Out-of-Sample Performance of Discrete-Time Spot Interest Rate Models

Yongmiao HONG
Department of Economics and Department of Statistical Sciences, Cornell University, Ithaca, NY 14850
Department of Economics, Tsinghua University, Beijing, China 100084 (*yh20@cornell.edu*)

Haitao LI
Johnson Graduate School of Management, Cornell University, Ithaca, NY 14850 (*hl70@cornell.edu*)

Feng ZHAO
Department of Economics, Cornell University, Ithaca, NY 14850 (*fz17@cornell.edu*)

We provide a comprehensive analysis of the out-of-sample performance of a wide variety of spot rate models in forecasting the probability density of future interest rates. Although the most parsimonious models perform best in forecasting the conditional mean of many financial time series, we find that the spot rate models that incorporate conditional heteroscedasticity and excess kurtosis or heavy tails have better density forecasts. Generalized autoregressive conditional heteroscedasticity significantly improves the modeling of the conditional variance and kurtosis, whereas regime switching and jumps improve the modeling of the marginal density of interest rates. Our analysis shows that the sophisticated spot rate models in the existing literature are important for applications involving density forecasts of interest rates.

KEY WORDS: Density forecast; Generalized autoregressive conditional heteroscedasticity; Generalized spectrum; Jumps; Parameter estimation uncertainty; Regime switching.

1. INTRODUCTION

The short-term interest rate plays an important role in many areas of asset pricing studies. For example, the instantaneous risk-free interest rate, or the so-called "spot rate," is the state variable that determines the evolution of the yield curve in the well-known term structure models of Vasicek (1977) and Cox, Ingersoll, and Ross (CIR) (1985). The spot rate is thus of fundamental importance to the pricing of fixed-income securities and the management of interest rate risk. Many interest rate models have been proposed, and over the last decade, a vast body of literature has been developed to rigorously estimate and test these models using high-quality interest rate data. (See, e.g., Chapman and Pearson 2001 for a survey of the empirical literature.)

Despite the progress that has been made in modeling interest rate dynamics, most existing studies have typically focused on the in-sample fit of historical interest rates and ignored out-of-sample forecasts of future interest rates. In-sample diagnostic analysis is important and can reveal useful information about possible sources of model misspecifications. In many financial applications, such as the pricing and hedging of fixed-income securities and the management of interest rate risk, however, what matters most is the evolution of the interest rate in the future, not in the past.

In general, there is no guarantee that a model that fits historical data well will also perform well out-of-sample due to at least three important reasons. First, the extensive search for more complicated models using the same (or similar) dataset(s) may suffer from the so-called "data-snooping bias," as pointed out by Lo and MacKinlay (1989) and White (2000). A more-complicated model can always fit a given dataset better than simpler models, but it may overfit some idiosyncratic features of the data without capturing the true data-generating process.

Out-of-sample evaluation will alleviate, if not eliminate completely, such data-snooping bias. Second, an overparameterized model contains a large number of estimated parameters and inevitably exhibits excessive sampling variation in parameter estimation, which in turn may adversely affect the out-of-sample forecast performance. Third, a model that fits a historical dataset well may not forecast the future well because of unforeseen structural changes or regime shifts in the data-generating process. Therefore, from both practical and theoretical standpoints, in-sample analysis alone is not adequate, and it is necessary to examine the out-of-sample predictive ability of spot rate models.

We contribute to the literature by providing the first comprehensive empirical analysis (to our knowledge) of the out-of-sample performance of a wide variety of popular spot rate models. Although some existing studies (e.g., Gray 1996; Bali 2000; Duffee 2002) have conducted out-of-sample analysis of interest rate models, what distinguishes our study is that we focus on forecasting the probability density of future interest rates, rather than just the conditional mean or the first few conditional moments. This distinction is important because interest rates, like most other financial time series, are highly non-Gaussian. Consequently, one must go beyond the conditional mean and variance to get a complete picture of interest rate dynamics. The conditional probability density characterizes the full dynamics of an interest rate model and thus essentially checks all conditional moments simultaneously (if the moments exist).

Density forecasts are important not only for statistical evaluation, but also for many economic and financial applications. For

example, the booming industry of financial risk management is essentially dedicated to providing density forecasts for important economic variables and portfolio returns, and to tracking certain aspects of distribution, such as value at risk (VaR), to quantify the risk exposure of a portfolio (e.g., Morgan 1996; Duffie and Pan 1997; Jorion 2000). More generally, modern risk control techniques all involve some form of density forecast, the quality of which has real impact on the efficacy of capital asset/liability allocation. In macroeconomics, monetary authorities in the United States and United Kingdom (the Federal Reserve Bank of Philadelphia and the Bank of England) have been conducting quarterly surveys on density forecasts for inflation and output growth to help set their policy instruments (e.g., inflation target). There is also a growing literature on extracting density forecasts from options prices to obtain useful information on otherwise unobservable market expectations (e.g., Fackler and King 1990; Jackwerth and Rubinstein 1996; Soderlind and Svensson 1997; Ait-Sahalia and Lo 1998).

In the recent development of time series econometrics, there is a growing interest in out-of-sample probability distribution forecasts and their evaluation, motivated from the context of decision-making under uncertainty. Important works in this area include those of Bera and Ghosh (2002), Berkowitz (2001), Christoffersen and Diebold (1996, 1997), Clements and Smith (2000), Diebold, Gunther, and Tay (1998), Diebold, Hahn, and Tay (1999), Diebold, Tay, and Wallis (1999), Granger (1999), and Granger and Pesaran (2000). One of the most important issues in density forecasting is to evaluate the quality of a forecast (Granger 1999). Suboptimal density forecasts will have real adverse impact in practice. For example, an excessive forecast for VaR would force risk managers and financial institutions to hold too much capital, imposing an additional cost. Suboptimal density forecasts for important macroeconomic variables may lead to inappropriate policy decisions (e.g., inappropriate level and timing in setting interest rates), which could have serious consequences for the real economy.

Evaluating density forecasts, however, is nontrivial, simply because the probability density function of an underlying process is not observable even ex post. Unlike for point forecasts, there are few statistical tools for evaluating density forecasts. In a pioneering contribution, Diebold et al. (1998) evaluated density forecast by examining the probability integral transforms of the data with respect to a density forecast model. Such a transformed series is often called the "generalized residuals" of the density forecast model. It should be iid U[0, 1] if the density forecast model correctly captures the full dynamics of the underlying process. Any departure from iid U[0, 1] is evidence of suboptimal density forecasts and model misspecification.

Diebold et al. (1998) used an intuitive graphical method to separately examine the iid and uniform properties of the "generalized residuals." This method is simple and informative about possible sources of suboptimality in density forecasts. Hong (2003) recently developed an omnibus evaluation procedure for density forecasts by measuring the departure of the "generalized residuals" from iid U[0, 1]. The evaluation statistic provides a metric of the distance between the density forecast model and the true data-generating process. The most appealing feature of this new test is its omnibus ability to detect a wide range of suboptimal density forecasts, thanks to the use of the generalized spectrum. The latter was introduced by Hong (1999) as an analytic tool for nonlinear time series and is based on the characteristic function in a time series context. There has been an increasing interest in using the characteristic function in financial econometrics (see, e.g., Chacko and Viceira 2003; Ghysels, Carrasco, Chernov, and Florens 2001; Hong and Lee 2003a,b; Jiang and Knight 1997; Singleton 2001).

Applying Hong's (2003) procedure, we provide a comprehensive empirical analysis of the density forecast performance of a variety of popular spot rate models, including single-factor diffusion, generalized autoregressive conditional heteroscedasticity (GARCH), regime-switching, and jump-diffusion models. The in-sample performance of these models has been extensively studied, but their out-of-sample performance, especially their density forecast performance is still largely unknown. We find that although previous studies have shown that simpler models, such as the random walk model, tend to provide better forecasts for the conditional mean of many financial time series, including interest rates and exchange rates (e.g., Duffee 2002; Meese and Rogoff 1983), more-sophisticated spot rate models that capture volatility clustering, and excess kurtosis and heavy tails of interest rates have better density forecasts. GARCH significantly improves the modeling of the dynamics of the conditional variance and kurtosis of the generalized residuals, whereas regime switching and jumps significantly improve the modeling of the marginal density of interest rates. Our analysis shows that the sophisticated spot rate models in the existing literature can indeed capture some important features of interest rates and are relevant to applications involving density forecasts of interest rates.

The article is organized as follows. In Section 2 we discuss the methodology for density forecast evaluation. In Section 3 we introduce a variety of popular spot rate models. In Section 4 we describe data, estimation methods, and the in-sample performance of each model. In Section 5 we subject each model to out-of-sample density forecast evaluation, and in Section 6 we provide concluding remarks. In the Appendix we discuss Hong's (2003) evaluation procedure for density forecasts.

2. OUT–OF–SAMPLE DENSITY FORECAST EVALUATION

The probability density function is a well-established tool for characterizing uncertainty in economics and finance (e.g., Rothschild and Stiglitz 1970). The importance of density forecasts has been widely recognized in recent literature due to the works of Diebold et al. (1998), Granger (1999), and Granger and Pesaran (2000), among many others. These authors showed that accurate density forecasts are essential for decision making under uncertainty when the forecaster's objective function is asymmetric and the underlying process is non-Gaussian. In a decision-theoretic context, Diebold et al. (1998) and Granger and Pesaran (2000) showed that if a density forecast coincides with the true conditional density of the data generating process, then it will be preferred by all forecast users regardless of their objective functions (e.g., risk attitudes). Thus testing the optimality of a forecast boils down to checking whether the density forecast model can capture the true data-generating process.

This is a challenging job, because we never observe an ex post density. So far there are few statistical evaluation procedures for density forecasts.

Diebold et al. (1998) used the probability integral transform of data with respect to the density forecast model to assess the optimality of density forecasts. Extending a result established by Rosenblatt (1952), they showed that if the model conditional density is specified correctly, then the probability integral transformed series should be iid U[0, 1].

Specifically, for a given model of interest rate r_t, there is a model-implied conditional density

$$\frac{\partial}{\partial r}P(r_t \leq r|I_{t-1}, \theta) = p(r, t|I_{t-1}, \theta),$$

where θ is an unknown finite-dimensional parameter vector and $I_{t-1} \equiv \{r_{t-1}, \ldots, r_1\}$ is the information set available at time $t - 1$. Suppose that we have a random sample $\{r_t\}_{t=1}^N$ of size N, which we divide into two subsets, an estimation sample $\{r_t\}_{t=1}^R$ of size R and a prediction sample $\{r_t\}_{t=R+1}^N$ of size $n = N - R$. We can then define the dynamic probability integral transform of the data $\{r_t\}_{t=R+1}^N$ with respect to the model conditional density,

$$Z_t(\theta) \equiv \int_{-\infty}^{r_t} p(r, t|I_{t-1}, \theta)\,dr, \qquad t = R+1, \ldots, N.$$

Suppose that the model is correctly specified in the sense that there exists some θ_0 such that $p(r, t|I_{t-1}, \theta_0)$ coincides with the true conditional density of r_t. Then the transformed sequence $\{Z_t(\theta_0)\}$ is iid U[0, 1].

For example, consider the discretized Vasicek model

$$\Delta r_t = \alpha_0 + \alpha_1 r_{t-1} + \sigma z_t,$$

where $\{z_t\} \sim$ iid N(0, 1). For this model, the conditional density of r_t given I_{t-1} is

$$p(r, t|I_{t-1}, \theta)$$
$$= \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}\left[r - \left(\alpha_0 + (1+\alpha_1)r_{t-1}\right)\right]^2\right\},$$
$$r \in (-\infty, \infty),$$

where $\theta = (\alpha_0, \alpha_1, \sigma)'$. Suppose that $p(r, t|I_{t-1}, \theta_0) = p_0(r, t|I_{t-1})$ for some θ_0, where $p_0(r, t|I_{t-1})$ is the true conditional density of r_t. Then

$$Z_t(\theta_0) = \int_{-\infty}^{r_t} p(r, t|I_{t-1}, \theta_0)\,dr = \Phi(z_t) \sim \text{iid U}[0, 1],$$

where $\Phi(\cdot)$ is the N(0, 1) cumulative distribution function (cdf).

As noted earlier, the transformed series $\{Z_t(\hat{\theta})\}$ is called the "generalized residual" of model $p(r, t|I_{t-1}, \theta)$, where $\hat{\theta}$ is a consistent estimator for θ_0. This series provides a convenient approach for evaluating the density forecast model $p(r, t|I_{t-1}, \theta)$. Intuitively, the U[0, 1] property characterizes the correct specification for the stationary distribution of $\{r_t\}$, and the iid property characterizes correct dynamic specification for $\{r_t\}$. If $\{Z_t(\theta)\}$ is not iid U[0, 1] for all θ, then the density forecast model $p(r, t|I_{t-1}, \theta)$ is not optimal, and there is room for further improvement.

To test the joint hypothesis of iid U[0, 1] for $\{Z_t(\theta)\}$ is nontrivial. One may suggest using the well-known Kolmogorov–Smirnov test, which unfortunately checks U[0, 1] under the iid assumption rather than checking iid U[0, 1] jointly. It will miss the alternatives for which $\{Z_t(\theta)\}$ is uniform, but not iid. Similarly, tests for iid alone will miss the alternatives for which $\{Z_t(\theta)\}$ is iid but not U[0, 1]. Moreover, parameter estimation uncertainty is expected to affect the asymptotic distribution of the Kolmogorov–Smirnov test statistic.

Diebold et al. (1998) used the autocorrelations of $Z_t^m(\hat{\theta})$ to check the iid property and the histogram of $Z_t(\hat{\theta})$ to check the U[0, 1] property. These graphical procedures are simple and informative. To compare and rank different models, however, it is desirable to use a single evaluation criterion. Hong (2003) recently developed an omnibus evaluation procedure for out-of-sample density forecast that tests the joint hypothesis of iid U[0, 1] and explicitly addresses the impact of parameter estimation uncertainty on the evaluation statistics. This is achieved using a generalized spectral approach. The generalized spectrum was first introduced by Hong (1999) as an analytic tool for nonlinear time series analysis. It is based on the characteristic function transformation embedded in a spectral framework. Unlike the power spectrum, the generalized spectrum can capture any kind of pairwise serial dependence across various lags, including those with zero autocorrelation [e.g., an ARCH process with iid N(0, 1) innovations].

Specifically, Hong (2003) introduced a modified generalized spectral density function that incorporates the information of iid U[0, 1] under the null hypothesis. Consequently, the modified generalized spectrum becomes a known "flat" function under the null hypothesis of iid U[0, 1]. Whenever $Z_t(\theta)$ depends on $Z_{t-j}(\theta)$ for some $j > 0$ or $\{Z_t(\theta)\}$ is not U[0, 1], the modified generalized spectrum will be nonflat as a function of frequency ω. Hong (2003) proposed an evaluation statistic (denoted by M_1) for out-of-sample density forecasts, by comparing a kernel estimator of the modified generalized spectrum with the flat spectrum.

The most attractive feature of M_1 is its omnibus property; it has power against a wide range of suboptimal density forecasts, thanks to the use of the characteristic function in a spectral framework. The M_1 statistic can be viewed as an omnibus metric measuring the departure of the density forecast model from the true data-generating process. A better density forecast model is expected to have a smaller value of M_1, because its generalized residual series $\{Z_t(\theta)\}$ is closer to having the iid U[0, 1] property. Thus M_1 can be used to rank competing density forecast models in terms of their deviations from optimality. (See the Appendix for more discussion.)

When a model is rejected using M_1, it is interesting to explore possible reasons for the rejection. The model generalized residuals $\{Z_t(\theta)\}$ contain much information on model misspecification. As noted earlier, Diebold et al. (1998) have illustrated how to use the histograms of $\{Z_t(\theta)\}$ and autocorrelograms $\{Z_t^m(\theta)\}$ to identify sources of suboptimal density forecasts. Although intuitive and convenient, these graphical methods ignore the impact of parameter estimation uncertainty in $\hat{\theta}$ on the asymptotic distribution of evaluation statistics, which generally exists even when $n \to \infty$. Hong (2003) provided a class of rigorous asymptotically N(0, 1) separate inference statistics $M(m, l)$ that measure whether the cross-correlations between $Z_t^m(\theta)$ and $Z_{t-|j|}^l(\theta)$ are significantly different from 0. Although the moments of the generalized

residuals $\{Z_t\}$ are not exactly the same as those of the original data $\{r_t\}$, they are highly correlated. In particular, the choice of $(m, l) = (1, 1), (2, 2), (3, 3)$, and $(4, 4)$ is very sensitive to autocorrelations in level, volatility, skewness, and kurtosis of $\{r_t\}$. Furthermore, the choices of $(m, l) = (1, 2)$ and $(2, 1)$ are sensitive to the "ARCH-in-mean" effect and the "leverage" effect. Different choices of order (m, l) can thus illustrate various dynamic aspects of the underlying process r_t. (Again, see the Appendix for more discussion.)

3. SPOT RATE MODELS

We apply Hong's (2003) procedure to evaluate the density forecast performance of a variety of popular spot rate models, including single-factor diffusion, GARCH, regime-switching, and jump-diffusion models. We now discuss these models in detail.

3.1 Single-Factor Diffusion Models

One important class of spot rate models is the continuous-time diffusion models, which have been widely used in modern finance because of their convenience in pricing fixed-income securities. Specifically, the spot rate is assumed to follow a single-factor diffusion,

$$dr_t = \mu(r_t, \theta)\, dt + \sigma(r_t, \theta)\, dW_t,$$

where $\mu(r_t, \theta)$ and $\sigma(r_t, \theta)$ are the drift and diffusion functions and W_t is a standard Brownian motion. For diffusion models, $\mu(r_t, \theta)$ and $\sigma(r_t, \theta)$ completely determine the model transition density, which in turn captures the full dynamics of r_t. Table 1 lists a variety of discretized diffusion models examined in our analysis, which are nested by Ait-Sahalia's (1996) nonlinear drift model,

$$\Delta r_t = \alpha_{-1}/r_{t-1} + \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2 + \sigma r_{t-1}^\rho z_t, \qquad (1)$$

where $\{z_t\} \sim$ iid N$(0, 1)$. To be consistent with the other models considered in this article, we study the discretized version of the single-factor diffusion and jump-diffusion models. The discretization bias for daily data that we use in this article is unlikely to be significant (e.g., Stanton 1997; Das 2002). In (1), we allow the drift to have zero, linear, and nonlinear specifications and allow the diffusion to be a constant or to depend on the interest rate level, which we refer to as the "level effect." We term the volatility specification in which the elasticity parameter ρ is estimated from the data the *constant elasticity volatility* (CEV). Density forecasts for the spot rate r_t are given by the model-implied transition density $p(r_t, t|I_{t-1}, \hat{\theta})$, where $\hat{\theta}$ is a parameter estimator using the maximum likelihood estimate (MLE) method. Because we study discretized diffusion models, the MLE method is suitable. For the continuous-time diffusion models, we can use the closed-form likelihood function, if available, or the approximated likelihood function via Hermite expansions (see Ait-Sahalia 1999, 2002).

3.2 GARCH Models

Despite the popularity of single-factor diffusion models, many studies (e.g., Brenner, Harjes, and Kroner 1996; Andersen and Lund 1997) have shown that these models are unreasonably restrictive by requiring that the interest rate volatility depend solely on the interest rate level. To capture the well-known persistent volatility clustering in interest rates, Brenner et al. (1996) introduced various GARCH specifications for volatility and showed that GARCH models significantly outperform single-factor diffusion models for in-sample fit.

To understand the importance of GARCH for density forecast, we consider six GARCH models as listed in Table 1, which are nested by the following specification:

$$\begin{cases} \Delta r_t = \alpha_{-1}/r_{t-1} + \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2 \\ \qquad\quad + \sigma r_{t-1}^\rho \sqrt{h_t} z_t \\ h_t = \beta_0 + h_{t-1}(\beta_2 + \beta_1 r_{t-1}^{2\rho} z_{t-1}^2) \\ \{z_t\} \sim \text{iid N}(0, 1). \end{cases} \qquad (2)$$

We consider three different drift specifications (zero, linear, and nonlinear drift) and two volatility specifications (pure GARCH and combined CEV–GARCH). These various GARCH models allow us to examine the importance of linear versus nonlinear drift in the presence of GARCH and CEV and the incremental contribution of GARCH with respect to CEV. For identification, we set $\sigma = 1$ in all GARCH models.

Another popular approach to capturing volatility clustering is the continuous-time stochastic volatility model considered by Andersen and Lund (1997), Gallant and Tauchen (1998), and many others. These studies show that adding a latent stochastic volatility factor to a diffusion model significantly improves the goodness of fit for interest rates. We leave continuous-time stochastic volatility models for future research, because their estimation is much more involved. On the other hand, we also need to be careful about drawing any implications from GARCH models for continuous-time stochastic volatility models. Although Nelson (1990) showed that GARCH models converge to stochastic volatility models in the limit, Corradi (2000) later showed that Nelson's results hold only under his particular discretization scheme. Other equally reasonable discretizations will lead to very different continuous-time limits for GARCH models.

3.3 Markov Regime-Switching Models

Many studies have shown that the behavior of spot rates changes over time due to changes in monetary policy, the business cycle, and general macroeconomic conditions. The Markov regime-switching models of Hamilton (1989) have been widely used to model the time-varying behavior of interest rates (e.g., Gray 1996; Ball and Torous 1998; Ang and Bekaert 2002; Li and Xu 2000; among others). Following most existing studies, we consider a class of two-regime models for the spot rate, where the latent state variable s_t follows a two-state, first-order Markov chain. We refer to the regime in which $s_t = 1$ (2) as the first (second) regime. Following Ang and Bekaert (2002),

Table 1. Spot Rate Models Considered for Density Forecast Evaluation

Model	Mean	Volatility
Discretized single-factor diffusion models		
Random walk	α_0	σ
Log-normal	$\alpha_1 r_{t-1}$	σr_{t-1}
Dothan	0	σr_{t-1}
Pure CEV	0	σr_{t-1}^{ρ}
Vasicek	$\alpha_0 + \alpha_1 r_{t-1}$	σ
CIR	$\alpha_0 + \alpha_1 r_{t-1}$	$\sigma r_{t-1}^{.5}$
CKLS	$\alpha_0 + \alpha_1 r_{t-1}$	σr_{t-1}^{ρ}
Nonlinear drift	$\alpha_{-1}/r_{t-1} + \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2$	σr_{t-1}^{ρ}
GARCH models		
No drift GARCH	0	$\sigma \sqrt{h_t}$
Linear drift GARCH	$\alpha_0 + \alpha_1 r_{t-1}$	$\sigma \sqrt{h_t}$
Nonlinear drift GARCH	$\alpha_{-1}/r_{t-1} + \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2$	$\sigma \sqrt{h_t}$
No drift CEV–GARCH	0	$\sigma r_{t-1}^{\rho} \sqrt{h_t}$
Linear drift CEV–GARCH	$\alpha_0 + \alpha_1 r_{t-1}$	$\sigma r_{t-1}^{\rho} \sqrt{h_t}$
Nonlinear drift CEV–GARCH	$\alpha_{-1}/r_{t-1} + \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2$	$\sigma r_{t-1}^{\rho} \sqrt{h_t}$
Markov regime-switching models		
No drift RS CEV	0	$\sigma(s_t) r_{t-1}^{\rho(s_t)}$
Linear drift RS CEV	$\alpha_0(s_t) + \alpha_1(s_t) r_{t-1}$	$\sigma(s_t) r_{t-1}^{\rho(s_t)}$
Nonlinear drift RS CEV	$\alpha_{-1}(s_t)/r_{t-1} + \alpha_0(s_t) + \alpha_1(s_t) r_{t-1} + \alpha_2(s_t) r_{t-1}^2$	$\sigma(s_t) r_{t-1}^{\rho(s_t)}$
No drift RS GARCH	0	$\sigma(s_t) \sqrt{h_t}$
Linear drift RS GARCH	$\alpha_0(s_t) + \alpha_1(s_t) r_{t-1}$	$\sigma(s_t) \sqrt{h_t}$
Nonlinear drift RS GARCH	$\alpha_{-1}(s_t)/r_{t-1} + \alpha_0(s_t) + \alpha_1(s_t) r_{t-1} + \alpha_2(s_t) r_{t-1}^2$	$\sigma(s_t) \sqrt{h_t}$
No drift RS CEV–GARCH	0	$\sigma(s_t) r_{t-1}^{\rho(s_t)}$
Linear drift RS CEV–GARCH	$\alpha_0(s_t) + \alpha_1(s_t) r_{t-1}$	$\sigma(s_t) r_{t-1}^{\rho(s_t)}$
Nonlinear drift RS CEV–GARCH	$\alpha_{-1}(s_t)/r_{t-1} + \alpha_0(s_t) + \alpha_1(s_t) r_{t-1} + \alpha_2(s_t) r_{t-1}^2$	$\sigma(s_t) r_{t-1}^{\rho(s_t)}$
Discretized jump-diffusion models		
No drift JD CEV	0	σr_{t-1}^{ρ}
Linear drift JD CEV	$\alpha_0 + \alpha_1 r_{t-1}$	σr_{t-1}^{ρ}
Nonlinear drift JD CEV	$\alpha_{-1}/r_{t-1} + \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2$	σr_{t-1}^{ρ}
No drift JD GARCH	0	$\sigma \sqrt{h_t}$
Linear drift JD GARCH	$\alpha_0 + \alpha_1 r_{t-1}$	$\sigma \sqrt{h_t}$
Nonlinear drift JD GARCH	$\alpha_{-1}/r_{t-1} + \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2$	$\sigma \sqrt{h_t}$
No drift JD CEV–GARCH	0	$\sigma r_{t-1}^{\rho} \sqrt{h_t}$
Linear drift JD CEV–GARCH	$\alpha_0 + \alpha_1 r_{t-1}$	$\sigma r_{t-1}^{\rho} \sqrt{h_t}$
Nonlinear drift JD CEV–GARCH	$\alpha_{-1}/r_{t-1} + \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2$	$\sigma r_{t-1}^{\rho} \sqrt{h_t}$

NOTE: For consistency, all models are estimated in a discrete time setting using the maximum likelihood method (MLE). The eight discretized single-factor diffusion models are nested by the following specification: $\Delta r_{t-1} = \alpha_{-1}/r_{t-1} + \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2 + \sigma r_{t-1}^{\rho} z_t$, where $\{z_t\} \sim$ iid $N(0,1)$. The six GARCH models are nested by the following specification: $\Delta r_t = \alpha_{-1}/r_{t-1} + \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2 + \sigma r_{t-1}^{\rho} \sqrt{h_t} z_t$, where $h_t = \beta_0 + h_{t-1}(\beta_2 + \beta_1 r_{t-1}^{2\rho} z_{t-1}^2)$ and $\{z_t\} \sim$ iid $N(0,1)$. The nine regime-switching models are nested by the following specification: $\Delta r_t = \alpha_{-1}(s_t)/r_{t-1} + \alpha_0(s_t) + \alpha_1(s_t) r_{t-1} + \alpha_2(s_t) r_{t-1}^2 + \sigma(s_t) r_{t-1}^{\rho(s_t)} \sqrt{h_t} z_t$, where $h_t = \beta_0 + \beta_1 E\{e_t | r_{t-2}, s_{t-2}\}^2 + \beta_2 h_{t-1}, e_t = [\Delta r_{t-1} - E(\Delta r_{t-1} | r_{t-2}, s_{t-1})]/\sigma(s_{t-1})$, $\{z_t\} \sim$ iid $N(0,1)$, and s_t follows a two-state Markov chain with the transition probability $\Pr(s_t = l | s_{t-1} = l) = (1 + \exp(-c_l - d_l r_{t-1}))^{-1}$ for $l = 1, 2$. The nine discretized jump-diffusion models are nested by the following specification: $\Delta r_t = \alpha_{-1}/r_{t-1} + \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2 + \sigma r_{t-1}^{\rho} \sqrt{h_t} z_t + J(\mu, \gamma^2) \Delta \pi_t(q_t)$, where $h_t = \beta_0 + \beta_1 [r_{t-1} - E(r_{t-1} | r_{t-2})]^2 + \beta_2 h_{t-1}$, $\{z_t\} \sim$ iid $N(0,1)$, $J \sim N(\mu, \gamma^2)$, and $\Delta \pi_t(q_t)$ is Bernoulli(q_t) with $q_t = (1 + \exp(-c - d r_{t-1}))^{-1}$.

we assume that the transition probability of $\{s_t\}$ depends on the once-lagged spot rate level,

$$\Pr(s_t = l | s_{t-1} = l) = \frac{1}{1 + \exp(-c_l - d_l r_{t-1})}.$$

Table 1 lists a variety of regime-switching models, all of which are nested by the following specification:

$$\begin{cases} \Delta r_t = \alpha_{-1}(s_t)/r_{t-1} + \alpha_0(s_t) + \alpha_1(s_t) r_{t-1} \\ \qquad + \alpha_2(s_t) r_{t-1}^2 + \sigma(s_t) r_{t-1}^{\rho(s_t)} \sqrt{h_t} z_t \\ h_t = \beta_0 + \beta_1 E\{e_t | r_{t-2}, s_{t-2}\}^2 + \beta_2 h_{t-1} \\ e_t = [\Delta r_{t-1} - E(\Delta r_{t-1} | r_{t-2}, s_{t-1})]/\sigma(s_{t-1}) \\ \{z_t\} \sim \text{iid } N(0, 1). \end{cases} \quad (3)$$

As in the aforementioned models, we consider three specifications of the conditional mean: zero, linear, and nonlinear drift. We also consider three specifications of the conditional variance: CEV, GARCH, and combined CEV–GARCH. Thus we have a total of nine regime-switching models.

Although Gray (1996) removed the path-dependent nature of GARCH models by averaging the conditional and unconditional variances over regimes at every time point, we use the same GARCH specification across different regimes. Confirming results of Ang and Bekaert (2002), we find a very unstable estimation of Gray's (1996) model. In contrast, our specification turns out to have much better convergence properties. Unlike many previous studies that set the elasticity parameter equal to .5, we allow it to be regime-dependent and

estimate it from the data. For identification, we set the diffusion constant $\sigma(s_t) = 1$ for $s_t = 1$ in the regime-switching models with GARCH effect.

The conditional density of the interest rate r_t in a regime-switching model is

$$p(\Delta r_t | I_{t-1}) = \sum_{l=1}^{2} p(\Delta r_t | s_t = l, I_{t-1}) \Pr(s_t = l | I_{t-1}),$$

where the ex ante probability that the data are generated from regime l at t, $\Pr(s_t = l | I_{t-1})$, can be obtained using Bayes's rule via a recursive procedure described by Hamilton (1989). The conditional density of regime-switching models is a mixture of two normal distributions, which can generate unimodal or bimodal distributions and allows for great flexibility in modeling skewness, kurtosis, and heavy tails.

3.4 Jump-Diffusion Models

There are compelling economic and statistical reasons to account for discontinuity in interest rate dynamics. Various economic shocks, news announcements, and government interventions in bond markets have pronounced effects on the behavior of interest rates and tend to generate large jumps in interest rate data. Statistically, Das (2002) and Johannes (2003), among others, have shown that diffusion models (even with stochastic volatility) cannot generate the excessive leptokurtosis exhibited by the changes of the spot rates and that jump-diffusion models are a convenient way to generate excessive kurtosis or, more generally, heavy tails.

We consider a class of discretized jump-diffusion models listed Table 1. As in the aforementioned models, we consider zero, linear, and nonlinear drift specifications and CEV, GARCH and combined CEV–GARCH specifications for volatility. The nine different jump-diffusion models are nested by the following specification:

$$\begin{cases} \Delta r_t = \alpha_{-1}/r_{t-1} + \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2 \\ \quad + \sigma r_{t-1}^{\rho} \sqrt{h_t} z_t + J(\mu, \gamma^2) \Delta \pi_t(q_t) \\ h_t = \beta_0 + \beta_1 [r_{t-1} - E(r_{t-1}|r_{t-2})]^2 + \beta_2 h_{t-1} \\ \{z_t\} \sim \text{iid N}(0, 1) \\ \{\Delta \pi_t(q_t)\} \sim \text{independent Bernoulli}(q_t) \\ J \sim \text{N}(\mu, \gamma^2), \end{cases} \quad (4)$$

where J is the jump size and q_t is the jump probability with $q_t = \frac{1}{1+\exp(-c-dr_{t-1})}$.

The conditional density of the foregoing jump-diffusion models can be written as

$$f(\Delta r_t | r_{t-1})$$

$$= (1 - q_t) \frac{1}{\sqrt{2\pi\sigma^2(r_{t-1})}} \exp\left\{ \frac{-[\Delta r_t - \mu(r_{t-1})]^2}{2\sigma^2(r_{t-1})} \right\}$$

$$+ q_t \frac{1}{\sqrt{2\pi[\sigma^2(r_{t-1}) + \sigma_J^2]}}$$

$$\times \exp\left\{ \frac{-[\Delta r_t - \mu(r_{t-1}) - \mu]^2}{2[\sigma^2(r_{t-1}) + \gamma^2]} \right\},$$

where $\mu(r_{t-1})$ and $\sigma^2(r_{t-1})$ are the conditional mean and variance of the diffusion part in (4). Similar to the regime-switching models, the conditional density of the jump-diffusion models is also a mixture of two normal distributions. Thus these two classes of models represent two alternative approaches to generating excess kurtosis and heavy tails, although the regime-switching models have more sophisticated specifications. For example, in (3) all drift parameters are regime-dependent, whereas in (4) only the intercept term is different in the conditional mean and variance. For the regime-switching models in (3), the state probabilities evolve according to a transition matrix with updated priors at every time point. Thus it can capture both very persistent and transient regime shifts. For the jump-diffusion models in (4), the state probabilities are assumed to depend on the past interest rate level.

The in-sample performances of the four classes of models that we study have been individually studied in the literature. However, no one has yet attempted a systematic evaluation of all of these models in a unified setup, particularly in the out-of-sample context, because of the difficulty in density forecast evaluation and because of different nonnested specifications between existing classes of models. In the next section we compare the relative performance of these four classes of models in out-of-sample density forecast using the evaluation method described in Section 2. This comparison can provide valuable information about the relative strengths and weaknesses of each model and should be important for many applications.

4. MODEL ESTIMATION AND IN–SAMPLE PERFORMANCE

4.1 Data and Estimation Method

In modeling spot interest rate dynamics, yields on short-term debts are often used as proxies for the unobservable instantaneous risk-free rate. These include the 1-month T-bill rates used by Gray (1996) and Chan, Karolyi, Longstaff, and Sanders (CKLS) (1992), the 3-month T-bill rates used by Stanton (1997) and Andersen and Lund (1997), the 7-day Eurodollar rates used by Ait-Sahalia (1996) and Hong and Li (2004), and the Fed funds rates used by Conley, Hansen, Luttmer, and Scheinkman (1997) and Das (2002). In our study, we follow CKLS (1992) and use the daily 1-month T-bill rates from June 14, 1961 to December 29, 2000, with a total of 9,868 observations. The data are extracted from the CRSP bond file using the midpoint of quoted bid and ask prices of the T-bills with a remaining maturity that is closest to 1 month (30 calendar days) from the current date. Most of these T-bills have maturities of 6 months or 1 year when first issued and all T-bills used in our study have a remaining maturity between 27 and 33 days. We then compute the annualized continuously compounded yield based on the average of the ask and bid prices.

Figure 1 plots the level and change series of the daily 1-month T-bill rates, as well as their histograms. There is obvious persistent volatility clustering, and in general, the volatility is higher at a higher level of the interest rate (e.g., the 1979–1982 period). The marginal distribution of the interest rate level is skewed to the right, with a long right tail. Most daily changes in the interest rate level are very small, giving a
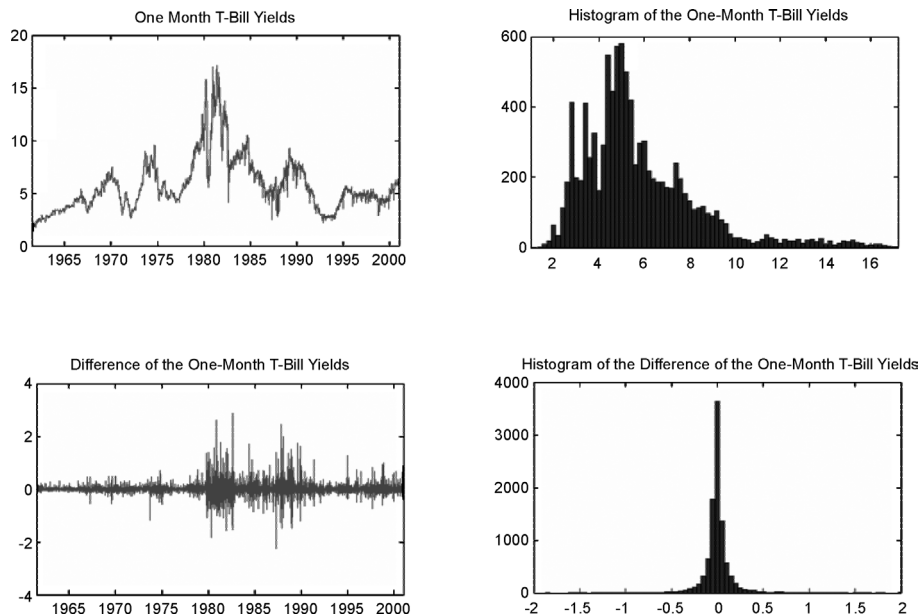
Figure 1. Daily 1-Month T-Bill Rates Between June 14, 1961 and December 29, 2000. This figure plots the level and change series of the daily 1-month t-bill rates, as well as their histograms. The data are extracted from the CRSP bond file using the midpoint of quoted bid and ask prices of the T-bills with a remaining maturity that is closest to 1 month (30 calendar days) from the current date.

sharp probability peak around 0. Daily changes of other spot interest rates, such as 3-month T-bill and 7-day Eurodollar rates, also exhibit a high peak around 0. But the peak is much less pronounced for weekly changes in spot rates. This, together with the long (right) tail, implies excess kurtosis, a stylized fact that has motivated the use of jump models in the literature (e.g., Das 2002; Johannes 2003).

We divide our data into two subsamples. The first, from June 14, 1961 to February 22, 1991 (with a total of 7,400 observations), is a sample used to estimate model parameters; the second, from February 23, 1991 to December 29, 2000 (with a total of 2,467 observations), is a prediction sample used to evaluate out-of-sample density forecasts. We first consider the in-sample performance of the various models listed in Table 1, all of which are estimated using the MLE method. The optimization algorithm is the well-known BHHH with STEPBT for step length calculation and is implemented via the constrained optimization code in GAUSS Windows Version 5.0. The optimization tolerance level is set such that the gradients of the parameters are less than or equal to 10^{-6}.

Tables 2–5 report parameter estimates of four classes of spot rate models using daily 1-month T-bill rates from June 14, 1961 to February 22, 1991, with a total of 7,400 observations. To account for the special period between October 1979 and September 1982, we introduce dummy variables to the drift, volatility, and elasticity parameters; that is, α_D, σ_D, and ρ_D equal 0 outside of the 1979–1982 period. We also introduce a dummy variable, D_{87}, in the interest rate level to account for the effect of 1987 stock market crash. D_{87} equals 0 except for the week right after the stock market crash. Estimated robust standard errors are reported in parentheses.

4.2   In-Sample Evidence

We first discuss the in-sample performance of models within each class and then compare their performance across different classes. Many previous studies (e.g., Bliss and Smith

1998) have shown that the period between October 1979 and September 1982, which coincides with the Federal Reserve experiment, has a big impact on model parameter estimates. To account for the special feature of this period, we introduce dummy variables in the drift, volatility, and elasticity parameters for this period. We also introduce a dummy variable in the interest rate level for the week right after 1987 stock market crash. Brenner et al. (1996) showed that the decline of the interest rates during this week is too dramatic to be captured by most standard models.

Table 2 reports parameter estimates with estimated robust standard errors and log-likelihood values for discretized single-factor diffusion models. The estimates of the drift parameters of Vasicek, CIR, and CKLS models all suggest mean-reversion in the conditional mean, with an estimated 6% long-run mean. For other models, such as the random walk, log-normal, and nonlinear drift models, most drift parameters are not significant. A comparison of the pure CEV, CKLS, and Ait-Sahalia (1996) nonlinear drift models indicates that the incremental contribution of nonlinear drift is very marginal, which is not surprising given that the drift parameter estimates in the nonlinear drift model are mostly insignificant. On the other hand, there is also a clear evidence of level effect; all estimates of the elasticity parameter are significant. Unlike previous studies (e.g., CKLS 1992), which found an estimated elasticity parameter close to 1.5, our estimate is about .25 (.70) outside (inside) the 1979–1982 period. Our results confirm previous findings of Brenner et al. (1996), Andersen and Lund (1997), Bliss and Smith (1998), and Koedijk, Nissen, Schotman, and Wolff (1997) that the estimate of the elasticity parameter is very sensitive to the choice of interest rate data, data frequency, sample periods, and specifications of the volatility function. The 1979–1982 period dummies suggest that the drift does not behave very differently (i.e., α_D is not significant), whereas volatility is significantly higher and depends more heavily on

Table 2. Parameter Estimates for the Single-Factor Diffusion Models

Parameters	RW	Log-normal	Dothan	Pure CEV	Vasicek	CIR	CKLS	Nonlinear drift
α_{-1}								.0289
								(.0751)
α_0	.0012				.0196	.0190	.0199	−.0098
	(.0019)				(.0056)	(.0049)	(.0053)	(.0485)
α_1		.0006			−.0033	−.0032	−.0034	.0045
		(.0004)			(.0010)	(.0009)	(.0009)	(.0095)
α_2								−.0006
								(.0006)
σ	.1523	.0304	.0304	.1011	.1522	.0658	.1007	.1007
	(.0013)	(.0003)	(.0003)	(.0041)	(.0013)	(.0006)	(.0041)	(.0041)
ρ				.2437			.2460	.2460
				(.0238)			(.0238)	(.0238)
α_D	−.0047	.0196			.0154	.0233	.0266	.0402
	(.0157)	(.0186)			(.0168)	(.0176)	(.0168)	(.0202)
σ_D	.2768	.0175	.0175		.2764	.0712		
	(.0112)	(.0013)	(.0013)		(.0112)	(.0036)		
ρ_D				.3692			.3684	.3680
				(.0132)			(.0132)	(.0132)
D_{87}	−.3988	−.2107	−.2079	−.3549	−.4003	−.3101	−.3574	−.3582
	(.0681)	(.0660)	(.0660)	(.0671)	(.0681)	(.0655)	(.0670)	(.0671)
Log-likelihood	2,649.6	2,187.0	2,184.9	2,654.5	2,655.5	2,672.0	2,661.7	2,662.5

NOTE: The models are nested by the following specification: $\Delta r_t = \alpha_{-1}/r_{t-1} + (\alpha_0 + \alpha_D) + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2 + (\sigma + \sigma_D) r_{t-1}^{(\rho + \rho_D)} z_t + D_{87}$, where $\{z_t\} \sim$ iid N(0, 1).

the interest rate level in this period (i.e., σ_D and ρ_D are significantly positive) than in other periods. For models with level effect, we introduce the dummy only for elasticity parameter ρ, to avoid the offsetting effect of dummies in ρ and σ, which would tend to generate unstable estimates. The stock market crash dummy is significantly negative, which is consistent with the findings of Brenner et al. (1996).

Estimation results of GARCH models in Table 3 show that GARCH significantly improves the in-sample fit of diffusion models (i.e., the log-likelihood value increases from

about 2,500 to about 5,700). Previous studies, such as those by Bali (2001) and Durham (2003), have also shown that for single-factor diffusion, GARCH, and continuous-time stochastic volatility models, it is more important to correctly model the diffusion function than the drift function in fitting interest rate data. This suggests that volatility clustering is an important feature of interest rates. All estimates of GARCH parameters are overwhelmingly significant. The sum of GARCH parameter estimates, $\hat{\beta}_1 + \hat{\beta}_2$, is slightly larger (smaller) than 1 without (with) level effect. Although the sum of GARCH parameters is

Table 3. Parameter Estimates for the GARCH Models

Parameters	No drift GARCH	Linear drift GARCH	Nonlinear drift GARCH	No drift CEV–GARCH	Linear drift CEV–GARCH	Nonlinear drift CEV–GARCH
α_{-1}			.1556			.1716
			(.0470)			(.0457)
α_0		.0053	−.1117		.0063	−.1233
		(.0025)	(.0323)		(.0025)	(.0311)
α_1		−.0003	.0263		−.0006	.0290
		(.0006)	(.0069)		(.0006)	(.0066)
α_2			−.0018			−.0020
			(.0004)			(.0004)
ρ				.1709	.1797	.1926
				(.0298)	(.0301)	(.0300)
β_0	9.7E−05	8.6E−05	8.6E−05	8.9E−05	8.2E−05	8.2E−05
	(1.2E−05)	(1.2E−05)	(1.6E−05)	(1.0E−05)	(.9E−05)	(.9E−05)
β_1	.1552	.1544	.1563	.0896	.0875	.0851
	(.0097)	(.0096)	(.0099)	(.0104)	(.0101)	(.0098)
β_2	.8707	.8723	.8712	.8583	.8582	.8553
	(.0065)	(.0064)	(.0065)	(.0074)	(.0076)	(.0078)
α_D		.0166	.1119		.0550	.1317
		(.0318)	(.0229)		(.0282)	(.0201)
σ_D	.1368	.1302	.1017			
	(.0327)	(.0331)	(.0334)			
ρ_D				.0151	.0064	−.0053
				(.0134)	(.0146)	(.0142)
D_{87}	−.5397	−.5426	−.5460	−.5288	−.5318	−.5352
	(.0774)	(.0777)	(.0771)	(.0753)	(.0751)	(.0741)
Log-likelihood	5,686.8	5,699.5	5,706.3	5,701.1	5,715.2	5,725.5

NOTE: The models are nested by the following specification: $\Delta r_t = \alpha_{-1}/r_{t-1} + (\alpha_0 + \alpha_D) + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2 + (1 + \sigma_D) r_{t-1}^{(\rho + \rho_D)} \sqrt{h_t} z_t + D_{87}$, where $h_t = \beta_0 + h_{t-1}(\beta_2 + \beta_1 r_{t-1}^{2\rho} z_{t-1}^2)$ and $\{z_t\} \sim$ iid N(0, 1).

Table 4. Parameter Estimates for the Markov Regime-Switching Models

Parameters	No drift CEV	Linear drift CEV	Nonlinear drift CEV	No drift GARCH	Linear drift GARCH	Nonlinear drift GARCH	No drift CEV–GARCH	Linear drift CEV–GARCH	Nonlinear drift CEV–GARCH
$\alpha_{-1}(1)$			.0607 (.3569)			.3321 (.1701)			.2646 (.1605)
$\alpha_0(1)$		.0973 (.0259)	.1517 (.2027)		.0457 (.0107)	−.1462 (.1009)		.0450 (.0105)	−.1086 (.0989)
$\alpha_1(1)$		−.0102 (.0031)	−.0288 (.0321)		−.0043 (.0019)	.0274 (.0181)		−.0042 (.0019)	.0209 (.0185)
$\alpha_2(1)$			.0011 (.0014)			−.0015 (.0010)			−.0011 (.0011)
$\sigma(1)$	.3361 (.0302)	.3171 (.0287)	.3106 (.0281)	1	1	1	1	1	1
$\rho(1)$	.1021 (.044)	.1277 (.0446)	.138 (.0448)	0	0	0	.1193 (.0452)	.1589 (.0460)	.1670 (.0487)
$\alpha_{-1}(2)$			−.0352 (.0329)			−.0267 (.0242)			−.0286 (.0183)
$\alpha_0(2)$		−.0002 (.0020)	.0141 (.0229)		−.0013 (.0017)	.0062 (.0153)		−.0020 (.0018)	.0062 (.0104)
$\alpha_1(2)$		−.0006 (.0004)	−.0012 (.0048)		−.0008 (.0004)	−.0001 (.0031)		−.0007 (.0004)	.0005 (.0020)
$\alpha_2(2)$			−.0001 (.0003)			−.0002 (.0002)			−.0002 (.0001)
$\sigma(2)$	.0143 (.0010)	.0141 (.0010)	.0138 (.0010)	.2566 (.006)	.2567 (.006)	.2574 (.006)	.2769 (.0275)	.3034 (.0303)	.3051 (.0311)
$\rho(2)$	.8122 (.0403)	.8179 (.0422)	.8257 (.0421)	0	0	0	.0844 (.0602)	.0727 (.0602)	.0800 (.0611)
β_0				3.83E−4 (7.11E−5)	3.18E−4 (6.14E−5)	2.96E−4 (5.79E−5)	3.21E−4 (.67E−4)	2.43E−4 (.50E−4)	2.25E−4 (.47E−4)
β_1				.0357 (.004)	.0336 (.0039)	.0335 (.0038)	.0265 (.0057)	.0259 (.0055)	.0252 (.0053)
β_2				.9027 (.009)	.9066 (.0089)	.9077 (.0086)	.8972 (.0099)	.8993 (.0098)	.9001 (.0094)
c_1	.3132 (.2734)	.3001 (.273)	.3171 (.2683)	−1.2515 (.3211)	−1.2007 (.3065)	−1.1293 (.3136)	−1.1954 (.3819)	−1.1429 (.3705)	−1.0012 (.3674)
d_1	.1081 (.0391)	.107 (.039)	.1029 (.0381)	.1303 (.044)	.1295 (.0426)	.1192 (.0447)	.1190 (.0594)	.1193 (.0582)	.0959 (.0593)
c_2	4.0963 (.2386)	4.0342 (.2393)	4.0089 (.2313)	1.8864 (.2048)	1.7868 (.1947)	1.7633 (.1909)	1.8227 (.2092)	1.7299 (.1983)	1.6947 (.2009)
d_2	−.2302 (.0354)	−.225 (.0354)	−.2225 (.0339)	−.0938 (.0276)	−.0806 (.0263)	−.0776 (.0257)	−.0906 (.0285)	−.0810 (.0276)	−.0757 (.0283)
Log-likelihood	6,493.1	6,507.9	6,512.8	7,079.0	7,132.8	7,138.2	7,082.8	7,138.6	7,144.1

NOTE: The models are nested by the following specification: $\Delta r_t = \alpha_{-1}(s_t)/r_{t-1} + \alpha_0(s_t) + \alpha_1(s_t)r_{t-1} + \alpha_2(s_t)r_{t-1}^2 + \sigma(s_t)r_{t-1}^{\rho(s_t)}\sqrt{h_t}z_t$, where $h_t = \beta_0 + \beta_1 E\{e_t|r_{t-2}, s_{t-2}\}^2 + \beta_2 h_{t-1}$, $e_t = [\Delta r_{t-1} - E(\Delta r_{t-1}|r_{t-2}, s_{t-1})]/\sigma(s_{t-1})$, $\{z_t\} \sim$ iid N(0, 1), and s_t follows a two-state Markov chain with transition probability $\Pr(s_t = l|s_{t-1} = l) = (1 + \exp(-c_l - d_l r_{t-1}))^{-1}$ for $l = 1, 2$.

slightly larger than 1, it is still possible that the spot rate model is strictly stationary (see Nelson 1991 for more discussion). We also find a significant level effect even in the presence of GARCH, but the estimated elasticity parameter is close to .20, which is much smaller than that of single-factor diffusion models. The specification of conditional variance also affects the estimation of drift parameters. Unlike those in diffusion models, most drift parameter estimates in linear drift GARCH models are insignificant. Among all GARCH models considered, the one with nonlinear drift and level effect has the best in-sample performance. In the 1979–1982 period, the interest rate volatility is significantly higher, but its dependence on the interest rate level is not significantly different from other periods. The 1987 stock market crash dummy is again significantly negative.

Parameter estimates of regime-switching models, given in Table 4, show that the spot rate behaves quite differently between two regimes. Although the sum of GARCH parameters is slightly larger than 1, it is still possible that the spot rate model is strictly stationary (see Nelson 1991 for more discussion). For models with a linear drift, in the first regime the spot rate has a high long-run mean (about 10%) and exhibits strong mean reversion (i.e., estimates of the mean-reversion speed parameter are significantly negative). The spot rate in the second regime behaves almost like a random walk, because most drift parameter estimates are close to 0 and insignificant. For models with a nonlinear drift, all drift parameters are insignificant. Volatility in the first regime is much higher, about three times that in the second regime. Our estimates show that level effect, although significant in both regimes, is much stronger in the second regime in the absence of GARCH. After including GARCH, the elasticity parameter estimate becomes insignificant in the second regime, but remains the same in the first regime. Apparently, regime switching helps capture volatility clustering; the sum of GARCH parameters, $\hat{\beta}_1 + \hat{\beta}_2$, is smaller in regime-switching models than in pure GARCH models. The estimated transition probabilities of the Markov state variable s_t show that the low-volatility regime is much more persistent than the high-volatility regime. Our results suggest that the spot rate evolves like a random walk with low volatility most of time, occasionally increasing and evolving with strong mean reversion and high volatility. Overall, the regime-switching model with a linear drift in each regime, CEV, and GARCH has the best in-sample performance.

Table 5. Parameter Estimates for the Jump-Diffusion Models

Parameters	No drift CEV	Linear drift CEV	Nonlinear drift CEV	No drift GARCH	Linear drift GARCH	Nonlinear drift GARCH	No drift CEV–GARCH	Linear drift CEV–GARCH	Nonlinear drift CEV–GARCH
α_{-1}			.0264 (.0293)			.0702 (.0384)			.0661 (.0380)
α_0		−.0011 (.0169)	−.0340 (.0203)		.0028 (.0019)	−.0506 (.0256)		.0026 (.0020)	−.0486 (.0253)
α_1		−.0007 (.0004)	.0101 (.0043)		−.0013 (.0004)	.0110 (.0053)		−.0012 (.0004)	.0107 (.0052)
α_2			−.0009 (.0003)			−.0008 (.0003)			−.0008 (.0003)
σ	.0127 (.0008)	.0126 (.0008)	.0124 (.0008)						
ρ	.8936 (.0390)	.8975 (.0397)	.9045 (.0400)				.0929 (.0518)	.0852 (.0510)	.0837 (.0512)
β_0				8.7E−05 (1.2E−05)	8.8E−05 (1.2E−05)	8.7E−05 (1.2E−05)	7.6E−5 (1.2E−5)	7.7E−5 (1.2E−5)	7.7E−5 (1.2E−5)
β_1				.0631 (.0076)	.0640 (.0073)	.0635 (.0073)	.0450 (.0101)	.0472 (.0102)	.0472 (.0102)
β_2				.8514 (.0139)	.8484 (.0134)	.8499 (.0134)	.8504 (.0140)	.8471 (.0135)	.8483 (.0134)
c	−2.7404 (.1608)	−2.7057 (.1605)	−2.6972 (.1614)	−3.7908 (.1982)	−3.7470 (.1970)	−3.7481 (.1976)	−3.6908 (.2097)	−3.6498 (.2084)	−3.6518 (.2093)
d	.1451 (.0275)	.1423 (.0275)	.1397 (.0275)	.2554 (.0307)	.2516 (.0305)	.2509 (.0307)	.2293 (.0343)	.2274 (.0340)	.2272 (.0341)
μ	.0233 (.0133)	.0385 (.0142)	.0321 (.0128)	.0700 (.0152)	.0887 (.0158)	.0895 (.0158)	.0679 (.0162)	.0892 (.0164)	.0897 (.0163)
γ	.4316 (.0117)	.4279 (.0120)	.4304 (.0104)	.3928 (.0197)	.3903 (.0189)	.3929 (.0187)	.4109 (.0199)	.4046 (.0182)	.4061 (.0178)
α_D		−.0094 (.0121)	.0296 (.0112)		−.0192 (.0117)	.0057 (.0154)		−.0197 (.0112)	.0055 (.0150)
σ_D				−.1027 (.1142)	−.1328 (.1054)	−.1551 (.1033)			
ρ_D	.1473 (.1079)	.1872 (.0703)	−.0697 (.0722)				−.1220 (.0769)	−.1186 (.0573)	−.1277 (.0564)
c_D	4.1604 (.6523)	4.3364 (.6535)	4.0036 (.5477)	4.0290 (.7495)	4.2505 (.7546)	4.1857 (.7421)	3.8622 (.6911)	4.0677 (.7129)	4.0272 (.7019)
d_D	−.2441 (.0808)	−.2721 (.0730)	−.1836 (.0534)	−.2748 (.0689)	−.2860 (.0682)	−.2791 (.0667)	−.2320 (.0685)	−.2487 (.0676)	−.2445 (.0662)
D_{87}	−.4283 (.0646)	−.4253 (.0641)	−.4266 (.0643)	−.4226 (.0826)	−.4198 (.0820)	−.4217 (.0815)	−.4233 (.0793)	−.4209 (.0793)	−.4225 (.0791)
Log-likelihood	6,306.3	6,317.2	6,326.3	6,794.1	6,810.8	6,813.9	6,796.3	6,812.9	6,816.1

NOTE: The models are nested by the following specification: $\Delta r_t = \alpha_{-1}/r_{t-1} + (\alpha_0 + \alpha_D) + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2 + (\sigma + \sigma_D) r_{t-1}^{(\rho+\rho_D)} \sqrt{h_t} z_t + J(\mu, \gamma^2) \Delta \pi_t(q_t + q_D) + D_{87}$, where $h_t = \beta_0 + \beta_1 [r_{t-1} - E(r_{t-1}|r_{t-2})]^2 + \beta_2 h_{t-1}$, $\{z_t\} \sim$ iid N(0, 1), $J \sim$ N(μ, γ^2), and $\Delta \pi_t(q_t + q_D)$ is Bernoulli$(q_t + q_D)$ with $q_t + q_D = (1 + \exp(-c - c_D - (d + d_D) \cdot r_{t-1}))^{-1}$.

Table 5 reports parameter estimates for discretized jump-diffusion models. There is some weak evidence of mean reversion, especially when there is GARCH. The contribution of nonlinear drift is again very marginal: most estimated drift parameters are insignificant. For jump-diffusion models, without GARCH, the elasticity parameter estimate is about .9, which is closer to the 1.5 found by CKLS. However, level effect is weakened (i.e., $\hat{\rho}$ becomes close to .1) after GARCH is introduced. It seems that CEV helps capture part of volatility clustering, but its importance diminishes in the presence of GARCH. We find that GARCH also significantly improves the performance of jump-diffusion models. Without GARCH, there is a high probability of small jumps, with an estimated mean jump size between 2% and 4%. With GARCH, the jump probability becomes smaller, but the mean jump size increases to about between 7% and 9%. This suggests that without GARCH, jumps can capture part of volatility clustering, but with GARCH, jumps mainly capture large interest rate movements. Of course, jumps also help explain volatility clustering; the sum of GARCH parameters is again much smaller than that in pure GARCH models. During the 1979–1982 period, other than the probability of jumps becoming substan-

tially higher, other aspects of the jump-diffusion models (e.g., the drift, volatility, or elasticity parameter) do not behave very differently.

To sum up, our in-sample analysis reveals some important stylized facts for the spot rate:

1. The importance of modeling mean reversion in mean is ambiguous. It seems that linear drift is adequate for this purpose once GARCH, regime switching, or jumps are included. The contribution of Ait-Sahalia's (1996) type of nonlinear drift is very marginal.

2. It is important to model conditional heteroscedasticity through GARCH or level effect, although GARCH seems to have much better in-sample performance than CEV.

3. Regime switching and jumps help capture volatility clustering and especially the excess kurtosis and heavy-tails of the interest rate.

4. Interest rates behave quite differently during the 1979–1982 period. The level and volatility of the interest rate, the dependence of the interest rate volatility on the interest rate level, and the probability of jumps are much higher during this period.

5. OUT–OF–SAMPLE DENSITY FORECAST PERFORMANCE

The foregoing in-sample analysis demonstrates that modeling volatility clustering through GARCH and heavy tails and excess kurtosis through regime switching or jumps significantly improves the in-sample fit of historical interest rate data. However, it is not clear whether these models will also perform well in out-of-sample density forecasts for future interest rates. Indeed, some previous studies have shown that more complicated models actually underperform the simple random walk model in predicting the conditional mean of future interest rates (e.g., Duffee 2002). In many financial applications, we are interested in forecasting the whole conditional distribution. We now apply the generalized spectral evaluation method described in Section 2 to evaluate density forecasts of the models under study. Among other things, we examine whether the features found to be important for in-sample performance remain important for out-of-sample forecasts, and whether the best in-sample performing models still perform best in out-of-sample forecasts.

For each model, we first calculate $\{Z_t(\hat{\theta})\}$, then model generalized residuals using the prediction sample, with model parameter estimates based on the estimation sample. Table 6 reports the out-of-sample evaluation statistic M_1 for each model. To compute M_1, we select a normal cdf $\Phi(\sqrt{12}u)$ for the weighting function $W(u)$ and a Bartlett kernel for the kernel function $k(z)$. To choose a lag order p, we use a data-driven method that minimizes the asymptotic integrated mean squared error of the modified generalized spectral density. This method involves choosing a preliminary lag order $\bar{p}$. We examine a wide range of $\bar{p}$ in our application. We obtain similar results for the preliminary lag order $\bar{p}$ between 10 and 30 and report results for $\bar{p} = 20$ in Table 6. For comparison, we also report the in-sample log-likelihood values obtained from the estimation sample.

The M_1 statistics show that all single-factor diffusion models are overwhelmingly rejected. One of the most interesting findings is that models that include a drift term (either linear or nonlinear) have much worse out-of-sample performance than those without a drift. Dothan's (1978) model and the pure CEV model have the best density forecast performance. Both models have a zero drift and level effect, the elasticity parameter $\rho = 1$ for the former and .25 (.60 for the 1979–1982 period) for the latter. On the other hand, although the CKLS model and Ait-Sahalia's (1996) nonlinear drift model have the highest in-sample likelihood values, they have the worst out-of-sample density forecasts in terms of M_1. This seems to suggest that for the purpose of density forecast, it is much more important to model the diffusion function than the drift function. Of course, this does not necessarily mean that the drift is not important.

Table 6. Out-of-Sample Density Forecast Performance of Spot Interest Rate Models

	Model	M_1	Log-likelihood
Single-factor diffusion models	Random walk	.108***	2,649.6
	Log-normal	.121***	2,187.0
	Dothan	.067**	2,184.9
	Pure CEV	.082**	2,654.5
	Vasicek	.213***	2,655.5
	CIR	.219***	2,672.0
	CKLS	.219***	2,661.7
	Nonlinear drift	.224***	2,662.5
GARCH models	No drift GARCH	.079**	5,686.8
	Linear drift GARCH	.276***	5,699.5
	Nonlinear drift GARCH	.339***	5,706.3
	No drift CEV–GARCH	.077**	5,701.1
	Linear drift CEV–GARCH	.291***	5,715.2
	Nonlinear drift CEV–GARCH	.358***	5,725.5
Regime-switching models	No drift RS CEV	.129***	6,493.1
	Linear drift RS CEV	.091***	6,507.9
	Nonlinear drift RS CEV	.162***	6,512.8
	No drift RS GARCH	.118***	7,079.0
	Linear drift RS GARCH	.047*	7,132.8
	Nonlinear drift RS GARCH	.055**	7,138.2
	No drift RS CEV–GARCH	.124***	7,082.8
	Linear drift RS CEV–GARCH	.056**	7,138.6
	Nonlinear drift RS CEV–GARCH	.067**	7,144.1
Jump-diffusion models	No drift JD CEV	.180***	6,306.3
	Linear drift JD CEV	.039*	6,317.2
	Nonlinear drift JD CEV	.076**	6,326.3
	No drift JD GARCH	.178***	6,794.1
	Linear drift JD GARCH	.056**	6,810.8
	Nonlinear drift JD GARCH	.067**	6,813.9
	No drift JD CEV–GARCH	.174***	6,796.3
	Linear drift JD CEV–GARCH	.053**	6,812.9
	Nonlinear drift JD CEV–GARCH	.065**	6,816.1

NOTE: This table reports the out-of-sample density forecast performance of the single-factor diffusion, GARCH, Markov regime-switching, and jump-diffusion models measured by the M_1 statistic. For convenience of comparison, we also report the in-sample log-likelihood value for each model. The in-sample parameter estimation is based on the observations from June 14, 1961 to February 2, 1991, and the out-of-sample density forecast evaluation is based on the observations from February 23, 1991 to December 29, 2000. The ratio between the size of the estimation sample and the forecast sample is about 3 : 1. In computing the out-of-sample M_1 statistics, a preliminary lag order $\bar{p}$ of 20 is used. Other lag orders give similar results. The asymptotic critical values for the out-of-sample evaluation statistic are .087, .051, and .037 at the 1%, 5%, and 10% levels and ***, **, and * indicate that the M_1 statistic is significant at the 1%, 5%, and 10% levels.

It might be simply because the drift specifications considered are severely misspecified and do not outperform the zero drift model. In fact, most drift parameter estimates have large standard errors and thus have excess sampling variation in parameter estimation. As a consequence, assuming zero drift may result in a smaller adverse effect on out-of-sample performance.

The generalized residuals $\{Z_t(\hat{\theta})\}$ contain much information about possible sources of model misspecification. Instead of being uniform, the histograms of the generalized residuals for all diffusion models (Fig. 2) exhibit a high peak in the center of the distribution, which indicates that the diffusion models cannot satisfactorily capture the marginal density of the spot rate. The separate inference statistics $M(m, l)$ in Table 7 also show that all the models fail to satisfactorily capture the dynamics of the generalized residuals. In particular, the large $M(2, 2)$ and $M(4, 4)$ statistics indicate that diffusion models perform rather poorly in modeling the conditional variance and kurtosis of the generalized residuals.

Similar to single-factor diffusion models, GARCH models with a zero drift have much better density forecasts than those models with a linear or nonlinear drift. The best GARCH and diffusion models have M_1 statistics roughly between .07

and .08. Comparing the best GARCH and single-factor diffusion models, we find that GARCH provides better density forecasts than CEV, and combining CEV and GARCH further improves the density forecasts. Diagnostic analysis of generalized residuals shows that GARCH models provide certain improvements over diffusion models. For example, the histograms of the generalized residuals of GARCH models have a less-pronounced peak in the center. The $M(m, l)$ statistics show that GARCH models significantly improve the fitting of even-order moments of the generalized residuals; the $M(2, 2)$ and $M(4, 4)$ statistics are reduced from close to 100 to single digits!

In summary, our analysis of single-factor diffusion and GARCH models demonstrates that both classes of models fail to provide good density forecasts for future interest rates. They fail to capture both the high peak in the center of the marginal density and the dynamics of different moments of the generalized residuals. GARCH models have a more uniform marginal density and significantly improve the ability to model the dynamics of even-order moments of the generalized residuals. For both classes of models, modeling the conditional variance through CEV or GARCH is much more important than modeling the conditional mean of the interest rate. Zero-drift models outperform either linear or nonlinear drift models in density forecast, although the $M(1, 1)$ statistics suggest that neither model adequately captures the dynamic structure in the conditional mean of the generalized residuals.

We next turn to regime-switching models. Regime-switching models generally have better density forecasts than single-factor diffusion and GARCH models: the best regime-switching models reduce the M_1 statistics of the best diffusion and GARCH models from above .07 to about .05. Although modeling the drift does not improve density forecast for diffusion and GARCH models, the best regime-switching models in density forecast have a linear drift in each regime. As pointed out by Ang and Bekaert (2002) and Li and Xu (2000), a regime-switching model with a linear drift in each regime actually implies nonlinear conditional mean dynamics for the interest rate. Thus our results suggest that perhaps a specific nonlinear drift is needed for out-of-sample density forecasts. GARCH is more important than CEV for density forecasts, and combining CEV with GARCH does not further improve model performance. Overall, the regime-switching GARCH model with a linear drift in each regime provides the best out-of-sample forecasts among all of the regime-switching models. Diagnostic analysis shows that regime-switching models provide a much better characterization of the marginal density of interest rates; the histograms of the generalized residuals are much closer to U[0, 1]. The $M(2, 2)$ and $M(4, 4)$ statistics of regime-switching models are slightly higher than those of GARCH models, however, and the $M(1, 1)$ and $M(3, 3)$ statistics of regime-switching models are actually higher than those of diffusion and GARCH models. Therefore the advantages of regime-switching models seem to come from better modeling of the marginal density, rather than from the dynamics of the generalized residuals.

The density forecast performance of jump-diffusion models shares many common features with that of regime-switching models. First, they have comparable performance; the M_1 statistics of the best jump-diffusion models are also around .05. Second, for jump-diffusion models, the linear drift specifi-
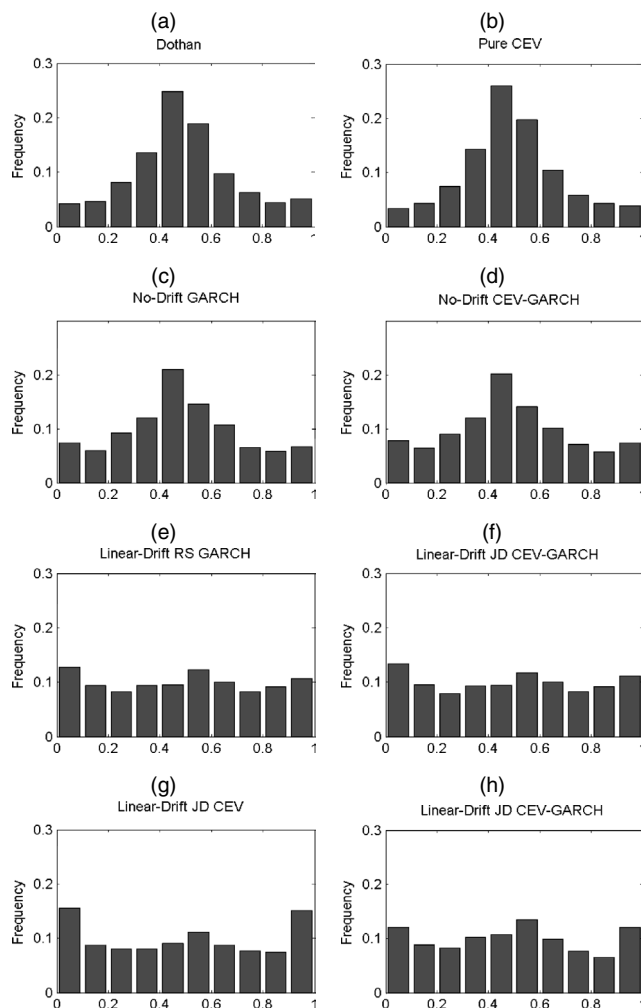

Figure 2. Histograms of the Generalized Residuals of Four Classes of Spot Rate Models. This figure plots the histograms of the generalized residuals of the single-factor diffusion, GARCH, regime-switching, and jump-diffusion spot rate models.

Table 7. Separate Inference Statistics for Out-of-Sample Density Forecast Performance

	Model	M(1, 1)	M(1, 2)	M(2, 1)	M(2, 2)	M(3, 3)	M(4, 4)
Single-	Random walk	40.08	6.35	40.48	99.73	21.50	80.12
factor	Log-normal	46.52	8.77	39.91	110.10	26.98	92.51
diffusion	Dothan	46.04	8.07	42.62	108.90	27.02	91.03
models	Pure CEV	41.62	6.63	41.55	96.81	22.85	80.42
	Vasicek	40.40	8.13	35.55	100.20	21.43	81.79
	CIR	43.61	8.68	36.30	98.57	24.39	85.02
	CKLS	42.02	8.55	35.75	97.89	22.80	82.77
	Nonlinear drift	42.40	8.37	35.93	98.47	22.77	82.98
GARCH	No drift GARCH	40.73	3.75	28.21	7.68	17.57	2.53
models	Linear drift GARCH	41.07	4.66	25.11	7.53	17.00	2.35
	Nonlinear drift GARCH	40.71	4.87	24.70	7.02	16.63	2.09
	No drift CEV–GARCH	41.30	4.10	28.40	7.46	19.24	2.74
	Linear drift CEV–GARCH	41.68	5.11	25.06	7.09	18.73	2.44
	Nonlinear drift CEV–GARCH	41.43	5.45	24.32	6.45	18.54	2.17
Regime-	No drift RS CEV	53.16	3.13	11.31	7.05	35.02	10.21
switching	Linear drift RS CEV	53.25	4.56	11.16	6.57	34.31	8.83
models	Nonlinear drift RS CEV	53.75	5.12	9.84	6.56	34.50	8.49
	No drift RS GARCH	57.29	4.05	21.59	10.95	43.71	7.63
	Linear drift RS GARCH	55.46	4.56	21.09	12.38	40.71	9.14
	Nonlinear drift RS GARCH	55.69	4.50	21.05	12.14	41.13	8.93
	No drift RS CEV–GARCH	45.35	3.69	20.91	10.67	34.25	7.08
	Linear drift RS CEV–GARCH	44.10	4.86	20.75	11.99	32.12	8.87
	Nonlinear drift RS CEV–GARCH	44.36	4.84	20.55	11.40	32.46	8.32
Jump-	No drift JD CEV	47.38	5.32	16.76	85.10	43.01	99.31
diffusion	Linear drift JD CEV	46.32	4.48	19.63	85.09	42.26	99.82
models	Nonlinear drift JD CEV	46.50	4.85	18.31	85.10	42.48	99.29
	No drift JD GARCH	43.32	5.10	20.35	12.38	28.44	8.96
	Linear drift JD GARCH	42.57	4.60	21.48	12.19	28.22	8.84
	Nonlinear drift JD GARCH	42.62	4.67	21.42	12.34	28.24	8.95
	No drift JD CEV–GARCH	43.55	5.05	20.51	13.36	29.31	10.07
	Linear drift JD CEV–GARCH	42.74	4.54	21.69	13.00	28.93	9.76
	Nonlinear drift JD CEV–GARCH	42.78	4.62	21.57	13.07	28.94	9.81

NOTE: This table reports the separate inference statistics $M(m, l)$ for the four classes of spot rate models. The asymptotically normal statistic $M(m, l)$ can be used to test whether the cross-correlation between the mth and lth moments of $\{Z_t\}$ is significantly different from 0. The choice of $(m, l) = (1, 1)$, $(2, 2)$, $(3, 3)$, and $(4, 4)$ is sensitive to autocorrelations in mean, variance, skewness, and kurtosis of $\{r_t\}$. We report results only for a lag truncation order $p = 20$; the results for $p = 10$ and 30 are rather similar. The asymptotical critical values are 1.28, 1.65, and 2.33 at the 10%, 5%, and 1% levels.

cation outperforms those with no drift or nonlinear drift for density forecasts. Third, for jump-diffusion models, GARCH generally provides better density forecasts than CEV. Fourth, jump-diffusion models have similar advantages over single-factor diffusion and GARCH models. They have much better characterization of the marginal density of the generalized residuals, but they may not necessarily perform better in modeling the dynamics of the generalized residuals, as indicated by their large $M(m, l)$ statistics. The jump model with a linear drift and level effect, although it has the smallest M_1 statistic, apparently fails to capture the even-order moments of the generalized residuals, with $M(2, 2)$ and $M(4, 4)$ statistics significantly higher than other jump-diffusion models that include GARCH effects.

In summary, our out-of-sample density forecast evaluation reveals the following important features:

1. For out-of-sample density forecast, the importance of modeling mean reversion in the interest rate is ambiguous. For diffusion and GARCH models, a linear or nonlinear drift gives worse density forecasts than a zero drift. However, when there are regime switches or jumps, linear drift models outperform those with zero or nonlinear drift. The best density forecast models still fail to adequately capture the mean dynamics of the generalized residuals.
2. It is important to capture volatility clustering in interest rates for both in-sample and out-of-sample performance

via GARCH. Most of the best-performing models have a GARCH component, which significantly improves their ability to model the dynamics of the even-order moments of the generalized residuals.
3. It is also important to capture excess kurtosis and heavy tails of the interest rate for both in-sample and out-of-sample performance via regime switching or jumps. The advantages of regime-switching and jump models are mainly reflected in modeling the marginal density, rather than in the dynamics of the generalized residuals.

Overall, our out-of-sample analysis demonstrates that more-complicated models that incorporate conditional heteroscedasticity and heavy tails of interest rates tend to have a better density forecast. This result is quite different from the results for those that focus on conditional mean forecast. As widely documented in the literature, the predictable component in the conditional mean of the interest rate appears insignificant. As a result, random walk models tend to outperform more-sophisticated models in terms of mean forecast (e.g., Duffee 2002). However, density forecast includes all conditional moments, and as a result, those models that can capture the dynamics of higher-order moments tend to have better density forecasts. Our analysis suggests that the more-complicated spot rate models developed in the literature can indeed capture some important features of interest rates and are relevant in applications involving density forecasts of interest rates.

6. CONCLUSION

Despite the numerous empirical studies of spot interest rate models, little effort has been devoted to examining their out-of-sample forecast performance. We have contributed to the literature by providing the first (to our knowledge) comprehensive empirical analysis of the out-of-sample performance of a wide variety of popular spot rate models in forecasting the conditional density of future interest rates. Out-of-sample density forecasts help minimize the data snooping bias due to excessive searching for more-complicated models using similar datasets. The conditional density, which completely characterizes the full dynamics of interest rates, is also an essential input to many financial applications. Using a rigorous econometric evaluation procedure for density forecasts, we examined the out-of-sample performance of single-factor diffusion, GARCH, regime-switching, and jump-diffusion models.

Although previous studies have shown that simpler models, such as the random walk model, tend to have more accurate forecasts for the conditional mean of interest rates, we find that models that capture conditional heteroscedasticity and heavy tails of the spot rate have better out-of-sample density forecasts. GARCH significantly improves modeling of the conditional variance and kurtosis of the generalized residuals, whereas regime-switching and jumps help in modeling the marginal density of interest rates. Therefore, our analysis shows that the more-sophisticated interest rate models developed in the literature can indeed capture some important features of interest rate data and perform better than simpler models in out-of-sample density forecasts. Although we have focused on density forecasts of the spot rates here, there is evidence that the term structure of interest rates is driven by multiple factors (e.g., Dai and Singleton 2000). Extending our analysis to forecast the joint density of multiple yields is an interesting issue that we will address in future research.

ACKNOWLEDGMENTS

APPENDIX: GENERALIZED SPECTRAL EVALUATION FOR DENSITY FORECASTS

Here we describe Hong's (2003) omnibus evaluation procedure for out-of-sample density forecasts, as well as a class of separate inference procedures. (For more discussion and formal treatment, see Hong 2003.)

A.1 Generalized Spectrum

Generalized spectrum was proposed by Hong (1999) as an analytic tool to provide an alternative to the power spectrum and the higher-order spectrum (e.g., bispectrum) in time series analysis. Define the centered generalized residual

$$Z_t \equiv Z_t(\theta) = \int_{-\infty}^{r_t} p(r, t|I_{t-1}, \theta)\, dr - \frac{1}{2}, \qquad t = R+1, \ldots, T,$$

where $p(r, t|I_{t-1}, \theta)$ is a conditional density model for time series $\{r_t\}$. Suppose that the series $\{Z_t\}_{t=-\infty}^{\infty}$ has marginal characteristic function $\varphi(u) \equiv E(e^{iuZ_t})$ and pairwise joint characteristic function $\varphi_j(u, v) \equiv E[\exp\{i(uZ_t + vZ_{t-|j|})\}]$, where $i \equiv \sqrt{-1}$, $u, v \in (-\infty, \infty)$ and $j = 0, \pm 1, \ldots$. The basic idea of generalized spectrum is to transform the series $\{Z_t\}$ and then consider the spectrum of the transformed series. Define the generalized covariance function

$$\sigma_j(u, v) \equiv \mathrm{cov}(e^{iuZ_t}, e^{ivZ_{t-|j|}}), \qquad j = 0, \pm 1, \ldots.$$

Straightforward algebra yields $\sigma_j(u, v) = \varphi_j(u, v) - \varphi(u)\varphi(v)$. Because $\varphi_j(u, v) = \varphi(u)\varphi(v)$ for all (u, v) if and only if Z_t and $Z_{t-|j|}$ are independent (cf. Lukacs 1970), $\sigma_j(u, v)$ can capture any type of pairwise serial dependence in $\{Z_t\}$ over various lags, including those with zero autocorrelation in $\{Z_t\}$.

Under certain regularity conditions on the temporal dependence of $\{Z_t\}$, the Fourier transform of the generalized covariance function $\sigma_j(u, v)$ exists and is given by

$$f(\omega, u, v) \equiv \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \sigma_j(u, v) e^{-ij\omega}, \qquad \omega \in [-\pi, \pi].$$

Like $\sigma_j(u, v)$, $f(\omega, u, v)$ can capture any type of pairwise serial dependence in $\{Z_t\}$ across various lags. An advantage of spectral analysis is that $f(\omega, u, v)$ incorporates all lags simultaneously. Thus it can capture the dependent processes in which serial dependence occurs only at higher-order lags. This can arise from seasonality (e.g., calender effects) or time delay in financial markets. Moreover, $f(\omega, u, v)$ is particularly powerful in capturing the dependent alternatives whose serial dependence decays to 0 slowly as $j \to \infty$, such as slow mean-reverting or persistent volatility clustering processes. For such processes, $f(\omega, u, v)$ has a sharp peak at frequency 0.

The function $f(\omega, u, v)$ does not require any moment condition on $\{Z_t\}$. When the moments of $\{Z_t\}$ exist, however, we can differentiate $f(\omega, u, v)$ with respect to (u, v) at $(0, 0)$,

$$f^{(0,m,l)}(\omega, 0, 0) \equiv \frac{\partial^{m+l}}{\partial u^m \, \partial v^l} f(\omega, u, v)\Big|_{(u,v)=(0,0)}$$

$$= \frac{i^{m+l}}{2\pi} \sum_{j=-\infty}^{\infty} \mathrm{cov}(Z_t^m, Z_{t-|j|}^l) e^{-ij\omega}.$$

In particular, $f^{(0,1,1)}(\omega, 0, 0)$ is the well-known power spectral density. (For applications of the power spectral density in economics and finance, see, e.g., Granger 1969; Durlauf 1991; Watson 1993.) For this reason, $f(\omega, u, v)$ is called the "generalized spectral density" of $\{Z_t\}$. The parameters (u, v) provide much flexibility in capturing linear and nonlinear dependence. Because $f(\omega, u, v)$ can be decomposed as a weighted sum of

$f^{(0,m,l)}(\omega, 0, 0)$ over various (m, l) via a Taylor series expansion around $(0, 0)$, it contains information on all autocorrelations and cross-correlations in the various power terms of $\{Z_t\}$. Thus it can be used to develop an omnibus procedure against a wide range of dependent processes.

A.2 Omnibus Evaluation for Density Forecast

To evaluate out-of-sample density forecasts of model $p(r, t | I_{t-1}, \theta)$, Hong (2003) used the generalized spectrum to develop an omnibus test of iid $U[-\frac{1}{2}, \frac{1}{2}]$ for the centered generalized residual $\{Z_t\}$. To test $\mathbb{H}_0 : \{Z_t\} \sim$ iid $U[-\frac{1}{2}, \frac{1}{2}]$, we use the modified generalized covariance

$$\sigma_j^{\mathrm{U}}(u, v) \equiv \varphi_j(u, v) - \varphi^{\mathrm{U}}(u)\varphi^{\mathrm{U}}(v),$$

where $\varphi^{\mathrm{U}}(u) \equiv \sin(u/2)/(u/2)$ is the characteristic function of a $U[-\frac{1}{2}, \frac{1}{2}]$ random variable. Because $\sigma_j^{\mathrm{U}}(u, v)$ has incorporated the information of $U[-\frac{1}{2}, \frac{1}{2}]$, it can capture any pairwise serial dependence in $\{Z_t\}$ and any deviation from $U[-\frac{1}{2}, \frac{1}{2}]$. The associated generalized spectrum is

$$f^{\mathrm{U}}(\omega, u, v) \equiv \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \sigma_j^{\mathrm{U}}(u, v)e^{-ij\omega},$$

$$\omega \in [-\pi, \pi], u, v \in (-\infty, \infty).$$

Under $\mathbb{H}_0$, we have $\sigma_j^{\mathrm{U}}(u, v) = 0$ for all $j \neq 0$, and so $f^{\mathrm{U}}(\omega, u, v)$ becomes a known "flat" spectrum,

$$f_0^{\mathrm{U}}(\omega, u, v) \equiv \frac{1}{2\pi}\sigma_0^{\mathrm{U}}(u, v)$$

$$= [\varphi^{\mathrm{U}}(u + v) - \varphi^{\mathrm{U}}(u)\varphi^{\mathrm{U}}(v)], \qquad \omega \in [-\pi, \pi].$$

When Z_t and Z_{t-j} are not independent at some lag $j \neq 0$, or when Z_t is not $U[-\frac{1}{2}, \frac{1}{2}]$, we have $f^{\mathrm{U}}(\omega, u, v) \neq f_0^{\mathrm{U}}(\omega, u, v)$. Hence, to test $\mathbb{H}_0$, one can check whether a consistent estimator of $f^{\mathrm{U}}(\omega, u, v)$ is significantly different from $f_0^{\mathrm{U}}(\omega, u, v)$.

Suppose that $\hat{\theta}$ is an estimator for model parameter θ based on the estimation sample $\{r_t\}_{t=1}^R$. To estimate $f^{\mathrm{U}}(\omega, u, v)$ using the prediction sample $\{r_t\}_{t=R+1}^N$, we have to use the estimated proxies $\{\hat{Z}_t\}_{t=R+1}^N$, where

$$\hat{Z}_t \equiv Z_t(\hat{\theta})$$

$$= \int_{-\infty}^{r_t} p(r, t | I_{t-1}, \hat{\theta})\, dr - \frac{1}{2}, \qquad t = R+1, \ldots, N.$$

Define the empirical measure,

$$\hat{\sigma}_j^{\mathrm{U}}(u, v) \equiv \hat{\varphi}_j(u, v) - \varphi^{\mathrm{U}}(u)\varphi^{\mathrm{U}}(v), \qquad j = 0, \pm 1, \ldots, \pm(n-1),$$

where $n \equiv N - R$ is the size of the prediction sample, and the empirical pairwise joint characteristic function,

$$\hat{\varphi}_j(u, v) \equiv (n - |j|)^{-1} \sum_{t=R+|j|+1}^{N} \exp\{i(u\hat{Z}_t + v\hat{Z}_{t-|j|})\}.$$

We introduce a class of smoothed kernel estimators,

$$\hat{f}^{\mathrm{U}}(\omega, u, v) \equiv \frac{1}{2\pi} \sum_{j=1-n}^{n-1} (1 - |j|/n)^{1/2} k(j/p)\hat{\sigma}_j^{\mathrm{U}}(u, v)e^{-ij\omega},$$

where $k(\cdot)$ is a kernel and $p \equiv p(n)$ is a bandwidth/lag order. An example of $k(\cdot)$ is the Bartlett kernel

$$k(z) = \begin{cases} 1 - |z| & \text{if } |z| \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

For the choice of p, Hong (1999) suggested a data-driven method that delivers an asymptotically optimal bandwidth in terms of an integrated mean squared error criterion. This still involves a preliminary bandwidth $\bar{p}$, but the impact of choosing $\bar{p}$ is much smaller than the impact of choosing lag order p.

A convenient divergence measure for $\hat{f}^{\mathrm{U}}(\omega, u, v)$ and $f_0^{\mathrm{U}}(\omega, u, v)$ is the quadratic form

$$2\pi n Q(\hat{f}^{\mathrm{U}}, f_0^{\mathrm{U}})$$

$$\equiv 2\pi n \int\int\int_{-\pi}^{\pi} |\hat{f}^{\mathrm{U}}(\omega, u, v)$$

$$- f_0^{\mathrm{U}}(\omega, u, v)|^2\, d\omega\, dW(u)\, dW(v)$$

$$= n \int\int |\hat{\varphi}_0(u + v) - \varphi^{\mathrm{U}}(u + v)|^2\, dW(u)\, dW(v)$$

$$+ 2\sum_{j=1}^{n-1} k^2(j/p)(n - j) \int\int |\hat{\sigma}_j^{\mathrm{U}}(u, v)|^2\, dW(u)\, dW(v),$$

where the weighting function $W(\cdot)$ is positive and nondecreasing and the unspecified integrals are taken over the support of $W(\cdot)$. An example of this is

$$W(u) = \Phi(\sqrt{12}u),$$

where $\Phi(\cdot)$ is the $N(0, 1)$ cdf, which is commonly used in the characteristic function literature. The scale $\sqrt{12}$ matches the standard deviation of $U[-\frac{1}{2}, \frac{1}{2}]$, but this is not necessary.

Our evaluation statistic for $\mathbb{H}_0$ is a properly centered and scaled version of the quadratic form $Q(\hat{f}^{\mathrm{U}}, f_0^{\mathrm{U}})$,

$$M_1 \equiv \left[\sum_{j=1}^{n-1} k^2(j/p)\right]^{-1}$$

$$\times \int\int \sum_{j=1}^{n-1} k^2(j/p)(n - j)|\hat{\sigma}_j^{\mathrm{U}}(u, v)|^2\, dW(u)\, dW(v)$$

$$- \left[\int [1 - \varphi^{\mathrm{U}}(u)^2]\, dW(u)\right]^2.$$

The asymptotic distribution of M_1 was derived by Hong (2003). Under $\mathbb{H}_0$ and a set of regularity conditions,

$$M_1 \xrightarrow{d} \int |G(u, v)|^2\, dW(u)\, dW(v)$$

as $n \to \infty, R \to \infty$, and $n^{1+\delta}/R \to 0$ for any arbitrary small constant $\delta > 0$, where $G(u, v)$ is a complex-valued Gaussian process with mean 0 and a known covariance function,

$$\mathrm{cov}[G(u_1, v_1), G(u_2, v_2)]$$

$$= E[G(u_1, v_1)G(u_2, v_2)^*]$$

$$= \sigma_0^{\mathrm{U}}(u_1, -u_2)\varphi^{\mathrm{U}}(v_1)\varphi^{\mathrm{U}}(v_2) + \sigma_0^{\mathrm{U}}(u_1, -v_2)\varphi^{\mathrm{U}}(u_2)\varphi^{\mathrm{U}}(v_1)$$

$$+ \sigma_0^{\mathrm{U}}(v_1, -u_2)\varphi^{\mathrm{U}}(u_1)\varphi^{\mathrm{U}}(v_2)$$

$$+ \sigma_0^{\mathrm{U}}(v_1, -v_2)\varphi^{\mathrm{U}}(u_1)\varphi^{\mathrm{U}}(u_2).$$

The limit distribution of M_1 is nonstandard. However, because the Gaussian process $G(u, v)$ has mean 0 and a known, distribution-free covariance function, the limit distribution of M_1 can be easily tabulated. When $W(u) = \Phi(\sqrt{12}u)$ with a truncated support on $[-1, 1]$, the asymptotic critical values at the 10%, 5%, and 1% levels are .037, .051, and .087.

The most attractive feature of M_1 is its omnibus property; it has power against a wide range of suboptimal density forecasts, thanks to the use of the characteristic function in a spectral framework. The M_1 statistic can be viewed as an omnibus metric measuring the departure of the forecast model from the true data-generating process. A better density forecast model is expected to have a smaller value for M_1, because its generalized residual series $\{Z_t\}$ is closer to having the iid U[0, 1] property. Thus M_1 can be used to rank competing density forecast models in terms of their deviations from optimality. Moreover, the kernel function provides flexible weighting for various lags. Nonuniform weighting kernels usually discount higher-order lags. For most, if not all, economic and financial time series, today's behaviors are more affected by recent events than by remote events. Thus downward weighting for higher-order lags is expected to yield good power in finite samples.

A.3 Separate Inferences

When a model is rejected using M_1, it is interesting to explore possible reasons for the rejection. Hong (2003) provided a class of rigorous separate inference statistics that exploit the information contained in modeling generalized residuals $\{Z_t\}$ to understand possible reasons for model rejection. The $M(m, l)$ statistic is based on a kernel estimator of the generalized spectral density derivative $f^{(0,m,l)}(\omega, 0, 0)$, where m and l are the orders of the derivatives with respect to u and v. The statistic $M(m, l)$ is defined as

$$M(m, l) = \left[\sum_{j=1}^{n-1} (n-j)k^2(j/p)\hat{\rho}_j^2(m, l) \right.$$
$$\left. - \sum_{j=1}^{n-1} k^2(j/p) \right] \Big/ \left[2\sum_{j=1}^{n-2} k^4(j/p) \right],$$

where $\hat{\rho}_j(m, l)$ is the sample cross-correlation between $\hat{Z}_t^m$ and $\hat{Z}_{t-|j|}^l$, $t = R + 1, \ldots, N$. Under $\mathbb{H}_0$ and a set of regularity conditions,

$$M(m, l) \xrightarrow{d} N(0, 1)$$

for any given (m, l), as $n, R \to \infty$. The upper-tailed N(0, 1) critical values at the 10%, 5%, and 1% levels are 1.28, 1.65, and 2.33. As in the M_1 test, Hong (2003) considered a data-driven method for p that also involves the choice of a preliminary lag order $\bar{p}$.

Although the moments of the generalized residuals $\{Z_t\}$ are not exactly the same as those of the original data $\{r_t\}$, they are highly correlated. In particular, the choice of $(m, l) = (1, 1), (2, 2), (3, 3)$, and $(4, 4)$ is very sensitive to autocorrelations in level, volatility, skewness, and kurtosis of $\{r_t\}$. Furthermore, the choices of $(m, l) = (1, 2)$ and $(2, 1)$ are sensitive to the "ARCH-in-mean" effect and the "leverage" effect. Different choices of order (m, l) can thus allow examination of various dynamic aspects of the underlying process.

REFERENCES

Ait-Sahalia, Y. (1996), "Testing Continuous-Time Models of the Spot Interest Rate," *Review of Financial Studies*, 9, 385–426.
——— (1999), "Transition Densities for Interest Rate and Other Nonlinear Diffusions," *Journal of Finance*, 54, 1361–1395.
——— (2002), "Maximum-Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approach," *Econometrica*, 70, 223–262.
Ait-Sahalia, Y., and Lo, A. (1998), "Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices," *Journal of Finance*, 53, 499–547.
Andersen, T. G., and Lund, J. (1997), "Estimating Continuous Time Stochastic Volatility Models of the Short-Term Interest Rate," *Journal of Econometrics*, 77, 343–378.
Ang, A., and Bekaert, G. (2002), "Regime Switches in Interest Rates," *Journal of Business & Economic Statistics*, 20, 163–182.
Bali, T. (2000), "Modeling the Conditional Mean and Variance of the Short Rate Using Diffusion, GARCH and Moving Average Models," *Journal of Futures Markets*, 20, 717–751.
——— (2001), "Nonlinear Parametric Models of the Short-Term Interest Rate," unpublished manuscript, Baruch College.
Ball, C., and Torous, W. (1998), "Regime Shifts in Short-Term Riskless Interest Rates," unpublished manuscript, University of California Los Angeles.
Bera, A. K., and Ghosh, A. (2002), "Density Forecast Evaluation With Applications in Finance," unpublished manuscript, University of Illinois at Champaign, Dept. of Economics.
Berkowitz, J. (2001), "Testing the Accuracy of Density Forecasts, With Applications to Risk Management," *Journal of Business & Economic Statistics*, 19, 465–474.
Bliss, R., and Smith, D. (1998), "The Elasticity of Interest Rate Volatility: Chan, Karolyi, Longstaff, and Sanders Revisited," unpublished manuscript, Federal Reserve Bank of Atlanta.
Brenner, R., Harjes, R., and Kroner, K. (1996), "Another Look at Alternative Models of Short-Term Interest Rate," *Journal of Financial and Quantitative Analysis*, 31, 85–107.
Chacko, G., and Viceira, L. M. (2003), "Spectral GMM Estimation of Continuous-Time Processes," *Journal of Econometrics*, 116, 259–292.
Chan, K. C., Karolyi, G. A., Longstaff, F. A., and Sanders, A. B. (1992), "An Empirical Comparison of Alternative Models of the Short-Term Interest Rate," *Journal of Finance*, 47, 1209–1228.
Chapman, D., and Pearson, N. (2001), "Recent Advances in Estimating Models of the Term-Structure," *Financial Analysts Journal*, 57, 77–95.
Christoffersen, P., and Diebold, F. X. (1996), "Further Results on Forecasting and Model Selection Under Asymmetric Loss," *Journal of Applied Econometrics*, 11, 561–572.
——— (1997), "Optimal Prediction Under Asymmetric Loss," *Econometric Theory*, 13, 808–817.
Clements, M. P., and Smith, J. (2000), "Evaluating the Forecast Densities of Linear and Non-Linear Models: Applications to Output Growth and Unemployment," *Journal of Forecasting*, 19, 255–276.
Conley, T. G., Hansen, L. P., Luttmer, E. G. J., and Scheinkman, J. A. (1997), "Short-Term Interest Rates as Subordinated Diffusions," *Review of Financial Studies*, 10, 525–578.
Corradi, V. (2000), "Reconsidering the Continuous-Time Limit of the GARCH(1, 1) Process," *Journal of Econometrics*, 96, 145–153.
Cox, J. C., Ingersoll, J. E., and Ross, S. A. (1985), "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 385–407.
Dai, Q., and Singleton, K. (2000), "Specification Analysis of Affine Term Structure Models," *Journal of Finance*, 55, 1943–1978.
Das, S. R. (2002), "The Surprise Element: Jumps in Interest Rates," *Journal of Econometrics*, 106, 27–65.
Diebold, F. X., Gunther, T. A., and Tay, A. S. (1998), "Evaluating Density Forecasts With Applications to Financial Risk Management," *International Economic Review*, 39, 863–883.
Diebold, F. X., Hahn, J., and Tay, A. S. (1999), "Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High-Frequency Returns of Foreign Exchange," *Review of Economics and Statistics*, 81, 661–673.
Diebold, F. X., Tay, A. S., and Wallis, K. F. (1999), "Evaluating Density Forecasts of Inflation: The Survey of Professional Forecasters," in *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W. J. Granger*, eds. R. F. Engle and H. White, Oxford, U.K.: Oxford University Press, pp. 76–90.

Dothan, L. U. (1978), "On the Term Structure of Interest Rates," *Journal of Financial Economics*, 6, 59–69.

Duffee, G. (2002), "Term Premia and Interest Rate Forecasts in Affine Models," *Journal of Finance*, 57, 405–443.

Duffie, D., and Pan, J. (1997), "An Overview of Value at Risk," *Journal of Derivatives,* 4, 13–32.

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

Durlauf, S. (1991), "Spectral Based Testing of the Martingale Hypothesis," *Journal of Econometrics*, 50, 355–376.

Fackler, P. L., and King, R. P. (1990), "Calibration of Option-Based Probability Assessments in Agricultural Commodity Markets," *American Journal of Agricultural Economics*, 72, 73–83.

Gallant, A. R., and Tauchen, G. (1998), "Reprojection Partially Observed Systems With Applications to Interest Rate Diffusions," *Journal of the American Statistical Association*, 93, 10–24.

Ghysels, E., Carrasco, M., Chernov, M., and Florens, J. (2001), "Estimating Diffusions With a Continuum of Moment Conditions," unpublished manuscript, University of North Carolina.

Granger, C. (1969), "Testing for Causality and Feedback," *Econometrica,* 37, 424–438.

——— (1999), "The Evaluation of Econometric Models and of Forecasts," invited lecture at the Far Eastern Meeting of the Econometric Society, Singapore.

Granger, C., and Pesaran, M. H. (2000), "A Decision Theoretic Approach to Forecasting Evaluation," in *Statistics and Finance: An Interface*, eds. W. S. Chan, W. K. Li, and H. Tong, London: Imperial College, pp. 261–278.

Gray, S. (1996), "Modeling the Conditional Distribution of Interest Rates as a Regime-Switching Process," *Journal of Financial Economics*, 42, 27–62.

Hamilton, J. D. (1989), "A New Approach to the Econometric Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357–384.

Hong, Y. (1999), "Hypothesis Testing in Time Series via the Empirical Characteristic Function: A Generalized Spectral Density Approach," *Journal of the American Statistical Association*, 94, 1201–1220.

——— (2003), "Evaluation of Out-of-Sample Density Forecasts With Applications to Stock Prices," unpublished manuscript, Cornell University, Dept. of Economics and Dept. of Statistical Science.

Hong, Y., and Lee, T. H. (2003a), "Inference on the Predictability of Foreign Exchange Rate Changes via Generalized Spectrum and Nonlinear Models," *Review of Economics and Statistics*, 84, 1048–1062.

——— (2003b), "Diagnostic Checking for Nonlinear Time Series Models," *Econometric Theory*, 19, 1065–1121.

Hong, Y., and Li, H. (2004), "Nonparametric Specification Testing for Continuous-Time Models With Applications to Interest Rate Term Structures," *Review of Financial Studies*, forthcoming.

Jackwerth, J. C., and Rubinstein, M. (1996), "Recovering Probability Distributions From Option Prices," *Journal of Finance*, 51, 1611–1631.

Jiang, G., and Knight, J. (1997), "A Nonparametric Approach to the Estimation of Diffusion Processes With an Application to a Short-Term Interest Rate Model," *Econometric Theory*, 13, 615–645.

Johannes, M. (2003), "The Statistical and Economic Role of Jumps in Interest Rates," *Journal of Finance*, 59, 227–260.

Jorion, P. (2000), *Value at Risk: The New Benchmark for Managing Financial Risk*, New York: McGraw-Hill.

Morgan, J. P. (1996), *RiskMetrics—Technical Document* (4th ed.), New York: Morgan Guaranty Trust Company.

Koedijk, K. G., Nissen, F. G., Schotman, P. C., and Wolff, C. C. P. (1997), "The Dynamics of Short-Term Interest Rate Volatility Reconsidered," *European Finance Review*, 1, 105–130.

Li, H., and Xu, Y. (2000), "Regime Shifts and Short Rate Dynamics," unpublished manuscript, Cornell University.

Lo, A., and MacKinlay, A. C. (1989), "Data-Snooping Biases in Tests of Financial Asset Pricing Models," *Review of Financial Studies*, 3, 175–208.

Lukacs, E. (1970), *Characteristic Functions* (2nd ed.), London: Charles Griffin.

Meese, R., and Rogoff, K. (1983), "Empirical Exchange Rate Models of the Seventies: Do They Fit Out-of-Sample?" *Journal of International Economics*, 14, 3–24.

Nelson, D. (1990), "ARCH Models as Diffusion Approximation," *Journal of Econometrics*, 45, 7–39.

——— (1991), "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347–370.

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

Rothschild, M., and Stiglitz, J. (1970), "Increasing Risk. I. A Definition," *Journal of Economic Theory*, 2, 225–243.

Singleton, K. (2001), "Estimation of Affine Asset Pricing Models Using the Empirical Characteristic Function," *Journal of Econometrics, 102, 111–141.*

Soderlind, P., and Svensson, L. (1997), "New Techniques to Extract Market Expectations From Financial Instruments," *Journal of Monetary Economics*, 40, 383–429.

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

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

Watson, M. (1993), "Measures of Fit for Calibrated Models," *Journal of Political Economy*, 101, 1011–1041.

White, H. (2000), "A Reality Check for Data Snooping," *Econometrica*, 69, 1097–1127.