pi_0''(x) \int_{-1}^1 u^2k(u)du[1 + o(1)] = O(h^2), \end{aligned} \quad (10)$$

which vanishes as $h \rightarrow 0$. For a boundary point $x \in [0, h)$, however,

$$\begin{aligned} E\hat{\pi}_0(x) - \pi_0(x) &= \int_0^1 K_h(x - y)\pi_0(y)dy - \pi_0(x) \\ &= \int_{-x/h}^1 k(u)\pi_0(x + hu)du - \pi_0(x) \\ &= \pi_0(x) \left[\int_{-x/h}^1 k(u)du - 1 \right] + h\pi_0'(x) \int_{-x/h}^1 uk(u)du[1 + o(1)] \\ &= \pi_0(x) \left[\int_{-x/h}^1 k(u)du - 1 \right] + O(h). \end{aligned} \quad (11)$$

Because $\int_{-x/h}^1 k(u)du \neq 1$ for $x/h < 1$, $\hat{\pi}_0(x)$ is not asymptotically unbiased if $\pi_0(x) \geq \varepsilon > 0$, where ε is an arbitrarily small but fixed constant. A similar result holds for $x \in (1 - h, 1]$.

One may argue that although $\hat{\pi}_0(x)$ is not consistent in the boundary regions, these regions are asymptotically negligible and dominated by the interior region, because they are of size h and

⁶The boundary bias problem remains as long as the support of X_t is bounded or half-bounded (e.g., X_t has the support $[0, \infty)$, as is the case with interest rates). This occurs no matter whether the kernel function has bounded or unbounded support.

thus shrink to zero as $n \rightarrow \infty$. Unfortunately, as shown below, this is not the case for statistics such as the quadratic form M in (6). To see this, we decompose

$$M = \left[\int_0^h + \int_{1-h}^1 \right] \left[\hat{\pi}_0(x) - \hat{\pi}(x, \hat{\theta}) \right]^2 \hat{\pi}_0(x) dx + \int_h^{1-h} \left[\hat{\pi}_0(x) - \hat{\pi}(x, \hat{\theta}) \right]^2 \hat{\pi}_0(x) dx, \quad (12)$$

where the first term is the contribution from the kernel estimator in the boundary regions and the second from the interior region. Because $\hat{\theta}$ is \sqrt{n} -consistent for θ_0 , $\pi(x, \hat{\theta})$ converges to $\pi_0(x)$ under H_0 faster than $\hat{\pi}_0(x)$. Consequently, $\hat{\theta}$ has no impact on the asymptotic distribution of M , which solely depends on $\hat{\pi}_0(x)$. For an interior point $x \in [h, 1-h]$, the convergence rate of $\hat{\pi}_0(x)$ is $O_P(n^{-1}h^{-1} + h^4)$, where $O_P(n^{-1}h^{-1})$ is the variance of $\hat{\pi}_0(x)$ and $O_P(h^4)$ is the squared bias (cf. (10)).⁷ Thus, the convergence rate of the second term is $O_P(n^{-1}h^{-1} + h^4)$. For a boundary point $x \in [0, h)$ or $x \in (1-h, 1]$, the convergence rate of $\hat{\pi}_0(x) - \pi_0(x)$ is $O_P(n^{-1}h^{-1}) + \pi_0^2(x)O_P(1)$, where $O_P(n^{-1}h^{-1})$ is the variance of $\hat{\pi}_0(x)$ and $\pi_0^2(x)O(1)$ is the order of the squared bias. Consequently, the convergence rate of the first term in (12) is $hO_P(n^{-1}h^{-1} + 1) = O_P(n^{-1} + h)$ when $\pi(x) \geq \varepsilon > 0$ and thus dominates that of the second term in (12), although the boundary regions shrink to zero. This could be another reason for the poor performance of Ait-Sahalia's (1996) test even for relatively large sample sizes, as the asymptotic distribution theory developed there is essentially based on the second term only.⁸

To avoid the boundary bias problem of kernel estimation, one could simply ignore the data in the boundary regions and only use the data in the interior region. Such a trimming procedure is simple and popular in practice, but in the present context, it would lead to the loss of significant amount of information. If $h = sn^{-\frac{1}{5}}$ where $s^2 = \text{var}(X_t)$, for example, then about 23%, 20% and 10% of a $U[0, 1]$ sample will fall into the boundary regions when $n = 100, 500$, and $5,000$. For financial time series, one may be particularly interested in the tail distribution of the underlying process, which is exactly contained in (and only in) the boundary regions!

Another solution is to use a kernel that adapts to the boundary regions and can effectively eliminate the boundary bias. One possible choice is the so-called jackknife kernel, as used in Chapman and Pearson (2000).⁹ In the present context, the jackknife kernel, however, has some undesired features in finite samples. For example, it may generate negative density estimates in the boundary regions because the jackknife kernel can be negative in these regions.¹⁰ It also

⁷For the convergence rate of nonparametric kernel density estimators in a time series context, see (e.g.) Robinson (1983).

⁸The asymptotic theory for Ait-Sahalia's (1996) test is valid when X_t has unbounded support on \mathbb{R} . For bounded support (say, $[0, 1]$), Ait-Sahalia's (1996) asymptotic distribution theory still holds when $\pi_0(x) = o(x^{3/2})$ as $x \rightarrow 0$ and $\pi_0(x) = o((1-x)^{3/2})$ as $x \rightarrow 1$. For the half-bounded support $[0, \infty)$, it holds when $\pi_0(x) = o(x^{3/2})$ as $x \rightarrow 0$.

⁹See Härdle (1990) for further discussion on the jackknife kernel.

¹⁰Chapman and Pearson (2000) do not encounter this problem because they estimate regression functions for the drift and diffusion functions.

induces a relatively large variance in the boundary regions, adversely affecting the power of the test in finite samples. To avoid these problems, we use the following modified kernel. For $x \in [0, 1]$, define

$$K_h(x, y) \equiv \begin{cases} h^{-1}k\left(\frac{x-y}{h}\right) / \int_{-(x/h)}^1 k(u)du, & \text{if } x \in [0, h), \\ h^{-1}k\left(\frac{x-y}{h}\right), & \text{if } x \in [h, 1-h], \\ h^{-1}k\left(\frac{x-y}{h}\right) / \int_{-1}^{(1-x)/h} k(u)du, & \text{if } x \in (1-h, 1]. \end{cases} \quad (13)$$

The weighting functions in the denominators account for the asymmetric coverage of the kernel for the data in the boundary regions. They ensure that $\hat{\pi}_0(x)$ is asymptotically unbiased uniformly over the entire support $[0, 1]$. Moreover, it always produces nonnegative density estimates with a smaller variance in the boundary regions than a jackknife kernel. With this kernel, the convergence rate of the first term in (12) will be $hO_P(n^{-1}h^{-1} + h^2) = O_P(n^{-1} + h^3)$ when $\pi_0(x) \geq \varepsilon > 0$. This rate is still slower than the convergence rate of the second term in (12). However, we will work with a dynamic probability integral transformed series, which has a uniform density under H_0 . Consequently, the $O_P(h^3)$ term disappears because the uniform density is a constant and so has a zero bias. We thus do not encounter any problem in asymptotic analysis.

II. New Approach and Test Statistics

The limitations of marginal density-based tests prompt us to develop nonparametric tests based on the transition density, which can capture the full dynamics of the underlying diffusion process. Let $p_0(x, t|y, s)$ be the true transition density of the diffusion process X_t ; that is, the conditional density of $X_t = x$ given $X_s = y$, $s < t$. For a given pair of drift and diffusion models $\mu(X_t, t, \theta)$ and $\sigma(X_t, t, \theta)$, a certain family of transition densities $\{p(x, t|y, s, \theta)\}$ is characterized. When (and only when) H_0 in (2) coincides with the true data generating process, there exists some $\theta_0 \in \Theta$ such that $p(x, t|y, s, \theta_0) = p_0(x, t|y, s)$ almost everywhere for all $t > s$. Hence, the hypotheses of interest H_0 in (2) vs. H_A in (3) can be written as follows:

$$H_0 : p(x, t|y, s, \theta_0) = p_0(x, t|y, s) \text{ almost everywhere for some } \theta_0 \in \Theta \quad (14)$$

versus the alternative hypothesis

$$H_A : p(x, t|y, s, \theta) \neq p_0(x, t|y, s) \text{ for some } t > s \text{ and for all } \theta \in \Theta. \quad (15)$$

A natural approach to test H_0 vs. H_A would be to follow Ait-Sahalia (1996) and compare a model-implied transition density estimator $p(x, t|y, s, \hat{\theta})$ with a nonparametric transition estimator, say $\hat{p}_0(x, t|y, s)$.¹¹ From Pritsker's (1998) analysis, however, we expect that the size performance of such a nonparametric test might be even worse than Ait-Sahalia's (1996) marginal

¹¹In addition to the marginal density-based test, Ait-Sahalia (1996) also develops a nonparametric test based on transition density under the important assumption of stationarity.

density-based test, because the convergence rate of the nonparametric transitional density estimator $\hat{p}_0(x, t|y, s)$ is even slower than the nonparametric marginal density estimator $\hat{\pi}_0(x)$, due to the well-known “curse of dimensionality”.¹² Moreover, the finite sample distribution of the resulting test statistic is expected to be sensitive to the level of persistent dependence in data.

Instead of comparing $p(x, t|y, s, \hat{\theta})$ and $\hat{p}_0(x, t|y, s)$ directly, we first transform $\{X_{\tau\Delta}\}_{\tau=1}^n$ via a probability integral transform. Define a transformed sequence

$$Z_\tau(\theta) \equiv \int_{-\infty}^{X_{\tau\Delta}} p[x, \tau\Delta|X_{(\tau-1)\Delta}, (\tau-1)\Delta, \theta] dx, \tau = 1, 2, \dots, n. \quad (16)$$

Under (and only under) H_0 , there exists some $\theta_0 \in \Theta$ such that $p(x, \tau\Delta|X_{(\tau-1)\Delta}, (\tau-1)\Delta, \theta_0) = p_0(x, \tau\Delta|X_{(\tau-1)\Delta}, (\tau-1)\Delta)$ almost surely for all $\Delta > 0$. Consequently, the series $\{Z_\tau \equiv Z_\tau(\theta_0)\}_{\tau=1}^n$ is i.i.d. $U[0, 1]$ under H_0 . This result is first proven, in a simpler context, by Rosenblatt (1952), and more recently used for evaluation of out-of-sample density forecasts by Diebold, Gunther and Tay (1998) among others in discrete time series contexts. Intuitively, we may call Z_τ a “generalized residual” of the diffusion model $p(x, t|y, s, \theta)$. The i.i.d. $U[0, 1]$ property captures two important aspects of the model specification: i.i.d. characterizes the correct specification of the model dynamics, and uniformity characterizes the correct specification of the model marginal distribution.

To test H_0 , we can check whether $\{Z_\tau\}_{\tau=1}^n$ is i.i.d. $U[0, 1]$. This is not a trivial task, because i.i.d. $U[0, 1]$ is a joint composite hypothesis. One may suggest the well-known Kolmogorov-Simonov test. This test, however, checks $U[0, 1]$ under the i.i.d. assumption rather than tests i.i.d. and $U[0, 1]$ jointly. Consequently, it would miss the alternatives whose marginal distribution is uniform but not i.i.d.

Instead in this paper, we develop two nonparametric tests of the i.i.d. $U[0, 1]$ hypothesis by comparing a kernel estimator $\hat{g}_j(z_1, z_2)$ for the joint density $g_j(z_1, z_2)$ of $\{Z_\tau, Z_{\tau-j}\}$ with unity, the product of two $U[0, 1]$ densities. Our approach has at least two advantages. First, since there is no serial dependence in $\{Z_\tau\}$ under H_0 , nonparametric joint density estimators and related test statistics are expected to perform much better in finite samples. We expect that the finite sample distribution of the resulting tests will be robust to the level of persistent dependence in data. Second, there is no asymptotic bias for nonparametric density estimators under H_0 , because the conditional density of Z_τ given $\{Z_{\tau-1}, Z_{\tau-2}, \dots\}$ is uniform (i.e., a constant function).¹³

¹²Under certain regularity conditions, the optimal convergence rates of $\hat{\pi}_0(x)$ and $\hat{p}_0(x, t|y, s)$ are $O(n^{-2/5})$ and $O(n^{-1/3})$ respectively. See Silverman (1986) and Hardle (1990) for relevant discussion in the i.i.d. context, and Robinson (1983) in the time series context.

¹³Of course, the boundary bias problem still exists for kernel estimation based on the transformed series $\{Z_\tau\}$ if the kernel function is not modified in the boundary regions.

Our kernel joint density estimator is

$$\hat{g}_j(z_1, z_1) \equiv (n-j)^{-1} \sum_{\tau=j+1}^n K_h(z_1, \hat{Z}_\tau) K_h(z_2, \hat{Z}_{\tau-j}), \quad (17)$$

where $K_h(z_1, z_2)$ is as defined in (13), $\hat{Z}_\tau = Z_\tau(\hat{\theta})$, and $\hat{\theta}$ is a \sqrt{n} -consistent estimator for θ_0 . We permit but do not require the minimum distance estimator $\hat{\theta}$ in (7), which is used in Ait-Sahalia (1996). Any \sqrt{n} -consistent estimator $\hat{\theta}$ can be used. In practice, the choice of bandwidth h is much more important than the choice of the kernel $k(u)$. Similarly to Scott (1992), we choose $h = \hat{S}_Z n^{-\frac{1}{6}}$, where \hat{S}_Z is the sample standard deviation of $\{\hat{Z}_\tau\}_{\tau=1}^n$. This simple bandwidth rule attains the optimal rate for bivariate kernel density estimation, and it minimizes the asymptotic mean squared error of the kernel density estimator for a certain distribution of the underlying variable.

Analogous to Ait-Sahalia (1996), our first test is based on a quadratic form between $\hat{g}_j(z_1, z_2)$ and 1, the product of two $U[0, 1]$ densities:

$$\hat{M}_1(j) \equiv \int_0^1 \int_0^1 [\hat{g}_j(z_1, z_2) - 1]^2 dz_1 dz_2. \quad (18)$$

Note that the integration range is $[0, 1]$, because $\{Z_\tau\}$ has support $[0, 1]$. Also, we do not use any density estimator to weight the squared difference between $\hat{g}_j(x, y)$ and 1.

Our first test statistic is a properly centered and scaled version of $\hat{M}_1(j)$:

$$\hat{Q}(j) \equiv [(n-j)h\hat{M}_1(j) - A_h^0] / V_0^{1/2}, \quad (19)$$

where

$$A_h^0 \equiv (h^{-1} - 2) \int_{-1}^1 k^2(u) du + 2 \int_0^1 \int_{-1}^b k_b^2(u) du db, \quad (20)$$

$$V_0 \equiv 2 \left[\int_{-1}^1 \left[\int_{-1}^1 k(u+v)k(v) dv \right]^2 du \right]^2, \quad (21)$$

and $k_b(u) \equiv k(u) / \int_{-1}^b k(v) dv$. Note that the modification of the kernel $k(u)$ near the boundary regions affects the centering constant A_h^0 . Such impact is not negligible as $n \rightarrow \infty$. The asymptotic variance V_0 , however, does not depend on the modification of $k(u)$ in the boundary regions.

Under H_0 , we can show (see Theorem 1 in the Appendix) that as $n \rightarrow \infty$,

$$\hat{Q}(j) \rightarrow N(0, 1) \text{ in distribution.}$$

In the present context, the first lag $j = 1$ is the most informative and important. We also have under H_0

$$\text{cov} \left[\hat{Q}(i), \hat{Q}(j) \right] \rightarrow 0 \text{ in probability for } i \neq j$$

as $n \rightarrow \infty$. Thus, $\hat{Q}(i)$ and $\hat{Q}(j)$ are asymptotically independent under H_0 whenever $i \neq j$ (cf. Theorem 2 in the Appendix). Thus, we can use a set of test statistics $\{\hat{Q}(j)\}$ with different lags jointly. Our simulation study shows that in a simulation experiment mimicking the dynamics of U.S. interest rates via the Vasicek model, $\hat{Q}(j)$ has rather reasonable sizes for $n \geq 500$ (i.e., about two year daily data).

Under the alternative H_A , we can show that as $n \rightarrow \infty$,

$$\hat{Q}(j) \rightarrow \infty \text{ in probability}$$

whenever $\{Z_\tau, Z_{\tau-j}\}$ are not independent or $U[0, 1]$ (See Theorem 3 in the Appendix.)

The quadratic form test, though convenient and quite accurate when the true parameter θ_0 were known, might be adversely affected by any imprecise estimate for $\hat{\theta}$ in finite samples. This is particularly relevant in the present context because it is well-known that the drift parameters are difficult to estimate (e.g., Merton 1980). To alleviate this problem, we consider a second test based on the squared Hellinger metric

$$\hat{M}_2(j) \equiv \int_0^1 \int_0^1 \left[\sqrt{\hat{g}_j(z_1, z_2)} - 1 \right]^2 dz_1 dz_2, \quad (22)$$

which is a quadratic form between $\sqrt{\hat{g}_j(z_1, z_2)}$ and $\sqrt{1 \cdot 1} = 1$. The Hellinger metric is well-known in statistics for its robustness to outliers. The associated test statistic is

$$\hat{H}(j) \equiv \left[4(n-j)h\hat{M}_2(j) - A_h^0 \right] / V_0^{1/2}, \quad (23)$$

where A_h^0 and V_0 are the same as in (20) and (21). Under H_0 , this test has the same asymptotic distribution as $\hat{Q}(j)$ and is asymptotically equivalent to $\hat{Q}(j)$ in the sense that $\hat{Q}(j) - \hat{H}(j) \rightarrow 0$ in probability. This occurs because under H_0 , the quadratic form $\hat{M}_1(j)$ is the dominant term in a Taylor series expansion of $4\hat{M}_2(j)$. Under H_A , we also have $\hat{H}(j) \rightarrow \infty$ as $n \rightarrow \infty$ whenever $\{Z_\tau, Z_{\tau-j}\}$ are not independent or $U[0, 1]$. However, $\hat{Q}(j)$ and $\hat{H}(j)$ are not equivalent under H_A , because the quadratic norm and the squared Hellinger metric are not the same. This suggests that they may have different powers under H_A in finite samples, but the difference will vanish as $n \rightarrow \infty$.

We note that the Hellinger metric test $\hat{H}(j)$ may not dominate the quadratic form test $\hat{Q}(j)$ in finite samples, because its asymptotic distribution is based on the dominant quadratic form in a Taylor expansion, and the remaining asymptotically negligible higher order terms may have important impact in small and finite samples. Indeed, our simulation study shows that $\hat{H}(j)$ displays some overrejection in size when $n < 1,000$. It has reasonable sizes when $n \geq 1,000$ (i.e., about four years of daily data). The main reason to include $\hat{H}(j)$ is that with an extremely imprecise parameter estimate in $\hat{\theta}$, $\hat{H}(j)$ may have better sizes than $\hat{Q}(j)$ if n is sufficiently large,

because the Hellinger metric is more robust to parameter estimation uncertainty and the benefit from such robustness may overwhelm the effect of the asymptotically negligible higher order terms.

A technical issue that complicates our tests is that the closed-form solution for the transition density of a diffusion model is generally not available, which makes the probability integral transform $\{Z_\tau(\theta)\}$ in (16) difficult to compute. In empirical applications, fortunately, we can apply Ait-Sahalia's (2001) method to construct an accurate closed-form approximation of the model transition density via Hermite expansion with a sufficiently large order. Following Ait-Sahalia's (2001) proof, we could show that under proper regularity conditions, the use of a Hermite polynomial approximation has no impact on the asymptotic distribution of our test statistics. For the details on the Hermite approximation, see Ait-Sahalia (1999, 2001) or Egorov, Li and Xu (2002).

We summarize our testing procedures as follows: (i) estimate the continuous-time model using any method that yields a \sqrt{n} -consistent estimator $\hat{\theta}$; (ii) compute the dynamic probability integral transformed series $\{\hat{Z}_\tau = Z_\tau(\hat{\theta})\}_{\tau=1}^n$ via the model implied transition density, where $Z_\tau(\theta)$ is given in (16). If there is no closed-form solution for the model-implied transition density, use Ait-Sahalia's (2001) method to obtain an approximation for the model-implied transition density; (iii) compute the boundary-modified kernel joint density estimator $\hat{g}_j(z_1, z_2)$ in (17) for a prespecified lag j , using a kernel (e.g., the quartic kernel in (9)) and the bandwidth $h = \hat{S}_z n^{-\frac{1}{6}}$, where \hat{S}_z is the sample standard deviation of the transformed series $\{\hat{Z}_\tau\}_{\tau=1}^n$; (iv) compute the test statistics $\hat{Q}(j)$ in (19) and $\hat{H}(j)$ in (23); (v) compare the value of $\hat{Q}(j)$ or $\hat{H}(j)$ with the upper-tailed $N(0,1)$ critical value C_α at level α (e.g., $\alpha_{0.05} = 1.645$). If $\hat{Q}(j)$ or $\hat{H}(j)$ exceeds C_α , reject the null hypothesis H_0 at level α . The use of upper-tailed (rather than two-sided) $N(0,1)$ critical values is suitable because negative values of $\hat{Q}(j)$ and $\hat{H}(j)$ occur only under H_0 when n is sufficiently large. Both of the test statistics diverge to $+\infty$ when $\{Z_\tau, Z_{\tau-j}\}$ are not independent or $U[0, 1]$ under H_A .

It should be noted that some other nonparametric or semi-nonparametric specification tests for diffusion models or applicable to diffusion models have been available in the literature. Corradi and Swanson (2001) and Thompson (2001) consider tests for diffusion models using an empirical distribution function.¹⁴ The use of the empirical distribution function is computationally convenient, because there is no need to choose a smoothing parameter (e.g., bandwidth h). However, these tests are based on the marginal distribution function, and therefore cannot distinguish the diffusion processes that have the same marginal density but different transition densities. Moreover, the asymptotic distribution of these tests is data-dependent, and so no tabulation is available. To obtain the critical values, Thompson (2001) uses some upper bounds, and Corradi

¹⁴In addition to the test based on the empirical distribution function, Thompson (2001) also considers a test for serial dependence using periodogram. The periodogram can detect any autocorrelation, but it can easily miss any nonlinear sequence that has zero autocorrelation but not serially dependent. Such processes are not uncommon in nonlinear time series (cf. Granger and Terasvirta 1993).

and Swanson (2001) use a stationary bootstrap procedure. Use of upper bound critical values may be too conservative. The bootstrap procedure gives accurate sizes, but it is computationally demanding, especially for a large data set. Chen, Härdle and Kleinow (2000) also develop an empirical likelihood goodness of fit test for time series models. One advantage of this method is that the asymptotic distributions of the test statistic are free of unknown parameters.

III. Extensions

We emphasize that the nonparametric testing procedures just developed are a general methodology that can be applied to wide range of dynamic models used in economics and finance. This is because the idea of probability integral transform is very general: the transformed data via model-implied conditional density should be i.i.d. $U[0, 1]$ if the model is correctly specified. This fact should hold for not only the single-factor diffusion models as considered in Ait-Sahalia (1996), but also for other dynamic models, such as the discrete time series models, stochastic volatility models, and multi-factor term structure models. In this section, we explain how to apply our methods to other existing models in the literature.

In addition to single-factor diffusion models, our tests can be easily applied to many discrete time series models that have been proposed in the literature for spot interest rate. For example, GARCH models have been widely used in the literature to capture volatility clustering in interest rate data (e.g. Brenner, Harjes and Kroner (1996)); regime-switching models have also been proposed to capture the potential structural breaks in interest rate process (e.g. Gray (1996), Ang and Bekaert (1998), and Li and Xu (2000)); the importance of jumps in modeling the tail behavior of interest rate distribution has also been studied (e.g. Das (2002) and Johannes (2000)). These models are either discrete time in nature, such as the GARCH and regime-switching models, or discretized version of continuous-time models, such as the discretized jump-diffusion models considered in Das (2002). As a result, the specification of all these models can be summarized by $p(X_{t\Delta}|I_{(t-1)\Delta})$, the conditional density of $X_{t\Delta}$, given $I_{(t-1)\Delta}$, the information set at $(t-1)\Delta$. If the model is correctly specified, then the integral transformed data via $p(X_{t\Delta}|I_{(t-1)\Delta})$ should be i.i.d. $U[0, 1]$. The integral transform is actually much easier to conduct for discrete time series models than for diffusion models, as the conditional density is typically known in closed-form. As a result, most existing discrete time interest rate models can be examined by our tests.

Our tests can also be applied to partially observable nonlinear dynamic models, such as the stochastic volatility models. For example, Kim, Shephard, and Chib (1998) consider the following stochastic volatility model

$$\begin{aligned} X_{t\Delta} &= \beta e^{h_{t\Delta}/2} \epsilon_{t\Delta}, \quad t \geq 1 \\ h_{(t+1)\Delta} &= \mu + \phi(h_{t\Delta} - \mu) + \sigma_\eta \eta_{t\Delta}, \end{aligned}$$

$$h_1 \sim N\left(\mu, \frac{\sigma^2}{1 - \phi^2}\right),$$

where $X_{t\Delta}$ is the mean corrected return on holding an asset at time $t\Delta$, $h_{t\Delta}$ is the log volatility at time $t\Delta$ which is assumed to follow a stationary process ($|\phi| < 1$) with $h_{t\Delta}$ drawn from the stationary distribution, $\epsilon_{t\Delta}$ and $\eta_{t\Delta}$ are uncorrelated standard normal white noise shocks and $N(\cdot, \cdot)$ is the normal distribution. The model is estimated using Markov Chain Monte Carlo (hereafter MCMC) simulation method in a Bayesian framework. After obtaining model parameter estimates $\hat{\theta}$, i.e., $\hat{\theta} = (\hat{\beta}, \hat{\mu}, \hat{\phi}, \hat{\sigma}_\eta)$, Kim, Shephard and Chib (1998) show (see Section 4.2 on page 380) that the conditional density of $X_{(\tau+1)\Delta}$, given $X^{(\tau)} = (X_\Delta, X_{2\Delta}, \dots, X_{\tau\Delta})$ is given by

$$p\left(X_{(\tau+1)\Delta}|X^{(\tau)}, \hat{\theta}\right) = \int p\left(X_{(\tau+1)\Delta}|X^{(\tau)}, h_\tau, \hat{\theta}\right) p\left(h_\tau|X^{(\tau)}, \hat{\theta}\right) dh_\tau,$$

where $p\left(X_{(\tau+1)\Delta}|X^{(\tau)}, h_\tau, \hat{\theta}\right) = \int p\left(X_{(\tau+1)\Delta}|X^{(\tau)}, h_{(\tau+1)\Delta}, \hat{\theta}\right) p\left(h_{(\tau+1)\Delta}|X^{(\tau)}, h_\tau, \hat{\theta}\right) dh_{(\tau+1)\Delta}$, and $p\left(h_\tau|X^{(\tau)}, \hat{\theta}\right)$ can be obtained through a MCMC filtering algorithm as described in Kim, Shephard and Chib (1998). The whole integral in general can be obtained by simulation. Under the null hypothesis, the probability integral transformed data via $p\left(X_{(\tau+1)\Delta}|X^{(\tau)}, \hat{\theta}\right)$ should follow an i.i.d. $U[0, 1]$ distribution. While Kim, Shephard and Chib (1998) use this fact to conduct certain model diagnostic tests, they do not conduct a formal test of the joint hypothesis of i.i.d. and $U[0, 1]$. Our tests can be easily applied to the stochastic volatility model considered in Kim, Shephard and Chib (1998). In a more recent paper, Chib, Nardari, and Shephard (2001), using similar methodology, consider more complicated stochastic volatility models. For example they allow $\epsilon_{t\Delta}$ to follow a Student- t instead of standard normal distribution, to approximate the continuous-time Levy process; they also include jumps in the return process. The conditional density of these more complex stochastic volatility models (with jumps) can also be obtained in similar way as shown in the above equation and as a result, our nonparametric tests can be applied to those models.

Kim, Shephard and Chib (1998) consider only discretized versions of continuous-time stochastic volatility models. Anderson and Lund (1997), on the other hand, study a continuous-time stochastic volatility model for the spot interest rate:

$$\begin{aligned} dX_t &= \kappa_1(\mu - X_t)dt + \sigma_t X_t^\gamma dW_{1t}, \\ d\log \sigma_t^2 &= \kappa_2(\alpha - \log \sigma_t^2)dt + \eta dW_{2t}, \end{aligned}$$

where $\gamma \geq 0$, W_{1t} and W_{2t} are independent Brownian motions, and σ_t is the unobservable instantaneous volatility. Likelihood function for the above model is hard to obtain and as a result it is often estimated using the Efficient Method of Moments of Gallant and Tauchen (1997). After obtaining the model parameters, we can proceed to calculate the model transition density of

$X_{(\tau+1)\Delta}$ conditional on the past observable information, $X^{(\tau)} = (X_{\Delta}, X_{2\Delta}, \dots, X_{\tau\Delta})$. The reprojection technique of Gallant and Tauchen (1998) can be used to obtain the conditional density of σ_{τ} given all the past history of $X^{(\tau)}$, i.e., $p(\sigma_{\tau}|X^{(\tau)}, \hat{\theta})$. We can calculate the transition density of $X_{(\tau+1)\Delta}$ conditional only on past interest rate levels by integrating out the unobservable stochastic volatility; that is

$$p\left(X_{(\tau+1)\Delta}|X^{(\tau)}, \hat{\theta}\right) = \int p\left(X_{(\tau+1)\Delta}|X^{(\tau)}, \sigma_{\tau}, \hat{\theta}\right) p\left(\sigma_{\tau}|X^{(\tau)}, \hat{\theta}\right) d\sigma_{\tau},$$

where the transition density $p\left(X_{(\tau+1)\Delta}|X^{(\tau)}, \sigma_{\tau}, \hat{\theta}\right)$ in general does not have closed-form solution and has to be obtained through simulations. If the stochastic volatility model is correctly specified, then the probability integral transformed data with respect to $p\left(X_{(\tau+1)\Delta}|X^{(\tau)}, \hat{\theta}\right)$ will be i.i.d. $U[0, 1]$. Any deviation from i.i.d. $U[0, 1]$ will indicate model misspecification and can be detected using our testing procedures.

The models we have discussed so far focus on the time series behavior of the spot interest rate and are estimated using only the spot rate data. There are also many multi-factor diffusion models in the existing literature that focus on explaining the dynamics of the whole yield curve, such as the famous affine term structure models of Duffie and Kan (1995). In affine models, it is typically assumed that the spot rate is a linear (affine) function of the underlying state variables, i.e.,

$$R_t = \delta_0 + \sum_{i=1}^N \delta_i X_{it},$$

where R_t is the instantaneous riskfree borrowing and lending rate, and X_{it} , for $i = 1, \dots, N$ are the state variables and evolve over time as a multi-dimensional affine diffusion process

$$dX_t = \kappa(\Theta - X_t) dt + \Sigma \sqrt{S_t} dW_t,$$

where κ and Σ are $N \times N$ matrices, S_t is a $N \times N$ diagonal matrix whose i -th diagonal element is

$$[S_t]_{ii} = \alpha_i + \sum_{j=1}^N \beta_{ij} X_{jt},$$

and W_t is an N -dimensional vector of independent Brownian motions. In affine models, zero-coupon bond price is given by the following formula

$$\begin{aligned} P(X_t, t, T) &= E^Q \left[\exp \left\{ - \int_t^T R(s) ds \right\} | F_t \right] \\ &= \exp \left[A(t, T) + \sum_{i=1}^N B_i(t, T) X_{it} \right], \end{aligned}$$

where $E^Q[\cdot]$ is the expectation under the risk-neutral probability measure, and $A(t, T)$ and $B_i(t, T)$ can be solved from a system of ordinary differential equations. Therefore bond yields in

affine models are a linear (affine) function of the underlying state variables

$$\begin{aligned} Y(t, T) &= -\frac{1}{T-t} \ln P(X_t, t, T) \\ &= -\frac{1}{T-t} \left[A(t, T) + \sum_{i=1}^N B_i(t, T) X_{it} \right]. \end{aligned}$$

Consequently, affine models are typically estimated using time series observations of the yields of N zero coupon bonds with different maturities $\{Y(\tau\Delta, T_k)\}_{\tau=1}^n$, $k = 1, \dots, N$. These yields are assumed to be observed without error and can be used to infer the underlying state variables $\{X_{i,\tau\Delta}\}_{\tau=1}^n$, $i = 1, \dots, N$. The transition density of $X_{(\tau+1)\Delta}$ given $X_{\tau\Delta}$, $p(X_{(\tau+1)\Delta}|X_{\tau\Delta})$ in general is not known in closed form and the models are often estimated using the Efficient Method of Moments or Quasi-Maximum Likelihood (e.g. Dai and Singleton 1999, and Duffee 2001). Given estimates of model parameters $\hat{\theta}$, the probability integral transform of the yields $Y((\tau+1)\Delta, T_k)$, $k = 1, \dots, N$, can be calculated in the following way

$$\begin{aligned} Z((\tau+1)\Delta, T_k, \hat{\theta}) &= \int^{Y((\tau+1)\Delta, T_k)} p(y|X_{\tau\Delta}, \hat{\theta}) dy \\ &= \Pr \left[\sum_{i=1}^N B_i((\tau+1)\Delta, T_k) X_{i,(\tau+1)\Delta} \leq Y((\tau+1)\Delta, T_k) - A((\tau+1)\Delta, T_k) | X_{\tau\Delta}, \hat{\theta} \right], \end{aligned}$$

where $p(y|X_{\tau\Delta}, \hat{\theta})$ is the conditional density of $Y((\tau+1)\Delta, T_k)$ given $X_{\tau\Delta}$. This transition density and thus the above integral in general has to be obtained using simulation. If the affine term structure model is correctly specified, then $Z((\tau+1)\Delta, T_k, \hat{\theta})$ should follow an i.i.d. $U[0, 1]$ distribution, a fact can be easily tested using our nonparametric tests.

The above examples illustrate the wide applicability of our nonparametric tests. Our asymptotic theory directly applies to the models we have discussed, because we assume a general model specification in terms of the transition density rather than the stochastic differential equation. Of course, continuous-time stochastic volatility models and multi-factor term structure models are much more difficult to estimate and the probability integral transform is also computationally more demanding to implement. To keep the paper within reasonable length, in the application section, we only focus on testing the single-factor diffusion models and some discrete time series models and leave the more complicated continuous-time models for future research.

Gallant and Tauchen (1996), in an influential paper, propose a minimum chi-square test for both single-factor and multi-factor diffusion models. The basic idea is to check whether the following moment condition holds:

$$\int \frac{\partial \log f(x|y, \beta)}{\partial \beta} p(x, y, \theta) dx dy = 0,$$

where $p(x, y, \theta)$ is the model stationary density, and $f(x|y, \beta)$ is a semi-nonparametric density model, which is free of model misspecification in an asymptotic sense because the dimension β is

allowed to grow to infinity. Gallant and Tauchen (1996) use a Hermite polynomial approximation, with the dimension of β determined by such data-driven methods as AIC or BIC criteria. The above integration can be computed accurately using simulation methods.

Our tests complement Gallant and Tauchen’s EMM tests (1996). Both approaches are non-parametric: our tests are based on a nonparametric (kernel) transition density estimator, while Gallant and Tauchen (1996) use a semi-nonparametric (Hermite polynomial-based) transition density. However, while we use the nonparametric transition density estimator directly, Gallant and Tauchen (1996) use the derivative of the semi-nonparametric density estimator. As a consequence, at least in single-factor diffusion contexts, our tests are consistent against any model misspecification. However, the EMM tests may still have no power against some alternatives, because the semi-nonparametric density derivative may still have zero expectation under the distribution of a misspecified model. Indeed, as pointed out by Gallant and Tauchen (1998, p.), one still cannot conclude that a diffusion model is correctly specified even when the minimum chi-square EMM test statistic is insignificant. In addition, Gallant and Tauchen (1996) only consider stationary diffusion processes, while we allow for both time-homogeneous and inhomogeneous diffusion processes, thanks to our use of the probability integral transform.

When they reject a misspecified model, one may like to go further to explore what are possible sources for the rejection. For example, is the rejection due to misspecification in the drift, such as the ignorance of mean shifts or jumps? Or is it due to the ignorance of GARCH effects or stochastic volatility? Among other things, the greatest appeals of Gallant and Tauchen’s EMM method is that in addition to the minimum chi-square test, it also provides a spectrum of constructive individual t -statistics that are informative in revealing possible sources of model misspecification.

For our tests, the transformed data via the dynamic probability integral transform constitutes a “generalized residual” for the null diffusion model. Intuitively, our transition density-based omnibus tests contain information of the autocorregrams in all the moments of the generalized residuals. It is an omnibus test. To check possible sources for the rejection, we can use the generalized residual series to construct a spectrum of autocorregram-based tests in every moment of $Z_\tau(\theta)$, which are very informative. In particular, the autocorregram $\text{cov}(Z_\tau, Z_{\tau-j})$ can reveal information about the adequacy of the mean/drift model, and the autocorregram $\text{cov}(Z_\tau^2, Z_{\tau-j}^2)$ can reveal information about the adequacy of the diffusion or volatility model. Thus, our approach also shares the appeal of Gallant and Tauchen’s (1996) EMM method in providing constructive diagnostic tests for possible sources of model misspecification.

IV. Finite Sample Performance

We now study the finite sample performance of our tests via Monte Carlo experiments. Pritsker

(1998) has conducted a simulation study on Ait-Sahalia's (1996) marginal density-based test for time-homogeneous diffusion models. For comparison, we adopt similar simulation designs.

A. Size of the Tests

To examine the size of the tests, we follow Pritsker (1998) simulate data from Vasicek's (1977) model:

$$dX_t = \kappa(\alpha - X_t) dt + \sigma dW_t, \quad (24)$$

where α represents the long run mean and κ represents the speed of mean reversion to the long run mean.¹⁵ The smaller κ is, the higher the level of persistence of dependence in data, and consequently, the slower the convergence to the long run mean.

Like Pritsker (1998), we are particularly interested in the impact of the level of dependent persistence in the data generating process. Given that the finite sample performance of the tests may depend on both the marginal density and the persistence of dependence of the diffusion process, we follow Pritsker (1998) and keep the marginal density unchanged while varying the speed of mean reversion. This is achieved by changing κ and σ^2 in the same proportions. Thus, we can focus on the impact of persistent dependence. We consider the Vasicek model in (24) with both low and high levels of persistent dependence and adopt the same parameter values used in Pritsker (1998): $(\kappa, \alpha, \sigma^2) = (0.85837, 0.089102, 0.002185)$ for the low persistent dependence case, and $(\kappa, \alpha, \sigma^2) = (0.214592, 0.089102, 0.000546)$ for the high persistent dependence case.

For each parameterization of the Vasicek model, we simulate 1,000 data sets of a random sample $\{X_{\tau\Delta}\}_{\tau=1}^n$, where the sample size $n = 250, 500, 1,000, 2,500, 5,500$ respectively. These sample sizes correspond to about one year of daily data to twenty-two years of daily data. For each data set, we estimate the model parameters $\theta = (\kappa, \alpha, \sigma^2)'$ using the maximum likelihood method. After obtaining the probability integral transformation of the data $\{X_{\tau\Delta}\}_{\tau=1}^n$ by (quadrature) numerical integrations, we apply our tests to the transformed data. We consider the empirical rejection rates using the asymptotic critical values (1.28, 1.65, 2.33) at the 10%, 5% and 1% levels respectively.

Figure 1 reports the empirical sizes of the quadratic form test $\hat{Q}(j)$, as a function of lag order j from 1 to 20, for sample sizes $n = 250, 500, 1,000, 2,500$, and $5,500$. The first (second) column of Figure 1 is the rejection rates of $\hat{Q}(j)$ at the 10%, 5% and 1% levels under a correct Vasicek model with low (high) persistence of dependence. Overall, $\hat{Q}(j)$ has reasonable sizes for sample sizes as small as $n = 500$ (i.e., about two years of daily data), particularly at the 10% and 5%

¹⁵Put $V_E = \sigma^2/2\kappa$ and $\delta = t - s$. Then the Vasicek model has a marginal density: $\pi(x; \theta) = (2V_E)^{-1/2} \exp[-(x - \alpha)^2/2V_E]$ and a transition density

$$p_X(x, t|y, s; \theta) = \left[2\pi V_E(1 - e^{-\kappa\delta})\right]^{-1/2} \exp\left[-\frac{(x - (\alpha + (y - \alpha)e^{-\kappa\delta}))^2}{2V_E(1 - e^{-\kappa\delta})}\right].$$

levels. The most important difference from Ait-Sahalia's (1996) test is that the impact of the level of dependent persistence on the size of $\hat{Q}(j)$ is minimal. The sizes of $\hat{Q}(j)$ are virtually the same in both the low and high persistent case. There is no evidence of a poorer performance under a higher level of persistent dependence. This suggests that our probability integral transformation is indeed at work in achieving robustness to persistent dependence. In contrast, Pritsker (1998) finds that under the same simulation setting, Ait-Sahalia's (1996) test still shows strong overrejection under H_0 when $n = 5,500$, and it becomes worse when dependence becomes stronger.

The first (second) column of Figure 2 reports the rejection rates of the Hellinger metric test $\hat{H}(j)$ at the 10%, 5% and 1% levels under a correct Vasicek model with low (high) persistence of dependence. The $\hat{H}(j)$ test shows some overrejections when $n = 250, 500$. This is apparently due to the impact of the asymptotically negligible higher order terms of a Taylor expansion of $\hat{H}(j)$ that are absent in $\hat{Q}(j)$. However, $\hat{H}(j)$ becomes reasonable, particularly at the 10% and 5% levels, when $n \geq 1,000$ (about four years of daily data), indicating that the adverse impact of the asymptotically negligible higher order terms has diminished. Again, the performance of $\hat{H}(j)$ is essentially the same under both low and high persistence of dependence.

In summary, the $\hat{Q}(j)$ test has reasonable sizes for $n \geq 500$. The $\hat{H}(j)$ test show some overrejections for $n < 1,000$, but are reasonable and accurate for $n \geq 1,000$, particularly at the 10% and 5% levels. The performance of both the tests are robust to persistent dependence in data. Overall, the asymptotic theory for our tests provide reliable inferences in finite samples.

B. Power of the Tests

To investigate the power of the tests, we simulate data from the following four diffusion processes. Below, we denote $\delta \equiv t - s$ and $\phi(\cdot)$ for the standard normal density.

1. Cox, Ingersoll and Ross's (1985, CIR) Model:

$$dX_t = \kappa(\alpha - X_t) dt + \sigma\sqrt{X_t}dW_t. \quad (25)$$

Note that X_t is distributed on $\mathbb{R}^+ \equiv [0, +\infty)$ provided that $q \equiv 2\kappa\alpha/\sigma^2 - 1 \geq 0$. Its transition density $p_0^{CIR}(x, t|y, s) = ce^{-u-v} (v/u)^{q/2} I_q[2(uv)^{1/2}]$, where $c \equiv 2\kappa/\{\sigma^2(1 - e^{-\kappa\delta})\}$, $u \equiv cye^{-\kappa\delta}$, $v \equiv cx$, and I_q is the modified Bessel function of the first kind of order q (e.g., Abramowitz and Stegun 1970). In our simulation, we choose the same parameter values as in Pritsker (1998): $(\kappa, \alpha, \sigma^2) = (0.89218, 0.090495, 0.032742)$.

2. Ahn and Gao's (1999) Inverse-Feller Model:

$$dX_t = X_t [\kappa - (\sigma^2 - \kappa\alpha) X_t] dt + \sigma X_t^{3/2} dW_t. \quad (26)$$

Its transition density is $p_0(x, t|y, s) = (1/x^2) p_0^{CIR}(1/x, t|1/y, s)$. For this and next two models, we use the parameter values given in Table VI of Ait-Sahalia (1999), which are estimated from U.S. interest rate series. Here, we have $(\kappa, \alpha, \sigma^2) = (0.181, 15.157, 0.032742)$.

3. Chan, Karolyi, Longstaff and Sanders' (1992, CKLS) Model:

$$dX_t = \kappa(\alpha - X_t) dt + \sigma X_t^\rho dW_t. \quad (27)$$

We set $(\kappa, \alpha, \sigma^2, \rho) = (0.0972, 0.0808, 0.52186, 1.46)$.

4. Ait-Sahalia's (1996) Nonlinear Drift Model:

$$dX_t = (\alpha_{-1}X_t^{-1} + \alpha_0 + \alpha_1X_t + \alpha_2X_t^2) dt + \sigma X_t^\rho dW_t. \quad (28)$$

We set $(\alpha_{-1}, \alpha_0, \alpha_1, \alpha_2, \sigma^2, \rho) = (0.00107, -0.0517, 0.877, -4.604, 0.64754, 1.50)$.

For each of these four alternatives, we generate 500 realizations of a random sample $\{X_{\tau\Delta}\}_{\tau=1}^n$, where $n = 1,000, 2,500, 5,500$ respectively. For the CIR and Ahn and Gao's (1999) models, the transition density is a noncentral chi-square; we simulate data as the sum of i.i.d. normal random variables. For the CKLS and Ait-Sahalia' (1996) nonlinear drift models, the transition density has no a closed form solution; we simulate data using the convenient Milstein scheme:

$$X_{t+\delta} = X_t + \mu(X_t)\delta + \sigma(X_t)\sqrt{\delta}\varepsilon_t + \frac{1}{2}\sigma^2(X_t)\delta(\varepsilon_t^2 - 1),$$

where $\delta > 0$. While more accurate schemes are now available in the literature (e.g., Kloeden *et al.* 1991), the Miltstein scheme provides sufficient precision for our purpose. We simulate five observations each day but sample the data at only daily frequency ($\delta = \Delta/5$ and $\Delta = 1$).

To examine the power, we need not estimate the four alternative models. In our applications to interest rates in Section IV, we will estimate these models as well as the Vasicek model via MLE. For the CKLS and Ait-Sahalia's (1996) nonlinear drift models, we will use Ait-Sahalia's (2001) Hermite expansion to obtain an accurate closed-form approximation of the transition density:

$$p_0(x, t|y, s) = \frac{\phi(v)}{x^\rho\sqrt{\delta}} \sum_{k=0}^J \beta_k^{(m)} H_k(v),$$

where $v \equiv (x^{1-\rho} - y^{1-\rho})/\sqrt{\delta}(1 - \rho)\sigma$, $H_k(\cdot)$ is the Hermite polynomial of order k , $\beta_k^{(m)}$ is the coefficient of the Hermite expansion, J represents the number of Hermite polynomials included in the expansion, and m represents the highest order of δ in approximation of β_k .¹⁶ (see Ait-Sahalia 2001 for explicit expressions for $\beta_k^{(m)}$).

For each data set, we test the null hypothesis that the data is generated from the Vasicek model in (24). We fit the Vasicek model to the data and compute the two test statistics in

¹⁶The Hermite approximation of the transition densities of CKLS and Ait-Sahalia's nonlinear drift model are used later in this paper to estimate model parameters using maximum likelihood method. Previous studies such as Ait-Sahalia (1999, 2001) and Egorov, Li and Xu (2001) have shown that the approximation works very well for $J = 6$ and $m = 3$, which are also used in our paper. We also try $m = 1$ and 2 and obtain very similar test statistics.

(19) and (23). Because the results using empirical critical values are very similar for both $\hat{Q}(j)$ and $\hat{H}(j)$ when $n \geq 1,000$, we only report the power using asymptotic critical values, for $n = 1,000, 2,500, 5,500$.¹⁷ Figure 3 reports the power of our tests at the 5% level using asymptotic critical values. The power patterns at the 10% and 1% levels are very similar to those at the 5% level and thus are not reported for space. The first column of Figure 3 reports the rejection rates of the quadratic form test $\hat{Q}(j)$ at the 5% level. The test has all-round good power in detecting misspecification of the Vasicek model against the four alternatives. When $n = 5,500$, $\hat{Q}(j)$ rejects the CIR model at a rate of about 90%. For comparison, Pritsker (1998) reports that the size-corrected power of Ait-Sahalia's (1996) test in detecting the Vasicek model against the CIR alternative is about 38% when $n = 5,500$. Thus, our test is at least twice as powerful as Ait-Sahalia's (1996) test against the CIR model under the same simulation setting. The $\hat{Q}(j)$ test has virtually unit power against the other three alternatives when $n = 5,500$. It appears that our transition density-based test $Q(j)$ is more powerful than marginal density-based tests.

The second column of Figure 3 reports the rejection rates of the Hellinger metric test $\hat{H}(j)$ at the 5% level. The power is similar to that of $\hat{Q}(j)$.

Our simulation results show that with the help of the probability integral transform and the boundary bias correction for kernel density estimator, our transition density-based tests perform rather well even for highly persistently dependent data with sample sizes often encountered in empirical finance.

V. Application to Spot Interest Rates

The transition density-based nonparametric tests developed in the previous sections provide empirical researchers with a useful tool to study the dynamic behavior of important economic variables, such as stock prices, interest rates, exchange rates, commodity prices, and macroeconomic variables. As an illustration of the merits of our tests, in this section we apply them to test a wide variety of spot interest rate models.

Despite a large number of empirical studies, the evidence on the performance of different interest rate models is still not conclusive (see e.g. Chapman and Pearson 2001 for an excellent survey of the existing literature). In general, it is not easy to compare the performance of existing models, as different studies use different model specifications and estimation methods. For example, some models are specified in continuous time, while others in discrete time; some models are estimated using the maximum likelihood method, while others using GMM or nonparametric methods. The nonparametric tests just developed provide a convenient way to compare the performance of different models. Even though these models in general are not nested and are estimated using different methods, their performance can still be compared against each other

¹⁷The results using empirical critical values are available from the authors upon request.

using our nonparametric tests. By comparing the closeness of the distribution of the transformed data under each model to the i.i.d. $U[0, 1]$ distribution, we can determine which model provides better description of the data. We can also use the diagnostic tools associated with our tests to understand the possible sources of model misspecification.

In our study, we consider the single-factor diffusion models that have been studied in Ait-Sahalia (1996), and some popular discrete time series models such as the GARCH models, regime-switching models, and the discretized version of jump-diffusion models. For comparison with existing studies, we use the same interest rate data as that used in Ait-Sahalia (1996). The daily seven-day Eurodollar interest rate series used in Ait-Sahalia (1996) is from June 1, 1973 to February 25, 1995, with $n = 5,505$. Figure 4 displays the level and change series of Eurodollar rates, as well as their histograms. The Eurodollar rates display excess volatility before 1983. The marginal distribution of the level series is skewed to the right. There is a sharp peak around zero for the change series, indicating that small daily changes occur most of time.

A. Single-factor Diffusion Models

Ait-Sahalia (1996), using his marginal density-based test with asymptotic critical values, overwhelmingly rejects all the existing single-factor diffusion models for the spot rate. The main reason for the rejection is that contrary to the assumptions in most models, the drift of the spot rate appears to be a nonlinear function of the level of the interest rate: for the lower and middle ranges of the interest rate, the drift is almost zero, but the spot rate exhibits strong mean-reversion when the interest rate is high. Stanton (1997), using nonparametric kernel regression also reaches similar conclusions. Due to the boundary bias and the finite sample problems of the nonparametric methods used, however, the findings of nonlinear drift by Ait-Sahalia (1996) and Stanton (1997) have been challenged by Pritsker (1998) and Chapman and Pearson (2000). As noted earlier, Pritsker (1998) shows that Ait-Sahalia's (1996) test tends to strongly overreject a correct model using asymptotic critical values. With the empirical critical values provided in Pritsker (1998), Ait-Sahalia's (1996) test would fail to reject certain one-factor diffusion models for spot rate, such as the CKLS model and the nonlinear drift model.

Given the reasonable finite sample performance of our tests, we now apply them to the daily Eurodollar interest rates to re-examine the single-factor diffusion models considered in Ait-Sahalia (1996). Interestingly, we find that all the models are still overwhelmingly rejected by our new procedures. Unlike Ait-Sahalia (1996), we find that even the models with a nonlinear drift do not significantly improve the goodness of fit.

To compare with Ait-Sahalia's (1996) study, we consider five popular models in this section: the Vasicek, CIR, Ahn and Gao, CKLS and Ait-Sahalia's (1996) nonlinear drift models, as given in (24)–(28) respectively. For each model, we estimate parameters using the maximum likelihood method. The true likelihood function is used when the model has a closed-form transition density

(for the Vasicek, CIR and Ahn and Gao’s models); otherwise Ait-Sahalia’s (2001) Hermite expansion is used to obtain a closed-form approximation of the likelihood function (for the CKLS and Ait-Sahalia’s 1996 nonlinear drift models). Table 1 reports parameter estimates and their standard errors, which indicate that the drift parameter estimates are less precise than the diffusion parameter estimates, as is expected.

Figure 5 reports the values of the test statistics as a function of lag order j from 1 to 20. As shown in Figure 5(a), the $\hat{Q}(j)$ values for the five models range from 349.81 to 1,574.02. Compare to the upper-tailed $N(0,1)$ critical values, (e.g. 2.33 at the 1% level), the large Q -statistics are overwhelming evidence that all the five models are severely misspecified at any reasonable significance level. The Vasicek model performs the worst among the five models, with $\hat{Q}(j)$ values around 1,400 for all lags from 1 to 20. This is probably due to the fact that the Vasicek model assumes interest rate volatility to a constant which is obviously not consistent with the data. The CIR model dramatically reduces the $\hat{Q}(j)$ values to around 620 and the goodness of fit is further improved, in their order, by Ahn and Gao’s (1999) inverse Feller model, Ait-Sahalia’s (1996) nonlinear drift model, and the CKLS model. The CKLS model performs the best, with the $\hat{Q}(j)$ values around 370.

As shown in Figure 5(b), the $\hat{H}(j)$ tests also overwhelmingly reject all the five models at any conventional level. The $\hat{H}(j)$ values are different from the $\hat{Q}(j)$ values (as predicted by the asymptotic theory), though the relative ranking among the five models remains the same. The $\hat{H}(j)$ values are around 570 for the Vasicek model, and are dramatically reduced to around 460 for the CIR model. The CKLS model has the smallest $\hat{H}(j)$ statistics, which are around 360.

Although certain models perform better than others in the relative term, the extremely large test statistics of all five models indicate that none of them can adequately capture the dynamics of the spot rate. To include nonlinear drifts does not significantly improve the goodness of fit. There is obviously a long way to go before obtaining a correct model specification from any of these single-factor models.

Next we explore the possible reasons for the rejection of the single-factor diffusion models. Following Diebold, et al. (1998), we study the i.i.d. and $U[0, 1]$ properties of the generalized residuals $\{\hat{Z}_\tau\}_{\tau=1}^n$, which provide rich information about the possible sources of model misspecifications.

We compare the marginal densities of the five models with a nonparametric kernel estimator of the marginal density. Figure 6(a) shows that all five models underpredict the likelihood of small interest rate changes, or the values of the interest level around the mean: the model-implied probability of the interest rates around the mean is always smaller than the nonparametric counterpart. In contrast to most models for stock prices and exchange rates, all five models overpredict the tail probabilities of interest rates. In other words, all models overestimate the kurtosis of the interest rate level and underestimate the probabilities of the small movements of

interest rates around the mean. This could be the consequence of underestimating the speed of mean reversion (if any) or overestimating the magnitude of volatility. In Figure 6(b), we also plot the kernel density estimates for the transformed series $\{\hat{Z}_\tau\}_{\tau=1}^n$ for each of the five models. These density estimates all have peaks near 0.5, which indicate, in an alternative but equivalent way, that too many observations fall into the area near the mean of the interest rate level than predicted by each model. How to improve the goodness-of-fit in the tail and center of the distributions of interest rates should be a primary concern for future research. Clearly, the smoothed histogram of the generalized residuals $\{\hat{Z}_\tau\}_{\tau=1}^n$ is informative in revealing the aspects of model inadequacy, particularly the inadequacy of the model stationary or marginal density.

The $U[0, 1]$ property of the generalized residuals, however, does not provide all the information about model specification. To further understand the dynamic aspects of a diffusion model, we examine the autocorrelations of $\{Z_\tau^m\}_{\tau=1}^n$, for $m = 1, 2, 3, 4$, which reveal important information about the adequacy of the specification of the conditional mean, variance, skewness and kurtosis of the generalized residuals. As shown in Figure 7, the autocorrelations of the first four moments of $\{\hat{Z}_\tau\}_{\tau=1}^n$ in general are not zero, which indicate that none of the models can fully capture the dynamics of the data. All models exhibit similar behavior in $\text{corr}(\hat{Z}_\tau, \hat{Z}_{\tau-j})$, which shows that the introduction of nonlinear drift does not significantly improve the modeling of the conditional mean of the generalized residuals. In terms of modeling the conditional variance, skewness and kurtosis of the generalized residuals, the Vasicek model, which has a constant volatility, performs the worst. The autocorrelations of $\{Z_\tau^m\}_{\tau=1}^n$, for $m = 2, 3, 4$, show that the CIR model provides certain improvements, and the CKLS model provides the most significant improvements over the Vasicek model. The three models have the same drift specification, thus the improvements must have come from the difference in the way they model volatility. It appears that allowing volatility to depend on the level of interest rate is an important part in modeling interest rate data. Our results also confirm existing findings in the literature that the elasticity of volatility with respect to the interest rate level should be close to 1.5. The Ahn and Gao and Ait-Sahalia's nonlinear drift model provide certain improvements over the CIR model. The improvement, however, is mainly due to the specification of the volatility function, rather than the drift function, as the nonlinear drift models still underperform the CKLS model, which has a linear drift.

The inadequacy of the above single-factor diffusion models could be due to the reason that the interest rate simply does not follow a Markov process. We notice this possibility because the test statistics $\hat{Q}(j)$ and $\hat{H}(j)$ of all five models do not decline as lag order j increases. In other words, the generalized residuals $\{\hat{Z}_\tau\}_{\tau=1}^n$ display highly persistent dependence for each model. This, of course, may be due to the misspecification of the drift and/or diffusion functions. It may also suggest that using interest rate level alone seems insufficient to capture interest rate dynamics. In other words, the Markov property may not be a reasonable assumption. To check this, we

first estimate the transition density of data under the Markov assumption, using the following nonparametric density estimator:

$$\hat{p}_0(x, \tau\Delta|y, (\tau-1)\Delta) \equiv \frac{\sum_{\tau=2}^n K_h(x, X_{\tau\Delta})K_h(y, X_{(\tau-1)\Delta})}{\sum_{\tau=2}^n K_h(y, X_{(\tau-1)\Delta})}, \quad (29)$$

where $K_h(x, y)$ is as defined in (13), $h = \hat{S}_X n^{-\frac{1}{6}}$, and \hat{S}_X is the sample standard deviation of the sample $\{X_{\tau\Delta}\}_{\tau=1}^n$, which has been scaled to have support on the unit interval $[0, 1]$. Then we compute the probability integral transform:

$$\hat{Z}_\tau^0 \equiv \int_{x_0}^{X_{\tau\Delta}} \hat{p}_0(x, \tau\Delta|X_{(\tau-1)\Delta}, (\tau-1)\Delta) dx, \tau = 1, 2, \dots, n. \quad (30)$$

Under the Markov assumption, $\hat{p}_0(x, \tau\Delta|y, (\tau-1)\Delta)$ is consistent for the transition density $p_0(x, \tau\Delta|y, (\tau-1)\Delta)$ of X_t . As a consequence, the transformed series $\{\hat{Z}_\tau^0\}_{\tau=1}^n$ will be approximately i.i.d. $U[0, 1]$. Any deviation from i.i.d. $U[0, 1]$ will indicate violation of the Markov assumption. Figure 8 displays the kernel density estimates of $\{\hat{Z}_\tau^0\}_{\tau=1}^n$ and the sample autocorrelations of $\{\hat{Z}_\tau^0\}$. There appears significant evidence that $\{\hat{Z}_\tau^0\}$ is not uniformly distributed. In particular, there is a mode near 0.5, indicating that too many observations fall into the area near the mean of the interest rate than predicted by a Markovian model. Moreover, $\{\hat{Z}_\tau^0\}$ shows persistent autocorrelation. Thus, the Markov assumption does not hold for the data.

B. Discrete Time Series Models

The violation of the Markov property maybe due to the fact that we only consider single-factor diffusion models whose conditional density only depends on previous level of interest rate. As argued by many authors, the single-factor diffusion models are too simplistic to capture the rich dynamics of interest rate data. As a result, many more complex interest rate models have been developed in the literature.

For example, Brenner, Harjes, and Kroner (1996) and Andersen and Lund (1997) argue that it is too restrictive to assume that interest rate volatility solely depends on the level of the interest rate. It fails to capture those situations in which volatility is high but interest rate is low, or vice versa. It also ignores the obvious volatility clustering in interest rate data. They show that GARCH and/or stochastic volatility models provide significantly better fit of the data. Other authors, such as Gray (1996), Ang and Bekaert (1999), and Li and Xu (2000), point out that it is unrealistic to assume that the interest rate follow a time-homogeneous process over a long period of time, given the changes in business cycle and general macroeconomic conditions. These authors show that Markov regime switching models perform much better than the single-factor diffusion models. Das (2002) and Johannes (2000) also argue that surprises is a rule rather than an exception in financial markets. Interest rate, like other asset prices, exhibit infrequent jumps due to discrete release of important information. They show that jumps play an important role in capturing the marginal distribution of interest rate data.

To examine whether the violation of the Markov property is due to the ignorance of stochastic volatility, regime shifts, or jumps in interest rates, we consider the following discrete time non-Markovian interest rate models: GARCH models with nonlinear drift, regime-switching models with GARCH effects and state-dependent transition matrix, and the discretized jump-diffusion models with ARCH/GARCH effects.

We first consider a GARCH(1,1) model with both nonlinear drift and level effect:

$$\begin{aligned}\Delta X_{t\Delta} &= \alpha_{-1}X_{(t-1)\Delta}^{-1} + a_1X_{(t-1)\Delta} + \alpha_2X_{(t-1)\Delta}^2 + \sigma X_{(t-1)\Delta}^\rho \sqrt{h_{t\Delta}}\epsilon_{t\Delta}, \\ h_{t\Delta} &= \beta_0 + h_{(t-1)\Delta}(\beta_1 + \beta_2\epsilon_{(t-1)\Delta}^2),\end{aligned}$$

where $\epsilon_{t\Delta} \sim N(0, \Delta)$. This model is not Markov as $h_{(t-1)\Delta}$ depends on $X_{(t-2)\Delta}$. As shown by Nelson (1990), in continuous time limit a GARCH model converges to a stochastic volatility model which becomes Markov with the additional state variable, the unobservable stochastic volatility. Therefore GARCH models can be considered as a discrete-time approximation of a continuous-time stochastic volatility model, and Nelson (1990) shows that the approximation works well for daily data. Hence, our analysis of GARCH models also has implications for continuous-time stochastic volatility model, which is much more difficult to estimate (we leave the estimation and testing of continuous-time stochastic volatility models for interest rate for future research).

Next, we consider regime switching models with GARCH and level effect

$$\begin{aligned}\Delta X_{t\Delta} &= \alpha(s_{t\Delta}) + \beta(s_{t\Delta})X_{(t-1)\Delta} + \sigma(s_{t\Delta})X_{(t-1)\Delta}^{\rho(s_{t\Delta})} \sqrt{h_{t\Delta}}\epsilon_{t\Delta}, \\ h_{t\Delta} &= \beta_0 + h_{(t-1)\Delta}(\beta_1\epsilon_{(t-1)\Delta}^2 + \beta_2),\end{aligned}$$

where s_t follows a two-state Markov chain with time-varying transition matrix as specified in Ang and Bekaert (1998):

$$\begin{aligned}P(s_{t\Delta} = 1 | s_{(t-1)\Delta} = 1) &= \frac{\exp(a_{01} + a_{11}X_{(t-1)\Delta})}{1 + \exp(a_{01} + a_{11}X_{(t-1)\Delta})} \\ P(s_{t\Delta} = 0 | s_{(t-1)\Delta} = 0) &= \frac{\exp(a_{00} + a_{10}X_{(t-1)\Delta})}{1 + \exp(a_{00} + a_{10}X_{(t-1)\Delta})}.\end{aligned}$$

The above regime-switching model is slightly different from the specification used in Gray (1996). Unlike Gray (1996) who assumes $\rho = 0.5$, we estimate this parameter from the data and allow it to be regime dependent. Our GARCH specification is also different from that of Gray (1996), who removes the path-dependence nature of GARCH models by averaging over regimes the conditional and unconditional variances at every time point and allows the GARCH and ARCH parameters to be regime dependent. We find the estimation of Gray's (1996) model specification is very unstable: for instance, the sum of the GARCH parameters reaches 5. Ang and Bekaert (1998) also have similar experience that the estimates of Gray's model fail to converge. Our model specification

on the other hand, turns out to have much better convergence properties. Due to the GARCH effect and the fact that the ex ante probability $P(s_{t\Delta} = 1 | I_{(t-1)\Delta})$ depends on $I_{(t-2)\Delta}$, the above model is also not Markov.

Finally, we consider the discretized version of the following jump-diffusion model:

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t + J(X_t)d\pi[\lambda(X_t)],$$

where $J(X_t)$ is the jump size and $\lambda(X_t)$ is the jump intensity. Das (2002) considers the discretized version of the above model with nonlinear drift and ARCH effect. The only difference in our model is that we also introduce level and GARCH effect in jump-diffusion models. Especially we consider the following discretized jump-diffusion models:

$$\begin{aligned} \Delta X_{t\Delta} &= \alpha_{-1}X_{(t-1)\Delta}^{-1} + a_1X_{(t-1)\Delta} + \alpha_2X_{(t-1)\Delta}^2 + \sigma X_{(t-1)\Delta}^\rho \sqrt{h_{t\Delta}}\epsilon_{t\Delta} + J(\mu, \gamma^2)\Delta\pi(q), \\ h_{t\Delta} &= \beta_0 + \beta_1[X_{(t-1)\Delta} - E(X_{(t-1)\Delta} | X_{(t-2)\Delta})]^2, \text{ (for ARCH model)} \\ h_{t\Delta} &= \beta_0 + h_{(t-1)\Delta}(\beta_1 + \beta_2\epsilon_{(t-1)\Delta}^2) \text{ (for GARCH model)} \end{aligned}$$

where the jump size J follows an i.i.d. $N(\mu, \gamma^2)$ and $\Delta\pi(q)$ follows an i.i.d. Bernoulli (q) distribution. While we estimate the discretized version of the jump-diffusion model, Das (2002) has shown that the discretization bias is very small for daily data.

The above discrete time models can be easily estimated using the maximum likelihood method, as their conditional distributions are known in closed-form. Parameter estimates of each model are shown in Table 2. It is clear that the more complex models generally have higher likelihoods. We conduct the probability integral transform of the data under each model using the model-implied conditional probability density $p(X_{\tau\Delta} | I_{(\tau-1)\Delta})$, where $I_{(\tau-1)\Delta}$ is the information set at $\tau - 1$. After applying the nonparametric tests to the transformed data, we obtain test statistics under Q-test and H-test which are shown in Figure 9 (a) and (b).

It is interesting to point out that although the above models have been demonstrated in the literature to provide significant improvements over the single-factor diffusion models, they are still overwhelmingly rejected by our tests. The GARCH(1,1) model with nonlinear drift and level effect performs the worst among the four models, with Q-statistics above 1,000 and H-statistics close to 400. Introducing regime-shifts significantly improve the performance over GARCH models under the Q-test, and some improvements under the H-test. Jump-diffusion models with ARCH/GARCH effects perform the best among the four models. The jump-diffusion model with ARCH effect has the smallest Q-statistics, while the jump-diffusion model with GARCH effect has the smallest H-statistics. The difference in test statistics between the two jump-diffusion models is not big under both tests. We also compare the performance of the four discrete time series models with that of the single-factor diffusion models. The H-statistics of the jump-diffusion models with ARCH/GARCH effects are significantly smaller than that of the best performing single-factor

diffusion model, the CKLS model. While the improvements are obvious, the large test statistics show that the jump-diffusion models are still grossly misspecified. The Q-statistics of the jump-diffusion models, however, are not significantly different from that of the CKLS model, which maybe due to parameter estimation uncertainty.

Having demonstrated the failures of all the existing models, we examine the behavior of the probability integral transformed data to understand the possible sources of rejections. We first examine the marginal density of $\{\hat{Z}_\tau\}_{\tau=1}^n$ for the four different models in Figure 10. The improvement from single-factor diffusion models to GARCH models is not significant. But it is very clear that the introduction of regime shifts or jumps significantly improve the goodness of fit. The nonparametric density of $\{\hat{Z}_\tau\}_{\tau=1}^n$ becomes much more uniform for the regime-switching and the jump-diffusion models. This is consistent with the findings in Ang and Bekaert (1998), Das (2001) and Johannes (2000) that regime shifts and jumps are important factors in modeling the marginal density of interest rate data.

Then we examine in Figure 11 the autocorrelations of $\{\hat{Z}_\tau^m\}_{\tau=1}^n$ for $m = 1, 2, 3, 4$, under different model specifications. It is interesting to note that all four models behave similarly in capturing the conditional mean of the generalized residuals. This is consistent with the findings in Ang and Bekaert (1998) and Li and Xu (2000) that regime switching models with linear drift in each individual regime provide similar description of the conditional mean as a model with nonlinear drift. The four models also perform similarly in terms of capturing conditional variance, skewness and kurtosis of the residuals, although the jump-diffusion models provide some improvements over the other models. By comparing the autocorrelations of $\{\hat{Z}_\tau^m\}_{\tau=1}^n$ for $m = 1, 2, 3, 4$, of the discrete time models with that of the single-factor diffusion models, we can see that the discrete time models significantly improve the modeling of the conditional variance, skewness and kurtosis. For example, we find that the autocorrelations of \hat{Z}_τ^2 , \hat{Z}_τ^3 , and \hat{Z}_τ^4 are significantly different from zero for the five diffusion models, but are significantly reduced for the discrete time models. Our calculation of $Corr \left[\hat{Z}_\tau^2, \hat{Z}_{\tau-j} \right]$ and $Corr \left[\hat{Z}_\tau, \hat{Z}_{\tau-j}^2 \right]$, which measures the so-called leverage effect and the ARCH-IN-MEAN effect, also show the same results: the discrete time series models show improvements over the diffusion models.

Our analysis of the i.i.d. $U[0, 1]$ property of the transformed data $\{\hat{Z}_\tau\}_{\tau=1}^n$ of all the models reveal some interesting findings. First, we find that in general linear and nonlinear drift models perform similarly in modeling the conditional mean of $\{\hat{Z}_\tau\}_{\tau=1}^n$. Second, introducing level effect is important and models with the elasticity parameter closer to 1.5 perform better. Third, introducing GARCH effects, regime shifts and jumps improves the modeling of the conditional variance, skewness and kurtosis of the generalized residuals. Fourth, including regime shifts and jumps provide the most significant improvements in modeling the marginal density of interest rate data. Fifth, even the complicated models that allow for changing volatility, regime shifts, and jumps

are still not adequate to capture the rich dynamics of interest rate data.

VI. Conclusion

We have developed two nonparametric specification tests for continuous-time diffusion models. Our tests extend Ait-Sahalia's (1996) work in several directions. Instead of using the marginal density, we use the transition density, which can capture the full dynamics of a diffusion model. We employ a dynamic probability integral transform, which converts any highly persistently dependent data into an i.i.d. $U[0,1]$ sequence via the model implied transition density under correct model specification. Our tests compare a kernel estimator of the joint density of the transformed data with the product of uniform density and can detect any deviation from either i.i.d. or $U[0,1]$. As there is no serial dependence in the transformed data under correct model specification, nonparametric density estimators and related test statistics are expected to perform well in finite samples. Our omnibus tests are further supplemented by a spectrum of graphical smoothed histograms and autocorrelograms of the generalized residuals, which provide constructive information about the possible sources of model misspecification. Our tests, although developed for single-factor diffusion models, are applicable to a wide range of dynamic economic models, including discrete time series models, time-inhomogeneous diffusion models, stochastic volatility models, jump diffusion models and multi-factor term structure models. Simulation studies show that our tests perform rather well in finite samples even for data with highly persistent dependence. It has reasonable size and good power against a wide range of alternatives.

The probability integral transform requires the model-implied transition density, whose closed form solution is usually not available. For general diffusion models, one can use simulation methods developed in Pedersen (1995), Brandt and Santa-Clara (2001) and Elarian, Chib, and Shephard (2002) to estimate the model-implied transition density. Ait-Sahalia's (2001) recent important work provides a convenient and accurate closed-form approximation for the model-implied transition density for single-diffusion models.

When applied to U.S. interest rate data, our tests overwhelmingly reject some popular interest rate models in the existing literature, including single-factor diffusion models studied in Ait-Sahalia (1996) and discrete time models with GARCH effects, regime shifts and jumps. We find that models with nonlinear drifts do not significantly improve the goodness of fit, and the main reason for model inadequacy seems to be the violation of the Markov assumption. We further find that introducing GARCH effects, regime shifts and jumps help improve the performance of the models, although these more complicated models are also overwhelmingly rejected by our tests. Our study shows that nonparametric methods can provide a reliable and powerful tool for analyzing financial data. In future research, we would like to apply our tests to more complicated interest rate models, such as continuous-time stochastic volatility models with jumps,

continuous-time Levy processes and multi-factor term structure models.

REFERENCES

- Abramowitz, M. and I. A. Stegun, 1970, *Handbook of Mathematical Functions*. U.S. Govt. Print. Off., Washington, D.C.
- Ahn, D., and B. Gao, 1999, A Parametric Nonlinear Model of Term Structure Dynamics, *Review of Financial Studies* 12, 721-762.
- Ait-Sahalia, Y., 1996, Testing Continuous-Time Models of the Spot Interest Rate, *Review of Financial Studies* 9, 385-426.
- Ait-Sahalia, Y., 1999, Transition Densities for Interest Rate and Other Nonlinear Diffusions, *Journal of Finance* 54, 1361-1395.
- Ait-Sahalia, Y., 2001, Maximum-Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approach, forthcoming in *Econometrica*.
- Ang, A., and G. Bekaert, 1998, Regime Switches in Interest Rates, NBER Working Paper 6508.
- Black, F. and M. Scholes, 1973, The Pricing of Options and Corporate Liabilities, *Journal of Political Economy* 81, 637-654.
- Black, F., E. Derman, and W. Toy, 1990, A One-Factor Model of Interest Rates and its Application to Treasury Bond Options, *Financial Analysts Journal*, January-February, 33-39.
- Brandt, M., and P. Santa-Clara, 2001, Simulated Likelihood Estimation of Diffusions with an Application to Exchange Rate Dynamics in Incomplete Markets, *Journal of Financial Economics* forthcoming.
- Brenner, R., R. Harjes, and K. Kroner, 1996, Another Look at Alternative Models of Short-Term Interest Rate, *Journal of Financial and Quantitative Analysis* 31, 85-107.
- Brown, B. M. 1971, Martingale Limit Theorems, *Annals of Mathematical Statistics*, 42, 59-66.
- Campbell, J., A. Lo and C. MacKinlay, 1997, *The Econometrics of Financial Markets* (Princeton University Press, New Jersey).
- 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.
- Chapman, D., and N. Pearson, 2000, Is the Short Rate Drift Actually Nonlinear?, *Journal of Finance* 55, 355-388.
- Chapman, D., and N. Pearson, 2001, Recent Advances in Estimating Models of the Term Structure, forthcoming in *Financial Analysts Journal*.
- Chen, S., W. Härdle, and T. Kleinow, 2000, An Empirical Likelihood Goodness-of-Fit Test for Time Series, Working paper, Humboldt-Universität zu Berlin, Germany.
- Chib, S., F. Nardari, N. Shephard, 2001, Markov Chain Monte Carlo Methods for Stochastic Volatility Models, *Journal of Econometrics* forthcoming.
- 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. Swanson, 2001, Bootstrap Specification Tests with Dependent Observations and Parameter Estimation Error, Working Paper, Department of Economics, University of Exeter, U.K.
- Cox, J.C., J.E. Ingersoll and S.A. Ross, 1985, A Theory of the Term Structure of Interest Rates, *Econometrica* 53, 385-407.
- Derman, E., and I. Kani, 1994, The Volatility Smile and Its Implied Tree, Quantitative Strategies Research Notes (Goldman Sachs, New York).
- Das, S., 2002, The Surprise Element: Jumps in Interest Rates, *Journal of Econometrics*, 106, 27-65.
- Dai, Q., and K. Singleton, 2000, Specification Analysis of Affine Term Structure Models, *Journal of Finance*, Vol. LV, 1943-1978.
- Duffee, G., 2000, Term Premia and Interest Rate Forecasts in Affine Models, *Journal of Finance* forthcoming.
- Diebold, F. X., T. Gunther, and A. Tay, 1998, Evaluating Density Forecasts with Applications to Financial Risk Management, *International Economic Review* 39, 863-883.
- Duffie, D., and K. Singleton, 1993, Simulated Moments Estimation of Markov Models of Asset Prices, *Econometrica* 61, 929-952.
- Duffie, D., and R. Kan, 1996, A Yield-Factor Model of Interest Rates, *Mathematical Finance* 6, 379-406.
- Egorov, A., H. Li, and Y. Xu, 2001, Maximum Likelihood Estimation of Time-Inhomogeneous Diffusions, Working paper, Johnson School of Management, Cornell University.
- Elerian, O., S. Chib, and N. Shephard, 2001, Likelihood Inference for Discretely Observed Non-linear Diffusions, *Econometrica* forthcoming.
- Gallant, A.R. and G. Tauchen, 1996, Which Moments to Match?, *Econometric Theory* 12, 657-681.
- Gallant, A.R. and G. Tauchen, 1998, Reprojection Partially Observed Systems with Applications to Interest Rate Diffusions, *Journal of the American Statistical Association*, 93, 10-24.
- Gallant, R. and G. Tauchen, 2001, Efficient Method of Moments, Working paper, Department of Economics, University of North Carolina and Department of Economics, Duke University.
- Ghysels, E., A.C. Harvey and E. Renault, 1996, Stochastic Volatility, in *Statistical Methods in Finance* (eds. C.R. Rao and G.S. Maddala), 119-191, Amsterdam: North-Holland.
- Granger, C. and T. Teräsvirta, 1993, *Modelling Nonlinear Economic Relationships*. Oxford University Press: New York.
- Gray, S., 1996, Modeling the Conditional Distribution of Interest Rates as a Regime-Switching Process, *Journal of Financial Economics* 42, 27-62.
- Hall, P., 1984, Central Limit Theorem for Integrated Square Error of Multivariate Nonparametric

- Density Estimation, *Journal of Multivariate Analysis* 14, 1-16.
- Hall, P. and C. Hyde. 1980, *Martingale Limit Theory and Its Application*. Academic Press: New York.
- Hansen, L.P. and J.A. Scheinkman, 1995, Back to the Future: Generating Moment Implications for Continuous Time Markov Processes, *Econometrica* 63, 767-804.
- Härdle, W., 1990, *Applied Nonparametric Regression*, (Cambridge University Press: New York).
- Heath, D.C., Jarrow, R.A., Morton, A., 1992, Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claim Valuation, *Econometrica* 60, 77-105.
- Hong, Y. and H. White, 2001, Asymptotic Theory for Nonparametric Entropy Measures of Serial Dependence, Department of Economics, Cornell University and Department of Economics, University of California, San Diego.
- Hull, J. and A. White, 1990, Pricing Interest-Rate Derivative Securities, *Review of Financial Studies* 3, 573-592.
- Hull, J. and A. White, 1993, One-factor Interest-Rate Models and Valuation of Interest-Rate Derivative Securities, *Journal of Financial and Quantitative Analysis* 28, 235-254.
- Johannes, M., 2000, A Nonparametric View of the Role of Jumps in Interest Rates, Working paper, Columbia University.
- Kim, S., N. Shephard, and S. Chib, 1998, Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models, *Review of Economic Studies*, 65, 361-93.
- Kloeden, P.E., E. Platen, and H. Schurz, 1991, *Numerical Solution of SDE through Computer Experiments*, Springer.
- Lo, A.W., 1988, Maximum Likelihood Estimation of Generalized Ito Processes with Discretely Sampled Data, *Econometric Theory* 4, 231-247.
- Li, H., and Y. Xu, 2000, Short Rate Dynamics and Regime Shifts, Working paper, Cornell University.
- Lumsdaine, R., 1996, Asymptotic Properties of the Quasi-Maximum Likelihood Estimator in GARCH(1,1) and IGARCH(1,1) Models, *Econometrica* 64, 575-596.
- Melino, A., 1994, Estimation of Continuous-Time Models in Finance, in: C. Sims, ed., *Advances in Econometrics: Sixth World Congress*, Vol. 2 (Cambridge University Press, Cambridge).
- Merton, R.C., 1980, On Estimating the Expected Return on the Market: An Exploratory Investigation, *Journal of Financial Economics* 8, 141-183.
- Pedersen, A. R., 1995, A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations, *Scandinavian Journal of Statistics*, 22, 55-71.
- Pritsker, M., 1998, Nonparametric Density Estimation and Tests of Continuous Time Interest Rate Models, *Review of Financial Studies* 11, 449-487.
- Rice, J., 1986, Boundary Modification for Kernel Regression, *Communications in Statistics* 12,

1215-1230.

Robinson, P., 1983, Nonparametric Estimators for Time Series, *Journal of Time Series Analysis* 4, 184-207.

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

Rubinstein, M., 1994, Implied Binomial Trees, *Journal of Finance* 49, 771-818.

Silverman, B. W., 1986, *Density Estimation for Statistics and Data Analysis* (Chapman and Hall, London).

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

Sundaresan, S., 2001, Continuous-Time Methods in Finance: A Review and an Assessment, *Journal of Finance* 55, 1569-1622.

Tauchen, G., 1997, New Minimum Chi-Square Methods in Empirical Finance, in: D. Kreps and K. Wallis, eds., *Advances in Econometrics: Seventh World Congress* (Cambridge University Press, Cambridge).

Tauchen, G., 2001, Notes on Financial Econometrics, *Journal of Econometrics* 100, 57-64.

Thompson, T. (2001), Specification Tests for Continuous-time Models, Working Paper, Department of Economics, Harvard University.

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

White, H., 1984, *Asymptotic Theory for Econometricians*, Academic Press: Dan Diego.

White, H., 1994, *Estimation, Inference and Specification Analysis*, Cambridge University Press: New York.

Wong, E., 1964, The Construction of a Class of Stationary Markov Processes, in *Stochastic Processes in Mathematical Physics and Engineering*, Proceedings of Symposia in Applied Mathematics, 16, ed. by R. Bellman, American Mathematical Society, Providence, RI, 264-276.

Table I. Parameter Estimates of Single-factor Diffusion Models

This table contains parameter estimates for five popular spot rate models using daily Eurodollar interest rates from June 1, 1973 to February 25, 1995, as used in Ait-Sahalia (1996). For ease of comparison, we write each model as a special case of Ait-Sahalia's nonlinear drift model. Therefore we have, for Vasicek model, $dX_t = (\alpha_0 + \alpha_1 X_t) dt + \sigma dW_t$; for CIR model, $dX_t = (\alpha_0 + \alpha_1 X_t) dt + \sigma X_t^{0.5} dW_t$; for Ahn and Gao model, $dX_t = (\alpha_1 X_t + \alpha_2 X_t^2) dt + \sigma X_t^{1.5} dW_t$; for CKLS model, $dX_t = (\alpha_0 + \alpha_1 X_t) dt + \sigma X_t^\rho dW_t$; and for the nonlinear drift model, $dX_t = (\alpha_{-1} X_t^{-1} + \alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2) dt + \sigma X_t^\rho dW_t$. Parameter estimates are obtained using maximum likelihood method: the true likelihood function is used if available; otherwise the Hermite approximation of the likelihood function is used. Standard errors are given in the parentheses.

Model	α_{-1}	α_0	α_1	α_2	σ	ρ	Log-Likelihood
Vasicek	0.0	0.13 (3.81)	-1.59 (-4.18)	0.0	0.064 (104.40)	0.0	22503.6
CIR	0.0	0.096 (2.91)	-1.27 (-2.68)	0.0	0.19 (104.37)	0.5	23605.1
Ahn & Gao	0.0	0.0	0.94 (3.34)	-12.60 (-2.50)	2.17 (108.92)	1.5	24364.1
CKLS	0.0	0.04 (2.04)	-0.62 (-1.98)	0.0	1.48 (18.41)	1.35 (62.95)	24385.7
Nonlinear Drift	0.0001 (0.03)	-0.02 (-1.45)	1.47 (1.78)	-15.41 (-2.12)	1.50 (18.78)	1.36 (64.44)	24388.5

Table II. Parameter Estimates of Discrete Time Series Models

This table contains parameter estimates for four discrete time series spot rate models using daily Eurodollar interest rates from June 1, 1973 to February 25, 1995, as used in Ait-Sahalia (1996).

Panel A. Parameter Estimates for GARCH Model (with nonlinear drift and level effect) and Regime Switching Model (with GARCH and level effect)

Parameters	Estimates (GARCH)	Std. Error (GARCH)	Parameters	Estimates (RS)	Std. Error (RS)
α_{-1}	-0.0984	0.1249	α_0	1.5378	1.5378
α_0 (1e-02)	5.0494	6.3231	β_0	-1.0646	0.4207
α_1 (1e-03)	-4.4132	9.2876	α_1	-0.0013	0.0351
α_2	0.0000	0.0004	β_1	-0.0076	0.0484
			σ_1	0.3355	0.0483
ρ	1.0883	0.0408	ρ_0	0.3566	0.0693
			ρ_1	0.0064	0.0512
β_0 (1e-03)	0.0738	0.0119	b_0 (1e-03)	6.5126	1.9898
β_1 (1e-01)	6.4117	0.1359	b_1	0.0224	0.0034
β_2 (1e-01)	3.5260	0.2181	b_2	0.7810	0.0254
			a_{00}	0.2350	0.2192
			a_{01}	4.5398	0.2691
			a_{10}	0.0208	0.0184
			a_{11}	-0.2800	0.0296
Log-Likelihood	654.13		Log-Likelihood	2712.97	

Panel B. Parameter Estimates for Jump-Diffusion Model (with nonlinear drift, ARCH/GARCH and level effect)

Parameters	Estimates (ARCH)	Std. Error (ARCH)	Estimates (GARCH)	Std. Error (GARCH)
α_{-1}	0.2422	0.0776	0.2109	0.0768
α_0 (1e-01)	-1.3077	0.3865	-1.0865	0.3843
α_1 (1e-02)	2.0799	0.5655	1.6082	0.5719
α_2	-0.0010	0.0003	-0.0007	0.0003
ρ	0.7645	0.0576	0.3698	0.0445
β_0 (1e-03)	0.9009	0.0982	0.8765	0.0587
β_1	0.1041	0.0143	0.1488	0.0153
β_2			0.1747	0.0247
q	0.1015	0.0060	0.0789	0.0055
μ	0.0053	0.0263	-0.0138	0.0349
γ	0.7997	0.0307	0.7957	0.0379
Log-Likelihood	2482.39		2568.13	

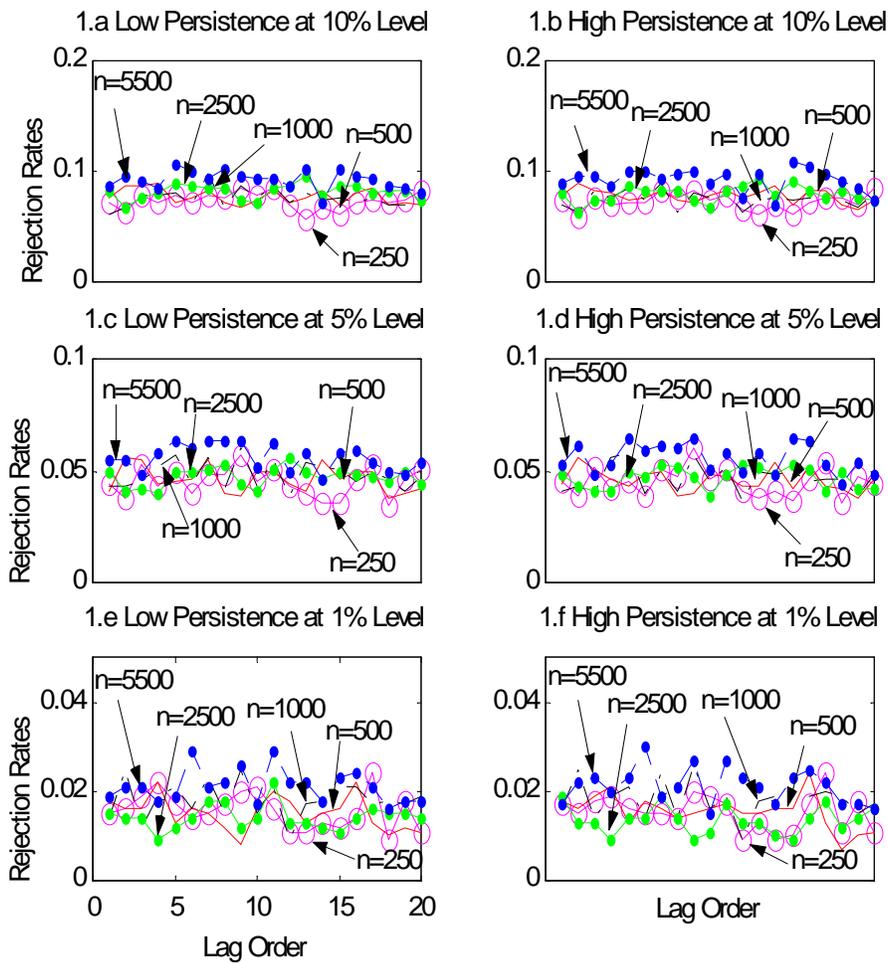


Figure 1: Size of the Q-Test for both high and low level of persistent dependence

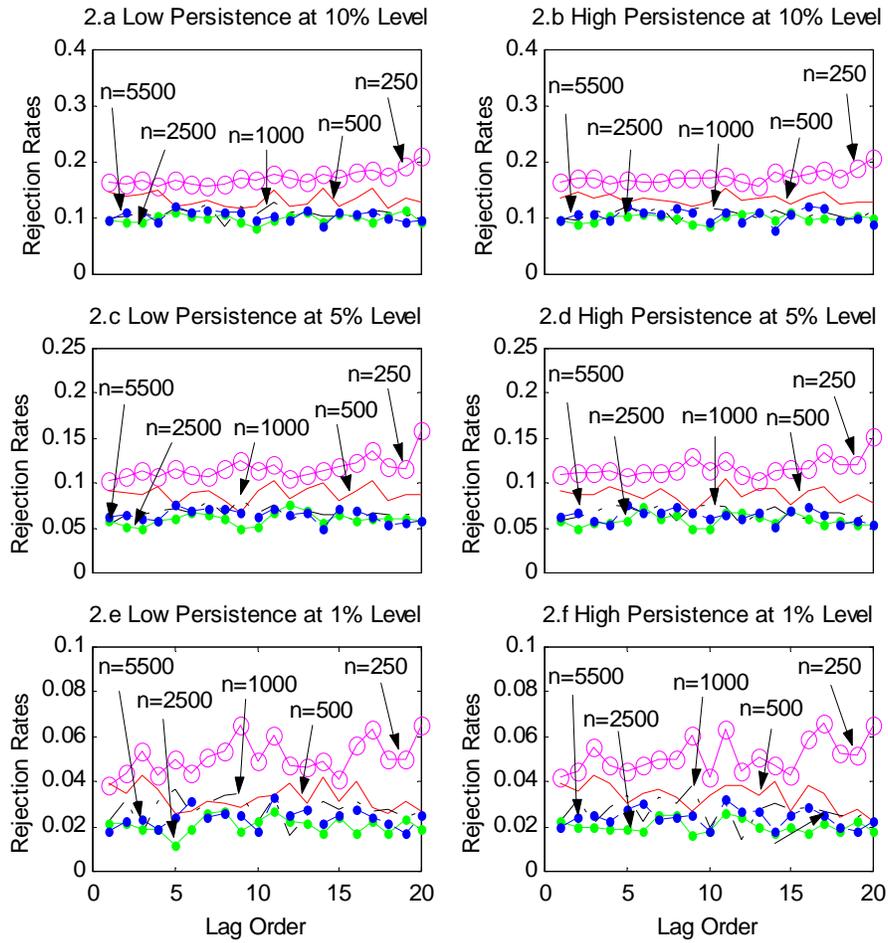


Figure 2: Size of the H-Test for both high and low level of persistent dependence

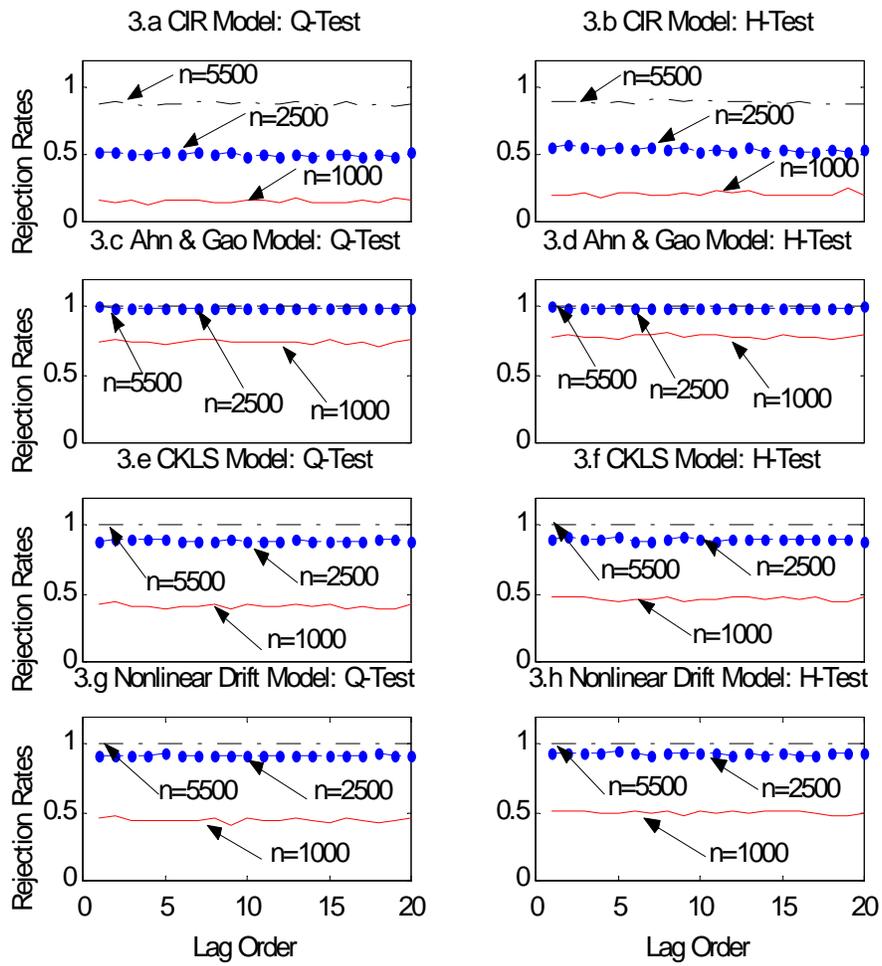


Figure 3: Power at 5% level of the Q-Test and H-Test against different alternatives

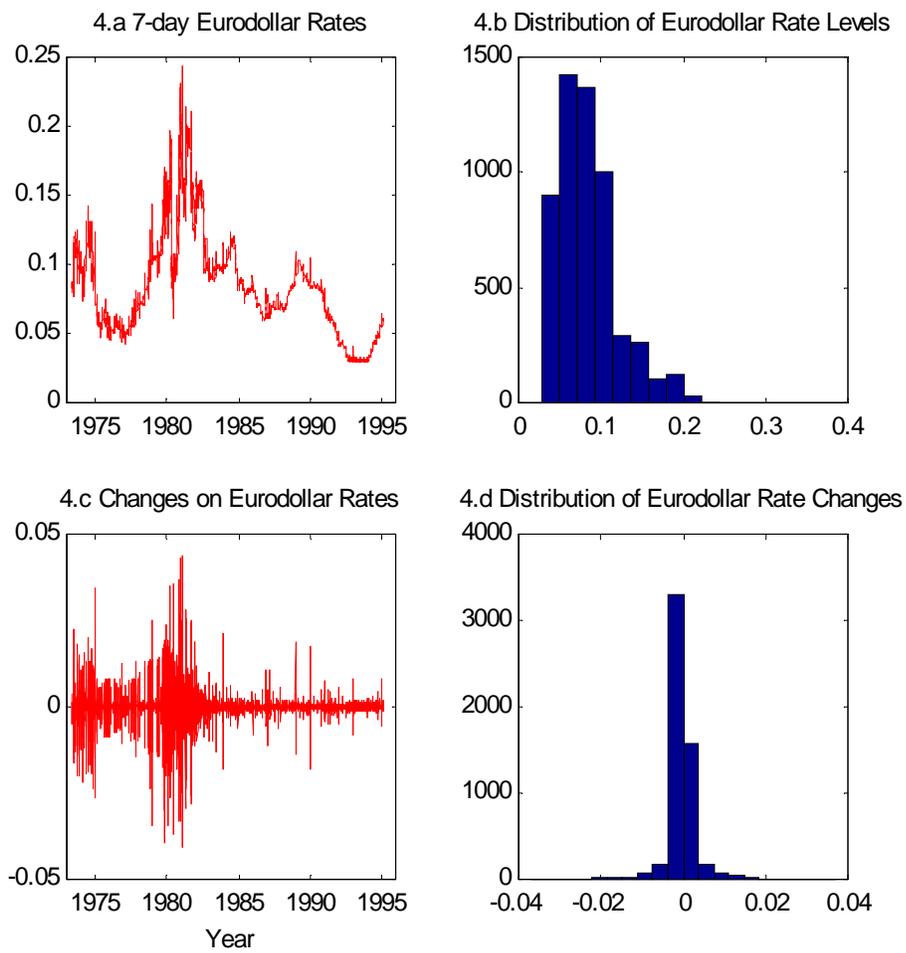


Figure 4: Level and change of 7-day Eurodollar interest rates

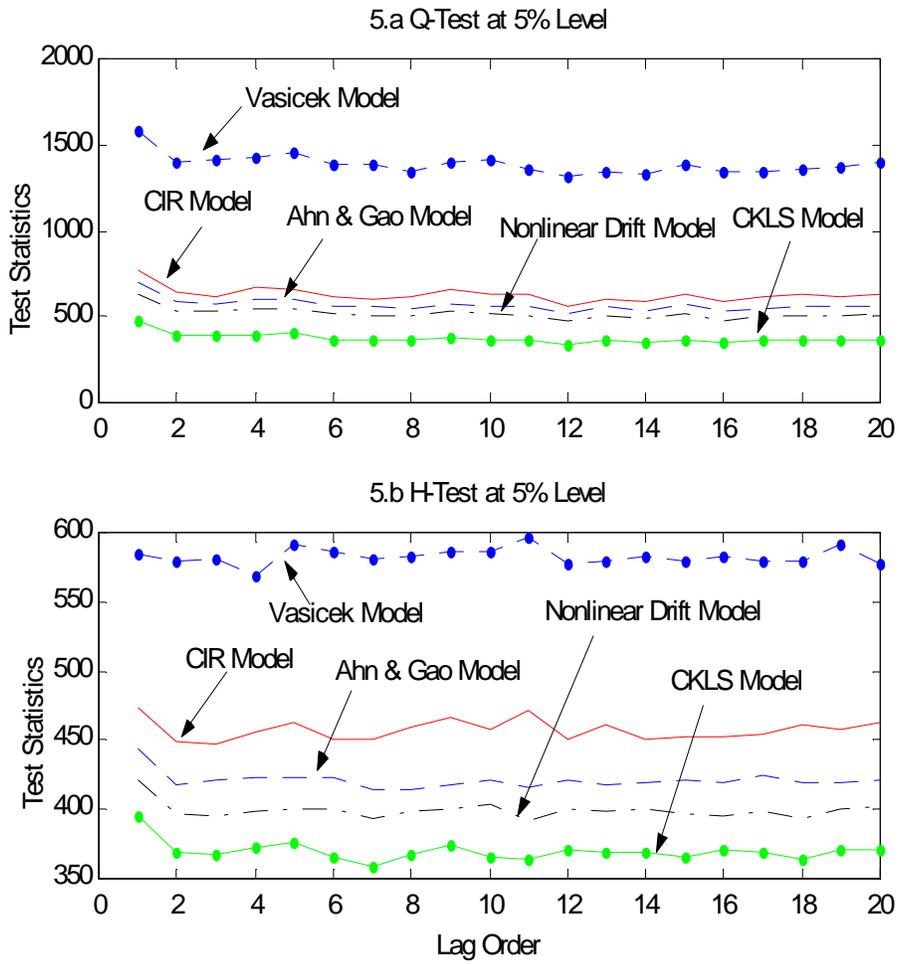
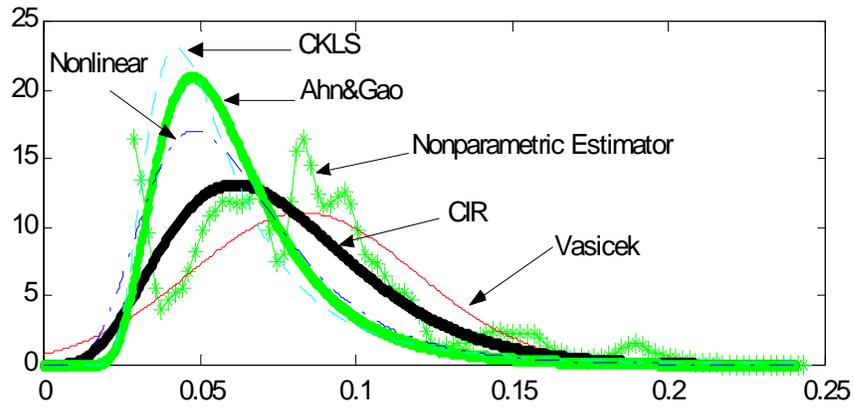


Figure 5: Test statistics under Q-Test and H-Test for five different spot rate models

6.a Parametric and Nonparametric Marginal Densities for the Eurodollar Interest Rates



6.b Nonparametric Densities for the PITs Under Five Different Models

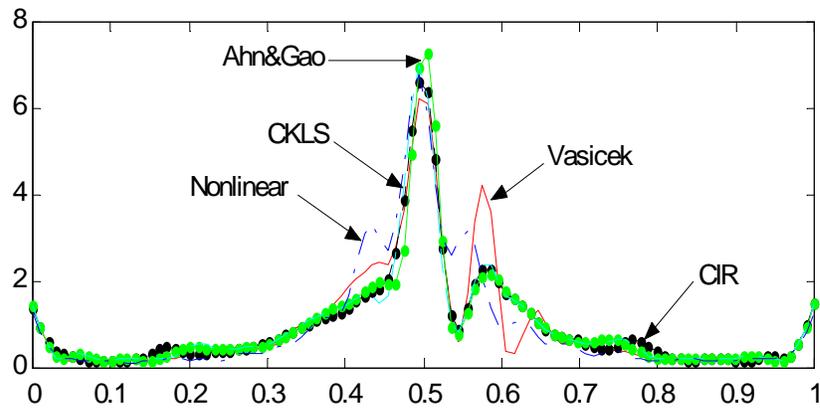


Figure 6: Marginal densities of the original and transformed data for five different parametric spot rate models

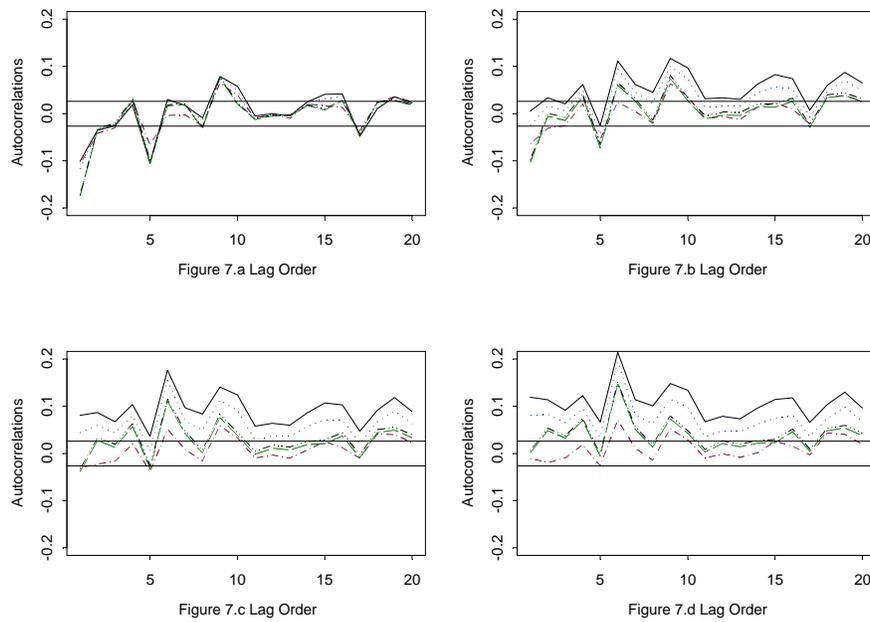
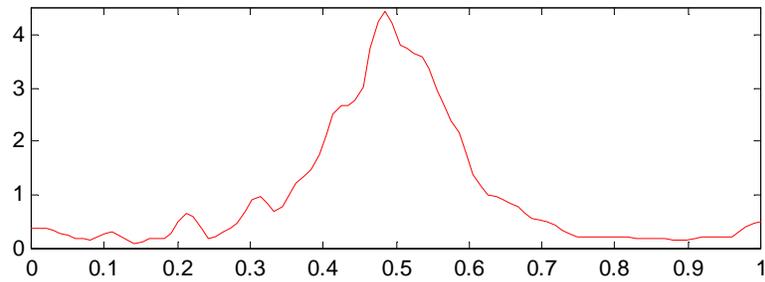


Figure 7: Autocorrelations of the first four moments of the “generalized residuals” $\left\{ \widehat{Z}_\tau^m \right\}_{\tau=1}^n$ for the five single-factor diffusion models. Subplots a, b, c, and c represents $Corr \left[\widehat{Z}_\tau^m, \widehat{Z}_{\tau-j}^m \right]$ for $m = 1, 2, 3,$ and 4 . In each plot, from top to bottom are Vasicek, CIR, Ahn and Gao, Nonlinear Drift, and CKLS models.

8.a Nonparametric Density of the PITs via Nonparametric Transition Density Under Markov Assumption



8.b Autocorrelation of the PITs via Nonparametric Density Under Markov Assumption

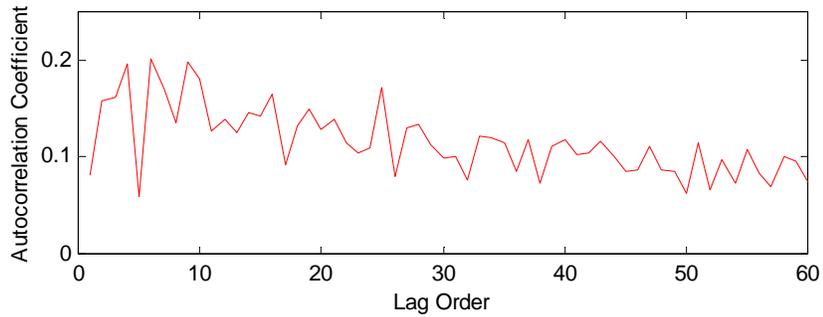


Figure 8: Marginal density and autocorrelation of the probability integral transforms (PITs) of original data via nonparametric transition density estimated under the Markov assumption

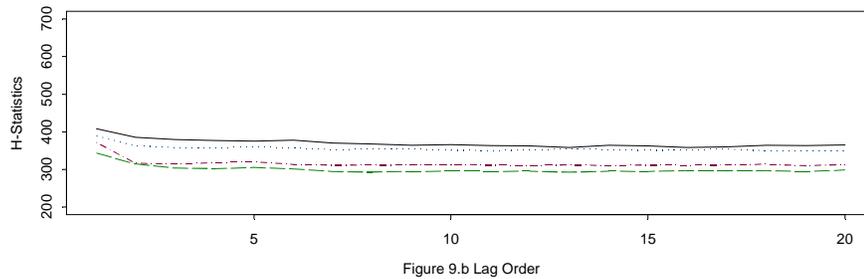
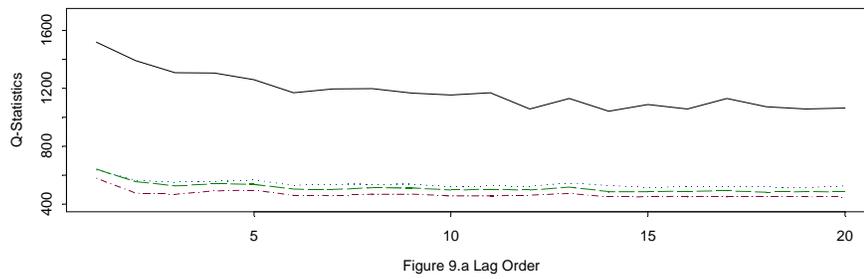


Figure 9: Test statistics under Q-test and H-test for the four discrete time series models. For Q-test, from top to bottom are GARCH, Regime switching, Jump-diffusion/GARCH, and Jump-diffusion/ARCH models. For H-test, from top to bottom are GARCH, Regime switching, Jump-diffusion/ARCH, and Jump-diffusion/GARCH models.

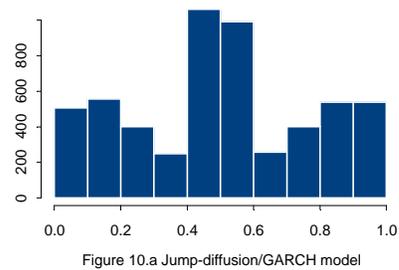
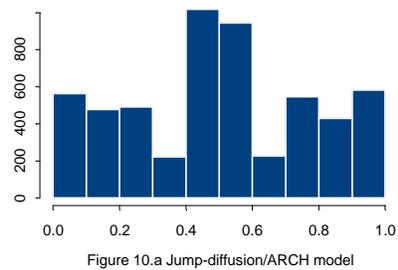
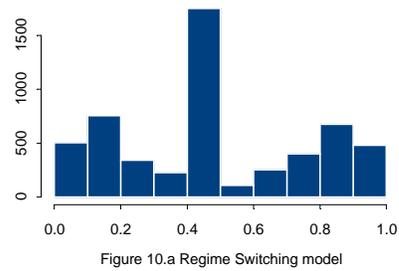
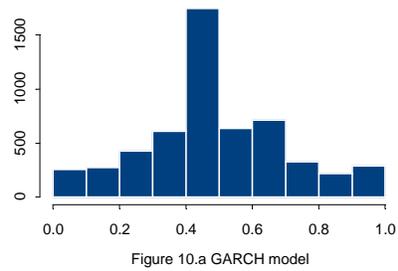


Figure 10: Histograms of the “generalized residuals” $\{Z_\tau\}_{\tau=1}^n$ for four discrete time series models

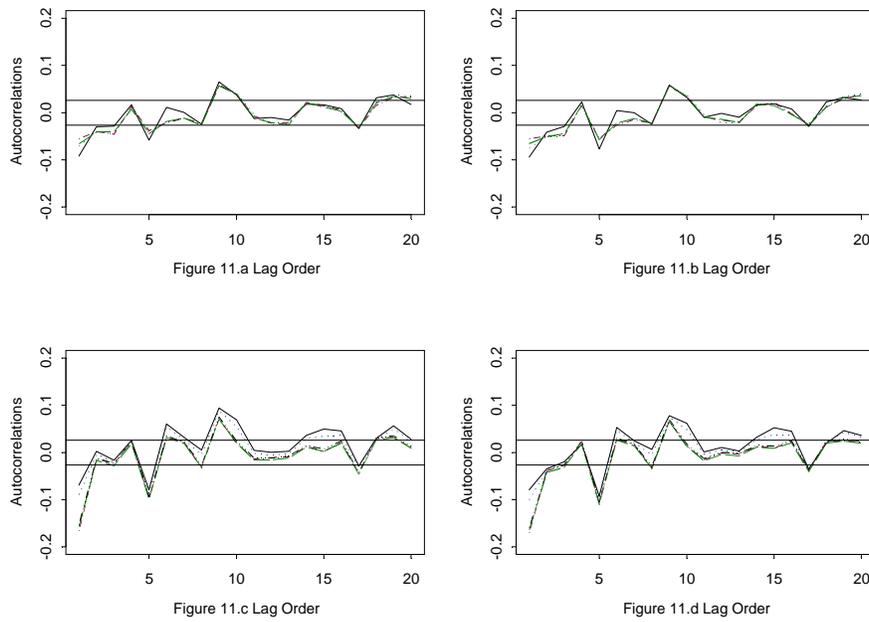


Figure 11: Autocorrelations of the first four moments of the “generalized residuals” $\left\{ \widehat{Z}_\tau^m \right\}_{\tau=1}^n$ for the four discrete time series models. Subplots a, b, c, and c represents $Corr \left[\widehat{Z}_\tau^m, \widehat{Z}_{\tau-j}^m \right]$ for $m = 1, 2, 3,$ and 4 . In each plot, from top to bottom are GARCH, Regime switching, Jump-diffusion/ARCH, Jump-diffusion/GARCH models.