Yield Curve and Volatility: Lessons from Eurodollar Futures and Options

Ruslan Bikbov
Federal Reserve Board

Mikhail Chernov
London Business School and CEPR

ABSTRACT
We evaluate the statistical and economic differences between affine term-structure models. Despite the voluminous literature on this subject, we have a limited understanding of those structural features of the models that are important in practice. Given that the key distinguishing characteristic of the affine models is the specification of the conditional volatility of the factors, we explore models that have critical differences in this respect: Gaussian (constant volatility) and stochastic volatility models. We estimate the models using the Eurodollar futures and options data as a basis. We subject these models to an exhaustive set of diagnostics. In particular, we develop a finite-sample version of the encompassing test for non-nested models. We find, based on the statistical tests and pricing errors, that there is little difference between the models when the models are estimated using only the yield curve information. Using options data enables us to separate the models very clearly. The stochastic volatility model is the most successful according to our diagnostics. (JEL: C1, G12)

KEYWORDS: asset pricing, econometrics, options, simulation, statistical analysis, term-structure models, volatility

We evaluate the statistical and economic differences between affine term-structure models (ATSM). Despite the voluminous literature on this subject, we have a lim-
ited understanding of those structural features of the models that are important in practice. Table 1 summarizes the major empirical contributions. Unfortunately, of the conclusions that have been drawn, there are few concerning which there is agreement.\(^1\) The three primary reasons for the divergent conclusions are different models being compared from paper to paper, differences in the frequency and type of data being used, and different evaluation metrics.

This paper addresses the shortcomings by determining the most promising models (on the basis of the results presented in Table 1) and comparing them by subjecting them to a comprehensive set of tests and diagnostics. Given that the key distinguishing characteristic of the affine models is the specification of the conditional volatility of the factors, we explore models that have critical differences in this respect: Gaussian (constant volatility) and stochastic volatility (SV) models. We refrain from comparing models that have more than three factors for reasons of parsimony.\(^2\)

We depart from most existing studies by complementing the yield curve data with option prices, which are by their nature sensitive to volatility specification. We use a rich sample that is derived from data for Eurodollar futures and options. These data have a number of unique qualities that make them more attractive than traditional data sources for term-structure studies. These markets are more mature than the related swaps markets and are more liquid (the daily dollar turnover in 2002 was $850 and $430 billion for futures and options, respectively, versus $120 and $50 billion for swaps and options on swaps, respectively). The quarterly maturity cycle of the futures allows us to use many points along the term structure.

Given that it is not difficult for any three-factor model to fit the cross-section of yields (see the principal component analysis of Litterman and Scheinkman 1991), we concentrate on the time-series performance of the models of interest. We relax the traditional tight link between the cross-section and the time series of yields by not imposing restrictions that arbitrarily require certain futures or option prices to be observed exactly.\(^3\) We use a quasi-maximum likelihood with a Kalman filter for estimating the models because the state variables cannot be inverted from prices in this setting.\(^4\)

Inference based on the asymptotic distribution of the estimated parameters and test statistics is difficult in our case. First, persistence of the interest rate can

---

\(^1\) There is one thing that many studies do seem to agree on: the models that have more than one stochastic volatility factor do not have a sufficiently flexible correlation structure to describe the yield curve.

\(^2\) Some authors prefer quadratic models to affine ones. However, as Cheng and Scaillet (2007) show, quadratic models are equivalent to restricted affine models that have many factors, for example, a typical three-factor quadratic model considered in the literature is equivalent to a nine-factor restricted affine model.

\(^3\) We illustrate that such an approach leads to drastically different estimation results.

\(^4\) The Kalman filter has undesirable asymptotic properties when state variables have stochastic volatility or asset prices have nonlinear dependence on state variables, as is the case with options. We conduct a series of comprehensive Monte Carlo studies, which confirm that the methodology performs well in practice in the term-structure setting.
Table 1 Reports major empirical findings in the ATSM literature. We summarize the findings of papers that evaluate the empirical performance of the affine term-structure models. By tradition, we denote them by $A_m(N)$, where $N$ represents the total number of factors, and $m$ is the number of “square-root” factors affecting the conditional volatility. $A^Q$ denotes a restricted affine model, which represents an equivalent affine representation of a linear-quadratic model (details may be found in Cheng and Scaillet 2007). The notation for the authors is as follows: DS1—Dai and Singleton (2000); DS2—Dai and Singleton (2002), Duffee—Duffee (2002); ADG—Ahn, Dittmar, and Gallant (2002); ADGG—Ahn et al. (2003); BC—Brandt and Chapman (2003); JK—Jacobs and Karoui (2009); LZ—Li and Zhao (2006); JKS—Jagannathan, Kaplin, and Sun (2003); Umantsev—Umantsev (2001); AGJ—Almeida, Graveline, and Joslin (2006). $J$-test denotes the GMM overidentifying restrictions test, $LPY$ denotes the expectation hypothesis regression coefficients pattern (see DS2 for details), and $CVY$ refers to the hump-shaped term structure of unconditional volatility. The asterisk (*) in the ADGG model list indicates that we omitted one specification (a hybrid of $A^Q_1(3)$ and an inverted model).

<table>
<thead>
<tr>
<th>Authors</th>
<th>Models</th>
<th>Data frequency</th>
<th>Data type</th>
<th>Primary evaluation metric</th>
<th>Preferred model/conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>DS1</td>
<td>$A_1(3), A_2(3)$</td>
<td>Weekly</td>
<td>Yields</td>
<td>Pricing errors, $CVY$</td>
<td>$A_1(3)$</td>
</tr>
<tr>
<td>Duffee</td>
<td>$A_0(3), A_1(3), A_2(3)$</td>
<td>Monthly</td>
<td>Yields</td>
<td>Forecasting errors</td>
<td>$A_0(3)$</td>
</tr>
<tr>
<td>DS2</td>
<td>$A_0(3), A_1(3), A_2(3), A_3(3)$</td>
<td>Monthly</td>
<td>Yields</td>
<td>$LPY$</td>
<td>$A_0(3)$</td>
</tr>
<tr>
<td>ADG</td>
<td>$A_2(3), A_3^Q(9)$</td>
<td>Monthly</td>
<td>Yields</td>
<td>$J$-test, filtered volatility</td>
<td>$A_3^Q(9)$</td>
</tr>
<tr>
<td>ADGG</td>
<td>$A_2(3), A_3^Q(4), A_3^Q(9)^*$</td>
<td>Monthly</td>
<td>Yields</td>
<td>$J$-test, filtered volatility</td>
<td>$A_3^Q(9)$</td>
</tr>
<tr>
<td>BC</td>
<td>$A_0(3), A_1(3), A_3^Q(9)$</td>
<td>Monthly</td>
<td>Yields</td>
<td>$J$-test (based on $LPY$, other “economic” moments)</td>
<td>$A_3^Q(9)$</td>
</tr>
<tr>
<td>JK</td>
<td>$A_m(3), m = 1, 2, 3$</td>
<td>Monthly</td>
<td>Yields</td>
<td>Corr between EGARCH and model-based conditional vol</td>
<td>$A_3(3)$ (mildly preferred)</td>
</tr>
<tr>
<td>LZ</td>
<td>$A_3^Q(9)$</td>
<td>Daily</td>
<td>Yields</td>
<td>Option hedging errors</td>
<td>Additional factors are required to capture the errors</td>
</tr>
<tr>
<td>JKS</td>
<td>$A_1(1), A_2(2), A_3(3)$</td>
<td>Weekly</td>
<td>Yields</td>
<td>Pricing errors</td>
<td>Options data do not improve option pricing</td>
</tr>
<tr>
<td>Umantsev</td>
<td>$A_1(2), A_1(3)$</td>
<td>Daily</td>
<td>Yields</td>
<td>Implied volatility fits</td>
<td>Options data improve option pricing</td>
</tr>
<tr>
<td>AGJ</td>
<td>$A_m(3), m = 0, 1, 2$</td>
<td>Weekly</td>
<td>Yields</td>
<td>Excess return fit</td>
<td>$A_1(3)$ and $A_2(3)$</td>
</tr>
</tbody>
</table>
affect the asymptotic properties of our tests. Second, the greatest challenge in comparing Gaussian and stochastic volatility models that have the same number of factors is that they are not nested. Encompassing tests (see Gourieroux and Monfort 1994 for a review) were designed to tackle this problem. However, the asymptotic distribution of the corresponding test statistics is complicated, which led to limited use of these tests in practice. We address both challenges by constructing finite-sample distributions of the test statistics. Specifically, we rely on parametric bootstrap to compute standard errors and implement tests throughout the paper (see Conley, Hansen, and Liu 1997). Thus, our paper is the first to explicitly test non-nested term-structure models using their likelihood.

From our analysis of the futures data, we find that there is virtually no difference between the two models in the results from a variety of diagnostic tests, such as the encompassing test, conditional mean and volatility fits, term structure of unconditional volatility, and kurtosis fits. We note that conditional volatility of the Gaussian model is constant. Thus, our results indicate that the futures data are not informative enough about the changing volatility of interest rates, and this is why we cannot distinguish statistically constant and stochastic volatility. Even if one doubts the power of these statistical tests, the economic implications of these models are unequivocal: the magnitudes of the pricing errors are similar across the models. This finding implies that one can use a simple Gaussian model to model the yield curve, at least within a single regime in a regime-switching model of interest rates. This conclusion is important because it is much easier to value assets and to implement estimation in the framework of this model than in other ATSM.

We next turn to the evidence that is afforded by the joint futures and options dataset. The results differ dramatically from the futures-only case. The encompassing test indicates that the Gaussian model does not encompass the SV model, that is, it cannot capture the characteristics of the data implicit in that model. The Gaussian model cannot capture the kurtosis of the short interest rate and the term structure of volatility. The pricing errors are larger than those of the SV model.

Incorporating the options data into estimation also reveals a tension between different models’ abilities to fit term structure versus options and to fit lower-order moments versus higher-order moments. One promising way to relax this tension is to use the unspanned stochastic volatility (USV) model of Collin-Dufresne and Goldstein (2002). The USV restrictions on the SV model simultaneously weaken the dependence of term structure on spot volatility and free up that same volatility factor to focus on option fitting. However, Bikbov and Chernov (2008) and Joslin (2006) formally reject the USV model in favor of the unrestricted SV model.

The remainder of the paper is organized as follows. Section 1 describes the institutional features of the Eurodollar markets, which we find to be important for our study. Section 2 develops affine models of Eurodollar futures and options prices. Section 3 presents the empirical approach. Section 4 discusses all of the

5 Compare, for instance, the models and the debate about intra-regime interest rate specification in Bansal, Tauchen, and Zhou (2004) and Dai, Singleton, and Yang (2003).
results. The final section makes concluding observations on the findings. Technical details are presented in three appendices.

1 THE EURODOLLAR MARKETS

In this paper, we follow a strategy regarding data choice that is different from that pursued in most studies by estimating models that use data for Eurodollar futures and options as a basis, rather than data for U.S. interest rate swaps and swaptions. The reason is that the Eurodollar markets have a number of important advantages. This section outlines these advantages and provides details concerning futures contracts that will be important for valuation in the following section.

Similarly to swap rates, Eurodollar futures are contracts written directly on the interest rate, as opposed to a traded asset. The underlying interest rate is the 90 day LIBOR, which we denote by $\ell_t(d)$, where $d = 90/360$—the LIBOR horizon in years according to the market conventions. The Eurodollar futures rate with maturity $\tau$ is quoted as

$$F_t(\tau) = 100 - f_t(\tau)$$

where $f_t(\tau)$ is the future LIBOR rate. The contract is settled in cash based on $100 - \ell_{t+t}(d)$.

Eurodollar futures are issued every quarter. However, when the Eurodollar contract was introduced in 1981, the available maturities, $\tau$, were no greater than two years. The CME gradually introduced longer maturities, and by the end of 1993, futures with maturities of up to 10 years were available. As a result, there are now effectively 40 points of the term structure, compared to six available swap rates.

One of the perceived advantages of swaps data is the simplicity of valuation because of the par-rate representation. This representation is based on the assumption that contractual payments have AA credit quality, which is the same as that of the underlying LIBOR. However, recently, Collin-Dufresne and Solnik (2001) and Johannes and Sundaresan (2007) argued that because of the swap contract collateralization, which removes counterparty risk, and because of the costs associated with posting and maintaining the collateral, the swap payments should be discounted at a rate not higher than the risk-free rate. This observation implies that the estimation of a AA-quality term-structure model based on the swap rates entails modeling the risk-free term structure, modeling the cost of collateral process, and estimation that is based on T-bond data in addition to the swaps data.

In contrast, the mark-to-market feature implies that futures do not have to be adjusted for the distortions that are implied by the need for holding the collateral. The futures price, $f_t(\tau)$, is a martingale under the risk-neutral measure $Q$, and valuation does not require any discounting:

$$f_t(\tau) = E_t^Q (\ell_{t+t}(d))$$
As a result, the modeling costs are reduced because we can focus exclusively on the AA-quality interest rate.

Yields on zero-coupon bonds are linear functions of the state variables in the affine framework, simplifying estimation of such models. However, in the case of swaps, the mapping between the observed rates and factors is nonlinear, even if the simple par rate representation is valid. Futures are claims on a single future payoff and are, therefore, similar to zero-coupon yields. Thus, one of the main benefits of using Eurodollar futures is that one does not have to bootstrap a zero-coupon yield curve. Therefore, ATSM-based futures prices have an exponentially linear relationship with the factors. Indeed, LIBOR is linked to bond prices in the following way:

\[ \ell_t(d) = \frac{1}{d} \left( \frac{1}{P_t(d)} - 1 \right) \]  

where \( P_t(d) \) is the price of a hypothetical AA-quality zero-coupon bond that matures \( d \) years from time \( t \). It is clear from expressions (2) and (3) that if the bond price has an exponentially linear form, the futures price has a similar expression, which we derive in the following section.

Derivatives on the Eurodollar futures are American puts and calls, which lend themselves to easy valuation in the affine framework after conversion to their European counterparts. Finally, options data have been available since 1985, in contrast to data for swaptions and caps, which have been available only since 1997. One of the drawbacks of using Eurodollar options instead of swaptions or caps is that there are no longer dated options because available expirations range from one month to two years.

2 AFFINE MODELS OF EURODOLLAR MARKETS

The foundation of our analysis is a hypothetical AA-quality zero-coupon bond with 90 days to maturity. In the absence of arbitrage opportunities, its price is given by

\[ P_t(d) = E_t^Q \left( e^{-\int_{t}^{t+d} r_s ds} \right) \]  

where \( d \) is time to maturity expressed in years, that is, 90/360, \( Q \) denotes the risk-neutral probability measure, and \( r \) is the AA-quality spot interest rate. We have to assume a dynamic model of \( r \) and the structure of \( Q \).

As Dai and Singleton (2000) point out, there is a trade-off between (i) the flexibility of modeling time-varying volatility with multiple square-root factors and (ii) the ability of a model to generate flexible conditional and unconditional

---

6The details of this simple procedure are provided in Section 3.1.
correlations between the state variables. Given this metric, it makes sense to consider only two out of the four possible three-factor affine term-structure models: a Gaussian model, often denoted as $A_0(3)$, and a model with one square-root stochastic volatility factor, $A_1(3)$. This conclusion is supported by the results in Dai and Singleton (2000) and Brandt and Chapman (2003) who find that affine models with multiple volatility factors perform worse than $A_1(3)$. It is important to point out that those studies estimate affine models based on yields data only and not on joint yields/options data. Almeida, Graveline, and Joslin (2006) estimate $A_0(3)$, $A_1(3)$, and $A_2(3)$ models using swap rates and cap option data and report that the $A_2(3)$ model estimated with yields and options outperform the $A_0(3)$ model (estimated with only yields) in capturing expected excess returns of swap rates.

As we mentioned in footnote 2, we do not consider quadratic models for reasons of parsimony. Cheng and Scaillet (2007) show that low-dimensional quadratic models have exactly the same pricing functions as high-dimensional restricted affine models. Therefore, the equivalence of these models follows by the same logic as that by which we group affine models into equivalency classes $A_m(N)$. For instance, the three-factor quadratic model studied by all of the papers reviewed in Table 1 is equivalent to a nine-factor restricted affine model. Having said this, we acknowledge that the properties of high-dimensional affine term-structure models have not been studied by empirical researchers. Thus, it is important to investigate whether the restrictions implicitly imposed by quadratic models hinder effectiveness of these models as compared to unrestricted specifications from the affine class. We leave this question for future research.

We fix the filtered probability space $(\Omega, \{F_t\}, F, P)$ and specify the considered models in the $Ar$ (affine in $r$, see Dai and Singleton 2000) form. This means that, to facilitate interpretation of the state variables, we select $r$ to be one of them. In this respect, we follow the tradition of many ATSM studies, which assign statistical interpretation to the latent factors. 

7 Consider a one-factor quadratic Gaussian model to understand the intuition behind the results. Introduce an auxiliary state variable that is equal to the square of the original Gaussian state variable. The new variable follows a restricted square-root process (see, for instance, Heston 1993). As a result, one obtains the interest rate as a linear function of two variables, one of which is Gaussian and the other follows a square-root process. Note that it is more appropriate to compare these models in terms of pricing functions (transforms) rather than in terms of state variables (see Chen, Filipovic, and Poor 2004 for the technical caveats). We are grateful to Peng Cheng, Pierre Collin-Dufresne, and Olivier Scaillet for discussions regarding these issues.

The three-factor Gaussian model, $A_0(3)$, is specified as follows:

\[
\begin{align*}
  \text{dr}_t &= \kappa_P^r \left( \theta_t + s_t - r_t \right) dt + \sigma_r dW_t^r(\mathbb{P}) + \sigma_{rt} \sigma_{\theta} dW_t^\theta(\mathbb{P}) + \sigma_{rs} dW_t^s(\mathbb{P}) \\
  \text{d} \theta_t &= \kappa_P^\theta \left( \tilde{\theta} - \theta_t \right) dt + \sigma_{\theta} dW_t^\theta(\mathbb{P}) + \sigma_{\theta s} dW_t^s(\mathbb{P}) \\
  \text{ds}_t &= -\kappa_P^s s_t dt + \sigma_s dW_t^s(\mathbb{P})
\end{align*}
\]

where $\theta$ traditionally represents the “central tendency” factor. We will refer to $s$ as the “spread” factor. This factor could be interpreted as the impact of the Federal Reserve interest rate targeting (see Attari 2001 and Piazzesi 2005 for related interpretations).

The stochastic volatility model, $A_1(3)$, is specified as follows:

\[
\begin{align*}
  \text{dr}_t &= \kappa_P^r \left( \theta_t - r_t \right) dt + \kappa_P^\theta \left( \tilde{\theta} - \theta_t \right) dt \\
  &\quad + \sqrt{\sigma_r^2 + \sigma_{rt}^2} dW_t^r(\mathbb{P}) + \sigma_{rt} \sqrt{\sigma_{\theta}^2 + \beta_{\theta \theta r} v_t} dW_t^\theta(\mathbb{P}) + \sigma_{rt} \sqrt{\sigma_{s}^2 + \beta_{s s \theta} v_t} dW_t^s(\mathbb{P}) \\
  \text{d} \theta_t &= \kappa_P^\theta \left( \tilde{\theta} - \theta_t \right) dt \\
  &\quad + \sigma_{\theta r} \sqrt{\sigma_r^2 + \sigma_{rt}^2} dW_t^r(\mathbb{P}) + \sqrt{\sigma_{\theta}^2 + \beta_{\theta \theta \theta r} v_t} dW_t^\theta(\mathbb{P}) + \sigma_{\theta s} \sqrt{\sigma_{s}^2 + \beta_{s s \theta} v_t} dW_t^s(\mathbb{P}) \\
  \text{dv}_t &= \kappa_P^s \left( \tilde{v} - v_t \right) dt + \sigma_s \sqrt{\sigma_t^2 + \beta_{s s} v_t} dW_t^s(\mathbb{P})
\end{align*}
\]

Here, $\theta$ is again the central tendency factor, and $v$ is the stochastic volatility factor.

We introduce notation associated with the probability measure transformation from $\mathbb{P}$ to $\mathbb{Q}$ in order to implement the bond formula (4).\(^9\) If we denote the vector of risk premia by $\Lambda$, then, according to the Girsanov theorem, we can link the two probability measures via

\[
W_t(\mathbb{Q}) = W_t(\mathbb{P}) + \int_0^t \Lambda_s ds
\]

We consider the essentially affine risk premia specification of Duffee (2002). In the case of $A_0(3)$, we have

\[
\begin{align*}
  \Lambda_t^r &= \lambda_r + \lambda_{rr} r_t + \lambda_{r\theta} \theta_t + \lambda_{rs} s_t \\
  \Lambda_t^\theta &= \lambda_\theta + \lambda_{\theta r} r_t + \lambda_{\theta \theta} \theta_t + \lambda_{\theta s} s_t \\
  \Lambda_t^s &= \lambda_s + \lambda_{sr} r_t + \lambda_{s\theta} \theta_t + \lambda_{ss} s_t.
\end{align*}
\]

---

\(^9\) We ensure stationarity of both models under both probability measures by requiring the positivity of the real part of all the eigenvalues of the matrices constructed from the respective $\kappa$s (the Routh–Hurwitz criterion). We also impose the generalized Feller condition (condition A of Duffie and Kan 1996) to guarantee strict positivity of volatility in $A_1(3)$. 
If the model is $A_1(3)$, then

$$\Lambda_t^f = \frac{1}{\sqrt{\sigma_r^2 + \nu_t}} \left( \lambda_r \sigma_r^2 + \lambda_{rr} r_t + \lambda_{r\theta} \theta_t + (\lambda_r + \lambda_{rv}) \nu_t \right)$$

(15)

$$\Lambda_t^\beta = \frac{1}{\sqrt{\sigma_\theta^2 + \beta_0 \nu_t}} \left( \lambda_\theta \sigma_\theta^2 + \lambda_{\theta r} r_t + \lambda_{\theta \theta} \theta_t + (\lambda_\theta \beta_\theta + \lambda_{\theta v}) \nu_t \right)$$

(16)

$$\Lambda_t^v = \lambda_v \sqrt{\nu_t}.$$  

(17)

Cheredito, Filipovic, and Kimmel (2007) and Liptser and Shiryaev (2001) show that essentially affine risk premia specifications can be extended to cases in which the price of risk is inversely proportional to the square-root volatility factor, as in, for example, (15) with $\sigma_r = 0$. For this reason, we will not constrain $\sigma_r$ and $\sigma_\theta$ to be greater than zero at the estimation stage.\(^\text{10}\) However, we still omit one possible additional term in (17) because its value binds a regularity condition based on the results reported in Cheredito, Filipovic, and Kimmel (2007). Further exploration of the contribution of the additional risk premia terms is an interesting direction for future research.

We have made all the assumptions necessary to compute the bond price (4). The affine specification implies

$$P_t(d) = e^{A^P(d) + B^P(d)r_t + B^\beta(d)\theta_t + B^v(d)\nu_t} = e^{A^P(d) + B^P(d)\cdot S_t}$$

(18)

where $\phi$ generically denotes either $s$ in $A_0(3)$ or $v$ in $A_1(3)$, and, for compactness, $S_t = (r_t, \theta_t, \phi_t)'$. The coefficients $A^P$ and $B^P$ solve well-known Ricatti ODEs. For completeness, Appendix 5 describes these ODEs. The boundary conditions are $A^P(0) = 0, B^P(0) = 0$.

We now have enough structure to price instruments from the Eurodollar markets. We combine the definitions of the Eurodollar futures (2), the LIBOR rate (3), and the LIBOR bond price (18) to obtain the futures LIBOR rate:

$$f_t(\tau) = \frac{1}{d} F_t^Q \left( e^{-A^P(d) - B^P(d)r_t - B^\beta(d)\theta_t - B^v(d)\nu_t} - 1 \right)$$

$$= \frac{1}{d} e^{-A^P(d)} F_t^Q \left( e^{-B^P(d)r_t - B^\beta(d)\theta_t - B^v(d)\nu_t} \right) - \frac{1}{d}$$

(19)

$$= \frac{1}{d} e^{-A^P(d) + A^f(\tau) + B^f(\tau)r_t + B^\beta(\tau)\theta_t + B^v(\tau)\nu_t} - \frac{1}{d} \equiv \frac{1}{d} e^{-A^P(d) + A^f(\tau) + B^f(\tau)\cdot S_t} - \frac{1}{d}$$

where $A^f$ and $B^f$ solve almost the same Ricatti ODEs with slightly different boundary conditions than (18): $A^f(0) = 0, B^f(0) = -B^P(d)$.

\(^\text{10}\)Note that in order to satisfy the canonical representation in Dai and Singleton (2000), the parameters $\sigma_r$ and $\sigma_\theta$ in $A_1(3)$ cannot be zero. On the other hand, a model with these parameters set at zero is admissible and identifiable (see Duffie and Kan 1996 for the regularity conditions). Aït-Sahalia and Kimmel (2002) make a similar observation.
It is natural to work with yields, as opposed to bond prices. The zero-coupon bond yield is

\[ y_t^P(d) = -\frac{1}{d} \log P_t(d) = -\frac{A^P(d)}{d} - \frac{B^P(d)}{d} \cdot S_t \]  

(20)

By analogy with the bond (also see (3)), we can construct an object based on the futures price, which we will refer to as the futures yield:

\[ y_t^f(\tau) = \frac{1}{d} \log(1 + d \cdot f_t(\tau)) \]

\[ = \frac{-A^P(d) + A^f(\tau)}{d} + \frac{B^f(\tau)}{d} \cdot S_t \equiv g_f(S_t, \Theta, \tau_f), \]  

(21)

where the last expression introduces notation.

A European call option on Eurodollar futures can be evaluated via similar techniques:

\[ C_t(\tau_f, \tau_c, K) = E_t^Q \left( e^{-\int_{t+\tau_c}^{t+\tau_f} r_s ds} \left( F_{t+\tau_c}(\tau_f - \tau_c) - K \right)^+ \right) \]

\[ = E_t^Q \left( e^{-\int_{t+\tau_c}^{t+\tau_f} r_s ds} \left( k - f_{t+\tau_c}(\tau_f - \tau_c) \right)^+ \right) \equiv g_c(S_t, \Theta, \tau_c), \]  

(22)

where \( k = 100 - K \) and the last expression introduces notation. Hence, a call on Eurodollar futures rate can be considered as a put on the LIBOR futures rate.\(^{11}\)

We follow Duffie, Pan, and Singleton (2000) in computing the call price. Appendix 5 provides the details. The Eurodollar options traded on CME are American. We use a very accurate early exercise premium approximation, which is described in Section 3.1.

## 3 EMPIRICAL SETUP

The implementation of our models requires a number of additional assumptions, which take into an account the data and factor structure of the models. Hence, we start with a brief description of the dataset and then describe the estimation methodology.

---

\(^{11}\)Note that we are using the AA-quality instantaneous rate \( r_t \) to discount the option payoff at maturity. Given that options on Eurodollar futures are traded on CME and are subject to the usual regulations, in principle, one should discount the payoff at the risk-free rate. Alternatively, one could think that trading options involves a small counterparty risk, which is comparable to the default risk of AA-rated companies. Further modeling of these fine features lies beyond the scope of this paper.
3.1 Data

We use weekly Eurodollar futures and options data.\textsuperscript{12} Recall that CME gradually introduced futures of longer maturities, beginning with a maximum maturity of two years in 1981 and ending with a maximum maturity of ten years in 1993. While it is possible to handle varying longest maturity (e.g. Jegadeesh and Pennacchi 1996), we would like to use the full range of maturities to compare them with the existing interest rate swaps studies and start our sample in January 1994. The sample ends in June 2001. As a result, we have $T = 392$ weeks in our sample.

One concern is that our sample of yields does not span a sufficiently long period to address issues related to volatility. First, we have to consider the yields versus options data trade-off because the derivatives have a shorter time span in virtually every market. Our options sample is the longest out of the ones considered in the literature (see, e.g. Jagannathan, Kaplin, and Sun 2003; Heidari and Wu 2002; Umantsev 2001). Second, there is strong evidence for thinking that an adequate model of term structure needs to incorporate regime switches if it is to explain a long historical period beginning in the late 1960s/early 1970s (e.g. Bansal and Zhou 2002; Dai, Singleton, and Yang 2003). Thus, the remaining issue is to find an appropriate model within a regime.\textsuperscript{13} Arguably, our sample period is excellent for a one-regime study because it represents the stability in the monetary policy conduct (the Federal Reserve Bank started announcing the interest rate target in 1994) and in the U.S. economy.

In order to reduce the computational burden associated with using all 40 available points along the yield curve, we select only 11 maturities. The first maturity is six months, and the rest are annual maturities from one to 10 years. Note that futures, in contrast to the swap rates, do not have constant maturities. Therefore, when we mention a maturity of $n$ years, we mean the maturity closest to $n$ years. We track the same “$n$-year” contract until it has 10 business days left before it switches into the “$n - 1$-year” category. Therefore, the actual number of years to maturity is not a random number but follows a predictable seesaw pattern.

We also narrow down the available options data. Options have maturities of up to two years. Those that have maturities of up to one year are the most liquid. Moreover, CME offers both standard quarterly ($\tau_c = \tau_f$) and serial ($\tau_c < \tau_f$) options. Hence, quarterly options are effectively options on the cash Eurodollar rates. As a result, it is easier to value these options. In addition, quarterly options are more liquid. For these reasons, we use only quarterly options of maturities equal to six months and one year.

Given that the focus of our paper is on learning about interest rate volatility, and not accurate option valuation \textit{per se}, we further limit our choice of options.

\textsuperscript{12} We obtained the data from the Institute for Financial Markets.

\textsuperscript{13} Indeed, Bansal, Tauchen, and Zhou (2004) note: “Given the nature of yields data, it would seem that allowing for within regime volatility to be stochastic is quite important. It remains to be seen if the specification which assumes a constant within regime volatility can account for the observed time-varying volatility and conditional cross-correlation of yields.”
We compute options’ moneyness as the ratio of the strike to the futures price and select options that have a moneyness as close to 1 as possible. As a result, 90% of all the moneynesses are between 0.98 and 1.02, with the median being exactly 1; the smallest moneyness is 0.97 and the largest is 1.03.

Given that the Eurodollar options are American, we have to adjust our data for the early exercise premium. We anticipate that this premium will not be large for our at-the-money short-maturity options. Indeed, Flesaker (1993) reports that “the early exercise premium for options not substantially in-the-money is generally less than a single basis point, which is the tick size in this market.” Nonetheless, we use a very accurate approximation from Broadie, Chernov, and Johannes (2007) to compute the early exercise premium. First, we compute the Black-implied volatility using the binomial tree, which takes into the account the early exercise option. Second, we obtain the European option price by plugging the implied volatility into the European version of the Black formula. In terms of the Black-implied volatility, the median premium is 0.05% and the largest premium is 0.16% (as a reference, the smallest implied volatility in our sample is 6.7%).

3.2 Measurement Errors

The standard assumption made at the ATSM estimation stage is that three yields are observed exactly and that, therefore, the latent factors can be inverted from these yields. Since the yields have a linear relationship with the factors, the volatility factor $v$ can be computed as a linear combination of the three yields (Collin-Dufresne, Goldstein, and Jones 2009 and Piazzesi 2010 make related points). Therefore, it is not clear whether $v$ computed from the cross-section of yields is guaranteed to be the instantaneous variance of the spot rate $r$. If a model is misspecified, assuming exactly observed yields could lead to severe errors, a negative $v$ in particular. These implications call for a careful distinction between time series and cross-sectional information in the term-structure data.

A natural way to relax the tension between the time series and cross-sectional implications of affine models is to assume that all prices are observed with an error:

$$y^f_t(\tau_{ft}) = g_f(S_t, \Theta, \tau_{ft}) + \epsilon_t(\tau_{ft}), \quad \tau_{ft} = 0.5_t, 1_t, 2_t, \ldots, 10_t$$

$$C_t(\tau_{ct}, \tau_{ct}, K) = g_c(S_t, \Theta, \tau_{ct}) + \tilde{\epsilon}_t(\tau_{ct}), \quad \tau_{ct} = 0.5_t, 1_t$$

where $S$ follows the state Equation (5)–(7) or (8)–(10) depending on whether we estimate the $A_0(3)$ or $A_1(3)$ models, and the functions $g_f(\cdot)$ and $g_c(\cdot)$ are given

---

14 In principle, in order to be absolutely accurate, we should have computed the exercise premium for the particular model under consideration, that is, $A_1(3)$. However, the impact of the correct model on the computation is an order of magnitude smaller than the premium captured by the simple adjustment via the Black model. Given that we find this adjustment to be very small already, it would be impractical to use more complicated models.
in Equations (21) and (22), respectively. The number of years to maturity with subindex \( t \) emphasizes the fact that actual maturity is different on each day.

The terms \( \epsilon \) and \( \tilde{\epsilon} \) are often referred to as measurement errors. Indeed, it is natural to assume that because of the bid-ask spread and other market frictions, the traded price does not represent the “equilibrium” price prescribed by the model. We expect the errors to have small variance if the model is correctly specified. We assume that the errors are independent across various maturities, and all have the same variance, \( \sigma^2_{\epsilon} \), irrespective of maturity. We allow a different variance, \( \tilde{\sigma}^2_{\epsilon} \), for the option pricing errors. Such a restrictive specification puts more pressure on our original model to fit the data. It will be more difficult to detect that a model is misspecified if the error terms are flexible.\(^{15}\)

All parameters are identified in our models. This conclusion may be surprising, in the light of DAi and Singleton’s (2000) comments about the unidentifiability of some risk premia parameters associated with Gaussian factors. What helps us in identifying the premia is the availability of more prices than state variables (see de Jong 2000).\(^{16}\) Related identification analysis is conducted in Collin-Dufresne, Goldstein, and Jones (2009).

The presence of measurement errors in each price calls for filtering-based estimation methods. We chose to rely on the Kalman filter, which is discussed in the next section. At this point, it is natural to wonder whether the difference in estimation approaches (inversion versus filtering) makes a practical difference. The data presented in Table 2 show that the estimation method can affect the outcome drastically. We estimate the \( A_1(3) \) stochastic volatility model on the basis of futures data only using the Kalman filtering technique or assuming that the six month, three year and seven year futures prices (the most liquid contracts) are observed without error. It is easy to see that the results are quite different. For example, notice that persistence \( \kappa_P \) of \( v \), which is the only factor that drives instantaneous volatility of the interest rate \( r \) in the model, changes from 0.71 to 0.04. Also note the difference in standard errors, depending on the method.

### 3.3 Econometric Method

The preferred estimation methodology is maximum likelihood (ML). However, with the exception of the \( A_0(3) \) model estimated on yields only, the likelihood is

---

\(^{15}\) Heidari and Wu (2002) criticize existing studies of LIBOR and swap rates, and swap-based derivatives in the affine framework because the pricing formulas implicitly assume that the contracts will be settled on the basis of the theoretical LIBOR rates, that is, they do not take into account the measurement error. Our methodology is not subject to their criticism because both futures and quarterly options contracts are settled on the basis of the actual LIBOR rate at maturity. Given that, in contrast to all other studies, we do not use the LIBOR data in our study, we can simply assume that actual LIBOR rates are observed without an error without sacrificing internal consistency.

\(^{16}\) We conducted additional identification checks to verify this conclusion. The results are available upon request.
Table 2 compares the MLE estimates of the $A_1(3)$ model, based on the exact inversion and the Kalman filter. The model $A_1(3)$ dynamics is given by

$$
\begin{align*}
\frac{\text{d}r_t}{\text{d}t} &= \kappa_t^P (\theta_t - r_t) \text{d}t + \kappa_t^P (\vartheta - v_t) \text{d}t + \sqrt{\sigma_t^2 + \bar{\vartheta} v_t} \text{d}W_t^P (\mathbb{P}) \\
&+ \sigma_r \sqrt{\sigma_t^2 + \bar{\vartheta} v_t} \text{d}W_t^r (\mathbb{P}) \\
\frac{\text{d}\theta_t}{\text{d}t} &= \kappa_\theta (\bar{\theta} - \theta_t) \text{d}t + \sigma_{\theta r} \sqrt{\sigma_t^2 + \bar{\vartheta} v_t} \text{d}W_t^r (\mathbb{P}) + \sqrt{\sigma_\theta^2 + \bar{\theta} v_t} \text{d}W_t^\theta (\mathbb{P}) \\
&+ \sigma_{\theta \vartheta} \sqrt{\bar{\theta} v_t} \text{d}W_t^\vartheta (\mathbb{P}) \\
\frac{\text{d}v_t}{\text{d}t} &= \kappa_v (\bar{v} - v_t) \text{d}t + \sigma_v \sqrt{\bar{v} v_t} \text{d}W_t^v (\mathbb{P})
\end{align*}
$$

and the risk premia are

$$
\begin{align*}
\Lambda_t^r &= \frac{1}{\sqrt{\sigma_t^2 + \bar{v} v_t}} \left( \lambda_r \sigma_t^2 + \lambda_{rr} r_t + \lambda_{r\theta} \theta_t + (\lambda_r + \lambda_{rv}) v_t \right) \\
\Lambda_t^\theta &= \frac{1}{\sqrt{\sigma_\theta^2 + \bar{\theta} v_t}} \left( \lambda_\theta \sigma_\theta^2 + \lambda_{\theta r} r_t + \lambda_{\theta \theta} \theta_t + (\lambda_\theta \lambda_{\theta \theta} + \lambda_{\theta \vartheta}) v_t \right) \\
\Lambda_t^v &= \lambda_v \sqrt{\bar{v} v_t}
\end{align*}
$$

The “inverted” version of the model is estimated by assuming that the prices of futures with maturities six months, three years, and seven years are observed exactly. We report the asymptotic standard errors estimated using first (SE1) and second (SE2) derivatives of the information matrix.

<table>
<thead>
<tr>
<th>Method</th>
<th>Est.</th>
<th>Kalman std. err. (SE1, SE2)</th>
<th>Inversion std. err. (SE1, SE2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa_t^P$</td>
<td>0.34</td>
<td>(0.01, 0.02)</td>
<td>2.63</td>
</tr>
<tr>
<td>$\kappa_\theta$</td>
<td>0.74</td>
<td>(0.31, 0.36)</td>
<td>1.84</td>
</tr>
<tr>
<td>$\kappa_v$</td>
<td>0.71</td>
<td>(0.13, 0.09)</td>
<td>0.04</td>
</tr>
<tr>
<td>$\kappa_{\theta r} \times 10^{-2}$</td>
<td>-0.57</td>
<td>(1.39, 1.35)</td>
<td>-21.36</td>
</tr>
<tr>
<td>$\vartheta \times 10^2$</td>
<td>3.39</td>
<td>(3.27, 2.16)</td>
<td>1.61</td>
</tr>
<tr>
<td>$\sigma_{\theta r} \times 10^4$</td>
<td>0.69</td>
<td>(0.14, 0.11)</td>
<td>0.82</td>
</tr>
<tr>
<td>$\sigma_v \times 10^2$</td>
<td>0.25</td>
<td>(0.07, 0.08)</td>
<td>0.22</td>
</tr>
<tr>
<td>$\sigma_r \times 10^2$</td>
<td>0</td>
<td>0</td>
<td>0.22</td>
</tr>
<tr>
<td>$\sigma_\theta \times 10^2$</td>
<td>2.81</td>
<td>(1.56, 1.34)</td>
<td>0</td>
</tr>
<tr>
<td>$\beta_\theta$</td>
<td>4.15</td>
<td>(8.02, 7.32)</td>
<td>1.58</td>
</tr>
<tr>
<td>$\sigma_{\theta \theta}$</td>
<td>0.11</td>
<td>(0.07, 0.04)</td>
<td>0.23</td>
</tr>
<tr>
<td>$\sigma_{\theta \vartheta}$</td>
<td>-0.21</td>
<td>(0.09, 0.08)</td>
<td>0.26</td>
</tr>
<tr>
<td>$\sigma_{\theta r}$</td>
<td>0</td>
<td>0</td>
<td>-0.18</td>
</tr>
<tr>
<td>$\sigma_{\theta v}$</td>
<td>5.06</td>
<td>(2.04, 3.01)</td>
<td>-1.09</td>
</tr>
<tr>
<td>$\lambda_r \times 10^2$</td>
<td>-3.31</td>
<td>(0.41, 0.45)</td>
<td>-276.96</td>
</tr>
<tr>
<td>$\lambda_\theta \times 10^{-2}$</td>
<td>-0.91</td>
<td>(1.31, 1.38)</td>
<td>0</td>
</tr>
<tr>
<td>$\lambda_v \times 10^{-2}$</td>
<td>0.36</td>
<td>(0.52, 0.35)</td>
<td>-0.19</td>
</tr>
<tr>
<td>$\lambda_{rr}$</td>
<td>0</td>
<td>0</td>
<td>-0.97</td>
</tr>
<tr>
<td>$\lambda_{\theta r}$</td>
<td>0.32</td>
<td>(0.03, 0.04)</td>
<td>1.83</td>
</tr>
<tr>
<td>$\lambda_{rr} \times 10^4$</td>
<td>0</td>
<td>2.89</td>
<td>(37.82, 2.30)</td>
</tr>
<tr>
<td>$\lambda_{\theta \theta}$</td>
<td>0</td>
<td>0.31</td>
<td>(0.51, 0.05)</td>
</tr>
<tr>
<td>$\lambda_{\theta v} \times 10^{-2}$</td>
<td>-0.74</td>
<td>(0.31, 0.36)</td>
<td>-1.51</td>
</tr>
</tbody>
</table>

Log-likelihood 73.67 70.84
not available in analytical form. Therefore, we implement quasi-maximum likelihood (QML), which entails approximating non-Gaussian states by the Gaussian ones with the true mean and variance (see, for instance, Fisher and Gilles 1996).

An additional complication arises from the fact that the states $S$ are not observable in the presence of measurement errors. In the case of the $A_0(3)$ model and futures rates only, one can estimate the system with errors by using the Kalman filter (see, e.g. Hamilton 1994 for the details of the algorithm). Indeed, if the state evolves according to the Gaussian system and the function $g_f$ is linear, then the ML based on the Kalman filter is the optimal estimation methodology. However, if we consider a nonlinear function in the measurement equation, such as $g_c$, or non-Gaussian states such as the $v$ factor in $A_1(3)$, the exact filter is generally not computationally feasible, and the Kalman filter is no longer optimal.

When the measurement equation is nonlinear, one could use an extended Kalman filter (EKF), which is effectively the regular Kalman filter applied to the first-order Taylor expansion of the measurement functions ($g_c$). When the state is non-Gaussian, one could use a quasi-Kalman filter (QKF), which replaces such a state ($v$) with a Gaussian state that has identical first two moments (Duan and Simonato 1995; Chen and Scott 2003). Both filters are the best linear filters, but QML estimates are conditionally biased and inconsistent.

Despite its shortcomings, we expect our choice of estimation method to perform well. EKF applied to high signal-to-noise ratio systems is known to have good properties asymptotic in $\sigma$ (Picard 1991). Monte Carlo studies of the QKF applied to multifactor square-root processes in Chen and Scott (2003), de Jong (2000), Duan and Simonato (1995), and Duffee and Stanton (2004) show that, in practice, this procedure introduces minimal biases. Moreover, studies attempting to improve upon QKF in a theoretically sound way find little difference in practice (see, e.g. Lund 1997 and Frühwirth-Schnatter and Geyer 1998). Given that we have at most one square-root factor and only two out of 13 measurement equations are nonlinear, these results suggest that we might obtain reasonable estimates despite the undesirable asymptotic properties of the filters.

Given that linearizing option prices is a relatively novel approach (the only other study to use this technique is Heidari and Wu 2002), we conduct a Monte Carlo analysis to evaluate the degree of error introduced by such an approximation. A Gaussian model is perfect for such a study because the option price approximation is the only source of error in the implementation of the model. We simulate 800 paths out of our model and compute the corresponding futures and options prices on the assumption that the parameter values that we subsequently estimate are true. We reestimate the model along each path, either via Kalman filter using futures prices only or via EKF using futures and options prices. We can use the resulting bootstrapped distribution of the parameters to compute the finite-sample confidence intervals and bias.

---

17 Brandt and He (2002) propose a simulated ML approach combined with importance sampling as an alternative to QKF.
Table 3 reports the results. First, we note that consistent with other studies, we occasionally observe a substantive bias even when the method is consistent and efficient (the futures-only case). However, large biases typically occur for parameters that are difficult to identify in practice: risk premia and correlations.\footnote{These observations are consistent with the findings of Duffee and Stanton (2004).} With these observations in mind, we conclude that option price approximation in the EKF does not introduce any particularly striking features, such as unusually wide confidence bounds, or large biases. It appears that the EKF recovers the true values fairly well.

4 RESULTS

4.1 Preliminary Observations

We estimated two versions, depending on the dataset, of each of the two models: $\mathcal{A}_0(3)$ and $\mathcal{A}_1(3)$. We assign a superscript $f$ to the models that are estimated based on futures only and a superscript $fo$ to the ones estimated on the joint dataset. We initially estimated the maximal models. However, some of the parameters were insignificant according to the asymptotic standard errors. Following the strategy of Dai and Singleton (2002), we restricted some of these parameters to zero, provided that such a restriction did not lead to a notable decline in the value of the log-likelihood function. We report the resulting parameter estimates in Tables 3 and 4. The asymptotic standard errors may misrepresent the significance of the estimated parameters in the light of the well-known persistence of the interest rate. We follow Conley, Hansen, and Liu (1997) and construct finite-sample confidence intervals via the parametric bootstrap.

Specifically, we repeat the Monte Carlo procedure outlined in Section 3.3. Thus, our results not only produce finite-sample inferences but also represent a Monte Carlo analysis of the estimation method.\footnote{We do not report the finite-sample bias for the $\mathcal{A}_1(3)$ models, in the interests of conserving space. The results are largely consistent with the ones for the $\mathcal{A}_0(3)$ models and are available upon request.} We report the 95% confidence intervals, which turn out to be asymmetric relative to the estimates. One can appreciate the difference between the asymptotic and finite-sample confidence intervals by evaluating two types of asymptotic standard error computed for the model $\mathcal{A}_1(3)^f$ in Table 2.\footnote{The two types of standard error are estimated using either the first or second derivatives of the information matrix.} The implied confidence intervals are, of course, symmetric and are typically wider than their finite-sample counterparts. The last feature may counterfactually imply the insignificance of certain parameters, e.g. $\bar{\theta}, \sigma_{\theta}, \beta_{\theta}, \sigma_{\theta\theta}, \sigma_{\theta\theta}, \lambda_{\theta},$ and $\lambda_{\theta\theta}$. As noted by Dai and Singleton (2000), the significance of these parameters has important implications for the flexibility in the factor correlations and for the specification of the feasible risk premia.
Table 3 Reports estimated parameters for model $\mathcal{A}_0(3)$ and for a Monte Carlo study. The model $\mathcal{A}_0(3)$ dynamics is given by

$$
\begin{align*}
\frac{dr_t}{P} &= \kappa_r^P \left( \theta_t + s_t - r_t \right) dt + \sigma_r dW_t^\theta (\mathbb{P}) + \sigma_r \sigma_\theta dW_t^\theta (\mathbb{P}) + \sigma_r dW_t^\sigma (\mathbb{P}) \\
\frac{d\theta_t}{P} &= \kappa_\theta^P \left( \bar{\theta} - \theta_t \right) dt + \sigma_\theta dW_t^\theta (\mathbb{P}) + \sigma_\theta \sigma_t dW_t^\sigma (\mathbb{P}) \\
\frac{ds_t}{P} &= -\kappa_s^P s_t dt + \sigma_s dW_t^\sigma (\mathbb{P})
\end{align*}
$$

and the risk premia are

$$
\begin{align*}
\Lambda_r^t &= \lambda_r + \lambda_r r_t + \lambda_r \theta_t + \lambda_r s_t \\
\Lambda_\theta^t &= \lambda_\theta + \lambda_\theta r_t + \lambda_\theta \theta_t + \lambda_\theta s_t \\
\Lambda_s^t &= \lambda_s + \lambda_s r_t + \lambda_s \theta_t + \lambda_s s_t
\end{align*}
$$

The finite-sample relative bias is reported in parenthesis. The bootstrapped 95% confidence intervals are reported in square brackets. The results in this table also serve as a Monte Carlo study of the estimation method that was performed, assuming that the reported parameter values are the true values.

<table>
<thead>
<tr>
<th>Model</th>
<th>Estimate</th>
<th>$\mathcal{A}_0(3)^f$ Bias, %</th>
<th>Confidence interval</th>
<th>Estimate</th>
<th>$\mathcal{A}_0(3)^f_0$ Bias, %</th>
<th>Confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa_r^P$</td>
<td>0.90</td>
<td>(-4.68)</td>
<td>[0.59, 1.49]</td>
<td>1.18</td>
<td>(-4.72)</td>
<td>[0.48, 2.36]</td>
</tr>
<tr>
<td>$\kappa_\theta^P$</td>
<td>1.84</td>
<td>(-10.97)</td>
<td>[1.17, 3.63]</td>
<td>1.15</td>
<td>(-0.91)</td>
<td>[0.64, 1.76]</td>
</tr>
<tr>
<td>$\kappa_s^P$</td>
<td>0.16</td>
<td>(-7.11)</td>
<td>[0.07, 0.46]</td>
<td>0.88</td>
<td>(-0.40)</td>
<td>[0.41, 1.26]</td>
</tr>
<tr>
<td>$\theta \times 10^2$</td>
<td>0.30</td>
<td>(1.40)</td>
<td>[-0.00, 0.73]</td>
<td>3.43</td>
<td>(3.47)</td>
<td>[0.64, 5.89]</td>
</tr>
<tr>
<td>$\sigma_r \times 10^2$</td>
<td>0.87</td>
<td>(2.83)</td>
<td>[0.73, 0.95]</td>
<td>0.70</td>
<td>(1.43)</td>
<td>[0.58, 0.82]</td>
</tr>
<tr>
<td>$\sigma_\theta$</td>
<td>0.02</td>
<td>(2.07)</td>
<td>[0.01, 0.04]</td>
<td>0.02</td>
<td>(-16.29)</td>
<td>[0.01, 0.07]</td>
</tr>
<tr>
<td>$\sigma_s$</td>
<td>0.03</td>
<td>(-2.54)</td>
<td>[0.02, 0.05]</td>
<td>0.04</td>
<td>(1.92)</td>
<td>[0.02, 0.05]</td>
</tr>
<tr>
<td>$\sigma_{\theta \theta}$</td>
<td>0.06</td>
<td>(-23.30)</td>
<td>[0.01, 0.18]</td>
<td>-0.15</td>
<td>(5.19)</td>
<td>[-0.30, -0.05]</td>
</tr>
<tr>
<td>$\sigma_{rs} \times 10^2$</td>
<td>0.35</td>
<td>(-2.36)</td>
<td>[0.24, 0.48]</td>
<td>0.10</td>
<td>(-38.09)</td>
<td>[-0.05, 0.41]</td>
</tr>
<tr>
<td>$\sigma_{\theta s}$</td>
<td>-0.01</td>
<td>(-4.77)</td>
<td>[-0.03, 0.00]</td>
<td>-0.02</td>
<td>(0.54)</td>
<td>[-0.05, -0.00]</td>
</tr>
<tr>
<td>$\lambda_r$</td>
<td>-3.88</td>
<td>(-3.23)</td>
<td>[-4.91, -3.24]</td>
<td>-2.85</td>
<td>(10.09)</td>
<td>[-5.39, 2.11]</td>
</tr>
<tr>
<td>$\lambda_\theta$</td>
<td>0</td>
<td></td>
<td></td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_s$</td>
<td>-0.14</td>
<td>(-0.52)</td>
<td>[-0.16, 0.13]</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_{rr}$</td>
<td>-11.54</td>
<td>(-46.58)</td>
<td>[-81.78, 24.80]</td>
<td>12.27</td>
<td>(142.82)</td>
<td>[-109.95, 59.70]</td>
</tr>
<tr>
<td>Parameter</td>
<td>Estimate</td>
<td>t-value</td>
<td>CI lower</td>
<td>CI upper</td>
<td>Std. Error</td>
<td>t-value</td>
</tr>
<tr>
<td>-----------</td>
<td>----------</td>
<td>---------</td>
<td>----------</td>
<td>----------</td>
<td>------------</td>
<td>---------</td>
</tr>
<tr>
<td>$\lambda_{\theta\theta}$</td>
<td>-56.74</td>
<td>-22.32</td>
<td>[-168.65, -31.87]</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_{\theta s}$</td>
<td>0.86</td>
<td>24.19</td>
<td>[-5.24, 3.30]</td>
<td>7.40</td>
<td>3 (30.72)</td>
<td></td>
</tr>
<tr>
<td>$\lambda_{s r}$</td>
<td>0</td>
<td></td>
<td></td>
<td>14.71</td>
<td>20.64</td>
<td>[2.15, 25.79]</td>
</tr>
<tr>
<td>$\lambda_{s \theta}$</td>
<td>0</td>
<td></td>
<td></td>
<td>-20.64</td>
<td>2.15</td>
<td>[35.89, -6.11]</td>
</tr>
<tr>
<td>$\lambda_{s s}$</td>
<td>-5.00</td>
<td>-5.34</td>
<td>[-14.54, -2.66]</td>
<td>-35.45</td>
<td>2.02</td>
<td>[46.36, -22.06]</td>
</tr>
<tr>
<td>Log-likelihood</td>
<td>73.66</td>
<td></td>
<td></td>
<td>85.09</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4 Reports estimated parameters for model $A_1(3)$. The model $A_1(3)$ dynamics is given by
\[
\begin{align*}
\mathrm{d}r_t &= \kappa_r^P (\theta_t - r_t) \, \mathrm{d}t + \kappa_{vr}^p (\theta - v_t) \, \mathrm{d}t + \sqrt{\sigma_r^2 + \nu_t} \mathrm{d}W_t^r (\mathbb{P}) \\
&\quad + \sigma_r \sqrt{\sigma_r^2 + \beta_v v_t} \mathrm{d}W_t^r (\mathbb{P}) + \sigma_{rv} \sqrt{\nu_t} \mathrm{d}W_t^r (\mathbb{P}) \\
\mathrm{d}\theta_t &= \kappa_\theta^p (\theta - \theta_t) \, \mathrm{d}t + \sigma_{\theta r} \sqrt{\sigma_r^2 + \nu_t} \mathrm{d}W_t^r (\mathbb{P}) + \sqrt{\sigma_{\theta r}^2 + \beta_v v_t} \mathrm{d}W_t^r (\mathbb{P}) \\
&\quad + \sigma_{\theta v} \sqrt{\nu_t} \mathrm{d}W_t^r (\mathbb{P}) \\
\mathrm{d}v_t &= \kappa_v^p (\theta - v_t) \, \mathrm{d}t + \sigma_v \sqrt{\nu_t} \mathrm{d}W_t^r (\mathbb{P})
\end{align*}
\]
and the risk premia are
\[
\begin{align*}
\Lambda_t^r &= \frac{1}{\sqrt{\sigma_r^2 + \nu_t}} \left( \lambda_r \sigma_r^2 + \lambda_{rr} r_t + \lambda_{\theta r} \theta_t + (\lambda_r + \lambda_{rv}) v_t \right) \\
\Lambda_t^\theta &= \frac{1}{\sqrt{\sigma_{\theta r}^2 + \nu_t}} \left( \lambda_{\theta r} \sigma_r^2 + \lambda_{\theta r} r_t + \lambda_{\theta \theta} \theta_t + (\lambda_{\theta r} + \lambda_{\theta v}) v_t \right) \\
\Lambda_t^v &= \lambda_v \sqrt{\nu_t}
\end{align*}
\]
The bootstrapped 95% confidence bounds are reported in square brackets. An asterisk (*) denotes parameters significant at the 10% level.

<table>
<thead>
<tr>
<th>Model</th>
<th>$A_1(3)^{\dagger}$</th>
<th>$A_1(3)^{\ddagger}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa_r^P$</td>
<td>0.34</td>
<td>[0.32, 0.36]</td>
</tr>
<tr>
<td>$\kappa_\theta^p$</td>
<td>0.74</td>
<td>[0.50, 1.40]</td>
</tr>
<tr>
<td>$\kappa_v^p$</td>
<td>0.71</td>
<td>[0.59, 0.88]</td>
</tr>
<tr>
<td>$\kappa_{vr}^p \times 10^{-2}$</td>
<td>-0.57</td>
<td>[-10.17, 1.10]</td>
</tr>
<tr>
<td>$\bar{\theta} \times 10^2$</td>
<td>3.39</td>
<td>[0.10, 6.64]</td>
</tr>
<tr>
<td>$\sigma \times 10^4$</td>
<td>0.69</td>
<td>[0.52, 0.85]</td>
</tr>
<tr>
<td>$\sigma_v \times 10^2$</td>
<td>0.25</td>
<td>[0.09, 0.40]</td>
</tr>
<tr>
<td>$\sigma_r$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\sigma_{\theta r} \times 10^2$</td>
<td>2.81</td>
<td>[1.97, 3.51]</td>
</tr>
<tr>
<td>$\sigma_{\theta v}$</td>
<td>4.15*</td>
<td>[0.00, 11.98]</td>
</tr>
<tr>
<td>$\sigma_{\theta v}$</td>
<td>0.11</td>
<td>[0.07, 0.16]</td>
</tr>
<tr>
<td>$\sigma_{\theta v}$</td>
<td>-0.21</td>
<td>[-0.39, -0.01]</td>
</tr>
<tr>
<td>$\sigma_{\theta v}$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\sigma_{\theta v}$</td>
<td>5.06</td>
<td>[1.33, 8.04]</td>
</tr>
<tr>
<td>$\lambda_{rr} \times 10^{-2}$</td>
<td>-3.31</td>
<td>[-4.44, -2.74]</td>
</tr>
<tr>
<td>$\lambda_{\theta r} \times 10^{-2}$</td>
<td>-0.91</td>
<td>[-1.26, -0.51]</td>
</tr>
<tr>
<td>$\lambda_{\theta v} \times 10^{-2}$</td>
<td>0.36</td>
<td>[-0.28, 0.92]</td>
</tr>
<tr>
<td>$\lambda_{rr}$</td>
<td>0</td>
<td>-0.13</td>
</tr>
<tr>
<td>$\lambda_{\theta r}$</td>
<td>0.32</td>
<td>[0.29, 0.41]</td>
</tr>
<tr>
<td>$\lambda_{\theta r}$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\lambda_{\theta v}$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\lambda_{\theta v}$</td>
<td>-0.74</td>
<td>[-1.40, -0.50]</td>
</tr>
<tr>
<td>$\lambda_{\theta r} \times 10^{-2}$</td>
<td>16.98</td>
<td>[11.12, 22.79]</td>
</tr>
</tbody>
</table>

Log-likelihood 73.67 86.35
The parameter estimates allow us to take a first look at the models’ properties. We check whether the initial interpretation of the state variables corresponds to the actual role they play in the respective models. We correlate the factors, filtered conditional on the estimated parameters values, with some interesting observables, such as the six month futures yield, or the short rate; the 10 year futures yield, or the long rate; the difference between the two, or the slope of the term structure; the butterfly spread, which corresponds to a long position in the six month and 10 year futures and double short position in the two year futures; and, finally, the absolute implied variance.\(^{21}\) Table 5 reports the results. In many cases, there are two observables that are highly correlated with a particular state variable. We will focus, somewhat arbitrarily, on the observables that have the highest correlation.

The correlations of the factors \(r, \theta,\) and \(s\) with the observables are consistent across the models and make intuitive sense: they are correlated with either the short rate, the slope, or the long rate.\(^{22}\) The only exception is the factor \(\theta\) in the \(A_1(3)\) model, which is correlated with the butterfly spread. The most intriguing result is for the factor \(v:\) it is correlated with different observables for each implementation of the \(A_1(3)\) model.

The volatility factor is highly correlated with the butterfly in the case of \(A_1(3)\). This result is intuitive because the butterfly spread should be sensitive to volatility (see, for instance, Litterman, Scheinkman, and Weiss 1991). In the case of \(A_1(3)\), \(v\) is correlated with the slope, which, in combination with our observation regarding \(\theta\) in this model, raises a suspicion that the model may be misspecified. It appears that factor \(v\) has rotated into the central tendency, and factor \(\theta\) has rotated into the volatility (though the correlations are weaker than for the \(A_1(3)\) counterparts). However, \(\theta\) has almost no correlation with the implied volatility, while its counterpart \(v\) in \(A_1(3)\) has a respectable correlation of 0.24. It appears that when the \(A_1(3)\) model is confronted with the joint dataset, the factors deviate substantially from their intended roles in order to generate a realistic distribution of prices.\(^{23}\)

To summarize, we find that estimated latent state variables do not necessarily behave according to the initial model specification. This is particularly true for the

---

\(^{21}\) The absolute implied variance refers to the variance of absolute changes in the interest rates according to the Black model. The rationale for such volatility measure is that the interest rate in the Black model follows geometric Brownian motion with constant volatility. As a result, the implied volatility coefficient corresponds to the volatility of the relative interest rates. The notion of absolute volatility, or variance, is closer to the volatility factor \(v\) in our affine models.

\(^{22}\) The correlation of these factor with the respective observables deteriorates when the options data are added, which reflects the tension between the requirements to fit the term structure and option prices simultaneously.

\(^{23}\) As noted by Singleton (2005), this outcome is surprising in the light of Umantsev’s (2001) results, which indicate that adding options data results in the close match between the factor \(v\) and option-implied volatility in a \(A_1(3)\) model. It is likely that the divergence in results is explained by the divergence of the estimation strategies: Umantsev assumes that one of the options is priced exactly and inverts the volatility factor from options prices. This approach could lead to divergences in fit elsewhere. However, Umantsev does not explore this possibility.
Table 5 Displays correlations of state variables with observables. We want to determine which aspects of the term structure are reflected in the model state variables. Given this aim, we compute correlations between the model-implied state variables and their most plausible proxies in the data. Correlations in boldface denote the highest correlation for the particular factor.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Observables model</th>
<th>Short rate ( y^f (0.5) )</th>
<th>Long rate ( y^f (10) )</th>
<th>Slope ( y^f (10) - y^f (0.5) )</th>
<th>Butterfly ( y^f (0.5) + y^f (10) - 2y^f (2) )</th>
<th>Implied variance ( \sigma^{imp} (0.5) \cdot y^f (0.5)^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r )</td>
<td>( A_0 (3)^f )</td>
<td>0.94</td>
<td>0.29</td>
<td>-0.74</td>
<td>0.01</td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td>( A_0 (3)^f_o )</td>
<td>0.69</td>
<td>0.14</td>
<td>-0.61</td>
<td>0.45</td>
<td>-0.05</td>
</tr>
<tr>
<td></td>
<td>( A_1 (3)^f )</td>
<td>0.94</td>
<td>0.29</td>
<td>-0.74</td>
<td>0.02</td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td>( A_1 (3)^f_o )</td>
<td>0.75</td>
<td>0.14</td>
<td>-0.69</td>
<td>0.37</td>
<td>-0.08</td>
</tr>
<tr>
<td>( \theta )</td>
<td>( A_0 (3)^f )</td>
<td>0.32</td>
<td>-0.36</td>
<td><strong>0.69</strong></td>
<td>-0.52</td>
<td>-0.31</td>
</tr>
<tr>
<td></td>
<td>( A_0 (3)^f_o )</td>
<td>-0.02</td>
<td>-0.77</td>
<td>-0.68</td>
<td>-0.05</td>
<td>-0.57</td>
</tr>
<tr>
<td></td>
<td>( A_1 (3)^f )</td>
<td>0.30</td>
<td><strong>0.83</strong></td>
<td>0.45</td>
<td>-0.77</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>( A_1 (3)^f_o )</td>
<td>0.72</td>
<td>0.47</td>
<td>-0.34</td>
<td><strong>-0.77</strong></td>
<td>0.04</td>
</tr>
<tr>
<td>( s )</td>
<td>( A_0 (3)^f )</td>
<td>0.48</td>
<td><strong>0.99</strong></td>
<td>0.40</td>
<td>-0.38</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td>( A_0 (3)^f_o )</td>
<td>0.48</td>
<td><strong>0.98</strong></td>
<td>0.38</td>
<td>-0.32</td>
<td>0.55</td>
</tr>
<tr>
<td>( v )</td>
<td>( A_1 (3)^f )</td>
<td>0.02</td>
<td>0.41</td>
<td>0.36</td>
<td><strong>-0.93</strong></td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>( A_1 (3)^f_o )</td>
<td>-0.55</td>
<td>0.31</td>
<td><strong>0.89</strong></td>
<td>0.24</td>
<td>0.43</td>
</tr>
</tbody>
</table>
factor $v$. The evidence emphasizes the fact that an affine model does not necessarily have an intuitive relationship between the continuous-time properties of the factors and the discrete-time properties of the yields. However, these observations do not entail that the affine models fail to achieve their purpose. As pointed out by Collin-Dufresne, Goldstein, and Jones (2009), this outcome could be a result of nonuniqueness in the models specification. Individual factors may not be easy to interpret, but in combination, they could still produce a successful representation of the data.

4.2 Evidence from Futures

In the first step of our analysis, we evaluate the three models that are estimated on the basis of futures data only. This approach is consistent with the overwhelming majority of empirical studies (see Table 1). We also introduce many tests, which will be later applied in the context of the models that are estimated on the basis of the joint dataset.

Pricing errors. In order to evaluate the economic differences between the models, we report the pricing errors in Table 6 as a starting point for the evaluation of the models’ fit. We compute the average of the absolute differences between the model-implied and actual futures yields defined in (21). The futures errors are very small and similar across the models; it would be difficult to argue that there are important differences, even if a statistical test were able to distinguish them.

Perhaps the good pricing performance is not surprising because even two-factor models should be able to fit the cross-section of yields well, using the principal components results of Litterman and Scheinkman (1991) as a basis. Hence, an ability to simultaneously fit the time series well is more relevant for the model evaluation. Our estimation framework, which relaxes the connection between the cross-section and the time series, allows us to implement the time-series diagnostics.

Models’ fit. In order to distinguish the time-series properties of the models more effectively, we test hypotheses that are associated with the relative performance of the models using the conditional likelihood. Typically, a test of nested models is quite informative in locating a parsimonious model that represents the data well. In our case, the models from the $A_0(3)$ and $A_1(3)$ classes are non-nested. Moreover, usual tests assume that one of the models is correct. We suspect that all our models are misspecified, and we would like to take this complication into an account.

The concept of encompassing (see Appendix 5) is very useful in our situation. In the context of this paper, the model $A_0(3)$ encompasses a model $A_1(3)$ if the former is able to explain certain characteristics, that is, statistics, or likelihood values, of the latter. To implement the test, we essentially check whether the $A_1(3)$
Table 6 Reports the average absolute error (reported in basis points) in futures yields $|y_f(\tau, \text{data}) - y_f(\tau, \text{model})|$ and in absolute options-implied volatilities (reported in percent) $|[\sigma^{\text{imp}}(\tau, \text{data}) - \sigma^{\text{imp}}(\tau, \text{model})] \cdot y_f(\tau, \text{data})|$. The line “PC1” reports the percentage of the variation in the errors explained by the first principle component.

<table>
<thead>
<tr>
<th>Maturity</th>
<th>$A_0(3)^f$</th>
<th>$A_1(3)^f$</th>
<th>$A_0(3)^o$</th>
<th>$A_1(3)^o$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Futures, b.p.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>3.91</td>
<td>3.84</td>
<td>7.43</td>
<td>5.32</td>
</tr>
<tr>
<td>1</td>
<td>6.24</td>
<td>6.25</td>
<td>10.18</td>
<td>7.99</td>
</tr>
<tr>
<td>2</td>
<td>3.70</td>
<td>3.74</td>
<td>6.76</td>
<td>5.37</td>
</tr>
<tr>
<td>3</td>
<td>2.92</td>
<td>3.02</td>
<td>4.17</td>
<td>4.58</td>
</tr>
<tr>
<td>4</td>
<td>3.22</td>
<td>3.22</td>
<td>4.97</td>
<td>4.89</td>
</tr>
<tr>
<td>5</td>
<td>3.57</td>
<td>3.55</td>
<td>4.85</td>
<td>4.58</td>
</tr>
<tr>
<td>6</td>
<td>3.44</td>
<td>3.43</td>
<td>3.77</td>
<td>3.53</td>
</tr>
<tr>
<td>7</td>
<td>2.40</td>
<td>2.43</td>
<td>2.06</td>
<td>2.01</td>
</tr>
<tr>
<td>8</td>
<td>1.33</td>
<td>1.37</td>
<td>2.66</td>
<td>2.63</td>
</tr>
<tr>
<td>9</td>
<td>2.84</td>
<td>2.78</td>
<td>5.56</td>
<td>5.34</td>
</tr>
<tr>
<td>10</td>
<td>5.06</td>
<td>5.00</td>
<td>7.83</td>
<td>7.69</td>
</tr>
<tr>
<td>Average</td>
<td>3.51</td>
<td>3.51</td>
<td>5.48</td>
<td>4.90</td>
</tr>
<tr>
<td>Options, %</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.1598</td>
<td>0.1303</td>
<td>0.0353</td>
<td>0.0359</td>
</tr>
<tr>
<td>1</td>
<td>0.1723</td>
<td>0.1218</td>
<td>0.0484</td>
<td>0.0381</td>
</tr>
<tr>
<td>PCI</td>
<td>55.27%</td>
<td>56.33%</td>
<td>47.44%</td>
<td>40.31%</td>
</tr>
</tbody>
</table>

model that is estimated on the basis of data simulated from the $A_0(3)$ model is substantively different from the $A_1(3)$ model that is estimated on the basis of the actual data. It is important to emphasize that this test in no way tells us that one model is “better” than the other one when the null hypothesis is rejected. The rejection shows that the two models are sufficiently different from each other, given the dataset.

It is informative to see whether $A_0(3)$ encompasses $A_1(3)$, given that it is perceived by many researchers as a simpler model because volatility of the short interest rate is not stochastic. Table 7 presents results of the test. The Gaussian model clearly encompasses the stochastic volatility one. This result provides preliminary formal evidence that the $A_0(3)$ model can capture term-structure data as well as the more sophisticated $A_1(3)$. This finding has very important modeling implications. We will explore this further in subsequent sections.

We would like to understand whether the results of the encompassing test for the Gaussian model could be attributed to the lack of power of the test in certain dimensions. Therefore, in the sequel, instead of testing the whole likelihood function, we will focus on particular, practically relevant, aspects of the data distribution and see if we can find differences and similarities between the models.
Table 7 Displays results of encompassing tests. We test whether a model from the $A_0(3)$ branch encompasses a model from the $A_1(3)$ branch, using the finite-sample likelihood-ratio test. The inference is made by simulating 800 samples from the $A_0(3)$ model and estimating the respective $A_1(3)$ model using these samples as a basis. This procedure allows us to compute the LR statistic and its distribution.

<table>
<thead>
<tr>
<th>Test</th>
<th>$LR$</th>
<th>Conf. int.</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_0(3)^f \mathcal{E} A_1(3)^f$</td>
<td>0.008</td>
<td>[-0.011,0.047]</td>
<td>0.490</td>
</tr>
<tr>
<td>$A_0(3)^f \mathcal{E} A_1(3)^f$</td>
<td>1.385</td>
<td>[1.408,7.304]</td>
<td>0.025</td>
</tr>
</tbody>
</table>

Conditional mean and volatility. We now explore abilities of the models to fit conditional means and volatilities. This exercise is motivated by the challenges, highlighted by Dai and Singleton (2003), to the ability of the term-structure models to fit these moments. Typically, the ability to fit the conditional mean is cast in terms of the expectation hypothesis puzzle—challenge $LPY$. The second challenge, $CVY$, is related to a model’s ability to replicate (i) the time-varying volatility of the yields and (ii) the hump-shaped term structure of the unconditional volatility.

In our sample, the $LPY$ metric is not useful because of the high statistical uncertainty: the confidence bounds are so wide that it is not possible to distinguish the implications of different models. As an alternative, we estimate a descriptive model, which simultaneously captures conditional means and volatilities of the yield curve components. Subsequently, we can see how well our no-arbitrage models capture the patterns generated by the descriptive model. Thus, we will be able to address the $LPY$ and part (i) of the $CVY$ challenges. We will consider part (ii) of $CVY$ in the next section.

In theory, one can evaluate $CVY$-(i) simply by observing that the $A_0$ models can generate only constant conditional volatility. For instance, Jacobs and Karoui (2009) who compare three-factor affine models with respect to their ability to capture conditional volatilities of interest rates do not even consider the $A_0$ case because it produces constant conditional volatilities. Also, Almeida, Graveline, and Joslin (2006) do consider the $A_0(3)$ model when analyzing conditional volatilities but attach a constant value to its conditional volatility. As a result, with respect to this metric, the $A_0(3)$ model should be classified as misspecified.

However, here we ask whether the futures-only dataset provides information that is sufficiently powerful to distinguish between conditional volatility generated by the $A_0(3)$ model (constant in theory) and by the $A_1(3)$ model (stochastic in theory). Specifically, the recent trend in the literature is to use the GARCH as a descriptive, or “model-free,” estimate of conditional volatility (see, for instance, Collin-Dufresne, Goldstein, and Jones 2009; Dai and Singleton 2003; Umantsev

\[24\] Our results are available upon request. Collin-Dufresne, Goldstein, and Jones (2009) report qualitatively similar results.
2001). The idea is to estimate the GARCH on the actual data, and on the model-implied series, and to compare the implied properties of the conditional volatilities. We seek to determine if such a diagnostic can tell constant and stochastic volatility models apart in finite samples.

Because GARCH is still a model, one has to be careful when interpreting the results. To streamline the discussion, consider the basic GARCH(1,1) model:

\[ Y_t = \mu_{t-1} + \sigma_t e_t \]  \hspace{1cm} (25)
\[ \sigma_t^2 = c + \alpha \sigma_{t-1}^2 e_{t-1}^2 + \beta \sigma_{t-1}^2 \]  \hspace{1cm} (26)

where \( Y \) generically denotes the series of interest. The estimated values of parameters that govern the volatility dynamics (\( c, \alpha, \) and \( \beta \)) will be sensitive to the estimated conditional mean \( \mu \). Collin-Dufresne, Goldstein, and Jones (2009) focus directly on the dynamics of the state variable \( r \), that is, \( Y_t = r_t - r_{t-1} \), and compute \( \mu \) from the estimated ATSM model. As a result, the GARCH volatility becomes model dependent: different specifications of the affine models invariably produce different estimates of \( r \) and \( \mu \). It appears that it is more appropriate to focus on the volatility of the yields themselves because this leads to a truly ATSM-independent estimates of conditional volatility. For instance, Dai and Singleton (2003) estimate a univariate GARCH model for five year yields.

Despite its shortcomings, the advantage of the Collin-Dufresne, Goldstein, and Jones (2009) approach is in the potentially rich specification of \( \mu \), which should help in addressing the challenge LPY and leads to a more accurate estimate of the noise term \( e_t \). We combine this flexibility with the robustness of the actual yields in the role of \( Y \) by estimating conditional means and volatilities from a VAR(1)-GARCH(1,1) model applied to a trivariate series, which include the short rate, the slope, and the butterfly spread. These three series are highly correlated with the first three principal components and therefore contain an informative summary of the entire term structure.

The results of a Portmanteau test offer the strongest evidence for heteroscedasticity of residuals, \( Y_t - \mu_{t-1} \), in the slope, and the weakest one in the short rate. Nonetheless, we were able to estimate the GARCH models for all three series. Table 8 reports correlations between the means and volatilities that are implied by the data and those that are implied by the models.

We see that the \( A_0(3) \) and \( A_1(3) \) models produce almost identical estimates of the conditional means and volatilities. Moreover, both models fit the first two moments quite well, especially with respect to the level and slope. Interestingly, despite a clear difficulty in identifying the role of the instantaneous variance of the

\[ ^{25} \text{In spirit, such a diagnostic is similar to the reprojection technique of Gallant and Tauchen (1998).} \]
Table 8 Reports correlations between the conditional mean and volatility computed from the data and the respective moments from the models.

<table>
<thead>
<tr>
<th>Level</th>
<th>Slope</th>
<th>Curvature</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>VAR(1) GARCH(1,1)</td>
</tr>
<tr>
<td>$A_0(3)^f$</td>
<td>0.9882</td>
<td>0.9001</td>
</tr>
<tr>
<td>$A_1(3)^f$</td>
<td>0.9878</td>
<td>0.9010</td>
</tr>
<tr>
<td>$A_0(3)^fo$</td>
<td>0.9701</td>
<td>0.8030</td>
</tr>
<tr>
<td>$A_1(3)^fo$</td>
<td>0.9811</td>
<td>0.8133</td>
</tr>
</tbody>
</table>
short rate (see our discussion of Table 5), the models that were estimated on the
futures do a remarkable job in explaining the variance of the yields, irrespective of
their factor structure.

One explanation of such a finding is that the degree of the variation in the
conditional volatility is so small that we cannot reject the \( A_0 \) in favor of the \( A_1 \)
models based the GARCH series. An alternative view is that the similar perfor-
mance is built in because pricing errors are small for all models and the conditional
mean and variance are, in essence, functions of yields.\(^{26}\)

In order to address the first possibility, we implement additional diagnostics,
which are described in Appendix 5. The results indicate that, indeed, the Gauss-
ian and stochastic volatility models cannot be distinguished on the basis of the
low-order conditional moments of the term structure. This implies that for many
purposes, such as intra-regime modeling or the study of low-frequency data, the
Gaussian models are sufficient.\(^{27}\) We emphasize that Gaussian models are suffi-
cient not because they are well specified but because of the grounds of parsimony.
The conditional variation of futures, as reflected in the futures-only data, is con-
taminated with too much noise to warrant a more complex model. This finding
has important practical implications because it is easier to compute asset prices
and to estimate in the framework of Gaussian models.

In general, the second possibility could not be true because we estimate
the models by taking into an account the time series and cross-sectional char-
acteristics of the data simultaneously. Therefore, time-series implications cannot
follow from the cross-sectional properties alone. However, it might be that the
likelihood overweighs the information in the cross-section, which leads to the ob-
served performance.

Our findings complement the CVY analysis in Dai and Singleton (2003). These
authors find that a term-structure model requires at least three factors to generate
a sufficient degree of time variation in the volatility. We conclude that it does not
matter which three factors are used.

**Term structure of volatility.** As we mentioned, the CVY-(ii) challenge requires
a term-structure model to generate the hump-shaped term structure of the uncon-
ditional volatilities observed after 1983. The upper left panel of Figure 1 shows the
unconditional volatilities of yield changes, computed as standard deviations di-
rectly from the data (with bootstrapped confidence bounds), and population val-
ues that are implied by various models.

In the data, we see the hump that was documented for the swaps and Treas-
ury rates. The \( A_0(3)/f \) and \( A_1(3)/f \) models generate very similar term structures,

\(^{26}\)We are grateful to Chris Jones and Ken Singleton for an extensive discussion and suggestions on this
issue.

\(^{27}\)In particular, this finding addresses the concern expressed in Bansal, Tauchen, and Zhou (2004) regarding
which have magnitudes on a par with the data; however, the hump is much less pronounced. Again, this similarity is not an indication that $A_0(3)$ is well specified. Rather, it is an indication that the futures data are not informative enough to distinguish this model from its more sophisticated counterpart.

It is clear that we cannot distinguish the Gaussian and stochastic volatility term structures based on the first two moments. The next natural step is to evaluate the ability of the models to fit higher-order moments.

**Higher-order moments.** The stochastic volatility models are unconditionally not normal. Therefore, we might be able to distinguish them from the Gaussian models if we focus on their ability to match skewness and kurtosis. We continue with the same three series from the previous subsection: level, slope, and curvature. We compute their skewness and kurtosis in the sample and their finite-sample distribution via the parametric bootstrap. The skewness of all three series is very close to zero, which is consistent with what all models generate. Therefore, we do not report the skewness results and focus on kurtosis in Table 9. It turns out

![Figure 1](https://example.com/figure1.png)

Figure 1 Plots the population standard deviation and average implied volatility, evaluated at the estimates of the respective models. We compare these with the standard deviation (with bootstrapped confidence bounds) and the average implied volatility computed from our sample.
Table 9 Reports results of kurtosis tests. We test whether the estimated models can match the unconditional kurtosis of changes in the principal components of the log-futures ($y_f(\tau)$). We construct finite-sample 95% confidence intervals by simulating 800 sample paths from each model and evaluating the kurtosis along each path. We compare the intervals with the sample values. Statistics presented in bold face indicate a failure to reject the null hypothesis that a model can match a statistic.

<table>
<thead>
<tr>
<th>Sample model</th>
<th>Level $y_f(0.5)$</th>
<th>Slope $y_f(10) - y_f(0.5)$</th>
<th>Curvature $y_f(0.5) + y_f(10) - 2y_f(2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_0(3)^f$</td>
<td>2.98 [2.38, 3.51]</td>
<td>2.99 [2.38, 3.56]</td>
<td>2.98 [2.36, 3.53]</td>
</tr>
<tr>
<td>$A_1(3)^f$</td>
<td>3.02 [2.60, 3.59]</td>
<td>3.00 [2.61, 3.58]</td>
<td>2.99 [2.58, 3.55]</td>
</tr>
<tr>
<td>$A_0(3)^{fo}$</td>
<td>3.00 [2.60, 3.54]</td>
<td>2.98 [2.58, 3.55]</td>
<td>2.98 [2.58, 3.54]</td>
</tr>
<tr>
<td>$A_1(3)^{fo}$</td>
<td><strong>3.46 [2.77, 4.77]</strong></td>
<td>3.17 [2.66, 4.03]</td>
<td>3.00 [2.58, 3.60]</td>
</tr>
</tbody>
</table>

that it is extremely difficult for all models, even non-Gaussian ones, to generate the relatively high degree of kurtosis observed in the term structure.

4.3 Evidence from Options

We found that we cannot distinguish the Gaussian and stochastic volatility models using either economic or statistical criteria when only yield data are involved. Given that options prices are sensitive to volatility, we should be able to achieve greater success in distinguishing the different models. We proceed following the same outline as in the futures-only case.

Pricing errors. Table 6 reports pricing errors for yields and options. The latter are measured by the absolute difference between absolute Black-implied volatilities. Hence, before proceeding with the options-based estimation, we can evaluate how well the models estimated using futures-only value options as a basis. $A_1(3)^f$ appears to price options better than $A_0(3)^f$.

Proceeding with the options-based estimation, we note that the futures pricing performance is still quite similar across the models. However, we can see that $A_1(3)^{fo}$ improves upon $A_0(3)^{fo}$ for short maturities. The futures errors generally increase relative to the models that were estimated on futures only. This outcome is not surprising, given the increased pressure on the same model to fit a more
elaborate dataset.\textsuperscript{28} The increase in futures errors is associated with an impressive improvement in option pricing. The pricing errors decline three-to-four-fold.\textsuperscript{29}

**Models’ fits.** The encompassing test results in Table 7 are striking. While in the futures-only case the Gaussian model encompasses the stochastic volatility one, the null hypothesis is rejected when the options data are added. The result perfectly matches our initial intuition that options data will be helpful in distinguishing between the models. We now know that the models generate different patterns in practice, but the encompassing test does not tell us which one is more realistic. We will implement other diagnostics to establish this.

**Conditional mean and volatility.** Table 8 shows that the $A_0(3)$ and $A_1(3)$ models again produce almost identical estimates of the conditional means and volatilities. However, the correlations with the moments implied from the descriptive model, while still strong, decline relative to the futures-only counterparts. This outcome is consistent with the increase in the pricing errors observed earlier.

**Term structure of volatility.** The upper right panel of Figure 1 compares the unconditional term structure of standard deviation inferred from various models. We see that the $A_0(3)^{fo}$ term structure lies outside of the confidence bounds longer maturities (after two years), while the term structure implied by the $A_1(3)^{fo}$ model implies a realistic pattern.

Due to the fact that we are using options data, we can evaluate the term structure of implied volatilities as well. Backus, Li, and Wu (1998) document the hump-shaped term structure of cap-implied volatilities. They propose univariate ARMA models to capture this effect. Attari (2001) proposes four-factor affine models to replicate the upward-sloping term structure at short horizons. It would be interesting to see whether our three-factor models are able to generate the hump in implied volatilities.

The lower panels of Figure 1 show the population values of the average implied volatilities that are inferred from the models and average implied volatilities that are computed from the data (solid line). The data are not very informative because we used six month and one year options, which capture only the upward slope at the short horizons. Therefore, we can only see how the models differ from each other. The story is similar to the one for standard deviations. The $A_0(3)$ and $A_1(3)$ models are close to each other within each type of dataset and generate

\textsuperscript{28}This result is different from Jagannathan, Kaplin, and Sun (2003), who report no change in the swaps fit once the caps data are added. Given that their parameter values do not change, we suspect that they obtained a local optimum.

\textsuperscript{29}Umantsev (2001) also observes a dramatic improvement in the options fit of the $A_1(3)$ model once the options data is used for estimation. However, he does not acknowledge the deterioration of fit in the underlying security.
fairly reasonable volatility values. The models estimated using options generate
the hump (the one in $A_1(3)^{f_0}$ is more pronounced).

**Higher-order moments.** The kurtosis tests in Table 9 indicate that the stochastic
volatility model can fit the kurtosis of the level and that none of the models can fit
the kurtosis of the slope and curvature.

### 4.4 Futures versus Options

Two trends emerge from our analysis. First, information in option prices clearly
helps us distinguish the models better. There is virtually no difference between
the models when they are estimated on the basis of the futures data alone. This
finding has important implications for the yield curve modeling strategy: one can
use a simple Gaussian model, at least within a regime. On the other hand, we
can clearly say that the $A_1(3)$ model emerges as a leader on the basis of the joint
dataset. The model improves upon the $A_0(3)$ one in terms of pricing short-dated
contracts and has more flexibility to match the unconditional kurtosis of the short
rate and the volatility term structure.

Second, we observe a clear tension in the models’ ability to fit lower- versus
higher-order moments: the models fitted to the term structure do an excellent job
fitting $\text{VAR}(1)-\text{GARCH}(1,1)$, but they fail with kurtosis. The performance of the
models that are fitted to the joint dataset is just the opposite. Unfortunately, as
Bikbov and Chernov (2008) show, the USV model does not resolve this tension,
despite the fact that it was designed to do so.

### 5 CONCLUSION

We evaluated two affine term-structure models that differ in the volatility speci-
fication: constant versus stochastic. We used the prices of Eurodollar futures and
options to conduct comprehensive statistical and economic analyses. We found
that the futures-only data (the information in the yield curve) is not sufficient to
separate the models. Therefore, we conclude that the simplest, constant volatility,
specification is preferred. Adding options data allows us to separate the models
in a very clean fashion. The stochastic volatility model performs favorably, on the
basis of statistical tests and pricing errors.

**Appendix A: Bond Pricing ODEs**

We use the general setup of Duffie, Pan, and Singleton (2000). Assume that the
state $S$ is described by the following diffusion under the probability measure $Q$:

$$dS_t = \mu(S_t)dt + \sigma(S_t)dW_t(Q), \quad (A1)$$
where

\[ \mu(s) = \kappa_0^Q + \kappa_1^Q s \]  \hspace{1cm} (A2)

\[ (\sigma(s)\sigma(s)^\top)_{ij} = (\Sigma_0)_{ij} + (\Sigma_1)_{ij} \cdot s \]  \hspace{1cm} (A3)

and the spot interest rate

\[ r(s) = \rho_0 + \rho_1 \cdot s. \]  \hspace{1cm} (A4)

In the models considered in this paper, \( \text{dim}(S_t) = 3 \), \( \rho_0 = 0 \), and \( \rho_1 = (1, 0, 0)^\top \). Further, in the case of the \( A_0(3) \) model, Equations (5)–(7), (11), and (12)–(14) imply that:

\[
\begin{align*}
\text{dr}_t & = \left( \gamma^Q_r - \kappa^Q_r r_t - \kappa^Q_{r\theta} \theta_t - \kappa^Q_{rs} s_t \right) dt \\
& + \sigma_r dW^r_t(Q) + \sigma_{r\theta} dW^\theta_t(Q) + \sigma_{rs} dW^s_t(Q) \\
\text{d}\theta_t & = \left( \gamma^Q_\theta - \kappa^Q_{r\theta} r_t - \kappa^Q_{\theta\theta} \theta_t - \kappa^Q_{\theta s} s_t \right) dt + \sigma_\theta dW^\theta_t(Q) + \sigma_{\theta s} dW^s_t(Q) \\
\text{ds}_t & = \left( \gamma^Q_s - \kappa^Q_{s r} r_t - \kappa^Q_{s\theta} \theta_t - \kappa^Q_{s s} s_t \right) dt + \sigma_s dW^s_t(Q)
\end{align*}
\]  \hspace{1cm} (A5)

where

\[
\begin{align*}
\gamma^Q_r & = -\lambda_r \sigma_r - \lambda_\theta \sigma_\theta \sigma_\theta - \lambda_s \sigma_{rs} \\
\kappa^Q_r & = \kappa^P_r + \lambda_{rr} \sigma_r + \lambda_{r\theta} \sigma_\theta \sigma_\theta + \lambda_{rs} \sigma_{rs} \\
\kappa^Q_{r\theta} & = -\kappa^P_{r\theta} + \lambda_{r\theta} \sigma_r + \lambda_{\theta\theta} \sigma_\theta \sigma_\theta + \lambda_{\theta s} \sigma_{rs} \\
\kappa^Q_{rs} & = -\kappa^P_{rs} + \lambda_{rs} \sigma_r + \lambda_{\theta s} \sigma_\theta \sigma_\theta + \lambda_{ss} \sigma_{ss} \\
\gamma^Q_\theta & = \kappa^P_\theta - \lambda_\theta \sigma_\theta - \lambda_s \sigma_{\theta s} \\
\kappa^Q_{r\theta} & = \lambda_{r\theta} \sigma_\theta + \lambda_{s r} \sigma_{\theta s} \\
\kappa^Q_\theta & = \kappa^P_\theta + \lambda_{\theta\theta} \sigma_\theta + \lambda_{s\theta} \sigma_{\theta s} \\
\kappa^Q_{\theta s} & = \lambda_{\theta s} \sigma_\theta + \lambda_{ss} \sigma_{\theta s} \\
\gamma^Q_s & = -\lambda_s \sigma_s \\
\kappa^Q_{s r} & = \lambda_{s r} \sigma_s \\
\kappa^Q_{s\theta} & = \lambda_{s\theta} \sigma_s \\
\kappa^Q_s & = \kappa^P_s + \lambda_{ss} \sigma_s
\end{align*}
\]  \hspace{1cm} (A6)

Therefore, \( \kappa_0^Q = (\gamma^Q_r, \gamma^Q_\theta, \gamma^Q_s)^\top \) and

\[
\kappa_1^Q = - \begin{pmatrix}
\kappa^Q_r & \kappa^Q_{r\theta} & \kappa^Q_{rs} \\
\kappa^Q_{r\theta} & \kappa^Q_\theta & \kappa^Q_{\theta s} \\
\kappa^Q_{s r} & \kappa^Q_{s\theta} & \kappa^Q_s
\end{pmatrix}. \]  \hspace{1cm} (A20)
Finally, $\Sigma_1 = 0$ and

$$\Sigma_0 = \begin{pmatrix} \sigma_r & \sigma_{r\theta} & \sigma_{rs} \\ 0 & \sigma_\theta & \sigma_{\theta s} \\ 0 & 0 & \sigma_s \end{pmatrix} \begin{pmatrix} \sigma_r & \sigma_{r\theta} & \sigma_{rs} \\ 0 & \sigma_\theta & \sigma_{\theta s} \\ 0 & 0 & \sigma_s \end{pmatrix}^\top$$

(A21)

Similarly, Equations (8)–(10), (11), and (15)–(17) imply that under the measure $Q$, the $A_1(3)$ model evolves according to:

$$d r_t = \left( \gamma_r^Q - \kappa_r^Q r_t - \kappa_{r\theta}^Q \theta_t - \kappa_{rs}^Q \sigma_t \right) dt$$

$$+ \sqrt{\sigma_r^2 + \nu_t W_t^Q (Q)} + \sigma_{r\theta} \sqrt{\sigma_\theta^2 + \beta_\theta \nu_t W_t^\theta (Q)} + \sigma_{rs} \sqrt{\nu_t W_t^\sigma (Q)} (A22)$$

$$d \theta_t = \left( \gamma_\theta^Q - \kappa_{r\theta}^Q r_t - \kappa_{\theta \theta}^Q \theta_t - \kappa_{\theta \nu}^Q \nu_t \right) dt$$

$$+ \sigma_{\theta r} \sqrt{\sigma_r^2 + \nu_t W_t^Q (Q)} + \sqrt{\sigma_\theta^2 + \beta_\theta \nu_t W_t^\theta (Q)} + \sigma_{\theta \nu} \sqrt{\nu_t W_t^\sigma (Q)} (A23)$$

$$d \nu_t = \left( \kappa_\nu^P \nu_t - \kappa_\nu^Q \nu_t \right) dt + \sigma_\nu \sqrt{\nu_t W_t^\nu (Q)}$$

(A24)

where

$$\gamma_r^Q = \kappa_r^P \bar{\theta} - \lambda_{r \theta} \sigma_r^2 - \lambda_\theta \sigma_\theta^2 \sigma_{r \theta}$$

(A25)

$$\kappa_r^Q = \kappa_r^P + \lambda_{r r} + \lambda_{r \theta} \sigma_{r \theta}$$

(A26)

$$\kappa_{r \theta}^Q = - \kappa_r^P + \lambda_{r \theta} + \lambda_{r \theta} \sigma_{r \theta}$$

(A27)

$$\kappa_{r \nu}^Q = \kappa_{r \nu}^P + \lambda_r + \lambda_{r \nu} \sigma_{r \nu} + \lambda_\theta \sigma_{r \theta} + \lambda_{\theta \nu} \sigma_{\theta \nu}$$

(A28)

$$\gamma_\theta^Q = \kappa_\theta^P \bar{\theta} - \lambda_{r \theta} \sigma_r^2 \sigma_{r \theta} - \lambda_\theta \sigma_\theta^2$$

(A29)

$$\kappa_{\theta r}^Q = \lambda_{r r} \sigma_{r \theta} + \lambda_{r \theta}$$

(A30)

$$\kappa_\theta^Q = \kappa_\theta^P + \lambda_{r \theta} \sigma_{r \theta} + \lambda_\theta$$

(A31)

$$\kappa_{\theta \nu}^Q = \lambda_r \sigma_{r \theta} + \lambda_{r \nu} \sigma_{r \nu} + \lambda_\theta \beta_\theta + \lambda_{\theta \nu} \sigma_{\theta \nu}$$

(A32)

$$\kappa_\nu^Q = \kappa_\nu^P + \lambda_{r \nu} \sigma_{r \nu}$$

(A33)

Therefore, $\kappa_0^Q = (\gamma_r^Q, \gamma_\theta^Q, \kappa_\nu^Q) \top$ and

$$\kappa_1^Q = - \begin{pmatrix} \kappa_r^Q & \kappa_{r \theta}^Q & \kappa_{r \nu}^Q \\ \kappa_{r \theta}^Q & \kappa_\theta^Q & \kappa_{\theta \nu}^Q \\ 0 & 0 & \kappa_\nu^Q \end{pmatrix}.$$ 

(A34)

Finally,

$$\Sigma_0 = \begin{pmatrix} \sigma_r^2 + \sigma_{r \theta}^2 \sigma_{\theta \theta}^2 & \sigma_{r \theta}^2 \sigma_{r \theta} + \sigma_{r \theta} \sigma_{r \theta} & 0 \\ \sigma_{r \theta} \sigma_{r \theta} + \sigma_{r \theta} \sigma_{r \theta} & \sigma_{r \theta}^2 + \sigma_{r \theta}^2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

(A35)
and

$$
\Sigma_{1,ij3} = \begin{pmatrix}
1 + \sigma^2_{\theta_0} + \sigma^2_{\theta_0} & \sigma_{\theta r} + \sigma^2_{\theta_0} \sigma_{\theta_0} & \sigma_{\theta r} + \sigma^2_{\theta_0} \sigma_{\theta_0} \\
\sigma_{\theta r} + \sigma^2_{\theta_0} \sigma_{\theta_0} & \sigma^2_{\theta_0} + \beta^2_{\theta_0} + \sigma^2_{\theta_0} & \sigma_{\theta_0} \sigma_{\theta_0} \\
\sigma_{\theta r} + \sigma^2_{\theta_0} \sigma_{\theta_0} & \sigma_{\theta r} + \sigma^2_{\theta_0} \sigma_{\theta_0} & \sigma^2_{\theta_0}
\end{pmatrix}
$$

(A36)

and $\Sigma_{1,ij1} = \Sigma_{1,ij2} = 0$.

This setup implies that Equation (18) solves the following system of Ricatti ODEs:

$$
- \frac{\partial B^P(d)}{\partial d} = \rho_1 - K_1^{\top} B^P(d) - \frac{1}{2} B^{P\top}(d) \Sigma_1 B^P(d)
$$

$$
- \frac{\partial A^P(d)}{\partial d} = -K_0^Q \cdot B^P(d) - \frac{1}{2} B^{P\top}(d) \Sigma_0 B^P(d).
$$

Appendix B: Option Pricing Formula

First, we need an additional computation. Using (19), we get

$$
\left(k - f_{t+\tau_c}(\tau_f - \tau_c)\right)^+ = \left(k + \frac{1}{d} - \frac{1}{d} e^{-A^P(d) + A^f(\tau_f - \tau_c) + B^f(\tau_f - \tau_c) \cdot S_{t+\tau}}\right)^+
$$

(B1)

Substituting this back into (22), we obtain

$$
C_t(\tau_f, \tau_c, K) = \frac{1}{d} e^{-A^P(d) + A^f(\tau_f - \tau_c)} E_t^Q \left(e^{-\int_{t+\tau_c}^{T} r_s ds} \left(X - e^{B^f(\tau_f - \tau_c) \cdot S_{t+\tau}} \mathbf{1}\{B^f(\tau_f - \tau_c) \cdot S_{t+\tau} \leq \log X\} \right) \right)
$$

$$
= \frac{1}{d} e^{-A^P(d) + A^f(\tau_f - \tau_c)} \times \left(X G_0(\tau_f - \tau_c)(\log X, S_t, t + \tau_c) \right)
$$

(B2)

where

$$
G_{a,b}(y, S_t, T) = E_t^Q \left(e^{-\int_t^T r_s ds} e^{A^f S_{t+T}} \mathbf{1}\{b \cdot S_{t+T} \leq y\} \right)
$$

(B3)

which is computed via the extended characteristic function:

$$
\psi(u, S_t, t, T) = E_t^Q \left(e^{-\int_t^T r_s ds} e^{A^f S_T} \right) = e^{A(T-t) + B(T-t) \cdot S_t}
$$

(B4)
with boundary conditions \( A(0) = 0, B(0) = u \). Given \( \psi \),

\[
G_{a,b}(y, S_t, T) = \frac{\psi(a, S_t, t, T)}{2} - \frac{1}{\pi} \int_0^\infty \Im \left( \frac{\psi(a + ivb, S_t, t, T) e^{-ivy}}{v} \right) dv \quad (B5)
\]

Standard numerical integration techniques allow us to compute the integral.

**Appendix C: Encompassing**

We largely follow Gourieroux and Monfort (1994) in our exposition of encompassing. See Dhaene (1997) for more details.

Suppose the true data-generating process is associated with the log-likelihood \( L_d \). We are trying to evaluate the proximity between two non-nested and potentially misspecified models with log-likelihoods \( L_0 \), that is, \( A_0(3) \), and \( L_1 \), that is, \( A_1(3) \). The maximum likelihood estimation of the respective sets of parameters \( \Theta_0 \) and \( \Theta_1 \) using the sample \( y \) is equivalent to minimization of the Kullback–Leibler information criterion (KLIC):

\[
\hat{\Theta}_i = \arg \Theta_i \max E^d [L_d (y, \Theta_i)] = \arg \Theta_i \min E^d [L_i (y, \Theta_i)] \quad (C1)
\]

where superscript \( d \) emphasizes the true distribution. That is why this estimate is often referred to as the pseudo-true value.

A binding function links the two misspecified models in a manner similar to the way in which the pseudo-true value links a misspecified model to the true distribution:

\[
b_{10}(\hat{\Theta}_0) = \arg \Theta_1 \min E^0 [L_0 (y, \hat{\Theta}_0) - L_1 (y, \Theta_1)] \\
= \arg \Theta_1 \max E^0 [L_1 (y, \Theta_1)] \quad (C2)
\]

where superscript \( 0 \) emphasizes the distribution associated with \( A_0(3) \). In other words, a binding function is the value of parameter \( \Theta_1 \) such that the distribution of \( A_1(3) \) approximates the distribution of \( A_0(3) \) in the best way possible (according to KLIC), or, equivalently, this is an estimate of \( \Theta_1 \) if \( L_0 \) were the true distribution.

We say that the true distribution \( L_d \) is such that \( A_0(3) \) encompasses \( A_1(3) \), \( A_0(3) \) encompasses \( A_1(3) \), if and only if the pseudo-true value coincides with the binding function, \( \hat{\Theta}_1 = b_{10}(\hat{\Theta}_0) \). Put less formally, \( A_0(3) \) encompasses \( A_1(3) \) if the objects of interest that are associated with the latter behave as they should if \( L_0 \) were the true distribution. Thus, similar to nested hypothesis testing, the encompassing model is preferred if it is simpler.

It is natural to consider a likelihood-ratio test of encompassing. The implementation of the tests was hindered by complicated asymptotic distributions and the explicit knowledge of the binding functions. Given that we here rely on finite-sample inference, it is easy to avoid these complications. The finite-sample
inference is conducted by simulating 800 samples from the $A_0(3)$ models and estimating the respective $A_1(3)$ models using these samples as a basis. This procedure allows us to compute the LR statistic and its distribution.

### Appendix D: Distinguishing Different Models with Unobservable States

The futures-based results in Section 4 indicate that models with conditional constant and stochastic volatility specifications generate similar conditional volatility patterns. The natural question is whether one can distinguish the different models when the states are unobservable. Given that one has to rely on various data sources to estimate models and filter state variables, the answer to this question will depend on the informativeness of the particular dataset. As we have seen, futures data definitely do not furnish information that would allow us to distinguish the Gaussian and stochastic volatility models: the diagnostic is not powerful enough given the data. These conclusions are fully consistent with the results of encompassing tests.

We see two ways to resolve this problem. One way, at which our discussion hints, is to rely on additional data, which would enhance the power of our diagnostics. This is precisely the direction that we explore in this paper when we incorporate options data. Alternatively, we can explore whether a model can, in principle, generate the amount of heteroscedasticity found in the data by simulating samples taken from the estimated model. We proceed in the same fashion as with the encompassing tests. We simulate 800 samples from the respective models and reestimate GARCH along each of the paths. This procedure allows us to construct finite-sample confidence intervals for the GARCH coefficients.

First of all, one cannot estimate a GARCH model using the data simulated from the $A_0$ class. Given that conditional volatility is constant, parameters governing the GARCH dynamics in (26) are not identified: both sets $\alpha = \beta = 0$ and $c = \alpha = 0$ with $\beta = 1$ produce constant volatility. Therefore, we restrict $\beta$ to zero to allow the model to pick up some heteroscedasticity due to small sample noise via the ARCH component. We find that the 95% confidence bounds that were constructed on the basis of the 800 simulated paths cover the values of $\alpha$ obtained from the sample. However, they also cover zero. This finding is not surprising because we simulated futures rates from a constant volatility model.

A more interesting question is whether the $A_1(3)$ model can replicate the levels of GARCH in the data. Our finding is basically the same: the confidence intervals are wide enough to cover the value of the parameters obtained from the real data.

---

30 Due to the computational complexity, we estimated univariate GARCH models for each of the principal components. To be consistent, we also estimate a univariate GARCH in the data for the purposes of this exercise.

31 The values of $\alpha$ in the sample ranged from 0.07 for the curvature to 0.27 for the slope.
data and to cover zero. We believe that the intervals are so wide because the data are not sufficiently informative about this aspect of the model.

Received March 7, 2009; revised April 14, 2010; accepted April 14, 2010.

REFERENCES


