Identifying Term Structure Volatility from the LIBOR-Swap Curve

Samuel Thompson
Arrowstreet Capital

This paper proposes a new family of specification tests and applies them to affine term structure models of the London Interbank Offered Rate (LIBOR)-swap curve. Contrary to Dai and Singleton (2000), the tests show that when standard estimation techniques are used, affine models do a poor job of forecasting volatility at the short end of the term structure. Improving the volatility forecast does not require different models; rather, it requires a different estimation technique. The paper distinguishes between two econometric procedures for identifying volatility. The "cross-sectional" approach backs out volatility from a cross section of bond yields, and the "time-series" approach imputes volatility from time-series variation in yields. For an affine model, the volatility implied by the time-series procedure passes the specification tests, while the cross-sectionally identified volatility does not. This is surprising, since under correct specification, the "cross-sectional" approach is maximum likelihood. One explanation is that affine models are slightly misspecified; another is that bond yields do not span volatility, as in Collin-Dufresne and Goldstein (2002). (*JEL* C52, G13)

1. Introduction

The affine family of term structure models (see Duffie and Kan, 1996) is widely used as a framework for pricing interest rate derivatives and forecasting the term structure. Do affine models describe the dynamics of the LIBOR-swap curve? Using formal tests of overidentifying moment restrictions, Dai and Singleton (2000) concluded that the answer is yes. On the other hand, Piazzesi (2005) argued that affine models could be improved by including a model for the Federal Funds target rate. Collin-Dufresne and Goldstein (2002) show that affine models have counterfactual implications for option-implied volatilities. Neither of these papers carried out formal tests of model specification.

I am indebted to Thomas Rothenberg for his invaluable advice. This paper is based on my dissertation, which he supervised. I thank Donald Andrews, John Campbell, Gary Chamberlain, Kjell Doksum, Chris Foote, Petra Geraats, Angelo Melino, James Powell, Matt Pritsker, Paul Ruud, Neil Shephard, Ken Singleton, Richard Stanton, Jeremy Stein, Jim Stock, Tim Vogelsang, and Tuomo Vuolteenaho for useful comments. I am also grateful for the comments of seminar participants at Cornell University, the Helsinki School of Economics, the Massachusetts Institute of Technology, Northeastern University, Rice University, and Yale University. Advice from Yacine Ait-Sahalia and two anonymous referees improved the exposition of the paper. Many of the calculations were carried out using Mathematica code provided by Yacine Ait-Sahalia and Robert Kimmel, and by Gauss code provided by Yongmiao Hong and Haitao Li. Address correspondence to Samuel Thompson, Arrowstreet Capital, L.P., 44 Brattle Street, 5th Floor, Cambridge, MA 02138; telephone: (617) 547-9999; fax: (617) 547-0099; e-mail: sambthompson@gmail.com.

doi:10.1093/rfs/hhm082 Advance Access publication January 14, 2008

This paper proposes a general family of specification tests and applies them to affine models of the LIBOR-swap curve. The tests are alternatives to the tests of overidentifying moment restrictions proposed by Gallant and Tauchen (1996) and used by Dai and Singleton (2000). The tests support Dai and Singleton's (2000) conclusion that affine models do a good job of modeling the dynamics of longer maturity swaps. However, the tests strongly reject affine models at the short end of the term structure. The rejections are driven by misspecification of the conditional variance of LIBOR. These results are obtained using the same models and dataset used in Dai and Singleton (2000); thus, changing the specification test changes the conclusion.

The paper argues that modeling term structure volatility does not require a different class of models—rather, it requires a different estimation procedure. There are two econometric strategies for identifying volatility. The "cross-sectional" approach backs out volatility from a cross section of contemporaneously observed bond yields. The "time-series" approach imputes volatility from time-series variation in yields. While the cross-sectional approach dominates the empirical literature, the volatility implied by the time-series procedure passes the specification tests, but the cross-sectionally identified volatility does not. This is a surprising result. In a correctly specified model, the cross-sectional approach is the maximum likelihood. The time-series approach is less efficient, since it tries to infer bond volatility indirectly, rather than backing it out exactly from current yields.

There are at least two reasons to prefer the time-series approach. If the model is slightly misspecified, changing the moment conditions used to fit the model may improve the fit to certain aspects of the data. The time-series approach tries to fit recent variation in bond yields, and this seems to be a better predictor of future volatility than the current cross section of yields. Of course, changing the moment conditions means that we fit other aspects of the data less well. For example, the cross-sectional approach leads to slightly smaller pricing errors.

Collin-Dufresne and Goldstein (2002); and Collin-Dufresne, Goldstein, and Jones (2004) provide another reason to prefer the time-series approach: they provide examples of affine models, where volatility does not affect the cross section of bond yields. They label this phenomenon unspanned stochastic volatility (USV). We statistically reject the model restrictions imposed by USV, but this rejection seems to be driven by pricing errors, not an incorrect volatility model. The USV-implied volatility closely follows the volatility implied by the unrestricted affine model, so long as the time-series identification strategy is used.

The statistical tests proposed in this paper are based on a very old idea, dating at least to Segal (1938); and Rosenblatt (1952). Let Y_t denote a univariate bond yield generated from a continuous time model, and let $\{Y_{\tau\Delta}\}_{\tau=0}^{n}$ denote a discrete sample observed over a time span T at interval Δ, with $n = T/\Delta$. The model

implies the conditional distribution function

$$F_\tau(y, \theta_0) \equiv \Pr[Y_{\tau\Delta} \leq y | I_{(\tau-1)\Delta}, \theta_0], \tag{1}$$

where θ_0 is the true value of a parameter vector θ and $I_{(\tau-1)\Delta} = \{Y_0, Y_\Delta, \ldots, Y_{(\tau-1)\Delta}\}$ denotes the information set containing all past data. Under general conditions, the transformed variables $F_\tau(Y_{\tau\Delta}, \theta_0)$ are distributed independently and identically (iid) uniform on the interval [0, 1]. To evaluate the null hypothesis that the data were generated from F_τ, test whether the transforms $F_\tau(Y_{\tau\Delta}, \theta_0)$ are distributed as iid uniforms. While this testing approach is the simplest to explain for univariate data, this paper focuses on a multivariate extension to this idea, and the empirical application considers only multivariate models.

There are two technical challenges when translating this idea into a formal statistical test: calculating the transforms, and adjusting their distribution for parameter estimation. Calculation is difficult, because we do not know the conditional distribution function of many continuous time term structure models. Parameter estimation complicates the null distribution of most test statistics, because the distribution of $F_\tau(Y_{\tau\Delta}, \theta_0)$ is no longer iid uniform when evaluated at an estimate. The tests proposed here solve the parameter estimation problem. The paper also provides a number of algorithms for calculating the transforms. Calculating the transforms is especially difficult when volatility is identified by the time-series strategy, because the cross section of bond yields does not uniquely identify volatility. We approximate the continuous time-volatility model with a finite-state Markov model, which becomes arbitrarily close to the true model as we increase the number of states. The approximation leads to closed-form expressions for the likelihood function.

In contemporaneous work, Hong (1999, 2000); and Hong and Li (2005) have proposed a family of tests based on the transformation to uniforms, which correctly handle parameter estimation. Hong's tests rely on nonparametric kernel estimators, while the tests proposed here do not. Hong has shown formally that his tests are consistent against all alternatives. The tests in the present paper are not consistent, since they are essentially Generalized Method of Moments (GMM) tests, with the sample moments replaced by moment functions. If the moments are poorly chosen, they will not detect some alternative models. Therefore, Hong's tests have the potential to have power against alternatives that the tests in this paper miss. On the other hand, since kernel estimators converge at slower than usual rates, in large samples, the tests proposed here should have better power against some relevant alternatives. We carry out a small Monte Carlo experiment in which the tests in this paper have significantly more power than tests in Hong and Li (2005).

Fournie (1992); Inoue (1999); Bai (2003); and Corradi and Swanson (2001) also propose formal tests that correctly handle parameter estimation. Since their tests are based on the empirical distribution function, they detect departures

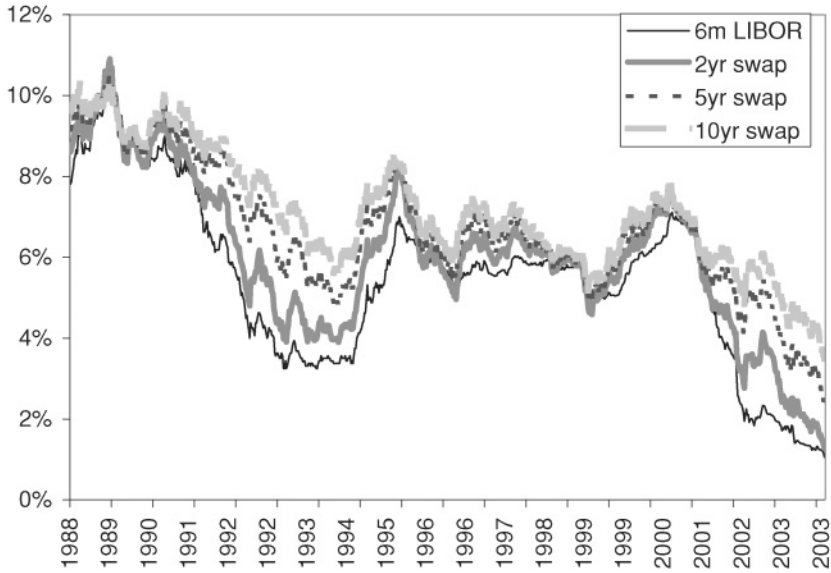

Figure 1
Annualized swap data from 10 June 1988 to 12 June 2003
The series are 6-month LIBOR, and 2-, 5-, and 10-year LIBOR swap yields.

from uniformity, but not from independence. Thus they do not have power against many forms of model failure that arise in time-series analysis. An exception to this is provided by Duan (2003), whose tests detect departures from both uniformity and independence.

2. Specification Tests for Multivariate Term Structure Models

The data $\{Y_{\tau\Delta}\}_{\tau=0}^{n}$ are discrete observations on the vectors $Y_t = (l_t, c_{t,2}, c_{t,10})'$, where l_t is 6-month LIBOR and $c_{t,2}$ and $c_{t,10}$ are the coupon rates for two-year and ten-year fixed-for-variable plain vanilla swaps. Later sections of the paper will add other swap rates to the data, but the main focus will be the two- and ten-year rates. There are $n = 783$ weekly observations over the sample period from June 1988 to June 2003. The data are from Datastream, which obtains LIBOR from the British Bankers Association and swap rates from ICAP Information Services. The data appear in Figure 1.

The London Interbank Offered Rate (LIBOR) is a median of interbank lending rates among approximately 15 large banks. Swaps allow investors to exchange risky payments at 6-month LIBOR for certain payments at the coupon rate. An h-year swap is formed at time t when two investors agree to exchange a certain payment of $c_{t,h}/2$ every six months for the uncertain payments $\exp(l_t/2) - 1, \exp(l_{t+.5}/2) - 1, \exp(l_{t+1}/2) - 1, \ldots, \exp(l_{t+h}/2) - 1$.

The goal of this paper is to find a model that fits both the cross section and time-series variation of the vector Y_t. To this end, I propose a family of specification tests to assess the goodness of fit of various models. Consider two conditional distribution functions for $Y_{i,\tau\Delta}$, the ith yield at time $\tau\Delta$

$$F^P_{i\tau}(z, \theta_0) \equiv \Pr[Y_{i,\tau\Delta} \leq z | I_{(\tau-1)\Delta}, \theta_0], \tag{2}$$

$$F^C_{i\tau}(z, \theta_0) \equiv \Pr[Y_{i,\tau\Delta} \leq z | Y_{1,\tau\Delta}, \ldots, Y_{i-1,\tau\Delta}, I_{(\tau-1)\Delta}, \theta_0]. \tag{3}$$

$F^P_{i\tau}$ conditions on *past* information, while $F^C_{i\tau}$ conditions on both past and *contemporaneous* information. F^C summarizes both the time-series and cross-sectional forecasts of the model, while F^P summarizes only the time-series forecast for a single bond yield. The null and alternative hypotheses are:

$$\begin{cases} H: Y_{i,\tau\Delta} \text{ is drawn from the conditional distribution } F^C_{i\tau}(z, \theta_0) \text{ for all } i \text{ and } \tau. \\ K: Y_{i,\tau\Delta} \text{ comes from a different distribution for some } i \text{ or } \tau. \end{cases}$$

The null hypothesis also implies that $Y_{i,\tau\Delta}$ is drawn from $F^P_{i\tau}(z, \theta_0)$ for all i and τ.

To test the hypothesis, I transform the data. Under the null, the transformed variables $F^C_{i\tau}(Y_{i,\tau\Delta}, \theta_0)$ are uniformly distributed and independent across both i and τ. The transformed variables $F^P_{i\tau}(Y_{i,\tau\Delta}, \theta_0)$ are distributed uniformly and independently across time, but not across yields i. To test both the cross-sectional and time-series implications of the model, we form the transformed variables $F^C_{i\tau}(Y_{i,\tau\Delta}, \widehat{\theta})$ using an estimate $\widehat{\theta}$ and test whether the terms are iid uniform across τ and i. To test only whether the model implies the correct time-series dynamics for each yield, we test whether each $F^P_{i\tau}(Y_{i,\tau\Delta}, \widehat{\theta})$ series is iid uniform across τ. For example, $F^P_{1\tau\Delta}(l_{\tau\Delta}, \widehat{\theta})$ tells us whether the model accurately forecasts LIBOR over time, while $\{F^C_{1\tau}(l_{\tau\Delta}, \widehat{\theta})$, $F^C_{2\tau}(c_{\tau\Delta,2}, \widehat{\theta})$, $F^C_{3\tau}(c_{\tau\Delta,10}, \widehat{\theta})\}$ is informative about a model's fit to the cross section of bond yields.

When constructing F^C, the ordering of the yields is arbitrary. We could set $\{Y_{1t}, Y_{2t}, Y_{3t}\} = \{l_t, c_{t,2}, c_{t,10}\}$, or $\{Y_{1t}, Y_{2t}, Y_{3t}\} = \{c_{t,2}, l_t, c_{t,10}\}$. No matter how we order the data, the conditional distribution functions summarize all implications of the model. However, different orderings lead to different values for the test statistics. In the empirical application, we report results for only one ordering, but in unreported results, we find very similar results for different orderings.

As an example, suppose we wish to test whether l_t comes from the autoregressive model

$$dl_t = K(\mu - l_t)\,dt + v dW_t, \tag{4}$$

where $\theta = \{K, \mu, \sigma\}$ are parameters and W_t is standard Brownian motion. The model implies that $l_{\tau\Delta}$ is conditionally normal with mean $m_{\tau\Delta} = \mu + \exp(-K\Delta)(l_{(\tau-1)\Delta} - \mu)$ and variance $\sigma^2_{\tau\Delta} = v^2(1 - \exp(-2K\Delta))/(2K)$. The transformations are $F^P_{1\tau}(l_{\tau\Delta}, \widehat{\theta}) = \Phi((l_{\tau\Delta} - \widehat{m}_{\tau\Delta})/\widehat{\sigma}_{\tau\Delta})$, where Φ is the cumula-

tive distribution function for a standard normal variable and the hatted variables denote that the mean and variance are evaluated at parameter estimates $\widehat{\theta}$. Since this is a univariate model, the F^P and F^C transforms are the same.

To gain intuition for this approach, think of $F^C_{1\tau}$ as a "generalized residual." A common approach to evaluating the model in Equation (4) is to calculate the regression residuals $\widehat{\varepsilon}_{\tau\Delta} = (l_{\tau\Delta} - \widehat{m}_{\tau\Delta})/\widehat{\sigma_{\tau\Delta}}$ and test whether they are normal and independent. A misspecified mean will lead to autocorrelated residuals. Unmodeled stochastic volatility will create correlations between absolute values of residuals. We can test for normality by evaluating the skewness and kurtosis of the residuals. The testing approach in this paper is quite similar. Instead of examining the residuals $\widehat{\varepsilon}_{\tau\Delta}$, we examine the generalized residuals $\Phi(\widehat{\varepsilon}_{\tau\Delta})$.

There are several reasons to transform to generalized residuals. As Hong and Li (2005) have argued, transforming the data may improve small sample test size. Pritsker (1998) has shown that the persistence exhibited by some continuous time interest rate models can significantly distort the size of a specification test proposed by Ait-Sahalia (1996). Since test statistics based on the transformed data are functions of approximately independent uniform terms, asymptotic approximations to the null may be accurate even when the underlying model is persistent.

Transforming the data may also prove useful when the null model is complicated. One common approach to testing term structure models makes use of tests of overidentifying moment restrictions. Many term structure models imply complicated, nonlinear dynamics, making it difficult to generate appropriate moment conditions. Consequently, authors such as Dai and Singleton (2000) use moment conditions generated by the overidentified semi-nonparametric model of Gallant and Tauchen (1996). Transforming to uniforms is an alternative way to generate a series of testable implications under the null. For example, suppose we evaluate a model for stochastic volatility

$$dl_t = k(\mu - l_t)dt + \sqrt{V_t}dW_{l,t} \tag{5}$$

$$dV_t = k_v(\mu_v - V_t)dt + \sigma_v\sqrt{V_t}dW_{v,t}. \tag{6}$$

$(W_{l,t}, W_{v,t})$ are correlated Brownian motions. l_t is observed, but the instantaneous variance V_t is not observed. How do we test this model? One sensible approach would be to form an estimate of the residual: $\varepsilon_{\tau\Delta} = \mathrm{Var}(l_{\tau\Delta}|I_{(\tau-1)\Delta})^{-1/2}(l_t - E(l_{\tau\Delta}|I_{(\tau-1)\Delta}))$, where $\mathrm{Var}(l_{\tau\Delta}|I_{(\tau-1)\Delta})$ is the volatility forecast conditional on the past observed data $\{l_{(\tau-1)\Delta}, l_{(\tau-2)\Delta}, \ldots, l_\Delta\}$, but not conditional on $V_{\tau\Delta}$, which is not observed. These residuals will have zero means and unit variances, but are not independent or normal. If the model fails because the random innovations contain jump processes rather than only Brownian motions, it is not obvious how this will affect the distribution of $\varepsilon_{\tau\Delta}$.

The cumulative distribution is

$$F^P_{1\tau}(z,\theta) = \int \Pr[l_{\tau\Delta} \le z | l_{(\tau-1)\Delta}, V_{(\tau-1)\Delta} = v] dP(v|I_{(\tau-1)\Delta}), \tag{7}$$

where $P(v|I_{(\tau-1)\Delta})$ is the distribution of $V_{(\tau-1)\Delta}$ conditional on the past observed data. Despite the complexity of this model, the generalized residual $F^P_{1\tau}(l_{\tau\Delta},\theta)$ has an independent and uniform distribution under the null of correct specification. The empirical application describes how to compute conditional distribution functions for similar models with unobserved stochastic volatility.

2.1 Test statistics

In this section, we provide test statistics to evaluate the uniformity and independence of the generalized residuals. The tests reject the null of correct specification for large values of

$$\mathcal{Q} = n \int_0^1 \mathbf{M}(s)' \mathbf{W}^{-1} \mathbf{M}(s)\, ds. \tag{8}$$

$\mathbf{M}(s) = (M_1(s), \ldots, M_L(s))'$ is an L-dimensional vector of moment functions that have zero expectation under the null. $\mathbf{W}$ is an $L \times L$ weight matrix that does not depend on s. To test the implications of the model for a single yield $Y_{i,\tau\Delta}$, we use moment functions that depend on the F^P residuals. In this case, all of the moment functions considered are subsets of the vector

$$\mathbf{M}(s) = \left(\frac{1}{n} \sum_{\tau=1}^{n} \left[1\left(F^P_{i\tau} \le s\right) - s \right],\ \sum_{j=1}^{n-1} \phi_j(s) \widehat{\rho}_{j,2}, \ldots, \sum_{j=1}^{n-1} \phi_j(s) \widehat{\rho}_{j,L} \right). \tag{9}$$

$\phi_j(s) = (j\pi)^{-1}\sqrt{2}\sin(j\pi s)$, and $\widehat{\rho}_{j,l}$ is the jth autocorrelation between functions of the residuals $g_l(F^P_{i\tau})$, where g_l is a function chosen to detect interesting dependencies. Specifically, $\widehat{\rho}_{j,l} = \widehat{\gamma}_{j,l}/\widehat{\gamma}_{j,0}$, with $\widehat{\gamma}_{j,l} = n^{-1}\sum_\tau (g_l(F^P_{i\tau}) - \overline{g}_l) g_l(F^P_{i,\tau-j})$, and $\overline{g}_l = n^{-1}\sum_\tau g_l(F^P_{i\tau})$. The test based on the $F^C_{i\tau}$ residuals takes exactly the same form, with the moment functions $\mathbf{M}(s)$ constructed from the ordered series $\{\{F^C_{i\tau}\}_{i=1}^3\}_{\tau=1}^n$.[1]

$\mathbf{W}$ is the asymptotic correlation matrix between elements of $\mathbf{M}(s)$. Since $\mathbf{W}$ is a correlation matrix (not a covariance matrix), the diagonals are all equal to 1. The $n^{-1/2}\sum[1(F_{i\tau} \le s) - s]$ function is independent of all the others, so $\mathbf{W}_{1k} = \mathbf{W}_{k1} = 0$ for all $k > 1$. The $\mathbf{W}_{ij}$ are equal to $(Corr[g_i(U), g_j(U)])^2$ for $i, j > 1$, where U is a uniform random variable. It is straightforward to show

[1] To be explicit, given i and τ, let $k = i + 3(\tau - 1)$, and define the alternative subscript notation $F^C_{[k]} \equiv F^C_{i\tau}$. The moment functions take the form

$$M_1(s) = \frac{1}{3n}\sum_{k=1}^{3n}\left[1\left(F^C_{[k]} \le s\right) - s\right], \text{ and } M_l(s) = \sum_{j=1}^{3n-1} \phi_j(s) \frac{\widehat{\gamma}_{j,l}}{\widehat{\gamma}_{j,0}} \text{ for } l > 1.$$

Here $\widehat{\gamma}_{j,l} = (3n)^{-1}\sum_{k=1}^{3n}(g_l(F^C_{[k]}) - \overline{g}^C_l) g_l(F^C_{[k-j]})$ and $\overline{g}_l = (3n)^{-1}\sum_{k=1}^{3n} g_l(F^C_{[k]})$.

that the **W** matrix does not depend on s. Notice that **W** is known and does not need to be estimated.

Specific form of the tests used in this paper. In the empirical application, we chose $L = 3$, with $g_2(x) = x$ and $g_3(x) = 2|x - .5|$. It is straightforward to show that for these choices of g, **W** is the identity matrix. Therefore, $\mathcal{Q} = n \int [M_1^2(s) + M_2^2(s) + M_3^2(s)]\, ds$. We also calculate results for the test statistics $\mathcal{T}_i = \int M_i^2(s)\, ds$ for $i = 1, 2, 3$. The $\mathcal{T}_i$ tests are special cases of the $\mathcal{Q}$ test. We motivate these choices below.

2.2 Interpreting the test statistics

A standard GMM test statistic is a quadratic form in a finite number of moments. The $\mathcal{Q}$ test is an integral over a quadratic form in *moment functions*. The reason we use moment functions is that the implications of the null are potentially infinite. To clarify this point, consider a test of uniformity based on the histogram. If the histogram has 25 bins, then under the null, we expect to see about 4% of the observations in each bin. This suggests a standard GMM test based on the 25 moment conditions $n^{-1} \sum_\tau 1(F_{i\tau}^P \leq j/25) - .04\, j$ for $j = 1, \ldots, 25$. Of course, the choice of the number of bins is arbitrary. Furthermore, if we only test using these 25 bins, we may miss departures from uniformity. So, we let the number of bins become very large, to arrive at the centered, normalized empirical distribution function

$$M_1(s) = n^{-1/2} \sum_\tau \left[1\left(F_{i\tau}^P \leq s\right) - s\right]. \tag{10}$$

Under the null hypothesis, $E\,1(F_{i\tau}^P \leq s) = s$ for all $s \in [0, 1]$. Thus we have a moment function $M_1(s)$, which implies an infinite number of moment conditions. We reject the null for large values of

$$\mathcal{T}_1 = \int_0^1 M_1^2(s)\, ds. \tag{11}$$

The $\mathcal{T}_1$ statistic, which is a special case of $\mathcal{Q}$, detects departures from uniformity. It is well known that if the $F_{i\tau}^P$ are iid but not uniform, the empirical distribution function will converge to the true distribution function. Thus M_1 will detect any departure from uniformity if the $F_{i\tau}^P$ are iid.

Next, consider the moment function $M_l(s) = n^{1/2} \sum_{j=1}^{n-1} \phi_j(s) \widehat{\rho}_{j,l}$, where $\widehat{\rho}_{j,l}$ is the correlation between a function g_l of the residuals. The M_l function clearly detects dependence between residuals. Why do we test for dependence using this function? A number of authors have assessed dependence between generalized residuals by plotting correlograms of functions of the residuals (e.g., Diebold, Gunther, and Tay, 1998; Kim, Shephard, and Chib, 1998; and Elerian, Chib, and Shephard, 2001). For the most part, the analysis has been graphical and has not relied on formal tests. The M_l function formalizes the graphical analysis; the function follows an old and well-studied approach to testing the statistical significance of sample correlations. Consider the sample

periodogram

$$I(\omega) = \frac{1}{2\pi}\left\{\widehat{\gamma}_0 + 2\sum_{j=1}^{n-1}\widehat{\gamma}_{j,l}\cos(\omega j)\right\}. \tag{12}$$

The sample periodogram and the estimated covariances contain the same information about the process— $\widehat{\gamma}_{j,l}$ can be calculated from $I(\omega)$ and vice versa. Thus all testable implications of uncorrelatedness can be evaluated by examining the shape of $I(\omega)$. Under the null, the correlations $\widehat{\gamma}_{j,l}$ are zero and the spectrum should not vary with ω. Following Durbin (1969); and Durlauf (1991), we base the test of uncorrelatedness on the normalized cumulative periodogram

$$M_l(s) = \sqrt{\frac{n}{2}}\left[\frac{\int_0^s I(\pi w)\,dw}{\int_0^1 I(\pi w)\,dw} - s\right] = n^{1/2}\sum_{j=1}^{n-1}\phi_j(s)\widehat{\rho}_{j,l}. \tag{13}$$

To test for dependence, we choose the g_l functions $g_2(x) = x$ and $g_3(x) = 2|x - .5|$. The g_2 function detects correlated residuals. The g_3 function follows Kim, Shephard, and Chib (1998), who point out that g_3 is useful for detecting types of dependence that will not show up as correlations between residuals. For example, the residuals could have a time-varying, predictable variance and still be serially uncorrelated. The predictable variance will show up as correlations between g_3 terms. The test statistics based on g_2 and g_3 are

$$\mathcal{T}_2 = \int_0^1 M_2^2(s)\,ds \qquad \text{and} \qquad \mathcal{T}_3 = \int_0^1 M_3^2(s)\,ds. \tag{14}$$

The $\mathcal{T}_2$ statistic detects correlated residuals, and the $\mathcal{T}_3$ statistic detects correlations between absolute values of centered residuals. Both statistics are special cases of the $\mathcal{Q}$ statistic. By choosing other g_l functions, we can incorporate most of the informal procedures proposed by various authors. For example, our testing approach incorporates Diebold, Gunther, and Tay's (1998) idea of looking at correlations between powers of $F_{i\tau}^P - .5$.

2.3 Critical values

If the true parameters θ_0 did not have to be estimated, the generalized residuals would be exactly iid uniform under the null. Thus, with known parameters, the critical values for each test statistic can be tabulated by simulation for any sample size. In practice, the parameters θ must be estimated. The choice of estimator affects the null distribution of the test statistics, even asymptotically. Various methods are available for calculating the critical value. One approach is the parametric bootstrap: simulate many datasets from the model evaluated at the estimate for θ. For each set of simulated data, the parameters are reestimated and the test statistics are calculated, resulting in a sample of bootstrap test statistics. The quantiles are approximated by the empirical quantiles of the bootstrap

sample. While the parametric bootstrap will probably lead to accurate critical values, in many situations, it is computationally very intensive. In particular, estimating the parameters of a continuous time model can be computationally expensive.

Various alternatives to the bootstrap are available. Bai (2003) proposed a transformation of the empirical cdf that rids the limiting null distribution of the effects of parameter estimation. While this approach has the potential to simplify our testing problem, it is only directly applicable to the test based on the empirical cdf. The extension to tests of independence is not obvious. Our approach is similar to Corradi and Swanson (2001), who showed how to simulate from the null distribution of the empirical cdf.

We calculate critical values by simulating from the asymptotic null of $\mathcal{Q}$, given in the following theorem. A discussion of the assumptions appears in the Appendix.

Theorem 1. *Suppose that $\mathcal{Q}$ is evaluated at the estimates $\widehat{\theta}$. Under the null, and under assumptions (1)–(5), $\mathcal{Q} \Rightarrow \sum_{j=1}^{\infty} \lambda_j^2 \chi_j^2$, where $\chi_1, \chi_2, \ldots, \chi_{LJ}$ are a series of independent standard normal variables, and $\lambda_1, \ldots, \lambda_{LJ}$ are eigenvalues of an $LJ \times LJ$ matrix $\mathbf{\Omega}$, discussed below.*

Recall that each $\mathcal{T}_l$ statistic is a special case of the $\mathcal{Q}$ statistic, so this theorem applies to all the tests considered here. L is the number of moment functions in $\mathbf{M}$, so $L = 1$ for the $\mathcal{T}_1, \mathcal{T}_2$, and $\mathcal{T}_3$ statistics, and $L = 3$ for $\mathcal{Q} = \mathcal{T}_1 + \mathcal{T}_2 + \mathcal{T}_3$. I obtain asymptotically exact quantiles of the null distribution of $\mathcal{Q}$ by simulating from the truncated sum $\sum_{j=1}^{LJ} \widehat{\lambda}_j^2 \chi_j^2$, where $\widehat{\lambda}_j$ are eigenvalues of a consistent estimate $\widehat{\mathbf{\Omega}}$ for $\mathbf{\Omega}$. In the proof of the theorem, I show that for large enough J the truncated sum is arbitrarily close to the infinite sum. Because of the simple form of the null, critical values can be obtained almost instantly. For all of the Monte Carlos and empirical applications in this paper, I choose J equal to 25 with 10,000 simulations. I have experimented with the various statistics and found that choosing $J \geq 25$ leads to a close approximation to the null.

Since all of the estimators in this paper are asymptotically efficient, I provide here a formula for $\mathbf{\Omega}$ that is valid only with an asymptotically efficient estimator.[2] The general version of $\mathbf{\Omega}$, which is valid for an inefficient GMM estimator, is given in the Appendix. If $\widehat{\theta}$ is efficient, we can consistently estimate $\mathbf{\Omega}$ with

$$\widehat{\mathbf{\Omega}} = (\mathbf{J} \otimes \mathbf{R})[\mathbf{I}_J \otimes \mathbf{W} - \widehat{\mathbf{C}}\widehat{\mathbf{H}}^{-1}\widehat{\mathbf{C}}'](\mathbf{J} \otimes \mathbf{R})'. \tag{15}$$

[2] By asymptotically efficient, I mean that $\sqrt{n}(\widehat{\theta} - \theta_0)$ converges in distribution to a normal vector with zero mean and variance equal to the inverse of the information matrix

$$\lim_{n\to\infty} \mathrm{Var}[\sqrt{n}(\widehat{\theta} - \theta_0)] = \left[-\lim_{n\to\infty} n^{-1} \sum_{\tau} E \frac{\partial^2 \log f_\tau(Y_{\tau\Delta}, \theta_0)}{\partial\theta\partial\theta'} \right].$$

Here, $f_\tau(x, \theta)$ is the density of $Y_{\tau\Delta}$, conditional on all past observations.

$\otimes$ indicates the Kronecker product, $\mathbf{R}'\mathbf{R} = \mathbf{W}^{-1}$, and $\mathbf{J}$ is the $J \times J$ diagonal matrix with jth diagonal equal to $(j\pi)^{-1}$. $\widehat{\mathbf{H}} = n^{-1}\sum_{\tau}(\mathbf{S}_\tau - \overline{\mathbf{S}})\mathbf{S}'_\tau$, where $\mathbf{S}_\tau$ denotes the p-dimensional score vector $\partial \log f_\tau(Y_{\tau\Delta}, \theta)/\partial\theta$, where f_τ is the density of $Y_{\tau\Delta}$ conditional on all past observations. $\widehat{\mathbf{C}}$ is an $LJ \times p$ matrix of terms $(\widehat{\mathbf{C}}'_1, \ldots, \widehat{\mathbf{C}}'_J)'$, where $\widehat{\mathbf{C}}_j$ is the $L \times p$ matrix

$$\widehat{\mathbf{C}}_j = n^{-1}\sum_{\tau}\widehat{\mathbf{G}}_{j,\tau}(\mathbf{S}_\tau - \overline{\mathbf{S}})'. \tag{16}$$

$\widehat{\mathbf{G}}_{j,\tau}$ has the same dimensionality as the vector of test moments $\mathbf{M}(s)$. For the $\mathcal{T}_1$ test, $\widehat{\mathbf{G}}_{j,\tau}$ is the one-dimensional $\widehat{\mathbf{G}}^1_{j,\tau} = \sqrt{2}\cos(j\pi F^P_{i\tau})$. For any $\mathcal{T}_l$ test with $l > 1$, $\widehat{\mathbf{G}}_{j,\tau}$ is the one-dimensional $\widehat{\mathbf{G}}^l_{j,\tau} = (g_l(F^P_{i\tau}) - \overline{g}_l)g_l(F^P_{i,\tau-j})/\widehat{\gamma}_{0,l}$. For the full $\mathcal{Q}$ test, $\widehat{\mathbf{G}}_{j,\tau}$ is the L-dimensional vector $(\widehat{\mathbf{G}}^1_{j,\tau}, \ldots, \widehat{\mathbf{G}}^L_{j,\tau})'$. These formulas for $\widehat{\mathbf{G}}_{j,\tau}$ apply to tests based on the transforms F^P. The formulas for the F^C-based tests take a very similar form.[3]

We should emphasize that this formula for $\widehat{\boldsymbol{\Omega}}$ is valid only if the estimator is asymptotically efficient. When the estimator arises from an inefficient GMM procedure, the more general formula given in the Appendix must be used. Asymptotically efficient estimators include the maximum likelihood estimator and the approximate maximum likelihood estimators of Lo (1988); and Ait-Sahalia (2002, 2007). The simulation-based approximate maximum likelihood estimators of Pedersen (1995); Brandt and Santa-Clara (2002); and Durham and Gallant (2002) are also available. The formula also covers the asymptotically efficient moment-based estimators of Gourieroux, Monfort, and Renault (1993); and Gallant and Tauchen (1996). Since Bayesian estimators are asymptotically efficient so long as the prior puts nonzero mass around the true parameter values, the Bayesian estimators of Eraker (2001); Elerian, Chib, and Shephard (2001); and Jones (2003b) may also be used.

3. Affine Factor Models

We apply the tests to the affine factor models of Duffin and Kan (1996); and Duffee (2002). Under the affine models, the observations Y_t follow

$$Y_t = \Upsilon(X_t, \theta), \qquad dX_t = [(K\mu) - KX_t]\,dt + \Sigma S_t dW_t. \tag{17}$$

Υ is a function that maps the observed yields Y_t to the unobserved factor vector X_t. The factors follow an affine diffusion, where W_t is a vector of independent

[3] Continuing the discussion from Footnote 1, we have

$$\widehat{\mathbf{C}}_j = (3n)^{-1}\sum_k E\widehat{\mathbf{G}}_{j,[k]}(\mathbf{S}_{[k]} - \overline{\mathbf{S}})',$$

where $\mathbf{S}_{[k]}$ is equal to $\mathbf{S}_{1+(k-i)/m}$, the score for the observation τ associated with index $[k]$.

standard Brownian motions, K and Σ are matrics, $(K\mu)$ is a vector, and S_t is a diagonal matrix with jth diagonal element $(\alpha_j + \beta_j' X_t)^{1/2}$.

We write the pricing function Υ in terms of bond prices. Let $P_{t,k}$ denote the time-t price of a zero coupon bond that pays 1 dollar in k periods. Following Duffie and Singleton (1997, 1999), if we equate the present value of payments on either side of a swap contract and interpret the interest rate as a default-adjusted discount rate, then swap rates can be expressed as functions of bond prices:

$$c_{t,k} = 2\frac{1 - P_{t,k}}{\sum_{j=1}^{2k} P_{t,.5j}}.$$

Thus we can write the pricing function Υ in terms of zero coupon bond prices:

$$\Upsilon(X_t, \theta) \equiv \begin{pmatrix} -2\log P_{t,.5} \\ 2(1 - P_{t,2}) \Big/ \sum_{j=1}^{4} P_{t,.5j} \\ 2(1 - P_{t,10}) \Big/ \sum_{j=1}^{20} P_{t,.5j} \end{pmatrix}. \tag{18}$$

$P_{t,k}$ is the time-t expected discounted value of 1 dollar paid at time $t + k$, given by

$$P_{t,k} \equiv E^Q \left[\exp\left(- \int_t^{t+k} r_s \, ds \right) \middle| X_t \right]. \tag{19}$$

$\exp(-\int_t^{t+k} r_s \, ds)$ discounts the time-$(t + k)$ payment to the present, and E^Q denotes that expectations are taken with respect to the risk-neutral pricing measure. The instantaneous interest rate r_t is a linear function of the factors

$$r_t = \delta_0 + \delta_1' X_t, \tag{20}$$

where δ_1 is a vector. Under the risk-neutral expectations operator E^Q, the factors follow the diffusion

$$\begin{aligned} dX_t &= [(K\mu)_Q - K_Q X_t]dt + \Sigma S_t d\widetilde{W}_t \\ &= [(K\mu) - K X_t - \Sigma S_t \Lambda_t]dt + \Sigma S_t d\widetilde{W}_t. \end{aligned} \tag{21}$$

Λ_t is a vector of real numbers, and $\widetilde{W}_t$ is a vector of Brownian motions. Each element of Λ_t denotes the market price of risk of the corresponding element of W_t. Λ_t takes the form

$$\Lambda_t = S_t \lambda_1 + S_t^{-} \lambda_2 X_t. \tag{22}$$

λ_1 is a vector, λ_2 is a matrix, and S_t^- is a diagonal matrix with jth diagonal element

$$S_{t(jj)}^- = \begin{cases} (\alpha_j + \beta_j' X_t)^{-1/2}, & \text{if } \inf(\alpha_j + \beta_j' X_t) > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{23}$$

These assumptions imply a convenient expression for zero coupon bond prices

$$P_{t,k} = \exp[A(k) - B(k)X_t]. \tag{24}$$

$(A(k), B(k))'$ is a vector that solves a pair of ordinary differential equations given in Dai and Singleton (2000). Since the equations admit convenient numerical solutions, the model provides an essentially closed-form expression for bond prices.

4. Identifying Term Structure Volatility

In this section, we discuss the estimation of the affine factor models, along with the calculation of the various test statistics. To focus the discussion, consider the $A_1(4)$ model, a four-factor affine model, where the conditional volatility depends on the single factor X_{1t}:

$$dX_t = \begin{pmatrix} K_{11} & 0 & 0 & 0 \\ K_{21} & K_{22} & K_{23} & K_{24} \\ K_{31} & K_{32} & K_{32} & K_{33} \\ K_{41} & K_{42} & K_{43} & K_{44} \end{pmatrix} \begin{pmatrix} \mu_1 - X_{1t} \\ -X_{2t} \\ -X_{3t} \\ -X_{4t} \end{pmatrix} dt + S_t dW_t, \tag{25}$$

with $S_{t(11)} = \sqrt{\beta_{11} X_{1t}}$, $S_{t(22)} = \sqrt{\alpha_2 + \beta_{21} X_{1t}}$, $S_{t(33)} = \sqrt{\alpha_3 + \beta_{31} X_{1t}}$, and $S_{t(44)} = \sqrt{\alpha_4 + \beta_{41} X_{1t}}$. The first row of λ_2 is zero, and the rest are unrestricted. $\delta_1 = (\delta_{11}, 1, 1, 1)'$, where δ_{11} is estimated to be 1 or -1. This is a "maximal" $A_1(4)$ model in the sense of Dai and Singleton (2000), meaning that it contains the most parameters that can be identified from a panel of zero coupon bond prices. How do we infer X_{1t}, the volatility factor? Consider two strategies: the cross-sectional identification strategy and time-series identification strategy.

Cross-sectional identification strategy: Use four yields to invert the pricing formula Υ to solve for the four factors, X_t. In the application below, we use 6-month LIBOR l_t, along with the 2-, 5-, and 10-year swap rates, $c_{t,2}$, $c_{t,5}$, $c_{t,10}$. This recovers the volatility factor X_{1t} exactly.

Time-series identification strategy: Do not use the 5-year swap rate $c_{t,5}$ to infer X_{1t}. Instead, use only the three yields l_t, $c_{t,2}$, and $c_{t,10}$, and use a filtering algorithm (see Jaquier, Polson, and Rossi, 1994; and Kim, Shephard, Chib, 1998) to infer the unknown distribution of the factors. We infer that volatility is high when l_t, c, and $c_{t,10}$ show an abnormal amount of time-series variation.

When the model is correctly specified, the cross-sectional strategy is preferred to the time-series strategy. The cross section of bond prices identifies

volatility exactly, while the time-series strategy recovers a distribution. Most papers take the cross-sectional approach. For example, Chen and Scott (1993); Pearson and Sun (1994); and Duffie and Singleton (1997) use cross-sectional identification, while Collin-Dufresne, Goldstein, and Jones (2004); and Bates (2006) seem to be the only papers that use time-series identification for continuous time bond models.

Suppose, however, that the model is misspecified, or that yields are measured with error. The model will imply an incorrect mapping from the cross section of yields to the time series of volatility. In this case, it is not obvious which identification strategy will be superior. It may be the case that, under misspecification, bond volatility is better predicted by recent yield variation than by the cross-sectionally implied volatility factor. In this case, the time-series strategy could be superior.

Furthermore, Collin-Dufresne and Goldstein (2002) argued that volatility may not be identified from the cross section of bond prices. They derived parameter restrictions for an $A_1(3)$ model, where the volatility factor does not affect bond prices. They label this phenomenon "unspanned stochastic volatility," since volatility is unidentified in the cross section. In a model with unspanned stochastic volatility, the volatility factor affects the variance of yields, but not their level. In terms of the pricing formula,

$$P_{t,k} = \exp[A(k) - B_1(k)X_{1t} - B_2(k)X_{2t} - B_3(k)X_{3t} - B_4(k)X_{4t}], \quad (26)$$

where volatility is unspanned when $B_1(k) = 0$. In this case, only the time-series strategy is available.

Thus we have two approaches for calculating the generalized residuals. The distribution function for $Y_{i,\tau\Delta}$ implied by the cross-sectional strategy is

$$\text{cross-sectional identification: } F^P_{i\tau}(z) = \Pr[Y_{i,\tau\Delta} \leq z | X_{(\tau-1)\Delta}]. \quad (27)$$

This holds because the vector of yields $Y_{(\tau-1)\Delta}$ uniquely identifies the lagged factor $X_{(\tau-1)\Delta}$, and X_t follows a Markov process. The time-series strategy does not uniquely identify the lagged factors, so the distribution function depends on an integral over the unknown factor:

$$\begin{aligned}\text{time-series identification: } F^P_{i\tau}(z) &= \Pr[Y_{i,\tau\Delta} \leq z | I_{(\tau-1)\Delta}] \\ &= \int \Pr[Y_{i,\tau\Delta} \leq z | X_{(\tau-1)\Delta}] dP(X_{(\tau-1)\Delta} | I_{(\tau-1)\Delta}), \quad (28)\end{aligned}$$

where $P(X_{(\tau-1)\Delta}|I_{(\tau-1)\Delta}, \theta)$ is the distribution function for $X_{(\tau-1)\Delta}$ conditional on $I_{(\tau-1)\Delta}$, the information set containing all past observations.

A potential criticism of the time-series strategy is that it will lead to larger cross-sectional pricing errors than the cross-sectional strategy. The cross-sectional strategy will fit four yields exactly, while the time-series strategy fits just three yields, then has a free factor to fit the time-series dynamics. This

leads us to use the same set of yields when applying either estimation strategy. We take the yields $Y_{\tau\Delta} = (l_{\tau\Delta}, c_{\tau\Delta,2}, c_{\tau\Delta,10})'$ to be observed without error, and the additional yields $\widetilde{Y}_{\tau\Delta} = (c_{\tau\Delta,3}, c_{\tau\Delta,4}, c_{\tau\Delta,5}, c_{\tau\Delta,7})'$ to be observed with pricing error ε_τ

$$Y_{\tau\Delta} = \Upsilon(X_{\tau\Delta}, \theta), \qquad \widetilde{Y}_{\tau\Delta} = \widetilde{\Upsilon}(X_{\tau\Delta}, \theta) + \varepsilon_\tau. \tag{29}$$

The pricing errors are independent across assets and across time, and have a $N(0, \sigma^2)$ distribution. By choosing a low value of σ^2, we penalize the time-series strategy for cross-sectional pricing errors. We use the 3, 4, 5, and 7-year swap rates, because they are available from Datastream for our entire sample period.

In the empirical application, we test both three-factor and four-factor affine models. We use the following identification schemes:

1. We test several three-factor models using the cross-sectional identification strategy only. We take the three yields $Y_t = (l_t, c_{t,2}, c_{t,10})'$ to be observed without error, and do not use any additional yields. This is the same approach taken by Dai and Singleton (2000), who evaluate affine models using tests of overidentifying restrictions. We want to mimic their application to demonstrate that the tests proposed here lead to different results.
2. We test the four-factor model in Equation (25) by the time-series strategy. We take $Y_t = (l_t, c_{t,2}, c_{t,10})'$ observed without error and $\widetilde{Y}_t = (c_{t,3}, c_{t,4}, c_{t,5}, c_{t,7})'$ observed with error.
3. We test the four-factor model in Equation (25) by the cross-sectional strategy. We take $Y_t = (l_t, c_{t,2}, c_{t,5}, c_{t,10})'$ observed without error and $\widetilde{Y}_t = (c_{t,3}, c_{t,4}, c_{t,7})'$ observed with error.
4. We test the four-factor USV model of Collin-Dufresne, Goldstein, and Jones (2004). This is a special case of the general four-factor model, with parameter restrictions[4] imposed, so that the pricing term $B_1(k)$ is equal to zero. We estimate the model by the time-series strategy, taking $Y_t = (l_t, c_{t,2}, c_{t,10})'$ observed without error and $\widetilde{Y}_t = (c_{t,3}, c_{t,4}, c_{t,5}, c_{t,7})'$ observed with error.

4.1 Inference for cross-sectional models

In this subsection, we describe how we estimate and test the three-factor models. There are many nonnested models in the affine class. We evaluate the $A_0(3)$,

[4] The USV model is

$$K_Q = \begin{pmatrix} K_{11} & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ -1 & 0 & 0 & -1 \\ -3\Sigma_{32} & K_{Q,42} & K_{Q,43} & K_{Q,44} \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \Sigma_{11} & 0 & 0 & 0 \\ \Sigma_{21} & 1 & \Sigma_{23} & \Sigma_{24} \\ \Sigma_{23}\Sigma_{21} & \Sigma_{23} & \Sigma_{33} & \Sigma_{34} \\ \Sigma_{23}^2\Sigma_{21} & \Sigma_{23}^2 & \Sigma_{43} & \Sigma_{44} \end{pmatrix}$$

with $K_{Q,42} = 2\Sigma_{32}^2(3\Sigma_{32} + K_{Q,44})$, $K_{Q,43} = -(7\Sigma_{32}^2 + 3\Sigma_{32}K_{Q,44})$. Other parameters are $\theta_Q = (\theta_{Q1}, 0, 0, 0)'$, and $\delta_1 = (0, 1, 0, 0)'$, with $S_{t(11)} = \sqrt{X_{1t}}$, $S_{t(22)} = \sqrt{\alpha_2 + X_{1t}}$, $S_{t(33)} = 1$, and $S_{t(44)} = 1$. $K_{12} = K_{13} = K_{14} = 0$, and the remaining elements of K are unrestricted. $\theta_1 = \theta_{Q1}$, and the remaining elements of θ are unrestricted.

$A_1(3)$, and $A_2(3)$ models, which are described in Dai and Singleton (2000); and Duffee (2002). The $A_0(3)$ model is

$$dX_t = \begin{pmatrix} K_{11} & 0 & 0 \\ K_{21} & K_{22} & 0 \\ K_{31} & K_{32} & K_{33} \end{pmatrix} \begin{pmatrix} -X_{1t} \\ -X_{2t} \\ -X_{3t} \end{pmatrix} dt + dW_t, \tag{30}$$

with λ_2 unrestricted. The $A_1(3)$ is

$$dX_t = \begin{pmatrix} K_{11} & 0 & 0 \\ K_{21} & K_{22} & K_{23} \\ K_{31} & K_{32} & K_{33} \end{pmatrix} \begin{pmatrix} \mu_1 - X_{1t} \\ -X_{2t} \\ -X_{3t} \end{pmatrix} dt + S_t dW_t, \tag{31}$$

with $S_{t(11)} = \sqrt{X_{1t}}$, $S_{t(22)} = \sqrt{1 + \beta_{21} X_{1t}}$, and $S_{t(33)} = \sqrt{1 + \beta_{31} X_{1t}}$. The first row of λ_2 is zero, and the rest are unrestricted. The $A_2(3)$ is

$$dX_t = \begin{pmatrix} K_{11} & K_{12} & 0 \\ K_{21} & K_{22} & 0 \\ K_{31} & K_{32} & K_{33} \end{pmatrix} \begin{pmatrix} \mu_1 - X_{1t} \\ \mu_2 - X_{2t} \\ -X_{3t} \end{pmatrix} dt + S_t dW_t, \tag{32}$$

with $S_{t(11)} = \sqrt{X_{1t}}$, $S_{t(22)} = \sqrt{X_{2t}}$, and $S_{t(33)} = \sqrt{1 + \beta_{31} X_{1t} + \beta_{32} X_{2t}}$. The first two rows of λ_2 are zero, and the third is unrestricted. These are the "canonical form" versions of the affine models. Dai and Singleton (2000) tested the "completely affine" version of each model, which imposes the restriction $\lambda_2 = 0$. Duffee (2002) suggested the "essentially affine" extension, which allows for nonzero λ_2. We test both fully and essentially affine models in this paper.

We estimate the three-factor models by maximum likelihood. Since the models are Markovian, the log-likelihood function is $\sum_{\tau=2}^{n} \log p(Y_{\tau\Delta} | Y_{(\tau-1)\Delta}, \theta)$, where p is the density of observation $Y_{\tau\Delta}$ conditional on $Y_{(\tau-1)\Delta}$. All inference in this paper is carried out conditional on the first observation, $Y_{1\Delta}$. The conditional density function is

$$p(Y_{\tau\Delta} | Y_{(\tau-1)\Delta}, \theta) = \det[J(x(Y_{\tau\Delta}, \theta), \theta)]^{-1} p_x(x(Y_{\tau\Delta}, \theta) | x(Y_{(\tau-1)\Delta}, \theta), \theta), \tag{33}$$

where $J(x, \theta)$ is the Jacobian matrix $\partial \Upsilon(x, \theta)/\partial x'$, and p_x is the density of $X_{\tau\Delta}$ given $X_{(\tau-1)\Delta}$. $x(Y, \theta)$ are the implied factors that solve $Y = \Upsilon(x(Y, \theta), \theta)$. $x(Y, \theta)$ is calculated numerically using the Newton-Raphson algorithm. To calculate p_x, I use the closed-form approximations developed in Ait-Sahalia and Kimmel (2002).[5] J may be calculated in closed form.

We calculate the generalized residuals with a simulation algorithm. For each pair of implied factors $x(Y_{\tau\Delta}, \theta)$ and $x(Y_{(\tau-1)\Delta}, \theta)$, break the time interval between them into M equally spaced steps of length $h = \Delta/M$. Simulate a

[5] I am grateful to Yacine Ait-Sahalia and Robert Kimmel for providing me with Mathematica code for carrying out these calculations.

multivariate Euler discretization to the continuous time process

$$\widehat{x}_{t+h} = \widehat{x}_t + K\,(\mu - \widehat{x}_t)\,h + \Sigma S_t \epsilon_{t+h}\sqrt{h}. \tag{34}$$

$\epsilon_{(\tau-1)\Delta}, \ldots, \epsilon_{\tau\Delta-h}, \epsilon_{\tau\Delta}$ are a series of iid standard normal vectors. The matrix S_t is diagonal with jth diagonal $\sqrt{\alpha_j + \beta_j'\widehat{x}_t}$. The initial condition is $\widehat{x}_{(\tau-1)\Delta} = x(Y_{(\tau-1)\Delta}, \theta)$. I carry out B simulations to obtain the draws $\widehat{x}^{(1)}_{\tau\Delta}, \ldots, \widehat{x}^{(B)}_{\tau\Delta}$.

$\widehat{x}^{(1)}_{\tau\Delta}, \ldots, \widehat{x}^{(B)}_{\tau\Delta}$ are iid draws from the conditional distribution of factors, so $\Upsilon(\widehat{x}^{(1)}_{\tau\Delta}, \theta), \ldots, \Upsilon(\widehat{x}^{(B)}_{\tau\Delta}, \theta)$ are iid draws from the conditional distribution of swap yields. Thus it is natural to approximate the distribution function with an average

$$\begin{aligned} F^P_{i\tau}(z, \theta) \;&=\; \Pr[Y_{i,\tau\Delta} \le z | X_{(\tau-1)\Delta}, \theta] \\ &\Rightarrow F^P_{\tau\Delta}(Y_{i,\tau\Delta}, \theta) \approx \frac{1}{B}\sum_{b=1}^{B} 1\big(\Upsilon_i\big(\widehat{x}^{(b)}_{\tau\Delta}, \theta\big) \le Y_{i,\tau\Delta}\big). \end{aligned} \tag{35}$$

Υ_i is the ith element of Υ. In the empirical applications and in the Monte Carlos, we used this formula with $M = 21$ discretization steps and $B = 10{,}000$ draws.

Now consider the calculation of the conditional transform F^C. By applying Ito's lemma to the Euler discretization, we can show that, conditional on $\widehat{x}_{\tau\Delta-h}$, the swaps $Y_{\tau\Delta}$ are distributed normally with mean $\mu_{\tau\Delta}$ and variance $\sigma^2_{\tau\Delta}h$, where

$$\begin{aligned} \mu_{\tau\Delta} &= \Upsilon(\widehat{x}_{\tau\Delta-h}, \theta) + J\,(\widehat{x}_{\tau\Delta-h}, \theta)\,K\,(\mu - \widehat{x}_{\tau\Delta-h})\,h + hH_{\tau\Delta-h}/2 \\ \sigma^2_{\tau\Delta} &= J\,(\widehat{x}_{\tau\Delta-h}, \theta)\,\Sigma S^2_{\tau\Delta-h}\Sigma' J\,(\widehat{x}_{\tau\Delta-h}, \theta)', \end{aligned} \tag{36}$$

with $[H_{\tau\Delta-h}]_j = trace(\Sigma S^2_{\tau\Delta-h}\Sigma'\partial[J(\widehat{x}_{\tau\Delta-h}, \theta)]_j/\partial\widehat{x}_{\tau\Delta-h})$. Thus, conditional on $\widehat{x}_{\tau\Delta-h}$, we have a closed-form normal density for $\Pr[c_{\tau\Delta,2} \le z | l_{\tau\Delta}, \widehat{x}_{\tau\Delta-h}, \theta]$ and $\Pr[c_{\tau\Delta,10} \le z | c_{\tau\Delta,2}, l_{\tau\Delta}, \widehat{x}_{\tau\Delta-h}, \theta]$. We approximate the F^C transforms by averaging over Gaussian conditional densities

$$F^C_{2,\tau\Delta}(z, \theta) \approx \frac{1}{B}\sum_{b=1}^{B} \Pr\big[c_{\tau\Delta,2} \le z | l_{\tau\Delta}, \widehat{x}^{(b)}_{\tau\Delta-h}, \theta\big] \tag{37}$$

$$F^C_{3,\tau\Delta}(z, \theta) \approx \frac{1}{B}\sum_{b=1}^{B} \Pr\big[c_{\tau\Delta,10} \le z | c_{\tau\Delta,2}, l_{\tau\Delta}, \widehat{x}^{(b)}_{\tau\Delta-h}, \theta\big]. \tag{38}$$

4.2 Cross-sectional inference for four-factor models

Here, we describe the cross-sectional estimation and testing of the four-factor model. The log-likelihood function is $\sum_{\tau=2}^{n} \log p(Y_{\tau\Delta}, \widetilde{Y}_{\tau\Delta} | Y_{(\tau-1)\Delta}, \theta)$, where p is the density of observation $(Y_{\tau\Delta}, \widetilde{Y}_{\tau\Delta})$ conditional on $Y_{(\tau-1)\Delta}$, the yields observed without error. Since the yield vector $Y_{\tau\Delta}$ can be inverted to obtain the factors $X_{\tau\Delta}$, the conditional density can be written as

$$p(Y_{\tau\Delta}, \widetilde{Y}_{\tau\Delta} | Y_{(\tau-1)\Delta}, \theta) = p(\widetilde{Y}_{\tau\Delta} | X_{\tau\Delta}, \theta) p(Y_{\tau\Delta} | Y_{(\tau-1)\Delta}, \theta). \tag{39}$$

$p(\widetilde{Y}_{\tau\Delta}|X_{\tau\Delta}, \theta)$ is the density of the yields observed with measurement error, and $p(Y_{\tau\Delta}|Y_{(\tau-1)\Delta}, \theta)$ is the density of the yields observed with no measurement error. Both densities are available in closed form. $\widetilde{Y}_{\tau\Delta}|X_{\tau\Delta} \sim N(\widetilde{\Upsilon}(X_{\tau\Delta}, \theta), \sigma^2\mathbf{I})$, where $\mathbf{I}$ is the identity matrix. $p(Y_{\tau\Delta}|Y_{(\tau-1)\Delta}, \theta)$ takes a form similar to expression (33):

$$p(Y_{\tau\Delta}|Y_{(\tau-1)\Delta}, \theta) = \det[J(x(Y_{\tau\Delta}, \theta), \theta)]^{-1} p_x(x(Y_{\tau\Delta}, \theta)|x(Y_{(\tau-1)\Delta}, \theta), \theta), \tag{40}$$

where $J(x, \theta)$ is the 4×4 Jacobian matrix $\partial\Upsilon(x, \theta)/\partial x'$, and p_x is the density of $X_{\tau\Delta,}$ given $X_{(\tau-1)\Delta}$. To calculate p_x, I use a four-dimensional extension to the closed-form approximations developed in Ait-Sahalia and Kimmel (2002).[6] We calculate the generalized residuals with the same simulation algorithms used to calculate the generalized residuals for the three-factor models.

4.3 Time-series inference for four-factor models

In this section, we provide details of our implementation of the time-series strategy. The log-likelihood function is $\sum_{\tau=2}^{n} \log p(Y_{\tau\Delta}, \widetilde{Y}_{\tau\Delta}|I_{(\tau-1)\Delta}, \theta)$, where p is the density of observation $(Y_{\tau\Delta}, \widetilde{Y}_{\tau\Delta})$ conditional on $I_{(\tau-1)\Delta} \equiv \{Y_{(\tau-1)\Delta}, \widetilde{Y}_{(\tau-1)\Delta}, \ldots, Y_{\Delta}, \widetilde{Y}_{\Delta}\}$, the information set containing all past observations. It is not obvious how to calculate p. We observe only three of the yields without error, so cannot back out the four factors $X_{\tau\Delta}$. Calculating the condition density requires solving a possibly intractable integral over the distribution of the unobserved factors.

We use an approximation to calculate the likelihood function. Integrating over the unobserved volatility state, $X_{1,t}$ is computationally difficult. Following Calvet and Fisher (2004), we approximate the underlying diffusion with a finite state Markov process. X_{1t} takes on a finite number of values $\{v_1, \ldots, v_d\}$. Since there are a finite number of grid points, integrals over the unknown volatility can be written as closed-form sums over probability-weighted functions of v_i. Our approximating process is chosen to closely mimic the dynamics of the underlying diffusion

$$dX_{1t} = K_{11}(\mu_1 - X_{1t})dt + \sqrt{\beta_{11}X_{1t}}dW_t^{(1)}. \tag{41}$$

The grid point v_i is the $(i - .5)/d$ quantile of the unconditional distribution for X_{1t}. This is available in closed form, since X_{1t} has an unconditional gamma distribution. The dynamics of X_{1t} are governed by the $d \times d$ Markov transition matrix $A = (a_{ij})_{1\le i,j\le d}$ with components $a_{ij} = \Pr[X_{1,\tau\Delta} = v_j|X_{1,(\tau-1)\Delta} = v_i]$. We choose the transition probabilities as

$$a_{ij} = \frac{p(X_{1,\tau\Delta} = v_j|X_{1,(\tau-1)\Delta} = v_i)((v_{j+1} - v_j)/2 - (v_j - v_{j-1})/2)}{\sum_{k=1}^{d} p(X_{1,\tau\Delta} = v_k|X_{1,(\tau-1)\Delta} = v_i)((v_{k+1} - v_k)/2 - (v_k - v_{k-1})/2)}. \tag{42}$$

[6] The Ait-Sahalia and Kimmel paper does not include approximations for four-factor models. Those were obtained directly from the authors.

$p(X_{1,\tau\Delta} = v_j | X_{1,(\tau-1)\Delta} = v_i)$ is Ait-Sahalia's (2002) approximation to the transition density of a square-root diffusion with weekly sampling. The endpoints are $v_0 = 0$ and $v_{d+1} = v_d + (v_d - v_{d-1})/2$. This is a completely closed-form approximation to the square-root diffusion. The approximation becomes closer to the true model as d becomes large.

With the discretization, the likelihood function can be calculated in closed form. The econometrician does not observe the four factors, only the probability of being in each state

$$\Pi_\tau^j = \Pr[X_{1,\tau\Delta} = v_j | I_{\tau\Delta}]. \tag{43}$$

Let $\mathbf{\Pi}_\tau = (\Pi_\tau^1, \ldots, \Pi_\tau^d)$ and $M_\tau = (p_{\tau,ij})_{1\le i,j\le d}$, where $p_{\tau,ij} = p(Y_{\tau\Delta}, \widetilde{Y}_{\tau\Delta} | X_{1,\tau\Delta} = v_j, X_{1,(\tau-1)\Delta} = v_i, I_{(\tau-1)\Delta}, \theta)$, the density conditional on the current and past volatility states. We can easily check that the density function is

$$p(Y_{\tau\Delta}, \widetilde{Y}_{\tau\Delta} | I_{(\tau-1)\Delta}, \theta) = \mathbf{\Pi}_{\tau-1}(M_\tau * A)\mathbf{1}, \tag{44}$$

where $*$ denotes element by element multiplication and $\mathbf{1}$ is a d-vector of 1's. The vector of probabilities $\mathbf{\Pi}_{\tau-1}$ are computed recursively by Bayesian updating. By Bayes's rule, the updating formula is

$$\mathbf{\Pi}_\tau = \frac{\mathbf{\Pi}_{\tau-1}(M_\tau * A)}{\mathbf{\Pi}_{\tau-1}(M_\tau * A)\mathbf{1}}. \tag{45}$$

The belief $\mathbf{\Pi}_\tau$ is, thus, a function of the observation $(Y_{\tau\Delta}, \widetilde{Y}_{\tau\Delta})$ and the prior probability $\mathbf{\Pi}_{\tau-1}$. The initial vector $\mathbf{\Pi}_0$ is chosen equal to the ergodic distribution of the Markov chain.

The density $p_{\tau,ij}$ is available in closed form; it can be factored as

$$\begin{aligned} p_{\tau,ij} = {} & p(\widetilde{Y}_{\tau\Delta} | X_{\tau\Delta}, X_{1,\tau\Delta} = v_j, \theta) \\ & \times p(Y_{\tau\Delta} | X_{(\tau-1)\Delta}, X_{1,\tau\Delta} = v_j, X_{1,(\tau-1)\Delta} = v_i, \theta). \end{aligned} \tag{46}$$

The conditional density for $\widetilde{Y}_{\tau\Delta}$ is given by $\widetilde{Y}_{\tau\Delta} | X_{\tau\Delta} \sim N(\widetilde{\Upsilon}(X_{\tau\Delta}, \theta), \sigma^2\mathbf{I})$, where $\mathbf{I}$ is the identity matrix. Letting $X_{\{2,3,4\},\tau\Delta} = (X_{2,\tau\Delta}, X_{3,\tau\Delta}, X_{4,\tau\Delta})'$, the remaining term can be written

$$\begin{aligned} p(Y_{\tau\Delta} | X_{(\tau-1)\Delta}, X_{1,\tau\Delta}, \theta) = {} & \det(\partial\Upsilon(X_{\tau\Delta}, \theta)/\partial X'_{\{2,3,4\},\tau\Delta})^{-1} \\ & \times \frac{p_x(x(Y_{\tau\Delta}, \theta) | X_{(\tau-1)\Delta}, \theta)}{p_{x1}(X_{1,\tau\Delta} | X_{(\tau-1)\Delta}, \theta)}, \end{aligned} \tag{47}$$

where p_{x1} is the density of $X_{1,\tau\Delta}$ conditional on the past. Both p_x and p_{x1} are calculated using extensions to the closed-form approximations developed in Ait-Sahalia and Kimmel (2002). Thus we have a completely closed-form approximation to the likelihood.

Table 1
Test results for three-factor models

	Joint test	6-month LIBOR			2-year swap			10-year swap		
Model	$\mathcal{Q}$	$\mathcal{T}_1$	$\mathcal{T}_2$	$\mathcal{T}_3$	$\mathcal{T}_1$	$\mathcal{T}_2$	$\mathcal{T}_3$	$\mathcal{T}_1$	$\mathcal{T}_2$	$\mathcal{T}_3$
Full affine $A_0(3)$	8.09**	2.78**	0.93**	6.98**	0.25**	0.51*	1.32**	0.20**	0.18	0.20
Full affine $A_1(3)$	6.84**	1.99**	1.23**	3.76**	0.20	0.45*	0.98**	0.19*	0.16	0.21
Full affine $A_2(3)$	6.91**	2.74**	0.92**	5.56**	0.54**	0.35	0.99**	0.33*	0.18	0.27
Ess. affine $A_0(3)$	7.85**	0.89**	0.89**	5.87**	0.18*	0.41*	1.52**	0.21*	0.20	0.29
Ess. affine $A_1(3)$	6.33**	1.92**	0.80**	2.52**	0.50**	0.41*	0.86**	0.53**	0.15	0.16
Ess. affine $A_2(3)$	4.84**	1.11**	1.08**	2.81**	0.53**	0.63*	1.55**	0.67**	0.17	0.15

The table shows goodness-of-fit statistics for three-factor models. The models are fit by a cross-sectional identification strategy, assuming $Y_{\tau\Delta} = (l_{\tau\Delta}, c_{\tau\Delta,2}, c_{\tau\Delta,10})'$ are observed without error, and no yields are observed with error. $\mathcal{T}_1$ assesses the shape of the conditional distribution, $\mathcal{T}_2$ the conditional mean, and $\mathcal{T}_3$ the conditional variance. $\mathcal{Q}$ is the joint test that assesses all three. The $\mathcal{T}_l$ statistics are computed from the marginal transform $F^P_{i\tau}$, and $\mathcal{Q}$ is computed from the conditional transform $F^C_{i\tau}$. We accept the model when the statistics are small and reject when they are large. Parameter estimates for the models appear in Table 3. The "completely affine" models impose $\lambda_2 = 0$ in expression (22), and the "essentially affine" models leave it unrestricted. * indicates rejection at the 5% level, and ** indicates rejection at the 1% level.

We calculate the generalized residuals with a simulation algorithm. The algorithm proceeds in several steps. We randomly draw a volatility state $X_{1,(\tau-1)\Delta}$ from the multinomial model with probabilities $\mathbf{\Pi}_{\tau-1}$. Given $X_{1,(\tau-1)\Delta}$, we invert from the three yields $Y_{(\tau-1)\Delta}$ to the implied factors $x(Y_{(\tau-1)\Delta}, \theta)$, which serve as the initial condition for the Euler discretization to the diffusion, given by expression (34). Simulating from the Euler approximation delivers random draws $\widehat{x}^{(1)}_{\tau\Delta}, \ldots, \widehat{x}^{(B)}_{\tau\Delta}$ from the distribution $p(X_{\tau\Delta}|I_{(\tau-1)\Delta}, \theta)$. The marginal transforms $F^P_{i\tau}$ are calculated using the average in Equation (35). The conditional transforms $F^C_{i\tau}$ are calculated by the averages in Equations (37) and (38).

5. Empirical Results

5.1 Empirical results for three-factor models

Table 1 reports specification test results for the three-factor affine models, and Table 2 reports results for four-factor models. Parameter estimates appear in Tables 3 and 4. Consider first the results for three-factor models. All of the models are strongly rejected, with p-values of essentially zero. The strongest rejections come from the $\mathcal{T}_3$ statistic applied to LIBOR. The models do a considerably better job of fitting 2- and 10-year swaps.

These results differ from those of Dai and Singleton (2000), who did not reject the completely affine $A_1(3)$ and $A_2(3)$ models. Since Dai and Singleton (2000) used weekly data observed over a different time period, I computed the tests for the dates in their paper and obtained essentially the same results. Thus, by changing the econometric procedure, we change the conclusion about three-factor affine models. A number of other authors have carried out formal

Table 2
Test results for four-factor models

Measurement		Evaluation of time-series fit										Mean absolute pricing errors, in bp							
		Joint test	6-month LIBOR			2-year swap			10-year swap										
error σ, in bp	log L	$\mathcal{Q}$	$\mathcal{T}_1$	$\mathcal{T}_2$	$\mathcal{T}_3$	$\mathcal{T}_1$	$\mathcal{T}_2$	$\mathcal{T}_3$	$\mathcal{T}_1$	$\mathcal{T}_2$	$\mathcal{T}_3$	1-year swap	3yr	4yr	5yr	6yr	7yr	8yr	9yr
Volatility identification by time-series strategy																			
10.0	40.52	5.60**	0.58**	1.01**	0.09	0.41*	0.79**	0.48*	0.33*	0.16	0.19	4.86	2.69	3.61	4.00	2.81	3.07	1.52	0.93
7.5	41.36	4.98**	0.63**	1.10**	0.19	0.40**	0.82**	0.47*	0.31**	0.16	0.19	4.94	2.69	3.60	4.00	2.79	3.05	1.49	0.91
5.0	42.11	4.08**	0.60**	1.05**	0.16	0.29*	0.81**	0.52*	0.25*	0.16	0.21	4.95	2.70	3.61	4.01	2.80	3.05	1.49	0.91
mle: 1.65	45.14	7.96**	2.50**	0.23	6.09**	0.37**	0.22	0.96**	0.23*	0.23	0.17	3.19	1.07	0.86	0.97	0.81	1.24	0.81	0.63
Model with unspanned stochastic volatility																			
mle: 7.13	40.23	4.09**	0.28	1.03**	0.24	0.19	0.79**	0.32	0.25	0.16	0.21	6.34	4.11	5.75	6.56	5.47	4.98	2.89	1.52
Volatility identification by cross-sectional strategy																			
10.0	41.62	7.16**	2.25**	0.24	4.51**	0.24**	0.38	0.66**	0.23**	0.21	0.22	2.98	1.59	1.51	–	0.67	1.43	0.82	0.63
7.5	42.43	7.42**	2.20**	0.24	4.78**	0.29**	0.33	0.58*	0.25**	0.21	0.18	3.12	1.55	1.47	–	0.65	1.43	0.78	0.61
5.0	44.81	7.60**	2.22**	0.24	4.92**	0.32**	0.28	0.65*	0.27*	0.21	0.17	3.35	1.49	1.42	–	0.65	1.44	0.76	0.61
mle: 2.07	44.89	7.75**	2.26**	0.24	5.22**	0.33**	0.32	0.69**	0.28*	0.21	0.15	3.36	1.49	1.42	–	0.66	1.45	0.78	0.62

The table shows goodness-of-fit statistics for four-factor models. The "time-series strategy" takes the three yields $Y_{\tau\Delta} = (l_{\tau\Delta}, c_{\tau\Delta,2}, c_{\tau\Delta,10})'$ to be observed without error, and the four yields $\widetilde{Y}_{\tau\Delta} = (c_{\tau\Delta,3}, c_{\tau\Delta,4}, c_{\tau\Delta,5}, c_{\tau\Delta,7})'$ to be observed with a normal measurement error with standard deviation of σ. The "cross-sectional strategy" takes the four yields $Y_{\tau\Delta} = (l_{\tau\Delta}, c_{\tau\Delta,2}, c_{\tau\Delta,5}, c_{\tau\Delta,10})'$ to be observed without error, and the three yields $\widetilde{Y}_{\tau\Delta} = (c_{\tau\Delta,3}, c_{\tau\Delta,4}, c_{\tau\Delta,7})'$ to be observed with error. The model with unspanned stochastic volatility is identified by the time-series strategy. $\mathcal{T}_1$ assesses the shape of the conditional distribution, $\mathcal{T}_2$ the conditional mean, and $\mathcal{T}_3$ the conditional variance. $\mathcal{Q}$ is the joint test that assesses all three. The $\mathcal{T}_l$ statistics are computed from the marginal transform $F^P_{i\tau}$, and the $\mathcal{Q}$ test is computed from the conditional transform $F^C_{i\tau}$. We accept the model when the statistics are small and reject when they are large. Parameter estimates for the models appear in Tables 5 and 6. * indicates rejection at the 5% level, and ** indicates rejection at the 1% level.

Table 3
Point estimates for three-factor and four-factor USV models

Completely affine $A_0(3)$	$K_{11} = 1.98$, $K_{21} = 2.82$, $K_{22} = 0.66$, $K_{31} = 3.15$, $K_{32} = 0.675$, $K_{33} = 0.011$, $\delta_0 = -0.30$, $\delta_1 = (0.0087, 0.0051, 0.0075)'$, $\lambda_1 = -(1.56, 1.52, 3.09)'$
Essentially affine $A_0(3)$	$K_{11} = 1.07$, $K_{21} = 2.17$, $K_{22} = 0.31$, $K_{31} = 3.56$, $K_{32} = 0.288$, $K_{33} = 0.064$, $\delta_0 = -.012$, $\delta_1 = (0.0040, .0037, .0050)'$, $\lambda_1 = -(1.07, 1.83, 3.35)'$, λ_2(row 1) $= (-.097, -.413, .622)$, λ_2(row 2) $= (0.441, .243, -.380)$, λ_2(row 3) $= (0.650, -.786, .786)$
Completely affine $A_1(3)$	$K_{11} = 0.443$, $K_{21} = 6.27$, $K_{22} = 0.388$, $K_{23} = 7.88$, $K_{31} = 0.377$, $K_{32} = 0.0119$, $K_{33} = 0.641$, $\mu_1 = 2.30$, $\beta_{21} = 5.84$, $\beta_{31} = 0.0155$, $\delta_0 = 0.00488$, $\delta_1 = (0.002, .002, .003)'$, $\lambda_1 = -(.443, .506, -.079)'$,
Essentially affine $A_1(3)$	$K_{11} = 0.944$, $K_{21} = 5.946$, $K_{22} = 0.069$, $K_{23} = 9.347$, $K_{31} = 0.334$, $K_{32} = 0.0068$, $K_{33} = 0.934$, $\mu_1 = 1.939$, $\beta_{21} = 6.633$, $\beta_{31} = 0.015$, $\delta_0 = 0.00664$, $\delta_1 = (0.002, .002, .0024)'$, $\lambda_1 = -(.944, .355, .430)'$, λ_2(row 2) $= (0.042, -.320, 1.43)$, λ_2(row 3) $= (-.046, -.003, .352)$
Completely affine $A_2(3)$	$K_{11} = 0.461$, $K_{12} = -.035$, $K_{21} = 0.000$, $K_{22} = 0.081$, $K_{31} = 15.899$, $K_{32} = -1.884$, $K_{33} = 0.685$, $\mu_1 = 2.76$, $\mu_2 = 21.50$, $\beta_{31} = 17.34$, $\beta_{32} = 0.0045$, $\delta_0 = -.016$, $\delta_1 = (0.054, .000, .0008)'$, $\lambda_1 = (21.91, .028, 10.12)'$
Essentially affine $A_2(3)$	$K_{11} = 0.639$, $K_{12} = 0.061$, $K_{21} = 3.443$, $K_{22} = 0.327$, $K_{31} = -2.778$, $K_{32} = -1.192$, $K_{33} = 0.882$, $\mu_1 = 10.042$, $\mu_2 = 14.760$, $\beta_{31} = 4.911$, $\beta_{32} = 0.000$, $\delta_0 = 0.192$, $\delta_1 = (0.0028, .0003, -.0022)'$, $\lambda_1 = (-.025, .032, -.643)'$, λ_2(row 3) $= (-.212, .160, -.348)$
USV $A_1(4)$	$K_{11} = 1.579$, $K_{21} = 2.70 \times 10^{-5}$, $K_{22} = -.0954$, $K_{23} = -.0516$, $K_{24} = 0.368$, $K_{31} = -.000243$, $K_{32} = 0.564$, $K_{33} = 0.731$, $K_{34} = -.984$, $K_{41} = -.000104$, $K_{42} = -1.214$, $K_{43} = -.644$, $K_{44} = 2.675$, $\theta_{Q,1} = 0.0000610$, $\theta_2 = -.0734$, $\theta_3 = 0.0144$, $\theta_4 = -.00830$, $K_{Q,44} = 2.484$, $\alpha_2 = 7.307 \times 10^{-12}$, $\delta_0 = 0.0756$, $\Sigma_{11} = 0.0138$, $\Sigma_{21} = 0.120$, $\Sigma_{23} = -.00883$, $\Sigma_{24} = 0.00148$, $\Sigma_{32} = -.118$, $\Sigma_{33} = 0.0511$, $\Sigma_{34} = -.00859$, $\Sigma_{43} = -.119$, $\Sigma_{44} = 0.0297$

The three-factor affine models are described in Section 4.1. The USV model is given in Footnote 5.

tests that reject the affine class.[7] However, those papers do not use swap data; rather they examine US government bonds observed over many decades.

To interpret the $\mathcal{T}_1$ statistic, examine Figure 2, which provides a histogram of the generalized residuals for the $A_2(3)$ model applied to LIBOR. Under the null, the histogram should look approximately uniform, but too many terms show up in the center and at the ends of the histogram. Since the $A_2(3)$ model is driven by a Brownian motion innovation, at short sampling frequencies the conditional distribution of LIBOR is approximately normal with conditional mean $m_{\tau\Delta}$ and variance $s^2_{\tau\Delta}$:

$$F^P_{l_\tau}(l_{\tau\Delta}, \widehat{\theta}) \approx \Phi\left(\left(l_{\tau\Delta} - m_{\tau\Delta}\right)/s_{\tau\Delta}\right). \tag{48}$$

$\Phi\left(\left(l_{\tau\Delta} - m_{\tau\Delta}\right)/s_{\tau\Delta}\right)$ is close to .5 when $l_{\tau\Delta}$ is close to $m_{\tau\Delta}$. The histogram shows that, relative to what the model predicts, interest rates move too little on most days and too much on some days. The data exhibit thick tails relative

[7] See Ahn, Dittmar, and Gallant (2002); Duffee (2002); Hong and Li (2005); Brandt and Chapman (2003); and Duarte (2004).

Table 4
Point estimates for maximal $A_1(4)$ models

	$A_1(4)$: TS, σ fixed		$A_1(4)$: TS, σ estimated		$A_1(4)$: CS, σ estimated	
	Est	StdErr	Est	StdErr	Est	StdErr
K_{11}	1.171	(0.337)	0.233	(0.00109)	0.166	(0.0513)
K_{21}	−1.055	(30.587)	0.428	(0.661)	2.012	(0.603)
K_{22}	0.0288	(0.0532)	−0.287	(0.421)	1.428	(0.532)
K_{23}	2.513	(0.531)	−0.0583	(0.112)	0.00805	(0.162)
K_{24}	2.157	(0.669)	−0.428	(0.571)	−0.777	(0.331)
K_{31}	−0.00152	(26.638)	−16.776	(1.878)	−6.62	(1.490)
K_{32}	−0.0265	(0.0525)	12.145	(0.87)	3.456	(1.367)
K_{33}	1.695	(0.551)	2.932	(0.231)	2.349	(0.316)
K_{34}	1.273	(0.595)	8.523	(1.781)	4.532	(0.867)
K_{41}	−0.000872	(19.301)	−0.931	(0.355)	−1.407	(1.146)
K_{42}	0.000255	(0.0294)	0.404	(0.21)	1.065	(0.939)
K_{43}	0.786	(0.356)	0.0688	(0.0555)	0.467	(0.298)
K_{44}	1.204	(0.429)	0.87	(0.319)	1.227	(0.553)
θ_1	9.80×10^{-5}	(0.000300)	0.00497	(1.39×10^{-5})	0.019	(0.00829)
α_2	1.51×10^{-8}	(2.87×10^{-6})	7.56×10^{-6}	(1.72×10^{-6})	4.04×10^{-5}	(2.44×10^{-5})
α_3	3.21×10^{-5}	(4.62×10^{-6})	1.00×10^{-13}	(1.91×10^{-5})	0	(0.00424)
α_4	1.65×10^{-5}	(2.65×10^{-6})	5.02×10^{-6}	(0.000128)	9.60×10^{-7}	(5.24×10^{-5})
β_{11}	0.000229	(0.000695)	0.00119	(9.48×10^{-6})	0.0034	(0.000913)
β_{21}	0.722	(2.198)	0.00117	(0.000234)	0.00061	(0.000349)
β_{31}	5.45×10^{-7}	(0.0212)	0.0181	(0.00244)	0.00627	(0.00698)
β_{41}	5.76×10^{-7}	(0.00816)	5.87×10^{-5}	(1.59×10^{-5})	0.00628	(0.00488)
δ_0	0.00604	(0.0126)	0.0949	(9.22×10^{-5})	0.0501	(0.0159)
$K_{Q,11}$	4.896	(34.211)	0.101	(0.00109)	0.0385	(0.00745)
$K_{Q,21}$	0.00602	(58.538)	−0.0349	(0.00169)	0.0349	(0.0200)
$K_{Q,22}$	0.216	(0.0422)	−0.113	(0.00145)	−0.0649	(0.0283)
$K_{Q,23}$	4.778	(0.231)	−0.11	(0.000349)	−0.0723	(0.0141)
$K_{Q,24}$	3.657	(0.246)	0.665	(0.00431)	0.11	(0.0222)
$K_{Q,31}$	0.207	(26.269)	−7.438	(0.0396)	−6.07	(0.219)
$K_{Q,32}$	−0.0176	(0.00991)	12.06	(0.022)	9.159	(0.560)
$K_{Q,33}$	1.554	(0.0697)	2.766	(0.00273)	3.687	(0.283)
$K_{Q,34}$	1.562	(0.0739)	1.954	(0.0511)	5.72	(0.146)
$K_{Q,41}$	−0.66	(19.147)	−0.123	(0.000402)	−0.801	(0.102)
$K_{Q,42}$	−0.0433	(0.0051)	−0.179	(0.000372)	−0.177	(0.167)
$K_{Q,43}$	0.203	(0.0649)	−0.0455	(0.000106)	0.0125	(0.0489)
$K_{Q,44}$	0.431	(0.0835)	0.164	(0.000701)	0.679	(0.0904)
$(K\theta)_{Q,2}$	0.0142	(0.000422)	0.000897	(1.52×10^{-5})	−0.000559	(3.17×10^{-5})
$(K\theta)_{Q,3}$	−0.00209	(0.00037)	−0.000448	(0.000259)	−0.129	(0.0610)
$(K\theta)_{Q,4}$	−0.00296	(0.000776)	−0.00174	(1.72×10^{-6})	−0.0174	(0.00779)
σ	0.0005	–	0.000165	(1.12×10^{-6})	0.000207	(1.98×10^{-6})

The model is described in Section 3 and in Equation (25). "TS, σ fixed" denotes time-series estimation with σ set equal to 5 basis points. "TS, σ estimated" denotes time-series estimation with σ estimated. "CS, σ estimated" denotes cross-sectional estimation with σ estimated. For the "TS, σ fixed" estimates, $\delta_1 = (1, 1, 1, 1)'$, and for the other estimates, $\delta_1 = (-1, 1, 1, 1)'$.

to the null model. One explanation for these patterns is that unexpected news, such as surprise Federal Funds target-rate movements, causes the short rate to jump. This has a large effect on short-term instruments like 6-month LIBOR, but mean reversion dampens the effect on longer-term instruments like swaps.

Other sources of thick tails are data problems, such as discreteness and rounding. Weekly changes in LIBOR tend to cluster around specific values too often to be drawn from a continuous distribution. For example, 131 of 782

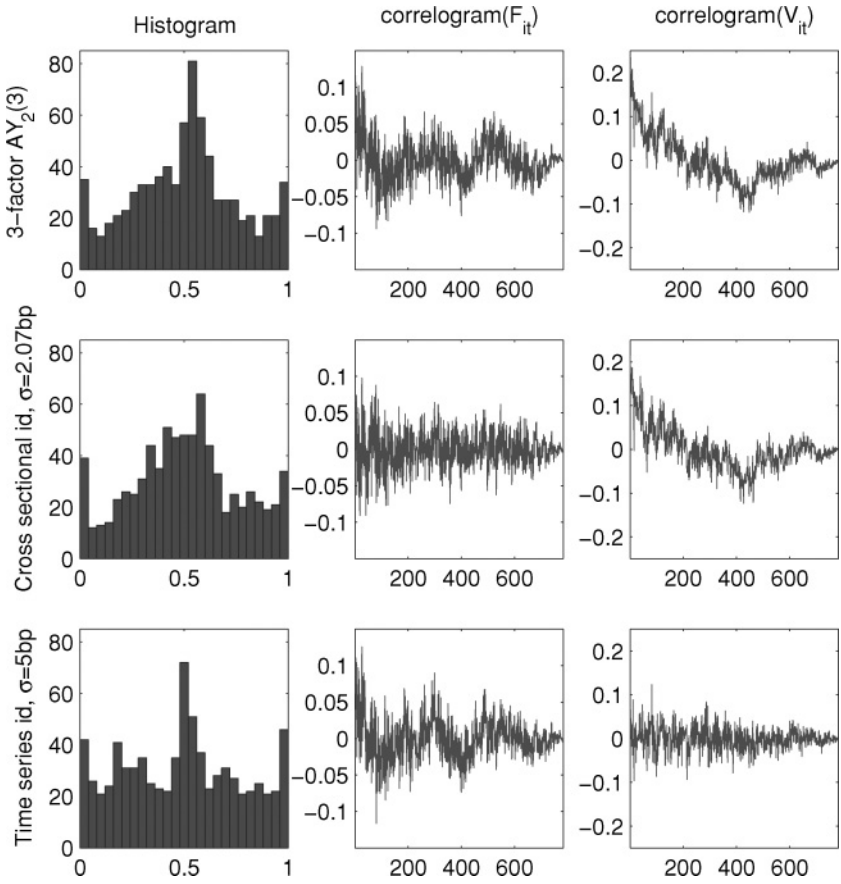

Figure 2
Goodness-of-fit evaluation of three models for 6-month LIBOR
Under the null of correct specification, the $F_{i\tau}^{P}$ terms should be uniform and independent. The $V_{i\tau}$ terms are $2|F_{i\tau}^{P} - .5|$. The row labeled "cross sectional id" refers to cross-sectional identification of the 4-factor $A_1(4)$ model in equation (25). The row labeled "time series id" refers to time-series identification of that model.

weekly changes in LIBOR are exactly zero. Simulations suggest that these data problems affect test size only slightly and do not explain the strong rejections.[8]

The $\mathcal{T}_2$ statistics provide weak evidence of positive correlation between generalized residuals. Figure 2 plots the correlations for the $A_2(3)$ model. Many of the correlations at short lags are positive, and the test statistics are significant at the 1% level. Since $F_{1\tau}^{P}(l_{\tau\Delta}, \widehat{\theta})$ is a monotonic function of $l_{\tau\Delta}$, the positive serial correlation between $F_{1\tau}^{P}$ terms suggests that the $l_{\tau\Delta}$ are positively

[8] We simulated 500 weekly datasets from the estimated $A_2(3)$ model with LIBOR rounded to increments of .0006. The F^P terms were calculated without reestimating the parameters. The 95% empirical quantile of the simulated statistics was .54. The maximum test statistic was 1.26. Both numbers are much lower than 2.74, the value in Table 1.

correlated beyond what the model predicts. There is unmodeled persistence: the yield reverts to its mean more slowly than is implied by the model.

The strongest rejections come from the $\mathcal{T}_3$ statistic, which detects correlations between absolute values of residuals. Figure 2 plots the correlations, which tend to be large and positive for short lags. The approximation $F^P_{1\tau}(l_{\tau\Delta}, \widehat{\theta}) \approx \Phi\left(\left(l_{\tau\Delta} - m_{\tau\Delta}\right)/s_{\tau\Delta}\right)$ implies that $|F^P_{1\tau} - .5|$ increases as $l_{\tau\Delta}$ moves farther away from the mean implied by the model. Positively correlated absolute values of residuals imply that large short-rate movements are too often followed by subsequent large movements. The nonzero correlations suggest the presence of unmodeled stochastic volatility.

5.2 Empirical results for four-factor models

Now consider the four-factor results in Table 2. The table reports test results for the various identification strategies, along with pricing errors over the entire swap curve. The first conclusion from the table is that the cross-sectional strategy does not identify volatility. The $\mathcal{T}_3$ specification test, which is sensitive to mismodeled volatility dynamics, strongly rejects the cross-sectional approach for LIBOR. The time-series strategy is much more successful at identifying volatility. As long as the standard deviation of measurement errors is at least 5 basis points, the time-series identification of volatility leads to statistically small values for the $\mathcal{T}_3$ tests.

Figure 2 provides a graphical analysis of the results. Both cross-sectional pictures display evidence of persistent, time-varying, and unmodeled stochastic volatility. The pictures change when we instead identify volatility by the time-series approach. Autocorrelations between $|F_{1\tau} - .5|$ terms are close to zero. The histogram also looks a bit more uniform, and correspondingly, the $\mathcal{T}_1$ statistic declines. With an improved volatility model, we also eliminate many of the outliers, and thus reduce the $\mathcal{T}_1$ statistic.

However, there are several caveats to the success of the time-series approach. The time-series approach leads to larger pricing errors. This is not surprising, since the cross-sectional approach fits four points on the yield curve exactly, while the time-series approach fits only three. Is the difference in pricing errors economically relevant? We compare the errors to the absolute value of the bid-offer spread, averaged over the sample period. For the 1-, 2-, . . . , 10-year swaps, the mean absolute bid-offer spreads are 3.0, 3.7, 3.8, 3.8, 3.0, 3.0, 3.0, and 3.9. Thus, for the cross-sectional strategy, the average pricing error is outside the average bid-offer spread for the 1-year yield, and inside the spread for all the other yields. For the time-series strategy, average pricing errors are outside the spread for the 1-, 3-, 5-, 6-, and 7-year yields.

The time-series results have a curious feature: at the maximum likelihood estimate of $\widehat{\sigma} = 1.65$, the time-series approach does not fit the volatility dynamics any better than the cross-sectional approach. As the standard deviation of the measurement error falls, the likelihood function assigns a larger penalty to pricing errors. For small values of σ, the parameters will be chosen to mini-

mize the pricing errors, rather than to properly fit the time series. So, it makes sense that small values of σ could lead to poor performance of the time-series approach. In unreported results, we found that the same thing happens to the cross-sectional approach: as we reduce σ, the $\mathcal{T}_l$ statistics become very large.

While it is clear that a small value for σ can lead to an ill-fitting time-series model, it is less clear why this would happen at the maximum likelihood estimate for σ. A potential explanation is that the measurement error processes are misspecified. For example, the pricing errors are positively serially correlated and contemporaneously cross-correlated, violating the assumption of independence. Bliss (1997) argues that this is not surprising, since many sources of mispricing, such as liquidity shortages for certain yields, probably exhibit persistence. Bliss (1997) has also pointed out that a Gaussian model for measurement errors implies a squared-error penalty function that penalizes the model, even when implied yields are inside the bid-offer spreads.

Aside from misspecification of the measurement error process, the underlying four-factor model is probably misspecified as well. The misspecification shows up in the statistically significant $\mathcal{T}_1$ and $\mathcal{T}_2$ statistics. To understand how misspecification could interact with pricing errors, consider the following thought experiment. In our time-series strategy, three yields are observed without error and four yields are observed with error. Suppose instead that 30 yields were observed with error. Then systematic mispricing will impose a large penalty on the likelihood function, and the parameters would be chosen to minimize the pricing errors, regardless of their time-series dynamics. Now suppose that a single yield is observed with error. Mispricing will impose a smaller penalty on the likelihood function, and the parameters will be chosen to better approximate the time-series dynamics. We found evidence in support of this view by reestimating the model using different numbers of yields observed with error. The $\mathcal{T}_l$ statistics increase as we add more yields.

Finally, we discuss the results for the model with unspanned stochastic volatility (USV). The USV model is a nested version of the maximal $A_1(4)$ model. We can test the USV restrictions using a likelihood ratio test. Reading the results off the table, the test statistic is $2(n-1)(45.14-40.23) = 7679.24$, which under the null has a chi-squared distribution with 9 degrees of freedom. The test implies a very strong rejection of the USV restrictions. The result is consistent with Bikbov and Chernov (2005), who also rejected the USV model in a dataset of swap derivatives. We should interpret these rejections carefully. The $\mathcal{T}_3$ statistics imply that the USV model does a much better job of fitting the volatility dynamics of the three yields. Rejection of the USV model comes from the larger pricing errors.

Figure 3 plots $(\mathrm{Var}_{(\tau-1)\Delta}(l_{\tau\Delta}))^{1/2}$, the square root of the volatility forecast for LIBOR implied by the various models. It also plots the implied volatility from a GARCH(1,1) model. The correlation between the time-series identified volatility and GARCH volatility is 93%, while for cross-sectionally identified volatility, the correlation is only 50%. Thus the time-series identified volatility

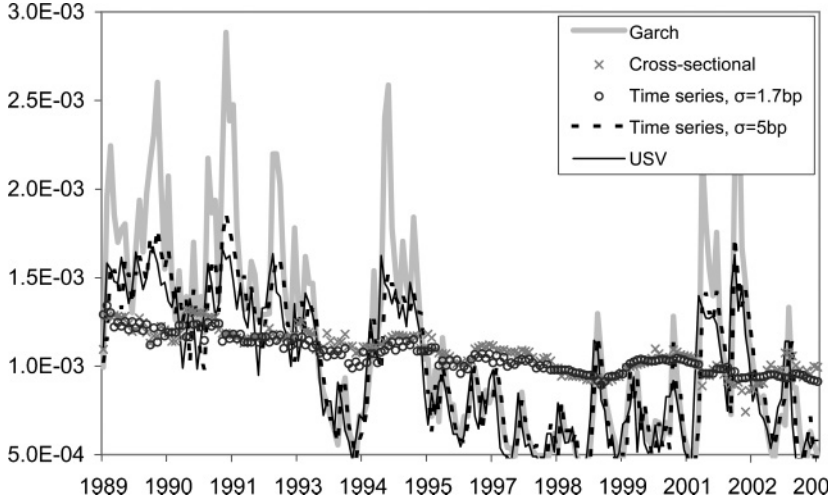

Figure 3
Model implied volatilities, $(\mathrm{Var}_{(\tau-1)\Delta} \mathbf{l}_{\tau\Delta})^{1/2}$, for 6-month LIBOR
When the standard deviation of yield measurement errors is at least 5 basis points, the time-series identification strategy tracks a GARCH model. The cross-sectionally identified volatility does not track GARCH, and is much less variable, implying more homoskedastic yields.

much more closely follows GARCH. The cross-sectionally identified volatility is much less variable than the time-series identified volatility: the standard deviation of volatility from the time-series approach is more than four times higher than for the cross-sectional approach. The cross-sectional strategy implies nearly homoskedastic yields over periods as long as five years.

We conclude that the evidence for the time-series approach is mixed. When the standard deviation of measurement errors is at least 5 basis points, the time-series approach does a good job of tracking volatility. The cross-sectional approach does not seem to identify volatility, but does a better job of fitting the yield curve. All of the models appear to be misspecified. Thus we have a tradeoff: we can choose an objective function that fits the volatility of yields, or we can choose an objective function that minimizes cross-sectional pricing errors.

5.3 Comparing the F^P and F^C results

Thus far, we have discussed the marginal generalized residuals F^P, which evaluate the time-series dynamics of each series separately. The marginal residuals do not use information in the contemporaneous joint distribution of yields. In most cases, the F^C transform, which tests the joint distribution, leads to the same conclusions as the marginal transforms. There is one case, which is not reported in the tables, where they lead to different conclusions. I separated the F^C transforms into three series: one for LIBOR, one for 2-year swaps, and one

for 10-year swaps. Under the null, each series is distributed approximately iid uniform. The $\mathcal{T}_1$ and $\mathcal{T}_3$ statistics calculated from the separate F^C series are similar to the statistics for the F^P transforms. The $\mathcal{T}_2$ statistics differ; the $\mathcal{T}_2$ statistic for 2-year swaps is at least 3 when computed from the F^C series, and much smaller when computed from the F^P transforms.

What explains the different values of $\mathcal{T}_2$? The statistic is computed from the autocorrelogram of $F^C_{2,\tau}(z,\theta) = \Pr[c_{\tau\Delta,2} \leq z \mid l_{\tau\Delta},\ Y_{(\tau-1)\Delta},\ \theta]$. For the $A_2(3)$ model, the first five autocorrelations are $-.21$, .16, .05, .05, and .03. For the unconditional transform $F^P_{2,\tau}(z,\theta) \equiv \Pr[2c_{\tau\Delta,2} \leq z | Y_{(\tau-1)\Delta}, \theta]$, the autocorrelations are $-.02$, .16, .05, .05, and .03. Since the F^C transform conditions on the contemporaneously observed $l_{\tau\Delta}$, the results suggest that the contemporaneous correlations between the variables are misspecified.

These results could be caused by asynchronously observed data. Each day, LIBOR is recorded at 11:00 a.m. London time and swap rates are recorded at 5:00 p.m. New York time. Following Dai and Singleton (2000); and Piazzesi (2005), I ignore the asynchronous nature of the data when estimating the models and forming the residuals. As Jones (2003a) has noted, intraday trading patterns or afternoon news events may induce correlation between residuals. Simulations indicate that these data problems could lead to the test results observed here.[9]

6. Monte Carlo Evaluation of Test Size and Power

One motivation for transforming to uniforms is to improve small sample test size. The usual asymptotic approximations assume that the underlying data are stationary. While interest rates may be stationary over the long run, it is difficult to reject the random-walk model in small samples. Persistence will cause the usual asymptotic approximations to be inaccurate in small samples. For example, Pritsker (1998) has shown that the persistence exhibited by some continuous time interest rate models significantly distorts the size of specification tests proposed by Ait-Sahalia (1996). In one Monte Carlo simulation, the test rejected a true null hypothesis in 50% of the trials.

Transforming to uniforms has the potential to improve small sample size. Since test statistics based on the transformed data are functions of approximately independent terms, asymptotic approximations to the null may be accurate, even when the underlying model is persistent.

Table 5 presents size results for sample sizes ranging from 50 to 5000 observations. For each sample size we simulate 5000 samples of weekly data from a Vasicek (1977) model with the parameters set to maximum likelihood

[9] We simulated 500 weekly datasets from the estimated fully affine $A_2(3)$ model, assuming that LIBOR is observed at 11:00 a.m. and swaps are observed at 5:00 p.m. We rounded each LIBOR observation to increments of .0006. The generalized residuals were calculated without reestimating the parameters. The F^C terms exhibited large negative first-order autocorrelations and inflated $\mathcal{T}_2$ statistics. Both the autocorrelations and the test statistics significantly declined when we computed the unconditional transform F^P.

Table 5
Rejection frequencies in small samples

T	$\mathcal{T}_1$	$\mathcal{T}_2$	$\mathcal{T}_3$	$\mathcal{Q}$
50	.045	.087	.068	.070
100	.053	.066	.057	.059
250	.051	.057	.054	.055
500	.054	.058	.050	.054
1000	.059	.049	.053	.053
5000	.059	.049	.053	.051

The table reports small sample rejection frequencies for the test statistics proposed in this paper. For each sample size, we simulate 5000 samples of weekly data from the Vasicek (1977) model $dr_t = .24(.038 - r_t)dt + .007dW_t$, which is estimated from the 1990–2002 data. For each sample, we test the true null that the data come from a Vasicek model.

estimates from the 1990 to 2002 subperiod. For each simulated dataset, we compute maximum likelihood estimates and test the true null, that the data come from a Vasicek model. The table demonstrates that we tend to not overreject a true null hypothesis, so long as the sample size is larger than 100 observations.

Hong and Li (2005) have also proposed specification tests based on the transformation to uniforms. They reject the null of correct specification for large values of

$$HL_j = h \int_0^1 \int_0^1 [\xi_j(z_1, z_2) - 1]^2 \, dz_1 dz_2, \tag{49}$$

where $\xi_j(z_1, z_2) = (n-j)^{-1}h^{-2}\sum_{\tau=j+1}^{n} k_h((F_{\tau\Delta} - z_1)/h)k_h((F_{(\tau-j)\Delta} - z_2)/h)$ and $k_h(x)$ is a density function.[10] $\xi_j(z_1, z_2)$ is a nonparametric kernel estimator of the joint density of $F_{\tau\Delta}$ and $F_{(\tau-j)\Delta}$, evaluated at z_1 and z_2. Under the null, $F_{\tau\Delta}$ and $F_{(\tau-j)\Delta}$ are uniformly distributed and independent of each other, so the joint density should be 1. The test statistic rejects the null when $\xi_j(z_1, z_2)$ is too far from 1.

The HL_j statistic has some advantages over other statistics proposed in this paper. It will detect any departure from independence and uniformity between $F_{\tau\Delta}$ and $F_{(\tau-j)\Delta}$, so it may have power against alternatives that the $\mathcal{T}_l$ test statistics miss. The properly normalized HL_j statistic converges to a standard normal variable, so obtaining asymptotically exact critical values is simple.

On the other hand, the HL_j statistic is based on a nonparametric kernel density estimator, which converges to the true density at less than the usual root-n rate. The $\mathcal{T}_l$ statistics are based on correlations and on the empirical distribution function, both of which converge to the truth at the usual root-n

[10] Hong and Li (2005) use the kernel

$$k_h((x-y)/h) = \begin{cases} q\left(\frac{x-y}{h}\right) \Big/ \int_{-(x/h)}^{1} q(u)du & \text{for } x \in [0, h) \\ q\left(\frac{x-y}{h}\right) & \text{for } x \in [h, 1-h] \\ q\left(\frac{x-y}{h}\right) \Big/ \int_{-1}^{(1-x)/h} q(u)du & \text{for } x \in (1-h, 1] \end{cases},$$

where $q(u) = (15/16)(1-u^2)^2 1(|u| < 1)$.

Table 6
Rejection frequencies for the $\mathcal{T}_l$ tests and Hong and Li (2005) tests

					HL_j statistics			
	$\mathcal{T}_1$	$\mathcal{T}_2$	$\mathcal{T}_3$	$\mathcal{Q}$	$j=1$	2	3	4
Monthly	.109	.084	.173	.195	.081	.085	.076	.092
Weekly	.395	.063	.485	.456	.271	.278	.280	.279
Daily	.819	.065	.880	.866	.712	.712	.701	.732

The table reports small sample power for the statistics proposed in this paper, along with the HL_j statistics proposed by Hong and Li (2005). We simulate 1000 datasets of 3045 daily observations, 608 weekly observations, and 140 monthly observations from Chen's (1996) model

$$\begin{aligned} dv_t &= k_v(\overline{v} - v_t)\,dt + \sigma_v\sqrt{v_t}dW_{1t} \\ d\mu_t &= k_\mu(\overline{\mu} - \mu_t)\,dt + \sigma_\mu\sqrt{\mu_t}dW_{2t} \\ dr_t &= k(\mu_t - r_t)\,dt + \sigma_v\sqrt{v_t}dW_{3t}, \end{aligned}$$

with market prices of risk $(\gamma_1\sqrt{v_t}, \gamma_2\sqrt{\mu_t}, \gamma_3\sqrt{v_t})'$. The parameters are $k_v = 0.477$, $\overline{v} = 0.000247$, $\sigma_v = 0.00802$, $k_\mu = 0.0577$, $\overline{\mu} = 0.00654$, $\sigma_\mu = 0.0420$, $k = 2.33$, $\gamma_1 = -2158$, $\gamma_2 = -101$, and $\gamma_3 = -261$. For each simulated dataset, we test the null hypothesis that six-month LIBOR comes from the Vasicek (1977) model $dr_t = k(\mu - r_t)dt + \sigma dW_t$.

rate. Thus the HL_j statistics do not have power against alternatives in a root-n neighborhood of the null, while the $\mathcal{T}_l$ statistics do. So in some cases, the $\mathcal{T}_l$ statistics should have greater power than the HL_j statistics.

Table 6 compares the power of the two approaches in a simple situation. The true data-generating process is Chen's (1996) three-factor term structure model

$$\begin{aligned} dv_t &= k_v(\overline{v} - v_t)\,dt + \sigma_v\sqrt{v_t}dW_{1t} \\ d\mu_t &= k_\mu(\overline{\mu} - \mu_t)\,dt + \sigma_\mu\sqrt{\mu_t}dW_{2t} \\ dr_t &= k(\mu_t - r_t)\,dt + \sigma_v\sqrt{v_t}dW_{3t}, \end{aligned} \tag{50}$$

with market prices of risk $(\gamma_1\sqrt{v_t}, \gamma_2\sqrt{\mu_t}, \gamma_3\sqrt{v_t})'$. Here, the short rate r_t mean reverts around an evolving mean μ_t and has a time-varying instantaneous variance v_t. We obtain maximum likelihood estimates from the weekly data over the 1990–2002 period, and simulate 1000 datasets of daily, weekly, and monthly observations from the estimated model. For each dataset, we test the null hypothesis, that the 6-month LIBOR comes from a Vasicek (1977) model.[11]

For weekly and monthly data, the $\mathcal{Q}$ statistics are almost twice as powerful as the HL_j statistics. This is consistent with the theoretical result that a root-n consistent test should be more powerful than a nonparametric test. The $\mathcal{T}_1$ and $\mathcal{T}_3$ tests are quite powerful against the Chen (1996) alternative, while the $\mathcal{T}_2$ statistic has virtually no power. The $\mathcal{T}_3$ statistic seems to detect the stochastic volatility, and the $\mathcal{T}_1$ statistic detects the departures from conditional normality, but the $\mathcal{T}_2$ statistic misses the time-varying mean. This may be due to the well-known result that as the sampling frequency increases, we gain more information about the distributional shape and instantaneous variance, but not

[11] I thank Yongmiao Hong and Haitao Li for providing me with Gauss code to calculate their test statistics.

about the mean function. To learn about the mean function, we have to observe the data over a long period of time. Perhaps 11 years of daily data contains a great deal of information about the variance of the process, but 11 years of data at any sampling frequency does not tell us much about the mean.

7. Conclusion

We have developed a family of diagnostic tests based on the transformation to independent uniforms. Transforming to independence removes the persistence in the data that led to size problems in Ait-Sahalia's (1996) tests. The Monte Carlo results indicate that the tests do not overreject a true null hypothesis, even in samples as small as 50 observations. The power of the tests was shown to be competitive with tests proposed by Hong and Li (2005).

The tests strongly reject all affine three-factor models. While the models do a good job with longer maturity swap yields, they do not generate volatility dynamics present at the short end of the term structure. These results are obtained using the same models and dataset used in Dai and Singleton (2000); thus, changing the econometric procedure changes the outcome.

The tests show that four-factor affine models can match the dynamics of term structure volatility, but this result is highly sensitive to how they are estimated. When we assume that four yields are observed without error and three with error, we can uniquely identify all four factors, including volatility. The implied volatility from this cross-sectional identification strategy does not track time-series variation in bond yields. However, when we assume that only three yields are observed without error and four with error, we cannot uniquely identify volatility. Instead, volatility is inferred from the time-series variation in yields. This time-series identification strategy appears to much more closely track volatility.

While this paper addresses the statistical significance of the volatility forecasts, it does not address their economic significance. A better forecast should lead to more accurate prices for interest rate futures and options. Bikbov and Chernov (2005) show that term structure models fit to zero coupon bond data do not accurately price interest rate derivatives. It will be interesting to see whether the time-series variation in bond prices identifies the factors needed to price options.

Appendix

A1 General formula for $\boldsymbol{\Omega}$

We will discuss the null distribution of the test statistics when an inefficient GMM estimator is used. The general formula for $\boldsymbol{\Omega}$ is

$$\boldsymbol{\Omega} = (\mathbf{J} \otimes \mathbf{R})[\mathbf{I}_J \otimes \boldsymbol{\Sigma} - \mathbf{C}\mathbf{D}' - \mathbf{D}\mathbf{C}' + \mathbf{C}\mathbf{V}_\theta\mathbf{C}'](\mathbf{J} \otimes \mathbf{R})', \tag{A1}$$

where $\mathbf{V}_\theta$ is $\lim_{T\to\infty} \mathrm{Var}[\sqrt{T}(\widehat{\theta} - \theta_0)]$, the variance of the normalized estimator. $\mathbf{C}$ and $\mathbf{D}$ denote the $LJ \times p$ matrices $(\mathbf{C}_1', \ldots, \mathbf{C}_J')'$ and $(\mathbf{D}_1', \ldots, \mathbf{D}_J')'$, where $\mathbf{C}_j$ and $\mathbf{D}_j$ are the $L \times p$ matrics

$$\begin{aligned}\mathbf{C}_j &= \lim_{n\to\infty} E n^{-1} \sum_\tau \widehat{\mathbf{G}}_{j,\tau}' \mathbf{S}_\tau(\theta_0),\\ \mathbf{D}_j &= \lim_{n\to\infty} E\left(n^{-1/2}\sum_\tau \widehat{\mathbf{G}}_{j,\tau}(\theta_0)\right)\sqrt{n}(\widehat{\theta} - \theta_0)'. \end{aligned} \tag{A2}$$

The notation $\widehat{\mathbf{G}}_{j,\tau}(\theta)$ indicates that we evaluate the transforms at the parameter vector θ.

To estimate $\mathbf{C}_j$ we replace the true parameter values with estimates, to obtain the estimate $n^{-1}\sum \widehat{\mathbf{G}}_{j,\tau}(\widehat{\theta})\mathbf{S}_\tau'(\widehat{\theta})$. In some cases, it is computationally difficult to compute scores of the likelihood function. Integration by parts leads to an equivalent formula for $\mathbf{C}_{j,i}$, which does not requires the scores

$$\mathbf{C}_j = -\lim_{n\to\infty} E n^{-1} \sum_\tau \frac{\partial \widehat{\mathbf{G}}_{j,\tau}(\theta_0)}{\partial \theta}. \tag{A3}$$

We can estimate this formula with $n^{-1}\sum \partial \widehat{\mathbf{G}}_{j,\tau}(\widehat{\theta})/\partial\theta$. The derivatives can be computed analytically in some cases, and by numerical differentiation generally.

The formulas for $\mathbf{V}_\theta$ and $\mathbf{D}_j$ depend on the form of the estimator. Under the usual assumptions, any GMM estimator admits the asymptotic approximation

$$\sqrt{n}(\widehat{\theta} - \theta_0) \approx -n^{-1/2}\sum_\tau \boldsymbol{\psi}_\tau(\theta_0). \tag{A4}$$

The function $\mathbf{D}_j$ is an estimate of the limiting covariance between $n^{-1/2}\sum_\tau \widehat{\mathbf{G}}_{j,\tau}(\theta_0)$ and $n^{-1/2}\sum_\tau \boldsymbol{\psi}_\tau(\theta_0)'$. When the $\boldsymbol{\psi}_\tau$ terms are serially uncorrelated, a consistent estimator is the sample covariance

$$\widehat{\mathbf{D}}_j = n^{-1}\sum_\tau (\widehat{\mathbf{G}}_{j,\tau}(\widehat{\theta}) - \overline{\mathbf{G}}_j)\boldsymbol{\psi}_\tau(\widehat{\theta})', \tag{A5}$$

where $\overline{\mathbf{G}}_j = n^{-1}\sum_\tau \widehat{\mathbf{G}}_{j,\tau}(\widehat{\theta})$ is the sample mean of the moments. When the $\boldsymbol{\psi}_\tau$ terms are serially correlated, we can use a heteroskedasticity and autocorrelation consistent estimator for $\mathbf{D}_j$; see Newey and West (1987); and Andrews (1991).

For the maximum likelihood estimator, $\boldsymbol{\psi}_\tau = \mathbf{V}_\theta \mathbf{S}_\tau(\theta_0)$ with $\mathbf{V}_\theta^{-1} = E[\mathbf{S}_\tau(\theta_0)\mathbf{S}_\tau'(\theta_0)]$, the $\boldsymbol{\psi}_\tau$ terms are uncorrelated. A general GMM estimator minimizes the quadratic form

$$\left(\sum_\tau \mathbf{m}_\tau(\theta)\right)' \mathbf{A}\left(\sum_\tau \mathbf{m}_\tau(\theta)\right), \tag{A6}$$

where $\mathbf{A}$ is a weight matrix. By the usual Taylor series expansions, the GMM estimator admits the asymptotic approximation

$$\sqrt{n}(\widehat{\theta} - \theta_0) \approx [\boldsymbol{\Lambda}(\theta_0)'\mathbf{A}\,\boldsymbol{\Lambda}(\theta_0)]^{-1}\left[\boldsymbol{\Lambda}\,(\theta_0)'\,\mathbf{A} n^{-1/2}\sum_\tau \mathbf{m}_\tau\,(\theta_0)\right], \tag{A7}$$

with $\boldsymbol{\Lambda}(\theta) = n^{-1}\sum_\tau \partial \mathbf{m}_\tau(\theta)/\partial\theta'$. This fits into our framework with

$$\boldsymbol{\psi}_\tau(\theta) = [\boldsymbol{\Lambda}(\theta_0)'\mathbf{A}\,\boldsymbol{\Lambda}(\theta_0)]^{-1}[\boldsymbol{\Lambda}(\theta)'\mathbf{A}\mathbf{m}_\tau(\theta)].$$

If the moment conditions $\mathbf{m}_\tau(\theta)$ are serially uncorrelated, we may use the simple formula in Equation (A5). Otherwise, we use an autocorrelation consistent estimator. The variance of the GMM estimator $\mathbf{V}_\theta$ is estimated by the usual methods (e.g., Cochrane, 2001).

When the estimator is asymptotically efficient, these formulas simplify. The formula for $\mathbf{D}_j$ can be expressed as $\mathbf{C}_j\mathbf{V}_\theta$, which leads to the simpler formula $\mathbf{\Omega} = (\mathbf{J}\otimes\mathbf{R})[\mathbf{I}_J\otimes\mathbf{W} - \mathbf{C}\mathbf{V}_\theta\mathbf{C}'](\mathbf{J}\otimes\mathbf{R})'$.

A2 Assumptions behind the theorem

To save space, I do not prove Theorem 1 here. The proof appears in an earlier version of this paper and is available by request from the author. The assumptions behind the proof appear below.

I assume that the nuisance parameters are well defined.

Assumption 1. $\mathbf{C}_j$ and $\mathbf{D}_j$ exist and are finite.

In order to accommodate Elerian, Chib, and Shephard's (2001) transformation $g_l(x) = 2|x - .5|$, I do not require the g_l functions to be differentiable; however, they must satisfy a Lipschitz condition.

Assumption 2. g_l satisfies the Lipschitz condition $|g_l(x) - g_l(y)| \le K|x - y|$ for all x and y, where K is a finite constant.

We need a series of assumptions about the smoothness of the conditional distribution function, the asymptotic representation for the estimator, and the degree of dependence between various terms of interest. Let $f^C_{i\tau}$ denote the density function based on the conditional transform $F^C_{i\tau}$, and let $\|A\| = [\sum_{j,k} A^2_{jk}]^{1/2}$ denote the usual Euclidean norm of a matrix A.

Assumption 3. (Smoothness of the conditional distribution) For all series i, there exists some $\mathcal{B}$, a fixed open ball centered at θ_0, so that for all $\theta_1, \theta_2 \in \mathcal{B}$ the following will hold:

(a) $\|\partial f^C_{i\tau}(y,\theta_1)/\partial\theta - \partial f^C_{i\tau}(y,\theta_2)/\partial\theta\| \le H_\tau(y)\|\theta_1 - \theta_2\|$, and $E\int_{-\infty}^{\infty} H_\tau(x)dx < K$.

(b) $E\sup_{\theta_1\in\mathcal{B}}\|\partial F^C_{i\tau}(y,\theta_1)/\partial\theta\|^4 < K^4$.

(c) $E\sup_{\theta_1,\theta_2\in\mathcal{B}}\int_{-\infty}^{\infty}\|\partial f^C_{i\tau}(y,\theta_2)/\partial\theta\|^r (f^C_{i\tau}(y,\theta_1))^{1-r}\,dy < K^r$ for some $r > 2$.

(d) For all $\varepsilon > 0$ there exists $\widetilde{K}(\varepsilon) < \infty$ so that

$$E\sup_{\varepsilon\le s\le 1-\varepsilon,\theta_1\in\mathcal{B}}\left(\frac{f^C_{i\tau}\left(F^{C-1}_{i\tau}(s,\theta_1),\theta_0\right)}{f^C_{i\tau}\left(F^{C-1}_{i\tau}(s,\theta_1),\theta_1\right)}\right)^2 \le (\widetilde{K}(\varepsilon))^2.$$

Assumption 4. (Asymptotic representation for the estimator) The normalized estimator satisfies $\sqrt{n}(\widehat{\theta} - \theta_0) = n^{-1/2}\sum_{\tau=1}^n \boldsymbol{\psi}_\tau + o_p(1)$, with $E_{\Delta(\tau-1)}\boldsymbol{\psi}_\tau = 0$ and $\mathrm{Var}[n^{-1/2}\sum\boldsymbol{\psi}_\tau] \to \mathbf{V}_\theta$. Also assume that $E\|\boldsymbol{\psi}_\tau\|^{2+2\epsilon} < K^{2+2\epsilon}$ for some $\varepsilon > 0$.

Assumption 5. (Degree of dependence) For each s, each element of the vector $(1(F^C_{i\tau} \le s), F^C_{i\tau}, \mathbf{S}'_\tau, \boldsymbol{\psi}'_\tau)'$ is near-epoch dependent in L_2 norm, of size -1, on an α-mixing (strong mixing) array of size $-(1+\epsilon)/\epsilon$. Furthermore, $n^{-1}\sum_{\tau=1}^n E\mathbf{S}_\tau 1(F^C_{i\tau} \le s)$ converges to a nonrandom function for each s.

Davidson (1994) (chapter 17) describes near-epoch dependence. Assumption (5) will be satisfied by any model that is stationary, mixing, and Markovian. This covers commonly used term structure models, such as the affine factor models. Near-epoch dependence also includes some non-Markovian, nonmixing stochastic volatility models. Hansen (1991); Sin and White (1996); and Carrasco and Chen (2002) have shown that while ARCH and GARCH processes are mixing only under restrictive assumptions, they are near-epoch dependent under much more general conditions.

Assumption (5) rules out autoregressive processes with unit roots, but simulations suggest that the critical values proposed in this paper are valid for those processes.

References

Ahn, D.-H., R. Dittmar, and A. R. Gallant. 2002. Quadratic Term Structure Models: Theory and Evidence. *The Review of Financial Studies* 15:243–88.

Ait-Sahalia, Y. 1996. Testing Continuous-Time Models of the Spot Interest Rate. *The Review of Financial Studies* 9(2):385–426.

Ait-Sahalia, Y. 2002. Maximum-Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approximation Approach. *Econometrica* 70(1):223–62.

Ait-Sahalia, Y. 2007. Closed-Form Expansions for Multivariate Diffusions. *Annals of Statistics*, forthcoming.

Ait-Sahalia, Y., and R. Kimmel. 2002. Estimating Affine Multifactor Term Structure Models Using Closed-Form Likelihood Expansions. Working Paper, Princeton University.

Andrews, D. W. 1991. Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica* 59(3):817–58.

Bai, J. 2003. Testing Parametric Conditional Distributions of Dynamic Models. *Review of Economics and Statistics* 85(3):531–49.

Bates, D. 2006. Maximum Likelihood Estimation of Latent Affine Processes. *Review of Financial Studies* 19:909–65.

Bikbov, R., and M. Chernov. 2005. Term Structure and Volatility: Lessons from the Eurodollar Markets. Working Paper, London Business School.

Bliss, R. 1997. Testing Term Structure Methods. *Advances in Futures and Options Research* 9:197–231.

Brandt, M., and D. Chapman. 2003. Comparing Multifactor Models of the Term Structure. Working Paper, Duke University.

Brandt, M. W., and P. Santa-Clara. 2002. Simulated Likelihood Estimation of Diffusions with an Application to Exchange Rate Dynamics in Incomplete Markets. *Journal of Financial Economics* 63:161–210.

Calvet, L., and A. Fisher. 2004. How to Forecast Long-Run Volatility: Regime-Switching and the Estimation of Multifractal Processes. *Journal of Financial Econometrics* 2:49–83.

Carrasco, M., and X. Chen. 2002. Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models. *Econometric Theory* 18:17–39.

Chen, L. 1996. Stochastic Mean and Stochastic Volatility—A Three-Factor Model of the Term Structure of Interest Rates and Its Applications in Derivatives Pricing and Risk Management. *Financial Markets, Institutions and Instruments* 5:1–88.

Chen, R. R., and L. Scott. 1993. Maximum Likelihood Estimation for a Multifactor Equilibrium Model of the Term Structure of Interest Rates. *Journal of Fixed Income* 3(3):14–31.

Cochrane, J. H. 2001. *Asset Pricing*. Princeton: Princeton University Press.

Collin-Dufresne, P., and R. Goldstein. 2002. Do Bonds Span the Fixed Income Markets? Theory and Evidence for Unspanned Stochastic Volatility. *Journal of Finance* 57(4):1685–730.

Collin-Dufresne, P., R. Goldstein, and C. S. Jones. 2004. Can Interest Rate Volatility Be Extracted from the Cross Section of Bond Yields? An Investigation of Unspanned Stochastic Volatility. NBER, Working Paper No. 10756.

Corradi, V., and N. R. Swanson. 2001. Bootstrap Specification Tests with Dependent Observations and Parameter Estimation Error. *Journal of Econometrics* 124(1):117–48.

Dai, Q., and K. J. Singleton. 2000. Specification Analysis of Affine Term Structure Models. *Journal of Finance* (55):1943–78.

Davidson, J. 1994. *Stochastic Limit Theory*. New York: Oxford University Press.

Diebold, F. X., T. A. Gunther, and A. S. Tay. 1998. Evaluating Density Forecasts, with Applications to Financial Risk Management. *International Economic Review* 39:863–83.

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

Duarte, J. 2004. Evaluating an Alternative Risk Preference in Affine Term Stucture Models. *Review of Financial Studies* 17:379–404.

Duffee, G. R. 2002. Term Premia and Interest Rate Forecasts in Affine Models. *Journal of Finance* 57:405–43.

Duffie, D., and R. Kan. 1996. A Yield-Factor Model of Interest Rates. *Mathematical Finance* 6(4):379–406.

Duffie, D., and K. J. Singleton. 1997. An Econometric Model of the Term Structure of Interest-Rate Swap Yields. *Journal of Finance* 52(4):1287–321.

Duffie, D., and K. J. Singleton. 1999. Modeling Term Structures of Defaultable Bonds. *Review of Financial Studies* 12(4):687–720.

Durbin, J. 1969. Tests for Serial Correlation in Regression Analysis Based on the Periodogram of Least-Squares Residuals. *Biometrika* 56(1):1–15.

Durham, G. B., and A. R. Gallant. 2002. Numerical Techniques for Maximum Likelihood Estimation of Continuous Time Diffusion Processes. *Journal of Business and Economic Statistics* 20:297–316.

Durlauf, S. 1991. Spectral Based Testing of the Martingale Hypothesis. *Journal of Econometrics* 50:355–76.

Elerian, O., S. Chib, and N. Shephard. 2001. Likelihood Inference for Discretely Observed Non-Linear Diffusions. *Econometrica* 69:959–93.

Eraker, B. 2001. MCMC Analysis of Diffusion Models with Application to Finance. *Journal of Business and Economic Statistics* 19(2):177–91.

Fournie, E. 1992. Un Test de Type Kolmogorov-Smirnov Pour Processus de Diffusion Ergodiques. Working Paper, No. 1696, INRIA, 1–29.

Gallant, A. R., and G. Tauchen. 1996. Which Moments to Match. *Econometric Theory* 12:657–81.

Gourieroux, C., A. Monfort, and E. Renault. 1993. Indirect Inference. *Journal of Applied Econometrics* 20:S85–S118.

Hansen, B. E. 1991. GARCH(1,1) Processes are Near Epoch Dependent. *Economics Letters* 36:181–86.

Hong, Y. 1999. Hypothesis Testing in Time Series via the Empirical Characteristic Function: A Generalized Spectral Density Approach. *Journal of the American Statistical Association* 94:448.

Hong, Y. 2000. Generalized Spectral Tests for Serial Dependence. *Journal of the Royal Statistical Society, Series B* 62:557–74.

Hong, Y., and H. Li. 2005. Nonparametric Specification Testing for Continuous-Time Models with Application to Spot Interest Rates. *Review of Financial Studies* 18:37–84.

Inoue, A. 1999. A Conditional Goodness-of-Fit Test for Time Series. Working Paper, University of British Columbia.

Jaquier, E., N. Polson, and P. Rossi. 1994. Bayesian Analysis of Stochastic Volatility Models. *Journal of Business and Economic Statistics* 12:371–417.

Jones, C. S. 2003a. Estimating Yield Curves from Asynchronous LIBOR and Swap Quotes. Working Paper, Marshall School of Business.

Jones, C. S. 2003b. Nonlinear Mean Reversion in the Short-Term Interest Rate. *Review of Financial Studies* 16:793–843.

Kim, S., N. Shephard, and S. Chib. 1998. Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *Review of Economic Studies* 65:361–93.

Lo, A. W. 1988. Maximum Likelihood Estimation of Generalized Ito Processes with Discretely Sampled Data. *Econometric Theory* 4:231–47.

Newey, W. K., and K. D. West. 1987. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* 55:703–8.

Pearson, N. D., and T.-S. Sun. 1994. Exploiting the Conditional Density in Estimating the Term Structure: An Application to the Cox, Ingersoll, and Ross Model. *Journal of Finance* 49(4):1279–304.

Pedersen, A. R. 1995. A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations. *Scandinavian Journal of Statistics* 22:55–71.

Piazzesi, M. 2005. Bond Yields and the Federal Reserve. *Journal of Political Economy* 113:311–44.

Pritsker, M. 1998. Nonparametric Density Estimation and Tests of Continuous Time Interest Rate Models. *Review of Financial Studies* 11(3):449–87.

Rosenblatt, M. 1952. Remarks on a Multivariate Transformation. *Annals of Mathematical Statistics* 23(3):470–72.

Segal, I. E. 1938. Fudicial Distribution of Several Parameters with Application to a Normal System. *Cambridge Philosophical Society Proceedings* 33:41–47.

Sin, C., and H. White. 1996. Information Criteria for Selecting Possibly Misspecified Parametric Models. *Journal of Econometrics* 71:207–25.

Vasicek, O. 1977. An Equilibrium Characterization of the Term Structure. *Journal of Financial Economics* 5:177–88.