

Term Structure and Volatility:  
Lessons from the Eurodollar Futures and Options\*

Ruslan Bikbov  
Columbia Business School<sup>†</sup>

Mikhail Chernov  
Columbia Business School<sup>‡</sup>

First Draft: August 2002

This Revision: November 23, 2005

---

\*We would like to thank Andrew Ang, Darrell Duffie, Steve Figlewski, Kris Jacobs, Mike Johannes, Sergei Levendorskii, Tano Santos, Ken Singleton as well as participants of workshops at Columbia, UNC, Bank of Canada for helpful comments.

<sup>†</sup>Division of Finance and Economics, 6T Uris Hall, 3022 Broadway, New York, NY 10027, USA, Email: rb2015@columbia.edu

<sup>‡</sup>Division of Finance and Economics, 413 Uris Hall, 3022 Broadway, New York, NY 10027, USA, Phone: (212) 854-9009, Fax: (212) 316-9180, Email: mc1365@columbia.edu

# Term Structure and Volatility: Lessons from the Eurodollar Futures and Options

## **Abstract**

We evaluate the statistical and economic differences between affine term structure models. Despite voluminous literature on this subject, we have a limited understanding of the models' structural features which are important in practice. Since the key distinguishing characteristic of the affine models is the specification of conditional volatility of the factors, we explore models which have critical differences in this respect: Gaussian (constant volatility), stochastic volatility, and unspanned stochastic volatility models. We estimate the models based on the Eurodollar futures and options data. We subject these models to an exhaustive set of diagnostics. We find that there is little difference between the models based on the statistical tests and pricing errors when the models are estimated using only the yield curve information. Options provide us with a very clear lens through which we separate the models. The stochastic volatility model is the most successful according to our diagnostics. Economically, we find that the stochastic volatility model generates average absolute pricing errors, in terms of the Black implied volatility, that are up to six times smaller than that of the unspanned stochastic volatility model. We find that small measurement errors help to account for the weak relationship between volatility and yields.

# 1 Introduction

We evaluate the statistical and economic differences between affine term structure models. Despite voluminous literature on this subject, we have a limited understanding of the models' structural features which are important in practice. Table 1 summarizes the major empirical contributions. Unfortunately, there are few conclusions, in which the studies do not diverge.<sup>1</sup> One can observe that the three primary reasons for the inconclusive evidence are different models being compared from paper to paper, different data frequency and type being used, and different evaluation metrics.

This paper addresses the shortcomings by comparing the most promising, judging by the results in Table 1, models and subjecting these models to a comprehensive set of tests and diagnostics. Since the key distinguishing characteristic of the affine models is the specification of conditional volatility of the factors, we explore models which have critical differences in this respect: Gaussian (constant volatility), stochastic volatility (SV), and unspanned stochastic volatility (USV) models. We abstract from comparing models with more than three factors for the reasons of parsimony.<sup>2</sup>

We depart from most existing studies by complementing the yield curve data with option prices, which are by their nature sensitive to volatility specification.<sup>3</sup> We use a rich dataset based on the Eurodollar futures and options. These data have a number of unique qualities which makes 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). Quarterly maturity cycle of the futures allows us to use many points along the term structure.

Since it is not hard 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

---

<sup>1</sup>There is one aspect that many studies seem to agree on: the models with more than one stochastic volatility factor do not have sufficiently flexible correlation to describe the yield curve.

<sup>2</sup>Some authors prefer the quadratic models to the affine ones. However, as Cheng and Scaillet (2005) show, quadratic models are equivalent to restricted affine models with many factors, e.g. a typical three-factor quadratic model considered in the literature is equivalent to a nine-factor restricted affine model.

<sup>3</sup>Dai and Singleton (2003) remark on "enormous potential for new insights from using derivatives data in model estimations." Casassus, Collin-Dufresne, and Goldstein (2004) conclude: "an open question is how well USV models of the term structure account for the joint dynamics of bond and fixed income derivatives prices."

series of yields by not imposing restrictions which arbitrarily require certain futures or option prices to be observed exactly.<sup>4</sup> We use quasi-maximum likelihood with Kalman filter for model estimation because the state variables cannot be inverted from prices in this setting. Kalman filter has undesirable asymptotic properties when state variables have stochastic volatility or asset prices have non-linear dependence on state variables as is the case with options. We conduct a series of comprehensive Monte Carlo studies which verify that the methodology performs well in practice at least in the term structure setting. Since persistence of the interest rate can affect the asymptotic properties of our tests, we are careful to conduct a finite sample inference via parametric bootstrap (Conley, Hansen, and Liu, 1997) to compute standard errors and implement tests throughout the paper.

We find, based on the futures data, that there is virtually no difference between the three models based on the variety of diagnostics, such as the likelihood ratio tests, conditional mean and volatility fits, unconditional volatility term structure and kurtosis fits.<sup>5</sup> Even if one doubts the power of these statistical tests, the economic implications of these models are unequivocal: the magnitudes of pricing errors are similar across the models. This finding implies that one can use a simple Gaussian model for the yield curve modelling, at least within a regime. This is very important because it is much easier to value assets and conduct estimation in the framework of this model.

We next turn to the evidence based on the joint futures and options dataset. The results are dramatically different from the futures-only case. The likelihood ratio tests indicate that the Gaussian model does not encompass the SV model, i.e. can not capture the characteristics of the data implicit in that model. The Gaussian model can not 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. Finally, the Gaussian model can not reproduce the relationship between the implied volatilities and the term structure observed in the data (the USV challenge, see Collin-Dufresne and Goldstein, 2002 – CDG henceforth).

Incorporating the options data into estimation also identifies a tension between 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 USV model. The USV restrictions on the SV model simultaneously weaken the term structure dependence on spot volatility, and free up that same volatility factor to focus on option fitting. We find, consistent with the motivation, that the factors of the USV model estimated from the joint futures and options dataset coincide with the intuitive interpretation

---

<sup>4</sup>We illustrate that such approach leads to drastically different estimation results.

<sup>5</sup>To the best of our knowledge, we are first to evaluate non-nested models via likelihood ratio test.

assigned to them at the specification stage. In particular, the stochastic volatility factor is highly correlated with option-implied volatility.<sup>6</sup> However, we formally reject the USV model in favor of the unrestricted SV model. We also show that USV models fail along a number of important dimensions, such as pricing errors, and ability to match conditional mean and variance.

In a recent paper, Collin-Dufresne, Goldstein, and Jones (2005) (CDGJ henceforth) discuss related issues by contrasting the SV and USV models. They compare these two models based on the swap data only. Consistently with our results, CDGJ find minimal differences between the USV and SV models. We show that one needs to incorporate information in options in order to see the clear differences between the two models.

Our results allow us to take first steps towards understanding the USV phenomenon – an apparent inability of bond prices to span option prices documented by CDG. First, we show that the theoretical relationship between option-implied volatilities and yields in all SV (including USV) models is nearly perfect. Second, if one allows for modest measurement errors, as is the case in virtually all implementations of term structure models, then for all (U)SV models the  $R^2$  of the USV regression declines to the low levels observed in the data. Therefore, measurement errors rather than particular model restrictions generate the USV effect.

The paper is organized into six sections. Section 2 describes the institutional features of the Eurodollar markets, which we find important for our study. Section 3 develops affine models of Eurodollar futures and options prices. Section 4 details the empirical approach. Section 5 discusses all of the results. The final section makes concluding observations on the findings. All technical details are delegated to four appendixes.

## 2 The Eurodollar Markets

In this paper, we follow a strategy regarding data choice that is different from most studies by estimating models based on Eurodollar futures and options instead of US 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 the futures contract details important for valuation in the following section.

Similarly to the swap rates, the Eurodollar futures are contracts written directly on the interest

---

<sup>6</sup>This result is not obvious since we do not invert factors from some exactly observed prices, but filter them.

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

$$F_t(\tau) = 100 - f_t(\tau) \tag{2.1}$$

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

Eurodollar futures are issued every quarter. However, when the Eurodollar contract was introduced in 1981, the available maturities,  $\tau$ , went out only up to two years. The CME gradually introduced longer maturities, and by the end of 1993 futures with maturities of up to ten years were available. As a result, there are now effectively forty points of the term structure, compared to six available swap rates.

Note, that the swap-based studies typically use a 6-month LIBOR to pin down the short end of the curve. In practice, this one additional point demands an additional factor in a term structure model (Duffie and Singleton, 1997, and Jagannathan, Kaplin, and Sun, 2003). Duffie and Singleton (1997) observe that the distinct institutional and credit characteristics of the LIBOR market could explain such a result. On the contrary, the Eurodollar futures market is homogeneous and therefore, one can be confident that one and the same data generating process is responsible for both short and long end of the curve.

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 Sundareshan (2003) argue 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 estimation of a AA-quality term structure model based on the swap rates entails modelling of the risk-free term structure, modelling the cost of collateral process, and estimation which 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 implied by the need for holding the collateral. The futures price,  $f_t(\tau)$ , is a martingale under the risk-neutral measure  $\mathbb{Q}$ , and valuation does not require any discounting:

$$f_t(\tau) = E_t^{\mathbb{Q}}(\ell_{t+\tau}(d)) \tag{2.2}$$

As a result, the modelling costs are reduced, as we can exclusively focus 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. The futures prices, similarly to zero-coupon bonds, have exponentially linear relationship with the factors. Indeed, LIBOR is linked to bond prices in the following way:

$$\ell_t(d) = \frac{1}{d} \left( \frac{1}{P_t(d)} - 1 \right) \quad (2.3)$$

where  $P_t(d)$  is the price of a hypothetical AA-quality zero-coupon bond which matures  $d$  years from time  $t$ . It is clear from expressions (2.2) and (2.3) that if the bond price has an exponentially linear form, the futures price also has a similar expression, which we derive in the following section.

Derivatives on the Eurodollar futures are American puts and calls, which lend themselves for an easy valuation in the affine framework after conversion to their European counterparts.<sup>7</sup> Finally, options data have been available since 1985 as opposed to swaptions and caps, which have been available only since 1997.

### 3 Affine Models of Eurodollar Markets

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

$$P_t(d) = E_t^{\mathbb{Q}} \left( e^{-\int_t^{t+d} r_s ds} \right) \quad (3.1)$$

where  $d$  is time to maturity expressed in years, i.e.  $90/360$ ,  $\mathbb{Q}$  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  $\mathbb{Q}$ .

As Dai and Singleton (2000) (henceforth DS1) point out, there is a trade-off between the flexibility of modelling time-varying volatility with multiple square-root factors, and the ability of a model to generate flexible conditional and unconditional correlations between the state variables. Based on 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  $\mathbb{A}_0(3)$ , and a model with one square-root stochastic volatility

---

<sup>7</sup>The details of this simple procedure are provided in section 4.1.

factor,  $\mathbb{A}_1(3)$ . This conclusion is supported by the results in DS1 and Brandt and Chapman (2003) who find that affine models with multiple volatility factors underperform the simpler ones.

As we mentioned in the introduction, we do not consider quadratic models for the reasons of parsimony. Cheng and Scaillet (2005) show that low-dimensional quadratic models have exactly the same pricing functions as high-dimensional restricted affine models. Therefore, the equivalency of these models follows by the same logic that we group affine models into equivalency classes  $\mathbb{A}_m(N)$ . 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.<sup>8</sup>

We fix the filtered probability space  $(\Omega, \{\mathcal{F}_t\}, \mathcal{F}, \mathbb{P})$  and specify the considered models in the  $Ar$  (affine in  $r$ , see DS1) form. It means that, for the ease of 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.<sup>9</sup> The advantage of this approach is that it provides an extra diagnostic tool for a model: if the estimated factors do not correspond to the assigned labels, the model could be misspecified.

The three-factor Gaussian model,  $\mathbb{A}_0(3)$ , is specified as follows:

$$dr_t = \kappa_r^{\mathbb{P}} (\theta_t + s_t - r_t) dt + \sigma_r dW_t^r(\mathbb{P}) + \sigma_{r\theta} \sigma_{\theta} dW_t^{\theta}(\mathbb{P}) + \sigma_{rs} dW_t^s(\mathbb{P}) \quad (3.2)$$

$$d\theta_t = \kappa_{\theta}^{\mathbb{P}} (\bar{\theta} - \theta_t) dt + \sigma_{\theta} dW_t^{\theta}(\mathbb{P}) + \sigma_{\theta s} dW_t^s(\mathbb{P}) \quad (3.3)$$

$$ds_t = -\kappa_s^{\mathbb{P}} s_t dt + \sigma_s dW_t^s(\mathbb{P}) \quad (3.4)$$

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, 2003 for related interpretations).

---

<sup>8</sup>Consider a one-factor quadratic Gaussian model to understand the intuition behind the results. Introduce an auxiliary state variable which 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 one 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, 2003 for the technical caveats). We are grateful to Peng Cheng, Pierre Collin-Dufresne and Olivier Scaillet for discussions regarding these issues.

<sup>9</sup>See, for instance, Balduzzi, Das, Foresi, and Sundaram (1996), Dai and Singleton (2000), Duffie, Pedersen, and Singleton (2003), Jegadeesh and Pennacchi (1996), Piazzesi (2003) as examples, and Piazzesi (2005) for a discussion. However, CDGJ highlight potential shortcomings of such an approach.

The stochastic volatility model,  $\mathbb{A}_1(3)$ , is specified as follows:

$$\begin{aligned} dr_t &= \kappa_r^{\mathbb{P}} (\theta_t - r_t) dt + \kappa_{rv}^{\mathbb{P}} (\bar{v} - v_t) dt \\ &+ \sqrt{\sigma_r^2 + v_t} dW_t^r(\mathbb{P}) + \sigma_{r\theta} \sqrt{\sigma_\theta^2 + \beta_\theta v_t} dW_t^\theta(\mathbb{P}) + \sigma_{rv} \sqrt{v_t} dW_t^v(\mathbb{P}) \end{aligned} \quad (3.5)$$

$$\begin{aligned} d\theta_t &= \kappa_\theta^{\mathbb{P}} (\bar{\theta} - \theta_t) dt \\ &+ \sigma_{\theta r} \sqrt{\sigma_r^2 + v_t} dW_t^r(\mathbb{P}) + \sqrt{\sigma_\theta^2 + \beta_\theta v_t} dW_t^\theta(\mathbb{P}) + \sigma_{\theta v} \sqrt{v_t} dW_t^v(\mathbb{P}) \end{aligned} \quad (3.6)$$

$$dv_t = \kappa_v^{\mathbb{P}} (\bar{v} - v_t) dt + \sigma_v \sqrt{v_t} dW_t^v(\mathbb{P}) \quad (3.7)$$

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

We introduce notations associated with the probability measure transformation from  $\mathbb{P}$  to  $\mathbb{Q}$  in order to implement the bond formula (3.1). If we denote the vector of risk premia by  $\Lambda$ , then, according to the Girsanov theorem, we can link the two probability measures via:

$$W_t(\mathbb{Q}) = W_t(\mathbb{P}) + \int_0^t \Lambda_s ds \quad (3.8)$$

We consider the essentially affine risk premia specification of Duffee (2002). In the case of  $\mathbb{A}_0(3)$  we have:

$$\Lambda_t^r = \lambda_r + \lambda_{rr} r_t + \lambda_{r\theta} \theta_t + \lambda_{rs} s_t \quad (3.9)$$

$$\Lambda_t^\theta = \lambda_\theta + \lambda_{\theta r} r_t + \lambda_{\theta\theta} \theta_t + \lambda_{\theta s} s_t \quad (3.10)$$

$$\Lambda_t^s = \lambda_s + \lambda_{sr} r_t + \lambda_{s\theta} \theta_t + \lambda_{ss} s_t \quad (3.11)$$

If the model is  $\mathbb{A}_1(3)$  then:

$$\Lambda_t^r = \frac{1}{\sqrt{\sigma_r^2 + v_t}} (\lambda_r \sigma_r^2 + \lambda_{rr} r_t + \lambda_{r\theta} \theta_t + (\lambda_r + \lambda_{rv}) v_t) \quad (3.12)$$

$$\Lambda_t^\theta = \frac{1}{\sqrt{\sigma_\theta^2 + \beta_\theta v_t}} (\lambda_\theta \sigma_\theta^2 + \lambda_{\theta r} r_t + \lambda_{\theta\theta} \theta_t + (\lambda_\theta \beta_\theta + \lambda_{\theta v}) v_t) \quad (3.13)$$

$$\Lambda_t^v = \lambda_v \sqrt{v_t} \quad (3.14)$$

For completeness, appendix A lists the dynamics of the models under the  $\mathbb{Q}$  measure.<sup>10</sup>

Chredito, Filipovic, and Kimmel (2003) and Liptser and Shiryaev (2001) show that essentially affine risk premia specifications can be extended to the cases where the price of risk is inversely proportional to the square-root volatility factor, as in, for example, (3.12) with  $\sigma_r = 0$ . For this reason we will

---

<sup>10</sup>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 Duffee and Kan, 1996) to guarantee strict positivity of volatility in  $\mathbb{A}_1(3)$ .

not constrain  $\sigma_r$  and  $\sigma_\theta$  to be greater than zero at the estimation stage.<sup>11</sup> However, we still omit one possible additional term in (3.14), because its value binds a regularity condition based on the results reported in Cheredito, Filipovic, and Kimmel (2003). 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 (3.1). The affine specification implies:

$$P_t^P(d) = e^{A^P(d) + B_r^P(d)r_t + B_\theta^P(d)\theta_t + B_\phi^P(d)\phi_t} \equiv e^{A^P(d) + B^P(d) \cdot \mathcal{S}_t} \quad (3.15)$$

where  $\phi$  generically denotes either  $s$  in  $\mathbb{A}_0(3)$ , or  $v$  in  $\mathbb{A}_1(3)$ , and, for compactness,  $\mathcal{S}_t = (r_t, \theta_t, \phi_t)'$ . The coefficients  $A^P$  and  $B^P$  solve well-known Ricatti ODEs (see DPS, for instance). The boundary conditions are:  $A^P(0) = 0, B^P(0) = 0$ .

It is possible to impose constraints on the  $\mathbb{Q}$  parameters of the  $\mathbb{A}_1(3)$  such that the volatility factor  $v$  disappears from the bond pricing equation (the USV feature).<sup>12</sup> More precisely, the markets are incomplete in the three-factor models if the loadings on the factors  $\theta$  and  $v$  are linearly dependent, i.e. one can find coefficients  $u_\theta$  and  $u_v$  (at least one of them should be non-zero) such that:

$$u_\theta B_\theta^P(d) + u_v B_v^P(d) = 0 \quad \forall d \quad (3.16)$$

CDG derive the parameter constraints implied by this requirement. We list these constraints in the appendix B.

We now have enough structure to price instruments from the Eurodollar markets. We combine the definitions of the Eurodollar futures (2.2), the LIBOR rate (2.3), and the LIBOR bond price (3.15) to obtain the futures LIBOR rate:

$$\begin{aligned} f_t(\tau) &= \frac{1}{d} E_t^{\mathbb{Q}} \left( e^{-A^P(d) - B_r^P(d)r_{t+\tau} - B_\theta^P(d)\theta_{t+\tau} - B_\phi^P(d)\phi_{t+\tau}} - 1 \right) \\ &= \frac{1}{d} e^{-A^P(d)} E_t^{\mathbb{Q}} \left( e^{-B_r^P(d)r_{t+\tau} - B_\theta^P(d)\theta_{t+\tau} - B_\phi^P(d)\phi_{t+\tau}} \right) - \frac{1}{d} \\ &= \frac{1}{d} e^{-A^P(d) + A^f(\tau) + B_r^f(\tau)r_t + B_\theta^f(\tau)\theta_t + B_\phi^f(\tau)\phi_t} - \frac{1}{d} \equiv \frac{1}{d} e^{-A^P(d) + A^f(\tau) + B^f(\tau) \cdot \mathcal{S}_t} - \frac{1}{d} \end{aligned} \quad (3.17)$$

where  $A^f$  and  $B^f$  solve almost the same Ricatti ODEs with slightly different boundary conditions as compared to (3.15):  $A^f(0) = 0, B^f(0) = -B^P(d)$ .

---

<sup>11</sup>Note that in order to satisfy the canonical representation in DS1, the parameters  $\sigma_r$  and  $\sigma_\theta$  in  $\mathbb{A}_1(3)$  can not be equal to zero. On the other hand, the model with these parameters set at zero is an admissible and identifiable model (see Duffie and Kan (1996) for the regularity conditions). Ait-Sahalia and Kimmel (2002) make a similar observation.

<sup>12</sup>The  $\mathbb{A}_0(3)$  model does not admit the USV property.

It is natural to work with yields, as opposed to bond prices. The zero-coupon bond yield is:

$$y_t^P(d) = -\frac{1}{d} \log P_t(d) = -\frac{A^P(d)}{d} - \frac{B^P(d)}{d} \cdot \mathcal{S}_t \quad (3.18)$$

By analogy with the bond (also see (2.3)), we can construct an object based on the futures price, which we will refer to as the futures yield:

$$y_t^f(\tau) = \frac{1}{d} \log(1 + d \cdot f_t(\tau)) = \frac{-A^P(d) + A^f(\tau)}{d} + \frac{B^f(\tau)}{d} \cdot \mathcal{S}_t \quad (3.19)$$

Note that the factor loadings on the bond and futures yields are different. Therefore, both types of assets might not simultaneously satisfy the USV condition (3.16).<sup>13</sup> Indeed, zero-coupon bond prices are directly related to the forward rates. The forward rates differ from the futures rates by the convexity adjustment, which is sensitive to volatility. Hence, the futures data are informative about volatility even in the presence of the USV. This feature allows us to identify the volatility risk premium, which would not be possible with the regular swap data (see CDGJ).<sup>14</sup>

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

$$\begin{aligned} C_t(\tau_f, \tau_c, K) &= E_t^{\mathbb{Q}} \left( e^{-\int_t^{t+\tau_c} r_s ds} (F_{t+\tau_c}(\tau_f - \tau_c) - K)^+ \right) \\ &= E_t^{\mathbb{Q}} \left( e^{-\int_t^{t+\tau_c} r_s ds} (k - f_{t+\tau_c}(\tau_f - \tau_c))^+ \right) \end{aligned} \quad (3.20)$$

where  $k = 100 - K$ . Hence, a call on Eurodollar futures rate can be considered as a put on LIBOR futures rate.<sup>15</sup> We follow Duffie, Pan, and Singleton (2000) in computing the call price. Appendix C 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 4.1.

## 4 Empirical Setup

The implementation of our model requires a bit of additional assumptions, which take into account the data and factor structure of the model. Hence, we start with a brief description of the dataset, and

---

<sup>13</sup>We have verified that our intuition is correct for the estimated models. The details are available upon request.

<sup>14</sup>Moreover, as we detail in the section 4.2, this sensitivity to volatility allows us to estimate this latent factor accurately, which would be much harder to do using the traditional data sources, such as bond yields or swap rates.

<sup>15</sup>Note that we are using the AA-quality instantaneous rate  $r_t$  to discount the option payoff at maturity. Since 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 small counterparty risk, which is comparable to the default risk of AA-rated companies. Further modelling of these fine features is beyond the scope of this paper.

then proceed with the estimation methodology. We conclude the section with the description of our parameter identification strategy.

## 4.1 Data

We use weekly Eurodollar futures and options data.<sup>16</sup> Recall that CME gradually introduced futures of longer maturities, beginning with the 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 be comparable 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  $T = 391$  weeks in our sample.

One concern is that our sample of yields does not span a sufficiently long period to address the volatility-related issues. First, we have to consider yields versus options data trade-off as the derivatives data have shorter time span virtually in 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 a strong evidence that an adequate model of term structure should incorporate regime switches if it were to explain a long historical period beginning in the late sixties/early seventies (e.g. Bansal and Zhou, 2002, Dai, Singleton, and Yang, 2003). Thus, the outstanding issue is finding an appropriate model within a regime.<sup>17</sup> Arguably, our sample period is excellent for a one-regime study as it represents the stability in the monetary policy conduct (the Fed started announcing the interest rate target in 1994) and in the US economy.<sup>18</sup>

In order to reduce the computational burden associated with using all forty available points along the yield curve, we select only eleven maturities. The first maturity is six months, and the rest are annual maturities from 1 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

---

<sup>16</sup>We obtained the data from the Institute for Financial Markets.

<sup>17</sup>Indeed, Bansal, Tauchen, and Zhou (2003) 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.”

<sup>18</sup>As we will see in the sequel, our futures-based results for the  $\mathbb{A}_1(3)$  and  $\mathbb{A}_1^{USV}(3)$  models are largely consistent with CDGJ who used the more typical set of swap rates beginning in 1988. This gives us additional confidence that our results are not driven by specifics of the sample.

to  $n$  years. We track the same “ $n$ -year” contract until it has ten business days left to switching 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. The ones with 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. Also, quarterly options are more liquid. For these reasons we use only quarterly options of maturities equal to six months and one year.

Since the focus of our paper is on learning about interest rate volatility, and not accurate option valuation per se, we further limit our choice of options. We compute option’s moneyness as ratio of the strike to the futures price, and select options with moneyness as close to 1 as possible. As a result, 90% of all the moneynesses is between 0.98 and 1.02 with median being exactly 1; the smallest moneyness is 0.97 and the largest is 1.03.

Since the Eurodollar options are American, we have to adjust our data for the early exercise premium. We anticipate that this premium is not 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 (2005) 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 equal to 0.05%, and the largest premium is 0.16% (as a reference, the smallest implied volatility in our sample is 6.7%).<sup>19</sup>

## 4.2 Measurement Errors

The standard assumption made at the ATSM estimation stage is that three yields are observed exactly, and, therefore, the latent factors can be inverted from these yields. Since the yields have a linear

---

<sup>19</sup>In principle, in order to be absolutely accurate, we should have computed the exercise premium for the particular model under consideration, i.e.  $\mathbb{A}_1(3)$ . 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. Since we find this adjustment to be very small already, it would be impractical to use more complicated models.

relationship with the factors, one can compute the volatility factor  $v$  as a linear combination of the three yields (CDGJ and Piazzesi, 2005 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:

$$y_t^f(\tau_{ft}) = g_f(\mathcal{S}_t, \Theta, \tau_{ft}) + \epsilon_t(\tau_{ft}), \quad \tau_{ft} = 0.5_t, 1_t, 2_t, \dots, 10_t \quad (4.1)$$

$$C_t(\tau_{ct}, \tau_{ct}, K) = g_c(\mathcal{S}_t, \Theta, \tau_{ct}) + \tilde{\epsilon}_t(\tau_{ct}), \quad \tau_{ct} = 0.5_t, 1_t \quad (4.2)$$

where  $\mathcal{S}$  follows the state equation (3.2)-(3.4) or (3.5)-(3.7) depending on whether we estimate the  $\mathbb{A}_0(3)$  or  $\mathbb{A}_1(3)$  models, and the functions  $g_f(\cdot)$  and  $g_c(\cdot)$  are given in the equations (3.19) and (3.20), 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 intentionally assume that the errors are independent across various maturities, and all have the same variance,  $\sigma_\epsilon^2$ , irrespective of maturity. We allow a different variance,  $\tilde{\sigma}_\epsilon^2$ , for the option pricing errors. Such a restrictive specification puts more pressure on our original model to fit the data. It will be harder to detect that a model is misspecified if the error terms are flexible.<sup>20</sup>

All parameters are identified in our models. This conclusion may be surprising in the light of DS1 comments about unidentifiability of some risk premia parameters associated with Gaussian factors. What helps us in identification of the premia is the availability of more prices than state variables (see

---

<sup>20</sup>Heidari and Wu (2002) critique 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 based on the theoretical LIBOR rates, i.e. they do not take into an account the measurement error. Our methodology is not subject to their criticism because both futures and quarterly options contracts are settled based on the actual LIBOR rate at maturity. Since, 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 the internal consistency.

de Jong, 2000).<sup>21</sup> Related identification analysis is conducted in CDGJ.

The presence of measurement errors in each price calls for filtering-based estimation methods. We chose to rely on 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. First, we note that the USV models, by their design, can not be estimated via exact inversion. Hence, filtering methods are a must.<sup>22</sup> Second, we illustrate in Table 2 that the estimation method can drastically affect the outcome. We estimate the  $\mathbb{A}_1(3)$  stochastic volatility model based on 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  $\kappa_v^{\mathbb{P}}$  of  $v$ , 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.

### 4.3 Econometric Method

The preferred estimation methodology is maximum likelihood. However, with the exception of the  $\mathbb{A}_0(3)$  model estimated on yields only, the likelihood is 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  $\mathcal{S}$  are not observable in the presence of measurement errors. In the case of the  $\mathbb{A}_0(3)$  model and futures rates only, one can estimate the system with errors via Kalman filter (see, e.g. Hamilton, 1994 for the details of the algorithm). Indeed, the state evolves according to the Gaussian system, the function  $g_f$  is linear, which implies that the ML based on Kalman filter is the optimal estimation methodology. However, if we consider non-linear function in the measurement equation, such as  $g_c$ , or non-Gaussian states such as the  $v$  factor in  $\mathbb{A}_1(3)$ , exact filter is generally not computationally feasible, and the Kalman filter is no longer optimal.

When the measurement equation is non-linear, one could use Extended Kalman filter (EKF), which is effectively the regular Kalman filter applied to the first order Taylor expansion of the measure-

---

<sup>21</sup>We have conducted additional identification checks to ascertain this conclusion. The results are available upon request.

<sup>22</sup>At a first glance, a problem with using Kalman filter might arise in the USV case. Since, by construction, yields contain no information about the volatility, our estimation strategy might result in very noisy estimates of the unspanned volatility factor. However, the futures prices (in contrast to bond prices) are sensitive to volatility even in the presence of USV due to the convexity effect, and, therefore, the Kalman filter produces accurate estimates.

ment functions ( $g_c$ ). When the state is non-Gaussian, one could use Quasi Kalman filter (QKF), which replaces such a state ( $v$ ) with a Gaussian state with identical first two moments (Duan and Simonato, 1995, and Chen and Scott, 2003).<sup>23</sup> Both filters are the best linear filters, but QML estimates are conditionally biased and inconsistent.

Despite the 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  $\sigma_\epsilon$  (Picard, 1991). Monte-Carlo studies of the QKF applied to multi-factor 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)). Since we have at most one square-root factor and only two out of thirteen measurement equations are non-linear, these results suggest that we might obtain reasonable estimates despite undesirable asymptotic properties of the filters.

Since linearizing option prices is a relatively novel approach (the only other paper using 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 model implementation. We simulate 800 paths out of our model and compute the corresponding futures and options prices under 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.

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 hard to identify in practice: risk premia and correlations.<sup>24</sup> With these observations in mind, we can 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 pretty well.

---

<sup>23</sup>Brandt and He (2002) propose a simulated maximum likelihood approach combined with importance sampling as an alternative to QKF.

<sup>24</sup>These observations are consistent with the findings of Duffee and Stanton (2004).

## 5 Results

### 5.1 Preliminary observations

We estimated two versions, depending on the dataset, of each of the three models:  $\mathbb{A}_0(3)$ ,  $\mathbb{A}_1(3)$ , and  $\mathbb{A}_1^{USV}(3)$ . We assign a superscript  $f$  to the models 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) (DS2 henceforth), we restricted some of these parameters to zero, if 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 4.3. Thus, our results not only produce finite-sample inference, but also represent a Monte Carlo analysis of the estimation method.<sup>25</sup> 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 errors computed for the model  $\mathbb{A}_1(3)^f$  in Table 2.<sup>26</sup> The implied confidence intervals are, of course, symmetric and are typically wider than their finite-sample counterparts. The last feature may counterfactually imply insignificance of certain parameters, e.g.  $\bar{\theta}$ ,  $\sigma_\theta$ ,  $\beta_\theta$ ,  $\sigma_{r\theta}$ ,  $\sigma_{\theta v}$ ,  $\lambda_\theta$ , and  $\lambda_{\theta v}$ . As noted by DS1, the significance of these parameters have important implications for the flexibility in the factor correlations and in the specification of the feasible risk premia.

The parameter estimates allow us to take a first glance at the models' properties. We check if the initial interpretation of the state variables corresponds to the actual role they play in 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 ten-year futures yield, or the long rate; the difference between the two, or the slope of the term structure; the butterfly spread,

---

<sup>25</sup>We do not report the finite-sample bias for the  $\mathbb{A}_1(3)$  models to conserve space. The results are largely consistent with the ones for the  $\mathbb{A}_0(3)$  models and are available upon request.

<sup>26</sup>The two types of standard errors are estimated using either the first or second derivatives of the information matrix.

which corresponds to a long position in the six-month and ten-year futures and double short position in the two-year futures; and, finally, the absolute implied variance.<sup>27</sup> 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 with 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.<sup>28</sup> The only exception is the factor  $\theta$  in the  $\mathbb{A}_1(3)^{fo}$  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  $\mathbb{A}_1(3)$  model.

The volatility factor is highly correlated with the butterfly in case of  $\mathbb{A}_1(3)^f$ . This result is intuitive because the butterfly spread should be sensitive to volatility (see, for instance, Litterman, Scheinkman, and Weiss (1991)). When we add the USV constraint, the factors rotate:  $v$  is correlated with the long rate in  $\mathbb{A}_1^{USV}(3)^f$  – the same observable that  $\theta$  is correlated with in this model. This puzzling result is explained by the USV feature. Formally, in the case of USV, a linear combination of two factors acts like one factor in the cross-section.<sup>29</sup> The dataset involving the futures only is not informative enough in this model to make a significant distinction between volatility and central tendency in the time series.

In the case of  $\mathbb{A}_1(3)^{fo}$ ,  $v$  is correlated with the slope, which, in combination with our observation regarding  $\theta$  in this model, raises the suspicion about the model misspecification. It appears that factor  $v$  has rotated into the central tendency, and factor  $\theta$  has rotated into the volatility (though correlations are weaker than for the  $\mathbb{A}_1(3)^f$  counterparts). However,  $\theta$  has virtually zero correlation with the implied volatility, while its counterpart  $v$  in  $\mathbb{A}_1(3)^f$  has a respectable correlation of 0.24. It appears that when the  $\mathbb{A}_1(3)$  model is confronted with the joint dataset, the factors substantially deviate from their intended roles in order to generate realistic distribution of prices.<sup>30</sup>

---

<sup>27</sup>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 interest rate in the Black model follows the 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.

<sup>28</sup>The correlation of these factor with the respective observables deteriorates when the options data is added reflecting the tension between the requirements to fit the term structure and option prices simultaneously.

<sup>29</sup>More precisely, this combination can be rotated into to the risk-neutral expected value of the spot interest rate (see CDGJ). Hence, high correlation with the long rate is natural.

<sup>30</sup>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  $\mathbb{A}_1(3)$  model. Most

$\mathbb{A}_1^{USV}(3)^{fo}$  makes perfect sense in terms of factor interpretations. The factor  $v$  is highly correlated with implied variance, and other factors have correlations similar to those of  $\mathbb{A}_1(3)^f$ . A combination of the USV restriction and options data cleanly separates the contribution of each of the factors.<sup>31</sup>

To summarize, we find that estimated latent state variables are not necessarily behaving according to the initial model specification. This fact is especially true for the 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 discrete-time properties of the yields. However, these observations do not mean a failure of the affine models. As pointed out by CDGJ, this outcome could be a result of non-uniqueness 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.

## 5.2 Evidence from Futures

As a first step of our analysis we evaluate the three models estimated based on the 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 estimated based on the joint dataset.

### Pricing Errors

In order to evaluate the economic differences between the models, we report pricing errors in Table 6 as a starting point of models' fit evaluation. We compute the average of the absolute differences between the model-implied and actual futures yields defined in (3.19). The futures errors are very small and similar across the models – it would be hard to argue that there are important differences even if a statistical test would be 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 based on the principal components results of Litterman and likely, 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 divergencies in fit elsewhere. However, Umantsev does not explore this possibility.

<sup>31</sup>However, Singleton (2005) remarks regarding a similar result in CDGJ based on the swap data, “since the  $\mathbb{A}_1^{USV}(3)$  model is literally nested in the unconstrained  $\mathbb{A}_1(3)$  model, this finding must be a consequence of misspecification of the  $\mathbb{A}_1(3)$  family: nested models such as these cannot be built up from distinct risk factors.” Hence it is important to formally test these models against each other. We will proceed with this in the following sections.

Scheinkman (1991). 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 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 perform hypotheses testing associated with relative models' performance based on the conditional likelihood. Typically, a test of nested models is quite informative in locating a parsimonious model which has a good data representation. In our case, the models from the  $\mathbb{A}_0(3)$  and  $\mathbb{A}_1(3)$  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 D) is very useful in our situation. In the context of this paper, the model  $\mathbb{A}_0(3)$  encompasses a model  $\mathbb{A}_1(3)$  if the former is able to explain characteristics, i.e. certain statistics, or likelihood values, of the latter. The encompassing principle applies to the nested models as well. For example, a restricted model  $\mathbb{A}_1^{USV}(3)$  encompasses a larger model  $\mathbb{A}_1(3)$  if the associated restrictions cannot be rejected. It is important to emphasize that these tests in no way tell us that one model is "better" than the other one when the null is rejected. The rejection shows that the two models are sufficiently different from each other given the dataset.

It is informative to see if  $\mathbb{A}_0(3)$  encompasses  $\mathbb{A}_1(3)$  since it is perceived by many researchers as a simpler model because volatility of the short interest rate is not stochastic. Table 7 presents the test. The Gaussian model clearly encompasses the stochastic volatility one. This result provides preliminary formal evidence that  $\mathbb{A}_0(3)$  model can capture term structure data as well as the more sophisticated  $\mathbb{A}_1(3)$ . This finding has very important modelling implications; therefore, we will explore this further in the subsequent sections.

We also want to check if the USV restrictions are supported by the data. Similarly to the  $\mathbb{A}_0(3)$  model, we test whether  $\mathbb{A}_1^{USV}(3)$  encompasses  $\mathbb{A}_1(3)$ . The results are reported Table 7 as well. In contrast to the Gaussian model, the null hypothesis of encompassing is rejected. Given worse pricing performance of the USV models, this test highlights problems with the USV restrictions.

We also directly check the significance of the individual USV restrictions on the parameters  $\kappa_{\theta}^{\mathbb{P}}$ ,  $\kappa_v^{\mathbb{P}}$ ,  $\kappa_{rv}^{\mathbb{P}}$ ,  $\beta_{\theta}$ , and  $\lambda_{\theta r}$  via the Wald-type test. Table 8 reports the results. We cannot reject the restriction on the risk premium  $\lambda_{\theta r}$  because of zero correlation between  $r$  and  $\theta$  (see (B.10)). The remaining USV

restrictions are rejected.<sup>32</sup>

## Conditional Mean and Volatility

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.

In this section, we explore abilities of the models to fit conditional means and volatilities. This exercise is motivated by the challenges, highlighted by Dai and Singleton (2003) (DS3 henceforth), to the ability of the term structure models to fit these moments. Typically, the ability to fit the conditional mean is cast in term of the expectation hypothesis puzzle – challenge *LPY*. The second challenge, *CVY*, is related to a model’s ability to replicate time-varying volatility of the yields and 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 one can not distinguish implications of different models.<sup>33</sup> 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 the first part of the *CVY* challenges. We will consider the second part of *CVY* in the next section.

Specifically, the recent trend in the literature is to use GARCH as a descriptive, or “model-free,” estimate of conditional volatility (see, for instance, CDGJ, DS3, and Umantsev, 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. In spirit, such a diagnostic is similar to the reprojection technique of Gallant and Tauchen (1998).

However, since GARCH is still a model, one has to be careful with the interpretation of results. To streamline the discussion, consider the basic GARCH(1,1) model:

$$Y_t = \mu_{t-1} + \sigma_t e_t \tag{5.1}$$

$$\sigma_t^2 = c + \alpha \sigma_{t-1}^2 e_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{5.2}$$

---

<sup>32</sup>The restriction (B.9),  $\beta_\theta = 0$  is rejected at the 10% level.

<sup>33</sup>Our results are available upon request. CDGJ report qualitatively similar results.

where  $Y$  generically denotes the series of interest. The estimated values of parameters governing the volatility dynamics ( $c$ ,  $\alpha$ , and  $\beta$ ) will be sensitive to the estimated conditional mean  $\mu$ . CDGJ focus directly on the dynamics of the state variable  $r$ , i.e.  $Y_t = r_t - r_{t-1}$ , and compute  $\mu$  from the estimated 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, DS3 estimate univariate GARCH model for 5-year yields.

Despite its shortcomings, the advantage of the CDGJ 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.

After accounting for the VAR(1) conditional mean, we find, based on the Portmanteau test, the strongest evidence for heteroscedasticity of residuals in the slope, and the weakest one in the short rate. Nonetheless, we were able to estimate the GARCH models for all three series. Figure 1 shows means and volatilities from the observed futures yields and the ones implied by our models. Table 9 complements the figure by reporting correlations between the means and volatilities implied from the data and those implied from the models.

We see that the  $\mathbb{A}_0(3)$  and  $\mathbb{A}_1(3)$  models produce nearly identical estimates of the conditional means and volatilities. Moreover, both models do a pretty good job in fitting the first two moments, especially for the level and slope. USV's fit deteriorates relative to the other two models. The fact that the USV model generates volatilities that are similar or worse than the other models contradicts CDGJ, who find that only USV volatility correlates well with the GARCH volatility. Most likely, the difference in the findings comes from the way GARCH volatility is estimated. Interestingly, despite a clear difficulty in identifying the role of the instantaneous variance of the short rate (see our discussion of Table 5), the models estimated on the futures, irrespective of their factor structure, do a remarkable job in explaining the variance of the yields.

This finding is quite surprising because the  $\mathbb{A}_0$  models can generate only constant conditional

volatility.<sup>34</sup> Perhaps, the degree of the conditional volatility variation is so small that we cannot reject the  $\mathbb{A}_0$  in favor of the  $\mathbb{A}_1$  models based the GARCH series. In order to address this issue, we implement additional diagnostics described in appendix E. The results indicate that, indeed, the Gaussian and stochastic volatility models are not distinguishable based on the low order conditional moments of the term structure. This implies that for many purposes, such as intra-regime modelling or studying of low frequency data, the Gaussian models are sufficient.<sup>35</sup> This finding has important practical implications as it is easier to compute asset prices and conduct estimation in the framework of Gaussian models.

Our findings complement the *CVY* analysis in DS3. These authors find that a term-structure model requires at least three factors to generate sufficient degree of the 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 second part of the *CVY* challenge requires a term-structure model to generate the hump-shaped term structure of the unconditional volatilities observed after 1983. The upper left panel of Figure 2 shows the unconditional volatilities of yield changes computed as standard deviation directly from the data (with bootstrapped confidence bounds) and population values implied by various models.

In the data, we see the hump that was documented for the swaps and Treasury rates. The  $\mathbb{A}_0(3)^f$  and  $\mathbb{A}_1(3)^f$  models generate very similar term structures which have magnitudes on par with the data, however the hump is much less pronounced. The USV models produces volatility values that fall outside of the confidence interval.

It is clear that we cannot distinguish the Gaussian and stochastic volatility ones based on the first two moments. Both are quite successful in reproducing the conditional and unconditional moments. The USV model fails to reproduce the stylized pattern. The next natural step is to evaluate the ability of the models to fit higher order moments.

---

<sup>34</sup>We are grateful to Ken Singleton for an extensive discussion and suggestions on this issue.

<sup>35</sup>In particular, this finding addresses the concern expressed in Bansal, Tauchen, and Zhou (2003) regarding the Dai, Singleton, and Yang (2003) regime-switching model with constant intra-regime volatility.

## 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, and this property is consistent with what all models generate. Therefore, we do not report the skewness results and focus on kurtosis in Table 10.

It turns out that it is extremely hard for all models, even non-Gaussian ones, to generate the relatively high degree of kurtosis observed in the term structure. Only the USV model succeeds in matching the kurtosis of the level and the slope.

## 5.3 Evidence from Options

We found that we can not distinguish the Gaussian and stochastic volatility models either using economic or statistical criteria when only yield data are involved. The USV model generally performs worse than the other two, however it is more successful in capturing the kurtosis of the level and slope of the term structure. Since options prices are sensitive to volatility, we should be more successful in distinguishing the different models. Moreover, the USV model was specifically designed to capture joint dynamics of yields and options, so it must be tested based on the joint dataset. 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 different models estimated based on futures only value options. The evidence is mixed.  $\mathbb{A}_1(3)^f$  appears to price options better than  $\mathbb{A}_0(3)^f$ . The  $\mathbb{A}_1^{USV}(3)^f$  model does particularly badly: its errors are roughly ten times larger than those of the other models. This outcome is puzzling, since the USV model was designed to capture the option prices better. Perhaps, its performance will improve once the model is estimated based on the options data.

Proceeding with the options-based estimation, we note that the futures pricing performance is still

quite similar across the models. However, we can notice that  $\mathbb{A}_1(3)^{fo}$  improves upon  $\mathbb{A}_0(3)^{fo}$  for short maturities. In contrast,  $\mathbb{A}_1^{USV}(3)^{fo}$  deteriorates relative to  $\mathbb{A}_0(3)^{fo}$  for long maturities. The futures errors generally increase relative to the models that were estimated on futures only. This outcome is not surprising because of the increased pressure on the same model to fit a more elaborate dataset.<sup>36</sup>

The increase in futures errors is associated with an impressive improvement in option pricing. The pricing errors decline three- to fourfold.<sup>37</sup> The option-pricing performance of  $\mathbb{A}_1^{USV}(3)^{fo}$  improves relative to the  $\mathbb{A}_1^{USV}(3)^f$  case, but still underperforms the other two models. This result is surprising as USV was designed to resolve the tension between the yield and option data. Instead the model performs poorly on both types of data.

## 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 telling the models apart. 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.

As regards the USV models, the encompassing tests are consistent with the futures only case.  $\mathbb{A}_1^{USV}(3)$  is rejected in favor of  $\mathbb{A}_1(3)$ . Only two of the USV individual restrictions,  $\kappa_v^{\mathbb{P}}$  and  $\kappa_{rv}^{\mathbb{P}}$ , cannot be rejected by the Wald-type test in Table 8.

## Conditional Mean and Volatility

Figure 3 and Table 9 show that the  $\mathbb{A}_0(3)$  and  $\mathbb{A}_1(3)$  models again produce nearly 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. The performance of the USV model is even worse than in the futures only case.

---

<sup>36</sup>This result is different from Jagannathan, Kaplin, and Sun (2003), who report no change in the swaps fit once the caps data are added. Since their parameter values do not change, we suspect that they obtained a local optimum.

<sup>37</sup>Umantsev (2001) also observes a dramatic improvement in the options fit of the  $\mathbb{A}_1(3)$  model once the options data is used for estimation. He, however, does not acknowledge the deterioration of fit in the underlying security.

## Term Structure of Volatility

The upper right panel of Figure 2 compares the unconditional term structure of standard deviation implied from various models. We see that the  $\mathbb{A}_0(3)^{fo}$  term structure is outside of the confidence bounds longer maturities (after two years), while the term structure implied by the other two models are not inconsistent with the data. The  $\mathbb{A}_1(3)^{fo}$  pattern is the most realistic.

Since 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 if our three-factor models are able to generate the hump in implied volatilities.

The lower panels of Figure 2 show the population values of average implied volatilities implied from the models and average implied volatilities computed from the data (solid line). The data are not very informative because we used only 6-month and 1-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  $\mathbb{A}_0(3)$  and  $\mathbb{A}_1(3)$  models are close to each other within each dataset type, and generate fairly reasonable volatility values. The models estimated using options generate the hump (the one in  $\mathbb{A}_1(3)^{fo}$  is more pronounced). The USV models generate drastically different patterns.

## Higher Order Moments

The kurtosis tests in Table 10 indicate that both stochastic volatility models can fit the kurtosis of the level and none of the models can fit the kurtosis of the slope and curvature.

### 5.4 Futures vs Options

It is worthwhile to review the accumulated evidence at this point. 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 estimated based on the futures data alone. This finding has important implications for the yield curve modelling strategy: one can use a simple Gaussian model, at least within a regime. On the other hand, we can clearly say that the  $\mathbb{A}_1(3)$  model emerges as a leader based on the joint dataset. The model improves upon the  $\mathbb{A}_0(3)$  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. Finally, the USV model is rejected in favor of the more flexible  $\mathbb{A}_1(3)$  model based both on (mostly options) pricing performance and statistical diagnostics.

Second, we observe a clear tension in models' ability to fit lower vs higher order moments: the models fitted to the term structure only do an excellent job fitting VAR(1)-GARCH(1,1), but they fail with kurtosis. The models that are fitted to the joint dataset have an opposite performance. Unfortunately, the USV model does not resolve the tension despite the fact that it was specifically designed to do so.

Our conclusions seem puzzling in the light of the apparent inability of affine models to explain option-implied volatility reported by CDG. Our models, at least based on theoretical considerations, must have drastically different option-spanning properties. Thus, it is natural to study the models' implications for this issue.

## 5.5 Reconciling the models' implications and the unspanned volatility

Before addressing the spanning issue, we make sure that we can replicate the CDG results. We want to check if the option-based volatility is related to the time series-based volatility, or, more generally, time series-based factors. We regress the volatilities implied from 6-month and 1-year options on the three principal components of the term structure. While such a regression is different from the one considered in CDG – they looked at straddles – we believe it speaks directly to the issue at hand. We report the  $R^2$  in Table 11. The statistics are fairly low and qualitatively consistent with the values reported by CDG. We view this result as a data pattern, which serves as a challenge to a term structure model in the same spirit as we view the challenges *LPY*, or *CVY*.

What could be the reason for the low  $R^2$ ? Two main explanations that we see are: (i) there is a weak linear relationship in the data; (ii) the principal components are measured with error (errors-in-variables problem). Both explanations can be easily explored within our framework as we can construct model-based implied volatility and we explicitly allow for measurement error (see section 4.2). We address the linearity issue by implementing the same regressions on samples simulated from our models under the assumption of no measurement errors. Then we evaluate the impact of the measurement errors in the simulated samples. We bootstrap the finite sample distribution by simulating 800 paths for each model. Table 11 reports the results.

In the 'no-errors' case, we find that overall  $\mathbb{A}_0(3)$  models generate  $R^2$  close to zero, while all  $\mathbb{A}_1$

models generate  $R^2$  close to one. The  $\mathbb{A}_1(3)$  result is consistent with the discussion in CDG. Most strikingly, the  $R^2$  are high even in the  $\mathbb{A}_1^{USV}(3)$  case despite the fact that it was designed to explain the USV pattern. The question is why do we see these different outcomes within the different models.

Figure 4 provides the answer. We plot how at-the-money absolute implied volatility changes in response to the changes in the state variables. As we can see, there is virtually no reaction to the changes in either  $r$  or  $\theta$  across all models. However, there is a roughly linear relationship between the factor  $v$  and implied volatility in the  $\mathbb{A}_1(3)$  models (even in the USV case), while there still no reaction to  $s$  – the third factor in the  $\mathbb{A}_0(3)$  models. Regressing implied volatility on the three principal components is the same as regressing it on the three factors in the affine models. For this reason, we find the reported relationship between implied volatility and the term structure. Hence, we conclude that the USV challenge is a powerful test allowing to discriminate between the Gaussian and stochastic volatility models.

In the ‘errors’ case, consistently with the previous discussion,  $\mathbb{A}_0(3)$  models fail in replicating the  $R^2$  in the data. However, the  $\mathbb{A}_1(3)$  models (including the USV ones) do a reasonable job. Especially successful are the  $fo$  models: pointwise, they closely match the data. Thus, we find that the  $\mathbb{A}_1(3)$  model can replicate the pattern in the data even without the USV restrictions in the presence of the measurement noise. Note that the measurement errors are modest (Table 6), hence our interpretation of the USV phenomenon does not hinge on implausible premises.

Given that the measurement errors evidently help to explain the USV phenomenon in the  $\mathbb{A}_1(3)$  case, it is natural to wonder whether these errors could be modelled via one additional factor. This question is particularly relevant in the light of the CDG findings: they report that a single factor explains 85% of variation in the USV regression errors combined across various options’ maturities. Our results associate the regression residuals with the measurement errors in our econometric setup and, thus, explain the commonality of the regression residuals in CDG.

In order to gauge the degree of association between the regression residuals in CDG and the measurement errors in our models, we conduct the principal components analysis of the pricing errors (which by construction coincide with the measurement errors). The percentage of variation in the errors explained by the first principal component is displayed in Table 6 (the line labeled ‘PC1’). The results are striking: the  $\mathbb{A}_1^{USV}(3)^f$  model is the only one where the first principle component explains as much as the first principal component of the residuals in CDG.

We interpret the evidence as follows. The measurement errors contaminate the otherwise strong

relationship between the yields and implied volatilities in the  $\mathbb{A}_1(3)$  model. As a result, the  $R^2$  declines from nearly 100% to a low number. Therefore, the regression residuals in the data must be almost linearly related to the measurement errors. The USV model proposed by CDG implies a measurement error structure which can be explained by one principal component. The regression residuals inherit this structure because of the aforementioned almost linear relationship. The final outcome is fully consistent with the CDG findings. However, as we saw in Table 11, this is not the only measurement error structure, which generates the low  $R^2$ . Which implicit measurement structure is more compelling? The analysis in this paper favors the  $\mathbb{A}_1(3)^{fo}$  model. In the framework of this model, the need for an extra factor which explains the remaining variation is less obvious as the first principal component explains only 40%, and one needs three more principal components to explain the same amount of variation as the first principal component does in  $\mathbb{A}_1^{USV}(3)^f$ . It appears that in the unrestricted model, the estimated measurement, or equivalently pricing, errors match the initial model specification better.

## 6 Conclusion

We evaluate three affine term structure models which differ in the volatility specification: constant, stochastic and unspanned. We use the prices of Eurodollar futures and options to conduct comprehensive statistical and economic analysis. We find that the futures-only data (the information in the term structure) 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 based on statistical tests and pricing errors. The latter are smaller than those in the unspanned volatility model by up to six times. We discover that the weak linear relationship between option-implied volatility and yields can be easily explained by small measurement errors that could be attributed to the market structure features.

## A Affine Models of the Interest Rate Under $\mathbb{Q}$

Equations (3.2)-(3.4), (3.8), and (3.9)-(3.11) imply that under the measure  $\mathbb{Q}$  the  $\mathbb{A}_0(3)$  model evolves according to:

$$dr_t = \left( \gamma_r^{\mathbb{Q}} - \underline{\kappa_r^{\mathbb{Q}}} r_t - \underline{\kappa_{r\theta}^{\mathbb{Q}}} \theta_t - \underline{\kappa_{rs}^{\mathbb{Q}}} s_t \right) dt + \sigma_r dW_t^r(\mathbb{Q}) + \sigma_{r\theta} \sigma_\theta dW_t^\theta(\mathbb{Q}) + \sigma_{rs} dW_t^s(\mathbb{Q}) \quad (\text{A.1})$$

$$d\theta_t = \left( \gamma_\theta^{\mathbb{Q}} - \boxed{\kappa_{\theta r}^{\mathbb{Q}}} r_t - \underline{\kappa_\theta^{\mathbb{Q}}} \theta_t - \boxed{\kappa_{\theta s}^{\mathbb{Q}}} s_t \right) dt + \sigma_\theta dW_t^\theta(\mathbb{Q}) + \sigma_{\theta s} dW_t^s(\mathbb{Q}) \quad (\text{A.2})$$

$$ds_t = \left( \gamma_s^{\mathbb{Q}} - \boxed{\kappa_{sr}^{\mathbb{Q}}} r_t - \boxed{\kappa_{s\theta}^{\mathbb{Q}}} \theta_t - \underline{\kappa_s^{\mathbb{Q}}} s_t \right) dt + \sigma_s dW_t^s(\mathbb{Q}) \quad (\text{A.3})$$

where

$$\gamma_r^{\mathbb{Q}} = -\lambda_r \sigma_r - \lambda_\theta \sigma_{r\theta} \sigma_\theta - \lambda_s \sigma_{rs} \quad (\text{A.4})$$

$$\kappa_r^{\mathbb{Q}} = \kappa_r^{\mathbb{P}} + \lambda_{rr} \sigma_r + \lambda_{\theta r} \sigma_{r\theta} \sigma_\theta + \lambda_{sr} \sigma_{rs} \quad (\text{A.5})$$

$$\kappa_{r\theta}^{\mathbb{Q}} = -\kappa_r^{\mathbb{P}} + \lambda_{r\theta} \sigma_r + \lambda_{\theta\theta} \sigma_{r\theta} \sigma_\theta + \lambda_{s\theta} \sigma_{rs} \quad (\text{A.6})$$

$$\kappa_{rs}^{\mathbb{Q}} = -\kappa_r^{\mathbb{P}} + \lambda_{rs} \sigma_r + \lambda_{\theta s} \sigma_{r\theta} \sigma_\theta + \lambda_{ss} \sigma_{rs} \quad (\text{A.7})$$

$$\gamma_\theta^{\mathbb{Q}} = \kappa_\theta^{\mathbb{P}} \bar{\theta} - \lambda_\theta \sigma_\theta - \lambda_s \sigma_{\theta s} \quad (\text{A.8})$$

$$\kappa_{\theta r}^{\mathbb{Q}} = \lambda_{\theta r} \sigma_\theta + \lambda_{sr} \sigma_{\theta s} \quad (\text{A.9})$$

$$\kappa_\theta^{\mathbb{Q}} = \kappa_\theta^{\mathbb{P}} + \lambda_{\theta\theta} \sigma_\theta + \lambda_{s\theta} \sigma_{\theta s} \quad (\text{A.10})$$

$$\kappa_{\theta s}^{\mathbb{Q}} = \lambda_{\theta s} \sigma_\theta + \lambda_{ss} \sigma_{\theta s} \quad (\text{A.11})$$

$$\gamma_s^{\mathbb{Q}} = -\lambda_s \sigma_s \quad (\text{A.12})$$

$$\kappa_{sr}^{\mathbb{Q}} = \lambda_{sr} \sigma_s \quad (\text{A.13})$$

$$\kappa_{s\theta}^{\mathbb{Q}} = \lambda_{s\theta} \sigma_s \quad (\text{A.14})$$

$$\kappa_s^{\mathbb{Q}} = \kappa_s^{\mathbb{P}} + \lambda_{ss} \sigma_s \quad (\text{A.15})$$

If the risk premia are completely affine, as in DS1, the parameters in boxes disappear, and underlined parameters become equal to  $\mathbb{P}$  parameters.

Equations (3.5)-(3.7), (3.8), and (3.12)-(3.14) imply that under the measure  $\mathbb{Q}$  the  $\mathbb{A}_1(3)$  model evolves according to:

$$\begin{aligned} dr_t &= \left( \gamma_r^{\mathbb{Q}} - \underline{\kappa_r^{\mathbb{Q}}} r_t - \underline{\kappa_{r\theta}^{\mathbb{Q}}} \theta_t - \underline{\kappa_{rv}^{\mathbb{Q}}} v_t \right) dt \\ &\quad + \sqrt{\sigma_r^2 + v_t} dW_t^r(\mathbb{Q}) + \sigma_{r\theta} \sqrt{\sigma_\theta^2 + \beta_\theta v_t} dW_t^\theta(\mathbb{Q}) + \sigma_{rv} \sqrt{v_t} dW_t^v(\mathbb{Q}) \\ d\theta_t &= \left( \gamma_\theta^{\mathbb{Q}} - \boxed{\kappa_{\theta r}^{\mathbb{Q}}} r_t - \underline{\kappa_\theta^{\mathbb{Q}}} \theta_t - \underline{\kappa_{\theta v}^{\mathbb{Q}}} v_t \right) dt \end{aligned} \quad (\text{A.16})$$

$$+ \sigma_{\theta r} \sqrt{\sigma_r^2 + v_t} dW_t^r(\mathbb{Q}) + \sqrt{\sigma_\theta^2 + \beta_\theta v_t} dW_t^\theta(\mathbb{Q}) + \sigma_{\theta v} \sqrt{v_t} dW_t^v(\mathbb{Q}) \quad (\text{A.17})$$

$$dv_t = \left( \kappa_v^{\mathbb{P}} \bar{v} - \kappa_v^{\mathbb{Q}} v_t \right) dt + \sigma_v \sqrt{v_t} dW_t^v(\mathbb{Q}) \quad (\text{A.18})$$

where

$$\gamma_r^{\mathbb{Q}} = \kappa_{rv}^{\mathbb{P}} \bar{v} - \lambda_r \sigma_r^2 - \lambda_\theta \sigma_\theta^2 \sigma_{r\theta} \quad (\text{A.19})$$

$$\kappa_r^{\mathbb{Q}} = \kappa_r^{\mathbb{P}} + \lambda_{rr} + \lambda_{\theta r} \sigma_{r\theta} \quad (\text{A.20})$$

$$\kappa_{r\theta}^{\mathbb{Q}} = -\kappa_r^{\mathbb{P}} + \lambda_{r\theta} + \lambda_{\theta\theta} \sigma_{r\theta} \quad (\text{A.21})$$

$$\kappa_{rv}^{\mathbb{Q}} = \kappa_{rv}^{\mathbb{P}} + \lambda_r + \lambda_{rv} + \lambda_\theta \beta_\theta \sigma_{r\theta} + \lambda_{\theta v} \sigma_{r\theta} + \lambda_v \sigma_{rv} \quad (\text{A.22})$$

$$\gamma_\theta^{\mathbb{Q}} = \kappa_\theta^{\mathbb{P}} \bar{\theta} - \lambda_r \sigma_r^2 \sigma_{\theta r} - \lambda_\theta \sigma_\theta^2 \quad (\text{A.23})$$

$$\kappa_{\theta r}^{\mathbb{Q}} = \lambda_{rr} \sigma_{\theta r} + \lambda_{\theta r} \quad (\text{A.24})$$

$$\kappa_\theta^{\mathbb{Q}} = \kappa_\theta^{\mathbb{P}} + \lambda_{r\theta} \sigma_{\theta r} + \lambda_{\theta\theta} \quad (\text{A.25})$$

$$\kappa_{\theta v}^{\mathbb{Q}} = \lambda_r \sigma_{\theta r} + \lambda_{rv} \sigma_{\theta r} + \lambda_\theta \beta_\theta + \lambda_{\theta v} + \lambda_v \sigma_{\theta v} \quad (\text{A.26})$$

$$\kappa_v^{\mathbb{Q}} = \kappa_v^{\mathbb{P}} + \lambda_v \sigma_v \quad (\text{A.27})$$

As in the  $\mathbb{A}_0(3)$  model, if the risk premia are completely affine, the parameters in boxes disappear, and underlined parameters become equal to  $\mathbb{P}$  parameters.

## B The USV Constraints

We use USV conditions (38)-(41) from CDGJ. In addition, CDGJ impose the condition  $\sigma_{r\theta}^2 \beta_\theta = 0$  to ensure that  $\mathbb{A}_1(3)$  does not reduce to a two-factor model. These conditions can be rewritten in terms of the  $\mathbb{Q}$ -parameters:

$$\kappa_{\theta r}^{\mathbb{Q}} + \kappa_r^{\mathbb{Q}} \sigma_{\theta r} + 2\kappa_{r\theta}^{\mathbb{Q}} \sigma_{\theta r}^2 = 0 \quad (\text{B.1})$$

$$3\kappa_{r\theta}^{\mathbb{Q}} \sigma_{\theta r} + 2\kappa_r^{\mathbb{Q}} - \kappa_\theta^{\mathbb{Q}} = 0 \quad (\text{B.2})$$

$$\kappa_{rv}^{\mathbb{Q}} \sigma_v - \kappa_{r\theta}^{\mathbb{Q}} (\sigma_{rv} \sigma_{\theta r} - \sigma_{\theta v}) = 0 \quad (\text{B.3})$$

$$\kappa_{rv}^{\mathbb{Q}} (\kappa_v^{\mathbb{Q}} - \kappa_\theta^{\mathbb{Q}}) + \kappa_{r\theta}^{\mathbb{Q}} \kappa_{\theta v}^{\mathbb{Q}} - 1 - \sigma_{rv}^2 = 0 \quad (\text{B.4})$$

$$\beta_\theta = 0 \quad (\text{B.5})$$

Next, we replace the  $\mathbb{Q}$ -parameters by the combination of the risk premia and the  $\mathbb{P}$ -parameters based on the expressions in the appendix A and obtain the following restrictions on the parameters to

be estimated:

$$\kappa_{\theta}^{\mathbb{P}} = 3\sigma_{\theta r}(-\kappa_r^{\mathbb{P}} + \sigma_{r\theta}\lambda_{\theta\theta}) + 2(\kappa_r^{\mathbb{P}} + \lambda_{rr} + \sigma_{r\theta}\lambda_{\theta r}) + 2\sigma_{\theta r}\lambda_{r\theta} - \lambda_{\theta\theta} \quad (\text{B.6})$$

$$\begin{aligned} \kappa_v^{\mathbb{P}} &= \kappa_{\theta}^{\mathbb{P}} + \sigma_{\theta r}\lambda_{r\theta} + \lambda_{\theta\theta} - \sigma_v\lambda_v \\ &+ \frac{1 + \sigma_{rv}^2 + (\kappa_r^{\mathbb{P}} - \lambda_{r\theta} - \sigma_{r\theta}\lambda_{\theta\theta})(\sigma_{\theta r}\lambda_r + \sigma_{\theta r}\lambda_{rv} + \lambda_{\theta v} + \sigma_{\theta v}\lambda_v)}{\kappa_{rv}^{\mathbb{P}} + \lambda_r + \lambda_{rv} + \sigma_{r\theta}\lambda_{\theta v} + \sigma_{rv}\lambda_v} \end{aligned} \quad (\text{B.7})$$

$$\kappa_{rv}^{\mathbb{P}} = \frac{1}{\sigma_v}(-\kappa_r^{\mathbb{P}} + \lambda_{r\theta} + \sigma_{r\theta}\lambda_{\theta\theta})(\sigma_{rv}\sigma_{\theta r} - \sigma_{\theta v}) - \lambda_r - \lambda_{rv} - \sigma_{r\theta}\lambda_{\theta v} - \sigma_{rv}\lambda_v \quad (\text{B.8})$$

$$\beta_{\theta} = 0 \quad (\text{B.9})$$

$$\lambda_{\theta r} = -\frac{\sigma_{\theta r}}{1 + \sigma_{\theta r}\sigma_{r\theta}} \left[ \kappa_r^{\mathbb{P}} + 2\sigma_{\theta r}(-\kappa_r^{\mathbb{P}} + \lambda_{r\theta} + \sigma_{r\theta}\lambda_{\theta\theta}) + 2\lambda_{rr} \right] \quad (\text{B.10})$$

## C Option Pricing Formula

First, we need an additional computation. Using (3.17) we get:

$$\begin{aligned} (k - f_{t+\tau_c}(\tau_f - \tau_c))^+ &= \left( k + \frac{1}{d} - \frac{1}{d} e^{-A^P(d) + A^f(\tau_f - \tau_c) + B^f(\tau_f - \tau_c) \cdot \mathcal{S}_{t+\tau_c}} \right)^+ \\ &= \frac{1}{d} e^{-A^P(d) + A^f(\tau_f - \tau_c)} \left( \underbrace{(dk + 1) e^{A^P(d) - A^f(\tau_f - \tau_c)} - e^{B^f(\tau_f - \tau_c) \cdot \mathcal{S}_{t+\tau_c}}}_X \right)^+ \end{aligned} \quad (\text{C.1})$$

Substituting this back into (3.20) we obtain:

$$\begin{aligned} C_t(\tau_f, \tau_c, K) &= \frac{1}{d} e^{-A^P(d) + A^f(\tau_f - \tau_c)} E_t^{\mathbb{Q}} \left( e^{-\int_t^{t+\tau_c} r_s ds} \left( X - e^{B^f(\tau_f - \tau_c) \cdot \mathcal{S}_{t+\tau_c}} \right) \mathbf{1}_{\{B^f(\tau_f - \tau_c) \cdot \mathcal{S}_{t+\tau_c} \leq \log X\}} \right) \\ &= \frac{1}{d} e^{-A^P(d) + A^f(\tau_f - \tau_c)} \\ &\times \left( X G_{0, B^f(\tau_f - \tau_c)}(\log X, \mathcal{S}_t, t + \tau_c) - G_{B^f(\tau_f - \tau_c), B^f(\tau_f - \tau_c)}(\log X, \mathcal{S}_t, t + \tau_c) \right) \end{aligned} \quad (\text{C.2})$$

where

$$G_{a,b}(y, \mathcal{S}_t, T) = E_t^{\mathbb{Q}} \left( e^{-\int_t^T r_s ds} e^{a \cdot \mathcal{S}_T} \mathbf{1}_{\{b \cdot \mathcal{S}_T \leq y\}} \right) \quad (\text{C.3})$$

which is computed via the extended characteristic function:

$$\psi(u, \mathcal{S}_t, t, T) = E_t^{\mathbb{Q}} \left( e^{-\int_t^T r_s ds} e^{u \cdot \mathcal{S}_T} \right) = e^{A(T-t) + B(T-t) \cdot \mathcal{S}_t} \quad (\text{C.4})$$

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

$$G_{a,b}(y, \mathcal{S}_t, T) = \frac{\psi(a, \mathcal{S}_t, t, T)}{2} - \frac{1}{\pi} \int_0^{\infty} \frac{\Im(\psi(a + ivb, \mathcal{S}_t, t, T) e^{-ivy})}{v} dv \quad (\text{C.5})$$

Standard numerical integration techniques allow to compute the integral.

## D Encompassing

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

Suppose the true data-generating process is associated with the log-likelihood  $\mathcal{L}_d$ . We are trying to evaluate the proximity between two non-nested and potentially misspecified models with log-likelihoods  $\mathcal{L}_0$ , i.e.  $\mathbb{A}_0(\mathfrak{Z})$ , and  $\mathcal{L}_1$ , i.e.  $\mathbb{A}_1(\mathfrak{Z})$ . 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):

$$\hat{\Theta}_i = \arg \max_{\Theta_i} E^d \mathcal{L}_i(y, \Theta_i) = \arg \min_{\Theta_i} E^d [\mathcal{L}_d(y) - \mathcal{L}_i(y, \Theta_i)] \quad (\text{D.1})$$

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 similar to how pseudo-true value links a misspecified model to the true distribution:

$$b_{10}(\hat{\Theta}_0) = \arg \min_{\Theta_1} E^0 [\mathcal{L}_0(y, \hat{\Theta}_0) - \mathcal{L}_1(y, \Theta_1)] = \arg \max_{\Theta_1} E^0 [\mathcal{L}_1(y, \Theta_1)] \quad (\text{D.2})$$

where superscript 0 emphasizes distribution associated with  $\mathbb{A}_0(\mathfrak{Z})$ . In other words, a binding function is the value of parameter  $\Theta_1$  such that the distribution of  $\mathbb{A}_1(\mathfrak{Z})$  approximates the distribution of  $\mathbb{A}_0(\mathfrak{Z})$  in the best way possible (according to KLIC), or, equivalently, this is an estimate of  $\Theta_1$  if  $\mathcal{L}_0$  were the true distribution.

We say that the true distribution  $\mathcal{L}^d$  is such that  $\mathbb{A}_0(\mathfrak{Z})$  encompasses  $\mathbb{A}_1(\mathfrak{Z})$ ,  $\mathbb{A}_0(\mathfrak{Z}) \mathcal{E} \mathbb{A}_1(\mathfrak{Z})$ , if and only if the pseudo-true value coincides with the binding function,  $\hat{\Theta}_1 = b_{10}(\hat{\Theta}_0)$ . Put less formally,  $\mathbb{A}_0(\mathfrak{Z})$  encompasses  $\mathbb{A}_1(\mathfrak{Z})$  if the objects of interests associated with the latter behave as they should were  $\mathcal{L}^0$  the true distribution. Thus, similar to nested hypothesis testing, the encompassing model is preferred if it is more simple.

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. Since in this paper we rely on finite sample inference, it is easy to avoid these complications. The finite sample inference is conducted by simulating 800 samples from the  $\mathbb{A}_0(\mathfrak{Z})$  models and estimating the respective  $\mathbb{A}_1(\mathfrak{Z})$  models based on these samples. This procedure allows to compute the  $LR$  statistic and its distribution.

## E Distinguishing different models with unobservable states

The futures-based results in section 5 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. Since, 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 with the encompassing test of the USV model, the data are informative enough to unequivocally show that the model is misspecified. However, it is definitely not enough to distinguish the Gaussian and stochastic volatility model: 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, as our discussion hints, is to rely on additional data, which would enhance the power of our diagnostics. We explore this direction in the subsequent sections 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 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.<sup>38</sup> This procedure allows us to construct finite sample confidence intervals for the GARCH coefficients.

First of all, one can not estimate a GARCH model on the data simulated from the  $\mathbb{A}_0$  class. Since conditional volatility is constant, parameters governing the GARCH dynamics in (5.2) are not identified: both sets  $\alpha = \beta = 0$ , and  $c = \alpha = 0$  with  $\beta = 1$  produce constant volatility. Therefore, we restricted  $\beta$  to zero to allow the model to pick up some heteroscedasticity due to small sample noise via the ARCH component. We found that the 95% confidence bounds constructed based on 800 simulated paths cover the values of  $\alpha$  obtained from the sample, however they also cover zero.<sup>39</sup> This finding is not surprising because we simulated from a constant volatility model.

The more interesting question is whether the  $\mathbb{A}_1(3)$  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 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.

---

<sup>38</sup>Because of the computational complexity, we estimated univariate GARCH models for each of the principal components. To be consistent, we also estimate univariate GARCH in the data for the purposes of this exercise.

<sup>39</sup>The values of  $\alpha$  in the sample ranged from 0.07 for the curvature to 0.27 for the slope.

## References

- Ahn, Dong-Hyun, Robert F. Dittmar, and A. Ronald Gallant, 2002, Quadratic term structure models: Theory and evidence, *Review of Financial Studies* 15, 243–288.
- , and Bin Gao, 2003, Purebred or hybrid?: Reproducing the volatility in term structure dynamics, *Journal of Econometrics* 116, 147–180.
- Aït-Sahalia, Yacine, and Robert Kimmel, 2002, Estimating Affine multifactor term structure models using closed-form likelihood expansions, Working paper, Princeton University.
- Attari, Mukkaram, 2001, Testing interest rate models: What does futures and options data tell us?, Working Paper, University of Wisconsin-Madison.
- Backus, David, Kai Li, and Liuren Wu, 1998, The “hump-shaped” mean term structure of interest rate derivative vols, Working Paper, New York University.
- Balduzzi, Pierluigi, Sanjiv Das, Silverio Foresi, and Rangarajan K. Sundaram, 1996, A simple approach to three-factor affine models of the term structure, *Journal of Fixed Income* 6, 43–53.
- Bansal, Ravi, George Tauchen, and Hao Zhou, 2003, Regime-shifts in term structure, expectations hypothesis puzzle, and the real business cycle, forthcoming, *Journal of Business and Economic Statistics*.
- Bansal, Ravi, and Hao Zhou, 2002, Term structure of interest rates with regime shifts, *Journal of Finance* 57, 1997–2043.
- Brandt, Michael, and David Chapman, 2003, Comparing multifactor models of the term structure, Working Paper, Duke University.
- Brandt, Michael W., and Ping He, 2002, Simulated likelihood estimation of affine term structure models from panel data, Working paper, Wharton.
- Broadie, Mark, Mikhail Chernov, and Michael Johannes, 2005, Model specification and risk premia: Evidence from futures options, *Journal of Finance*, forthcoming.
- Casassus, Jaime, Pierre Collin-Dufresne, and Robert Goldstein, 2004, Unspanned stochastic volatility and fixed income derivatives pricing, Working Paper, U.C. Berkeley.

- Chen, Li, Damir Filipovic, and H. Vincent Poor, 2003, Quadratic term structure models for risk-free and defaultable rates, forthcoming, *Mathematical Finance*.
- Chen, Ren-RAW, and Louis Scott, 2003, Multi-factor CIR models of the term structure: Estimates and tests from a state-space model using a Kalman filter, *Journal of Real Estate Finance and Economics* 27.
- Cheng, Peng, and Olivier Scaillet, 2005, Linear-quadratic jump-diffusion modeling with application to stochastic volatility, Working paper, HEC Geneve.
- Cheredito, Patrick, Damir Filipovic, and Robert Kimmel, 2003, Market price of risk specifications for affine models: Theory and evidence, Working Paper, Princeton University.
- Collin-Dufresne, Pierre, and Robert Goldstein, 2002, Do bonds span the fixed income markets? Theory and evidence for unspanned stochastic volatility, *Journal of Finance* 57, 1685–1730.
- , and Christopher Jones, 2005, Can interest rate volatility be extracted from the cross section of bond yields? an investigation of unspanned stochastic volatility, Working Paper, UC Berkeley.
- Collin-Dufresne, Pierre, and Bruno Solnik, 2001, On the term structure of default premia in the swap and LIBOR markets, *Journal of Finance* 56, 1095–1115.
- Conley, Timothy, Lars P. Hansen, and W.F. Liu, 1997, Bootstrapping the long run, *Macroeconomic Dynamics* 1, 279–311.
- Dai, Qiang, and Kenneth Singleton, 2000, Specification analysis of affine term structure models, *Journal of Finance* 55, 1943–1978.
- , 2002, Expectation puzzles, time-varying risk premia, and dynamic models of the term structure, *Journal of Financial Economics* 63, 415–441.
- , 2003, Term structure modeling in theory and reality, *Review of Financial Studies* 16, 631–678.
- , and Wei Yang, 2003, Regime shifts in a dynamic term structure model of U.S. Treasury bond yields, Working Paper, Stanford University.
- de Jong, Frank, 2000, Time series and cross-section information in affine term-structure models, *Journal of Business and Economic Statistics* 18, 300–314.

- Dhaene, Geert, 1997, *Encompassing: Formulation, Properties and Testing* (Springer: Heidelberg).
- Duan, Jin-Chuan, and Jean-Guy Simonato, 1995, Estimating and testing exponential-affine term structure models by Kalman filter, Working Paper, CIRANO.
- Duffee, Gregory, 2002, Term premia and interest rate forecasts in affine models, *Journal of Finance* 57, 405–443.
- , and Richard Stanton, 2004, Estimation of dynamic term structure models, Working paper, University of California at Berkeley.
- Duffie, Darrell, and Rui Kan, 1996, A yield-factor model of interest rates, *Mathematical Finance* 6, 379–406.
- Duffie, Darrell, Jun Pan, and Kenneth Singleton, 2000, Transform analysis and asset pricing for affine jump-diffusions, *Econometrica* 68, 1343–1376.
- Duffie, Darrell, Lasse Pedersen, and Kenneth Singleton, 2003, Modeling sovereign yield spreads: A case study of Russian debt, *Journal of Finance* 58, 119–160.
- Duffie, Darrell, and Kenneth Singleton, 1997, An econometric model of the term structure of interest-rate swap yields, *Journal of Finance* 52, 1287–1321.
- Fisher, Mark, and Christian Gilles, 1996, Estimating exponential-affine models of the term structure, Working Paper, Federal Reserve.
- Flesaker, Bjorn, 1993, Testing the Heath-Jarrow-Morton/Ho-Lee model of interest rate contingent claims pricing, *Journal of Financial and Quantitative Analysis* 28, 483–495.
- Frühwirth-Schnatter, Sylvia, and Alois L.J. Geyer, 1998, Bayesian estimation of econometric multifactor Cox-Ingersoll-Ross models of the term structure of interest rates via MCMC methods, Working Paper, Vienna University of Economics and Business Administration.
- Gallant, A. Ronald, and George Tauchen, 1998, Reprojecting partially observed systems with application to interest rate diffusions, *Journal of American Statistical Association* 93, 10–24.
- Gourieroux, Christian, and Alain Monfort, 1994, Testing non-nested hypothesis, in R. F. Engle, and D.L. McFadden, ed.: *Handbook of Econometrics, Vol. IV* (Elsevier Science: Amsterdam).

- Hamilton, James D., 1994, *Time Series Analysis* (Princeton University Press: Princeton, NJ).
- Heidari, Massoud, and Liuren Wu, 2002, Term structure of interest rates, yield curve residuals, and the consistent pricing of interest rates and interest rate derivatives, Working Paper, Baruch College.
- Heston, Steven, 1993, A closed-form solution for options with stochastic volatility with applications to bond and currency options, *Review of Financial Studies* 6, 327–343.
- Jagannathan, Ravi, Andrew Kaplin, and Steve Sun, 2003, An evaluation of multi-factor CIR models using LIBOR, swap rates, and cap and swaption prices, *Journal of Econometrics* 116, 113–146.
- Jegadeesh, Narasimhan, and George G. Pennacchi, 1996, The behavior of interest rates implied by the term structure of Eurodollar futures, *Journal of Money, Credit and Banking* 28, 426–446.
- Johannes, Michael, and Suresh Sundaresan, 2003, Collateralized swaps, Working Paper, Columbia University.
- Liptser, Robert S., and Albert N. Shiryaev, 2001, *Statistics of Random Processes: I. General Theory* (Springer: Berlin Heidelberg) 2nd edn.
- Litterman, Robert, and Jose Scheinkman, 1991, Common factors affecting bond returns, *Journal of Fixed Income* 3, 34–61.
- , and Laurence Weiss, 1991, Volatility and the yield curve, *Journal of Fixed Income* pp. 49–53.
- Lund, Jesper, 1997, Econometric analysis of continuous-time arbitrage-free models of the term-structure of interest rates, Working Paper, The Aarhus School of Business.
- Piazzesi, Monika, 2003, Bond yields and the Federal Reserve, forthcoming, *Journal of Political Economy*.
- , 2005, Affine term structure models, in Y. Aït-Sahalia, and L.P. Hansen, ed.: *Handbook of Financial Econometrics* (Elsevier Science: Amsterdam) forthcoming.
- Picard, Jean, 1991, Efficiency of the extended Kalman filter for nonlinear systems with small noise, *SIAM Journal on Applied Mathematics* 51, 843–885.
- Singleton, Kenneth, 2005, *Dynamic Asset Pricing Models: Econometric Specifications and Empirical Assessments* (Princeton University Press: Princeton, NJ).

Umantsev, Leonid, 2001, Econometric analysis of European LIBOR-based options within affine term-structure models, Ph.D. dissertation, Stanford University.

**Table 1 : Major Empirical Findings in the ATSM literature**

We summarize the findings of the papers evaluating empirical performance of the affine term structure models. By tradition we denote them by  $\mathbb{A}_m(N)$ , where  $N$  represents the total number of factors, and  $m$  is the number of “square-root” factors affecting the conditional volatility.  $\mathbb{A}^Q$  denotes a restricted affine model, which represents an equivalent affine representation of a linear-quadratic model (details are in Cheng and Scaillet, 2005).  $\mathbb{A}^{USV}$  denotes a restricted affine model, which represents an unspanned stochastic volatility model of Collin-Dufresne and Goldstein (2002). The notations for the authors are: DS1 – Dai and Singleton (2000); DS2 – Dai and Singleton (2002); Duffee – Duffee (2002); ADG – Ahn, Dittmar, and Gallant (2002); ADGG – Ahn, Dittmar, Gallant, and Gao (2003); BC – Brandt and Chapman (2003); CDGJ – Collin-Dufresne, Goldstein, and Jones (2005); JKS – Jagannathan, Kaplin, and Sun (2003); Umantsev – Umantsev (2001).  $J$ -test denotes the GMM overidentifying restrictions test,  $LR$ -test denotes the likelihood ratio test,  $LPY$  denotes the expectation hypothesis regression coefficients pattern (see DS2 for details),  $CVY$  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  $\mathbb{A}_1^Q(3)$  and an inverted model). The dagger (†) in the BC conclusions indicates that the paper is incomplete.

Authors	Models	Data Frequency	Data Type	Primary Evaluation Metric	Conclusions
DS1	$\mathbb{A}_1(3), \mathbb{A}_2(3)$	Weekly	Yields	Pricing errors, $CVY$	$\mathbb{A}_1(3)$ is preferred
Duffee	$\mathbb{A}_0(3), \mathbb{A}_1(3), \mathbb{A}_2(3)$	Monthly	Yields	Forecast errors	$\mathbb{A}_0(3)$ is preferred
DS2	$\mathbb{A}_0(3), \mathbb{A}_1(3), \mathbb{A}_2(3), \mathbb{A}_3(3)$	Monthly	Yields	$LPY$	$\mathbb{A}_0(3)$ is preferred
ADG	$\mathbb{A}_2(3), \mathbb{A}_3^Q(9)$	Monthly	Yields	$J$ -test, filtered volatility	$\mathbb{A}_3^Q(9)$ is preferred
ADGG	$\mathbb{A}_2(3), \mathbb{A}_2^Q(4), \mathbb{A}_3^Q(9)^*$	Monthly	Yields	$J$ -test, filtered volatility	$\mathbb{A}_3^Q(9)$ is preferred
BC	$\mathbb{A}_0(3), \mathbb{A}_1(3), \mathbb{A}_3^Q(9)$	Monthly	Yields	$J$ -test (based on $LPY$ , other ‘economic’ moments)	$\mathbb{A}_3^Q(9)$ is preferred†
CDGJ	$\mathbb{A}_1(2), \mathbb{A}_1(3), \mathbb{A}_1^{USV}(3), \mathbb{A}_1^{USV}(4)$	Weekly	Yields	$LR$ -test (for nested models), Pricing and forecast errors	$\mathbb{A}_1(3) \approx \mathbb{A}_1^{USV}(4)$
JKS	$\mathbb{A}_1(1), \mathbb{A}_2(2), \mathbb{A}_3(3)$	Weekly	Yields and Options	Pricing errors	Options data do not improve option pricing
Umantsev	$\mathbb{A}_1(2), \mathbb{A}_1(3)$	Daily	Yields and Options	Implied volatility fits	Options data improve option pricing

**Table 2 : Comparison of the MLE estimates of the  $\mathbb{A}_1(3)$  model based on the exact inversion and the Kalman filter**

The model  $\mathbb{A}_1(3)$  dynamics is given by

$$\begin{aligned} dr_t &= \kappa_r^{\mathbb{P}}(\theta_t - r_t) dt + \kappa_{rv}^{\mathbb{P}}(\bar{v} - v_t) dt + \sqrt{\sigma_r^2 + v_t} dW_t^r(\mathbb{P}) + \sigma_{r\theta} \sqrt{\sigma_\theta^2 + \beta_\theta v_t} dW_t^\theta(\mathbb{P}) + \sigma_{rv} \sqrt{v_t} dW_t^v(\mathbb{P}) \\ d\theta_t &= \kappa_\theta^{\mathbb{P}}(\bar{\theta} - \theta_t) dt + \sigma_{\theta r} \sqrt{\sigma_r^2 + v_t} dW_t^r(\mathbb{P}) + \sqrt{\sigma_\theta^2 + \beta_\theta v_t} dW_t^\theta(\mathbb{P}) + \sigma_{\theta v} \sqrt{v_t} dW_t^v(\mathbb{P}) \\ dv_t &= \kappa_v^{\mathbb{P}}(\bar{v} - v_t) dt + \sigma_v \sqrt{v_t} dW_t^v(\mathbb{P}) \end{aligned}$$

and the risk premia are:

$$\begin{aligned} \Lambda_t^r &= \frac{1}{\sqrt{\sigma_r^2 + v_t}} (\lambda_r \sigma_r^2 + \lambda_{rr} r_t + \lambda_{r\theta} \theta_t + (\lambda_r + \lambda_{rv}) v_t) \\ \Lambda_t^\theta &= \frac{1}{\sqrt{\sigma_\theta^2 + \beta_\theta v_t}} (\lambda_\theta \sigma_\theta^2 + \lambda_{\theta r} r_t + \lambda_{\theta\theta} \theta_t + (\lambda_\theta \beta_\theta + \lambda_{\theta v}) v_t) \\ \Lambda_t^v &= \lambda_v \sqrt{v_t} \end{aligned}$$

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

Method	Kalman		Inversion	
	Est.	std. err. ( <i>SE1</i> , <i>SE2</i> )	Est.	std. err. ( <i>SE1</i> , <i>SE2</i> )
$\kappa_r^{\mathbb{P}}$	0.34	(0.01, 0.02)	2.63	(2.53, 0.10)
$\kappa_\theta^{\mathbb{P}}$	0.74	(0.31, 0.36)	1.84	(2.06, 0.49)
$\kappa_v^{\mathbb{P}}$	0.71	(0.13, 0.09)	0.04	(0.04, 0.03)
$\kappa_{rv}^{\mathbb{P}} \times 10^{-2}$	-0.57	(1.39, 1.35)	-21.36	(39.97, 7.07)
$\bar{\theta} \times 10^2$	3.39	(3.27, 2.16)	1.61	(4.21, 0.80)
$\bar{v} \times 10^4$	0.69	(0.14, 0.11)	0.82	(0.59, 0.32)
$\sigma_v \times 10^2$	0.25	(0.07, 0.08)	0.22	(0.08, 0.06)
$\sigma_r \times 10^2$	0		0.22	(1.52, 0.19)
$\sigma_\theta \times 10^2$	2.81	(1.56, 1.34)	0	
$\beta_\theta$	4.15	(8.02, 7.32)	1.58	(2.08, 0.57)
$\sigma_{r\theta}$	0.11	(0.07, 0.04)	0.23	(1.11, 0.36)
$\sigma_{rv}$	-0.21	(0.09, 0.08)	0.26	(0.11, 0.03)
$\sigma_{\theta r}$	0		-0.18	(0.33, 0.10)
$\sigma_{\theta v}$	5.06	(2.04, 3.01)	-1.09	(1.29, 0.53)
$\lambda_r \times 10^{-2}$	-3.31	(0.41, 0.45)	-276.96	(3784.77, 189.88)
$\lambda_\theta \times 10^{-2}$	-0.91	(1.31, 1.38)	0	
$\lambda_v \times 10^{-2}$	0.36	(0.52, 0.35)	-0.19	(0.13, 0.08)
$\lambda_{rr}$	0		-0.97	(1.96, 0.45)
$\lambda_{r\theta}$	0.32	(0.03, 0.04)	1.83	(2.63, 0.65)
$\lambda_{rv} \times 10^{-4}$	0		2.89	(37.82, 2.30)
$\lambda_{\theta r}$	0	<b>39</b>	0.31	(0.51, 0.05)
$\lambda_{\theta\theta}$	-0.74	(0.31, 0.36)	-1.51	(2.07, 0.19)
$\lambda_{\theta v} \times 10^{-2}$	16.98	(13.14, 16.40)	0	
log-likelihood	73.67		70.84	

**Table 3 : Estimated parameters for model  $\mathbb{A}_0(3)$ /Monte Carlo Study**

The model  $\mathbb{A}_0(3)$  dynamics is given by

$$\begin{aligned} dr_t &= \kappa_r^{\mathbb{P}}(\theta_t + s_t - r_t) dt + \sigma_r dW_t^r(\mathbb{P}) + \sigma_{r\theta}\sigma_{\theta} dW_t^{\theta}(\mathbb{P}) + \sigma_{rs} dW_t^s(\mathbb{P}) \\ d\theta_t &= \kappa_{\theta}^{\mathbb{P}}(\bar{\theta} - \theta_t) dt + \sigma_{\theta} dW_t^{\theta}(\mathbb{P}) + \sigma_{\theta s} dW_t^s(\mathbb{P}) \\ ds_t &= -\kappa_s^{\mathbb{P}} s_t dt + \sigma_s dW_t^s(\mathbb{P}) \end{aligned}$$

and the risk premia are:

$$\begin{aligned} \Lambda_t^r &= \lambda_r + \lambda_{rr} r_t + \lambda_{r\theta} \theta_t + \lambda_{rs} s_t \\ \Lambda_t^{\theta} &= \lambda_{\theta} + \lambda_{\theta r} r_t + \lambda_{\theta\theta} \theta_t + \lambda_{\theta s} s_t \\ \Lambda_t^s &= \lambda_s + \lambda_{sr} r_t + \lambda_{s\theta} \theta_t + \lambda_{ss} s_t \end{aligned}$$

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

Model	$\mathbb{A}_0(3)^f$			$\mathbb{A}_0(3)^{fo}$		
	Estimate	Bias, %	Confidence interval	Estimate	Bias, %	Confidence interval
$\kappa_r^{\mathbb{P}}$	0.90	(-4.68)	[0.59, 1.49]	1.18	(-4.72)	[0.48, 2.36]
$\kappa_{\theta}^{\mathbb{P}}$	1.84	(-10.97)	[1.17, 3.63]	1.15	(-0.91)	[0.64, 1.76]
$\kappa_s^{\mathbb{P}}$	0.16	(-7.11)	[0.07, 0.46]	0.88	(-0.40)	[0.41, 1.26]
$\bar{\theta} \times 10^2$	0.30	(1.40)	[-0.00, 0.73]	3.43	(3.47)	[0.64, 5.89]
$\sigma_r \times 10^2$	0.87	(2.83)	[0.73, 0.95]	0.70	(1.43)	[0.58, 0.82]
$\sigma_{\theta}$	0.02	(2.07)	[0.01, 0.04]	0.02	(-16.29)	[0.01, 0.07]
$\sigma_s$	0.03	(-2.54)	[0.02, 0.05]	0.04	(1.92)	[0.02, 0.05]
$\sigma_{r\theta}$	0.06	(-23.30)	[0.01, 0.18]	-0.15	(5.19)	[-0.30, -0.05]
$\sigma_{rs} \times 10^2$	0.35	(-2.36)	[0.24, 0.48]	0.10	(-38.09)	[-0.05, 0.41]
$\sigma_{\theta s}$	-0.01	(-4.77)	[-0.03, -0.00]	-0.02	(0.54)	[-0.05, -0.00]
$\lambda_r$	-3.88	(-3.23)	[-4.91, -3.24]	-2.85	(10.09)	[-5.39, 2.11]
$\lambda_{\theta}$	0			0		
$\lambda_s$	-0.14	(-0.52)	[-0.16, -0.13]	0		
$\lambda_{rr}$	-11.54	(-46.58)	[-81.78, 24.80]	12.27	(142.82)	[-109.95, 59.70]
$\lambda_{r\theta}$	75.46	(-12.78)	[38.04, 148.32]	0		
$\lambda_{rs}$	67.21	(-9.69)	[39.89, 131.80]	34.37	(-16.98)	[4.91, 81.21]
$\lambda_{\theta r}$	0			-10.17	(37.86)	[-12.87, 3.69]
$\lambda_{\theta\theta}$	-56.74	(-22.32)	[-168.65, -31.87]	0		
$\lambda_{\theta s}$	0.86	(24.19)	[-5.24, 3.30]	7.40	(30.72)	[-12.52, 20.89]
$\lambda_{sr}$	0			14.71	(20.64)	[-2.15, 25.79]
$\lambda_{s\theta}$	0			-20.64	(2.15)	[-35.89, -6.11]
$\lambda_{ss}$	-5.00	(-5.34)	[-14.54, -2.66]	-35.45	(2.02)	[-46.36, -22.06]
log-likelihood	73.66			85.09		

**Table 4 : Estimated parameters for model  $\mathbb{A}_1(3)$**

The model  $\mathbb{A}_1(3)$  dynamics is given by

$$\begin{aligned} dr_t &= \kappa_r^{\mathbb{P}}(\theta_t - r_t) dt + \kappa_{rv}^{\mathbb{P}}(\bar{v} - v_t) dt + \sqrt{\sigma_r^2 + v_t} dW_t^r(\mathbb{P}) + \sigma_{r\theta} \sqrt{\sigma_\theta^2 + \beta_\theta v_t} dW_t^\theta(\mathbb{P}) + \sigma_{rv} \sqrt{v_t} dW_t^v(\mathbb{P}) \\ d\theta_t &= \kappa_\theta^{\mathbb{P}}(\bar{\theta} - \theta_t) dt + \sigma_{\theta r} \sqrt{\sigma_r^2 + v_t} dW_t^r(\mathbb{P}) + \sqrt{\sigma_\theta^2 + \beta_\theta v_t} dW_t^\theta(\mathbb{P}) + \sigma_{\theta v} \sqrt{v_t} dW_t^v(\mathbb{P}) \\ dv_t &= \kappa_v^{\mathbb{P}}(\bar{v} - v_t) dt + \sigma_v \sqrt{v_t} dW_t^v(\mathbb{P}) \end{aligned}$$

and the risk premia are:

$$\begin{aligned} \Lambda_t^r &= \frac{1}{\sqrt{\sigma_r^2 + v_t}} (\lambda_r \sigma_r^2 + \lambda_{rr} r_t + \lambda_{r\theta} \theta_t + (\lambda_r + \lambda_{rv}) v_t) \\ \Lambda_t^\theta &= \frac{1}{\sqrt{\sigma_\theta^2 + \beta_\theta v_t}} (\lambda_\theta \sigma_\theta^2 + \lambda_{\theta r} r_t + \lambda_{\theta\theta} \theta_t + (\lambda_\theta \beta_\theta + \lambda_{\theta v}) v_t) \\ \Lambda_t^v &= \lambda_v \sqrt{v_t} \end{aligned}$$

The bootstrapped 95% confidence bounds are reported in square brackets. An asterisk (\*) denotes parameters significant at the 10% level. A dagger (†) denotes parameters restricted by the USV conditions (B.6)-(B.10).

Model	$\mathbb{A}_1(3)^f$		$\mathbb{A}_1(3)^{fo}$		$\mathbb{A}_1^{USV}(3)^f$		$\mathbb{A}_1^{USV}(3)^{fo}$	
	Est.	Conf. int.	Est.	Conf. int.	Est.	Conf. int.	Est.	Conf. int.
$\kappa_r^{\mathbb{P}}$	0.34	[0.32, 0.36]	1.19	[0.94, 1.39]	0.26	[0.21, 0.63]	0.11	[0.11, 0.12]
$\kappa_\theta^{\mathbb{P}}$	0.74	[0.50, 1.40]	1.06	[0.78, 1.64]	0.40†	[0.39, 0.41]	0.20†	[0.09, 0.92]
$\kappa_v^{\mathbb{P}}$	0.71	[0.59, 0.88]	0.41	[0.24, 1.06]	0.25†	[0.14, 0.44]	1.03†	[0.65, 2.27]
$\kappa_{rv}^{\mathbb{P}} \times 10^{-2}$	-0.57	[-10.17, 1.10]	-4.99	[-41.98, 18.11]	0.44†	[0.05, 6.16]	0.87†	[0.06, 1.24]
$\bar{\theta} \times 10^2$	3.39	[0.10, 6.64]	4.37	[2.52, 5.08]	5.75	[0.56, 10.62]	3.90	[0.02, 9.16]
$\bar{v} \times 10^4$	0.69	[0.52, 0.85]	0.11	[0.04, 0.21]	6.50	[3.82, 11.43]	0.97	[0.65, 1.70]
$\sigma_v \times 10^2$	0.25	[0.09, 0.40]	0.31	[0.26, 0.39]	1.82	[1.55, 1.95]	1.35	[0.97, 2.18]
$\sigma_r$	0		0		0		0	
$\sigma_\theta \times 10^2$	2.81	[1.97, 3.51]	1.60	[1.42, 1.83]	2.76	[1.36, 3.86]	9.43	[5.60, 14.26]
$\beta_\theta$	4.15*	[0.00, 11.98]	5.28*	[0.00, 6.11]	0†		0†	
$\sigma_{r\theta}$	0.11	[0.07, 0.16]	-0.34	[-0.42, -0.26]	-0.15	[-0.41, -0.05]	-0.00	[-0.01, 0.01]
$\sigma_{rv}$	-0.21	[-0.39, -0.01]	0.16	[-2.01, 0.34]	0.29	[0.14, 0.44]	0.10	[-0.05, 0.30]
$\sigma_{\theta r}$	0		5.59	[3.60, 7.44]	0		0	
$\sigma_{\theta v}$	5.06	[1.33, 8.04]	-7.91	[-17.36, 0.14]	2.98	[0.00, 45.59]	-0.71	[-1.33, -0.06]
$\lambda_r \times 10^{-2}$	-3.31	[-4.44, -2.74]	-1.05	[-3.18, 0.72]	-0.10	[-0.27, 0.10]	0	
$\lambda_\theta \times 10^{-2}$	-0.91	[-1.26, -0.51]	2.76	[2.21, 3.48]	0.58	[0.26, 1.42]	0.02	[-0.01, 0.07]
$\lambda_v \times 10^{-2}$	0.36	[-0.28, 0.92]	0.15	[-1.87, 0.72]	-0.13	[-0.24, -0.08]	-0.76	[-1.29, -0.46]
$\lambda_{rr}$	0		-0.13	[-0.17, -0.10]	-0.06	[-0.46, -0.01]	0	
$\lambda_{r\theta}$	0.32	[0.29, 0.41]	0.05	[-0.01, 0.08]	0.05	[-0.08, 0.36]	0.07	[0.04, 0.09]
$\lambda_{rv}$	0		0		0		-81.82	[-112.96, 0.10]
$\lambda_{\theta r}$	0		0		0†		0†	
$\lambda_{\theta\theta}$	-0.74	[-1.40, -0.50]	-0.69	[-1.21, -0.37]	0		0.03	[-0.70, 0.14]
$\lambda_{\theta v} \times 10^{-2}$	16.98	[11.12, 22.79]	-19.57	[-34.34, -4.95]	-0.29	[-0.61, -0.02]	-0.66	[-1.13, -0.29]
log-likelihood	73.67		86.35		72.30		81.95	

**Table 5 : Correlations of state variables with observables**

We want to asses which aspects of the term structure are reflected in the model state variables. We, therefore, 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.

Factor	Observables	Short rate	Long rate	Slope	Butterfly	Implied Variance
	Model	$y^f(0.5)$	$y^f(10)$	$y^f(10) - y^f(0.5)$	$y^f(0.5) + y^f(10)$ $-2y^f(2)$	$(\sigma^{imp}(0.5) \cdot y^f(0.5))^2$
$r$	$\mathbb{A}_0(3)^f$	<b>0.94</b>	0.29	-0.74	0.01	-0.01
	$\mathbb{A}_0(3)^{fo}$	<b>0.69</b>	0.14	-0.61	0.45	-0.05
	$\mathbb{A}_1(3)^f$	<b>0.94</b>	0.29	-0.74	0.02	-0.01
	$\mathbb{A}_1(3)^{fo}$	<b>0.75</b>	0.14	-0.69	0.37	-0.08
	$\mathbb{A}_1^{USV}(3)^f$	<b>0.98</b>	0.38	-0.70	-0.15	0.02
	$\mathbb{A}_1^{USV}(3)^{fo}$	<b>0.98</b>	0.38	-0.70	-0.20	0.04
$\theta$	$\mathbb{A}_0(3)^f$	0.32	-0.36	<b>-0.69</b>	-0.52	-0.31
	$\mathbb{A}_0(3)^{fo}$	-0.02	<b>-0.77</b>	-0.68	-0.05	-0.57
	$\mathbb{A}_1(3)^f$	0.30	<b>0.83</b>	0.45	-0.77	0.45
	$\mathbb{A}_1(3)^{fo}$	0.72	0.47	-0.34	<b>-0.77</b>	0.04
	$\mathbb{A}_1^{USV}(3)^f$	0.42	<b>0.98</b>	0.45	-0.28	0.53
	$\mathbb{A}_1^{USV}(3)^{fo}$	0.04	<b>0.83</b>	0.72	-0.52	0.44
$s$	$\mathbb{A}_0(3)^f$	0.48	<b>0.99</b>	0.40	-0.38	0.53
	$\mathbb{A}_0(3)^{fo}$	0.48	<b>0.98</b>	0.38	-0.32	0.55
$v$	$\mathbb{A}_1(3)^f$	0.02	0.41	0.36	<b>-0.93</b>	0.24
	$\mathbb{A}_1(3)^{fo}$	-0.55	0.31	<b>0.89</b>	0.24	0.43
	$\mathbb{A}_1^{USV}(3)^f$	0.51	<b>0.96</b>	0.35	-0.16	0.50
	$\mathbb{A}_1^{USV}(3)^{fo}$	-0.02	0.45	0.44	0.22	<b>0.75</b>

**Table 6 : Pricing Errors**

We report 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)  $|\sigma^{imp}(\tau, \text{data}) - \sigma^{imp}(\tau, \text{model})| \cdot y^f(\tau, \text{data})$ . The line 'PC1' reports the percentage of the variation in the errors explained by the first principle component.

Maturity	$\mathbb{A}_0(3)^f$	$\mathbb{A}_1(3)^f$	$\mathbb{A}_1^{USV}(3)^f$	$\mathbb{A}_0(3)^{fo}$	$\mathbb{A}_1(3)^{fo}$	$\mathbb{A}_1^{USV}(3)^{fo}$
Futures, b.p.						
0.5	3.91	3.84	5.97	7.43	5.32	7.42
1	6.24	6.25	6.94	10.18	7.99	10.52
2	3.70	3.74	5.04	6.76	5.37	9.83
3	2.92	3.02	2.74	4.17	4.58	7.77
4	3.22	3.22	3.19	4.97	4.89	7.80
5	3.57	3.55	2.83	4.85	4.58	6.08
6	3.44	3.43	2.59	3.77	3.53	3.72
7	2.40	2.43	2.36	2.06	2.01	2.93
8	1.33	1.37	1.78	2.66	2.63	5.21
9	2.84	2.78	2.15	5.56	5.34	8.23
10	5.06	5.00	3.91	7.83	7.69	10.94
Average	3.51	3.51	3.59	5.48	4.90	7.31
Options, %						
0.5	0.1598	0.1303	1.5930	0.0353	0.0359	0.0873
1	0.1723	0.1218	1.5722	0.0484	0.0381	0.2448
PC1	55.27%	56.33%	87.34%	47.44%	40.31%	45.84%

**Table 7 : Encompassing Tests**

We test whether models from the  $\mathbb{A}_0(3)$  branch, or the  $\mathbb{A}_1^{USV}(3)$  branch encompass models from the  $\mathbb{A}_1(3)$  branch based on the finite sample likelihood-ratio test. The inference is conducted by simulating 800 samples from the  $\mathbb{A}_0(3)$ , or the  $\mathbb{A}_1^{USV}(3)$  models and estimating the respective  $\mathbb{A}_1(3)$  models based on these samples. This procedure allows to compute the  $LR$  statistic and its distribution.

Test	$LR$	Conf. int.	$p$ -value
$\mathbb{A}_0(3)^f \mathcal{E} \mathbb{A}_1(3)^f$	0.008	[-0.011, 0.047]	0.490
$\mathbb{A}_1^{USV}(3)^f \mathcal{E} \mathbb{A}_1(3)^f$	1.365	[0.002, 0.019]	0.000
$\mathbb{A}_0(3)^{fo} \mathcal{E} \mathbb{A}_1(3)^{fo}$	1.385	[1.408, 7.304]	0.025
$\mathbb{A}_1^{USV}(3)^{fo} \mathcal{E} \mathbb{A}_1(3)^{fo}$	4.415	[0.001, 0.071]	0.000

**Table 8 : The USV restrictions tests**

We test whether the USV conditions hold in the data. In order to do this, we have to check whether restrictions (B.6)-(B.10) hold in the model  $\mathbb{A}_1(3)$  estimated without imposing such restrictions. We simulated 800 paths, estimated the model for every path and constructed the 5% confidence bounds for the null hypothesis that LHS-RHS in (B.6)-(B.10) is equal to zero.

Restriction	$\mathbb{A}_1(3)^f$	$\mathbb{A}_1(3)^{fo}$
$\kappa_\theta^{\mathbb{P}}$	[-0.72, -0.65]	[8.26, 16.72]
$\kappa_v^{\mathbb{P}}$	[0.94, 5.11]	[-35.20, 48.68]
$\kappa_{rv}^{\mathbb{P}}$	[-1367.50, -233.55]	[-1036.00, 3344.30]
$\beta_\theta$	[0.00, 12.18]	[0.00, 6.14]
$\lambda_{\theta r}$	[0.00, 0.00]	[34.33, 122.43]

**Table 9 : Conditional mean and volatility correlations**

This table complements Figures 1 and 3 by reporting correlations between conditional mean and volatility computed from the data and the respective moments from the models.

	Level		Slope		Curvature	
	$y^f(0.5)$		$y^f(10) - y^f(0.5)$		$y^f(0.5) + y^f(10) - 2y^f(2)$	
	VAR(1)	GARCH(1,1)	VAR(1)	GARCH(1,1)	VAR(1)	GARCH(1,1)
$\mathbb{A}_0(3)^f$	0.9882	0.9001	0.9379	0.8974	0.8595	0.8663
$\mathbb{A}_1(3)^f$	0.9878	0.9010	0.9375	0.8952	0.8621	0.8665
$\mathbb{A}_1^{USV}(3)^f$	0.9786	0.8919	0.9062	0.8614	0.7763	0.7246
$\mathbb{A}_0(3)^{fo}$	0.9701	0.8030	0.9003	0.8478	0.8170	0.5048
$\mathbb{A}_1(3)^{fo}$	0.9811	0.8133	0.9082	0.7979	0.8590	0.5209
$\mathbb{A}_1^{USV}(3)^{fo}$	0.9723	0.7354	0.8087	0.7673	0.7282	0.2125

**Table 10 : Kurtosis Tests**

We test whether the estimated models can match 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. Boldfaced statistics indicate a failure to reject the null that a model can match a statistic.

Sample	Level		Slope		Curvature	
	$y^f(0.5)$		$y^f(10) - y^f(0.5)$		$y^f(0.5) + y^f(10) - 2y^f(2)$	
	4.31		4.70		5.95	
Model	Value	Conf. int.	Value	Conf. int.	Value	Conf. int.
$\mathbb{A}_0(3)^f$	2.98	[2.38, 3.51]	2.99	[2.38, 3.56]	2.98	[2.36, 3.53]
$\mathbb{A}_1(3)^f$	3.02	[2.60, 3.59]	3.00	[2.61, 3.58]	2.99	[2.58, 3.55]
$\mathbb{A}_1^{USV}(3)^f$	<b>4.05</b>	[2.93, 6.30]	<b>3.66</b>	[2.83, 5.21]	3.00	[2.59, 3.58]
$\mathbb{A}_0(3)^{fo}$	3.00	[2.60, 3.54]	2.98	[2.58, 3.55]	2.98	[2.58, 3.54]
$\mathbb{A}_1(3)^{fo}$	<b>3.46</b>	[2.77, 4.77]	3.17	[2.66, 4.03]	3.00	[2.58, 3.60]
$\mathbb{A}_1^{USV}(3)^{fo}$	<b>4.23</b>	[2.68, 5.17]	3.00	[2.59, 3.59]	2.98	[2.57, 3.52]

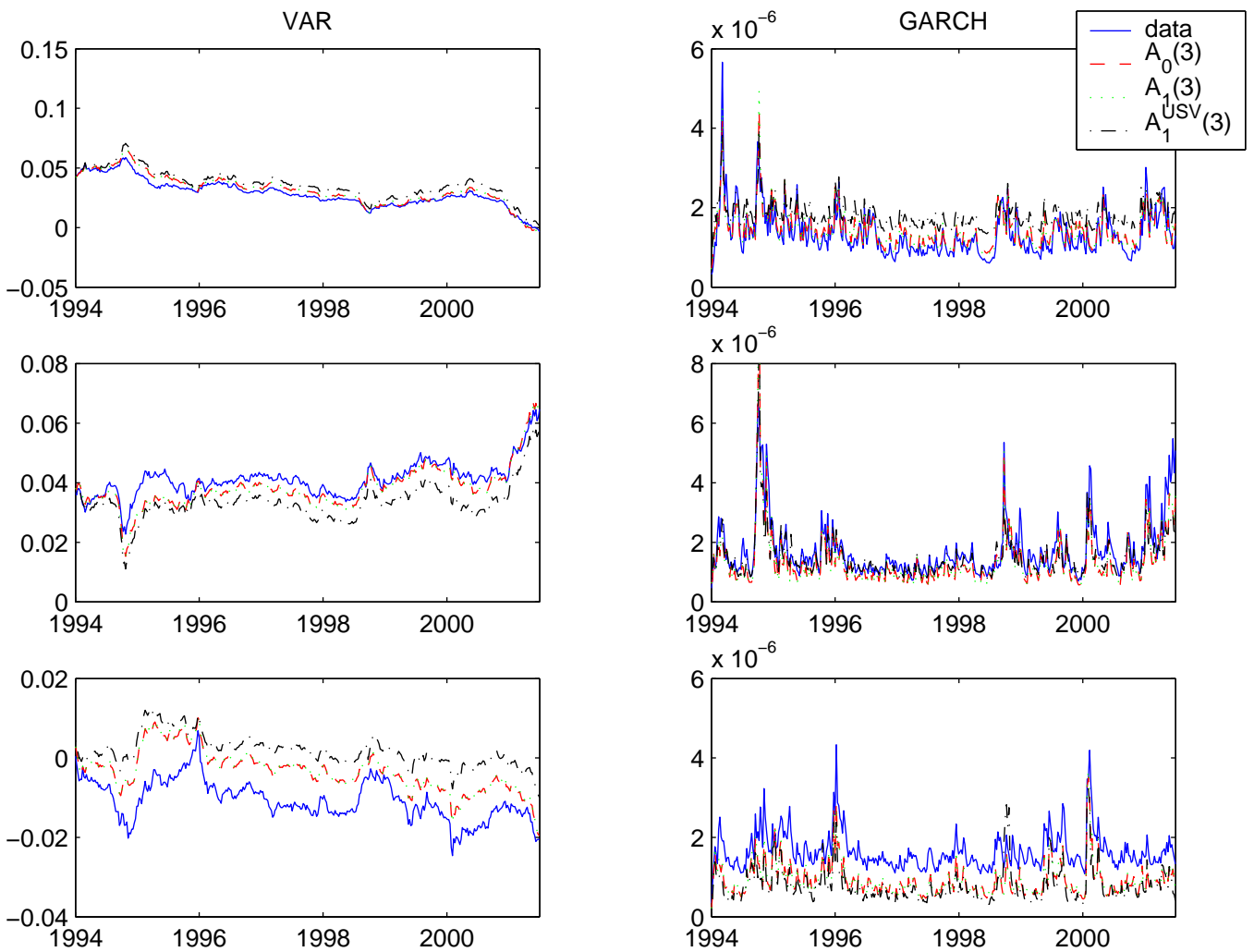
**Table 11 :  $R^2$  of regression of implied volatility on term structure principal components**

We compute the  $R^2$  of the regression of implied volatility on the three principle components. Then we construct finite sample distribution of this statistics by simulating 800 paths from each of the models, and recomputing the  $R^2$  along each path. Boldfaced values indicate a rejection of the null that a particular model can replicate the  $R^2$  observed in the data.

Option Maturity	6 months		1 year	
Sample $R^2$	0.34		0.46	
Model	$R^2$	Conf. int.	$R^2$	Conf. int.
No Errors				
$A_0(3)^f$	0.36	[0.00, 0.47]	<b>0.07</b>	[0.00, 0.35]
$A_0(3)^{fo}$	<b>0.02</b>	[0.01, 0.12]	<b>0.02</b>	[0.01,0.06]
$A_1(3)^f$	<b>0.97</b>	[0.71, 0.99]	<b>0.96</b>	[0.84,0.99]
$A_1(3)^{fo}$	<b>0.88</b>	[0.55, 0.93]	0.91	[0.34, 0.96]
$A_1^{USV}(3)^f$	<b>0.98</b>	[0.61, 0.99]	0.98	[0.27, 0.98]
$A_1^{USV}(3)^{fo}$	<b>0.90</b>	[0.78, 0.97]	<b>0.97</b>	[0.78, 0.97]
Measurement Errors				
$A_0(3)^f$	<b>0.03</b>	[0.00, 0.06]	<b>0.01</b>	[0.00 0.02]
$A_0(3)^{fo}$	<b>0.06</b>	[0.01, 0.12]	<b>0.00</b>	[0.00,0.02]
$A_1(3)^f$	0.16	[0.02, 0.42]	0.30	[0.14,0.58]
$A_1(3)^{fo}$	0.47	[0.22, 0.77]	0.45	[0.10, 0.83]
$A_1^{USV}(3)^f$	<b>0.81</b>	[0.38, 0.97]	0.74	[0.13, 0.97]
$A_1^{USV}(3)^{fo}$	0.39	[0.23, 0.70]	0.27	[0.08, 0.65]

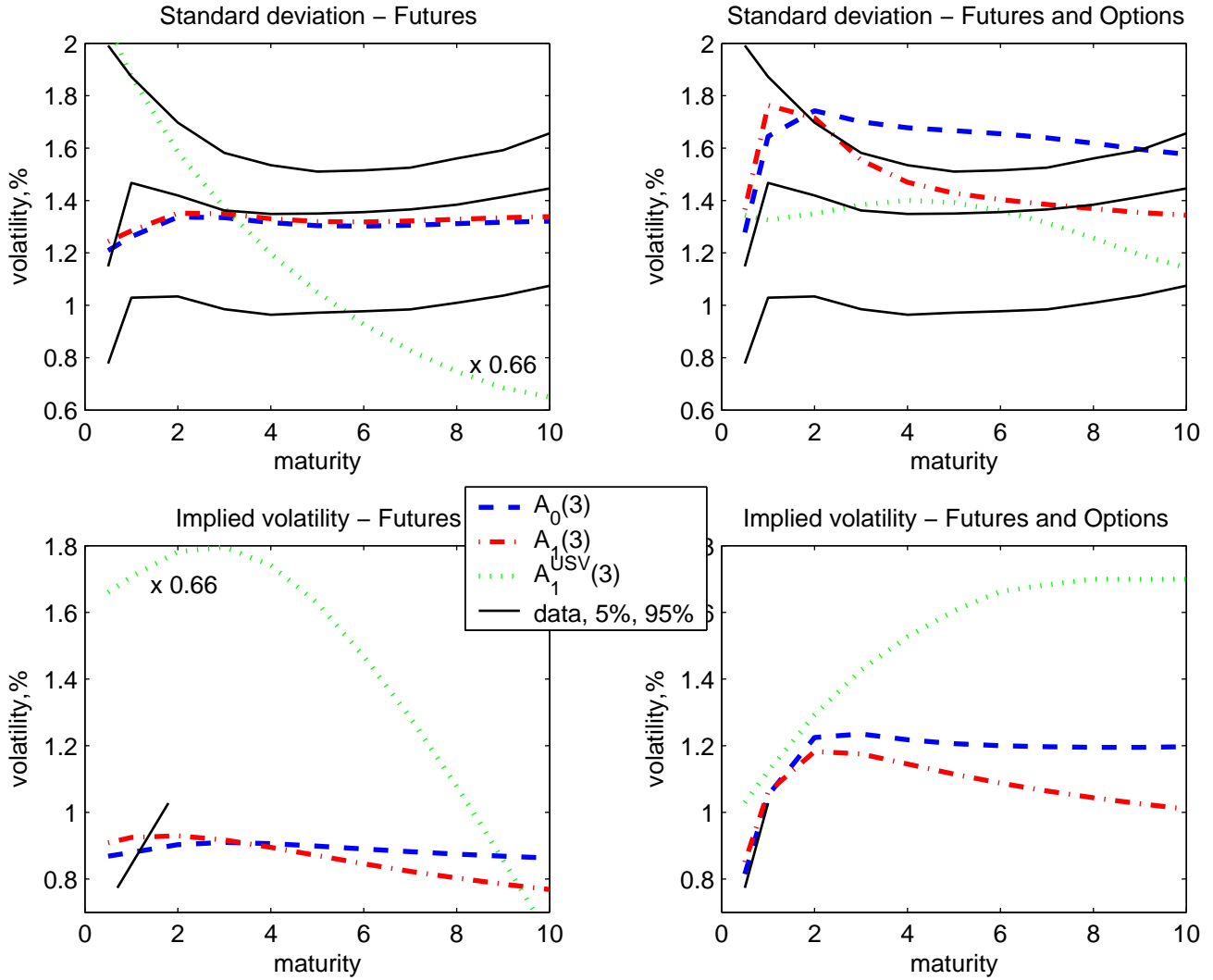
**Figure 1. Conditional means and variances of the principal components:  
futures-based estimation.**

We plot conditional means and variances from the VAR(1)-GARCH(1,1) model estimated on the three principal components constructed from the futures from the data and the models. The first row reports the results for the level,  $y^f(0.5)$ , the second row reports for the slope,  $y^f(10) - y^f(0.5)$ , and finally the last row reports for the butterfly spread  $y^f(0.5) + y^f(10) - 2y^f(2)$ .



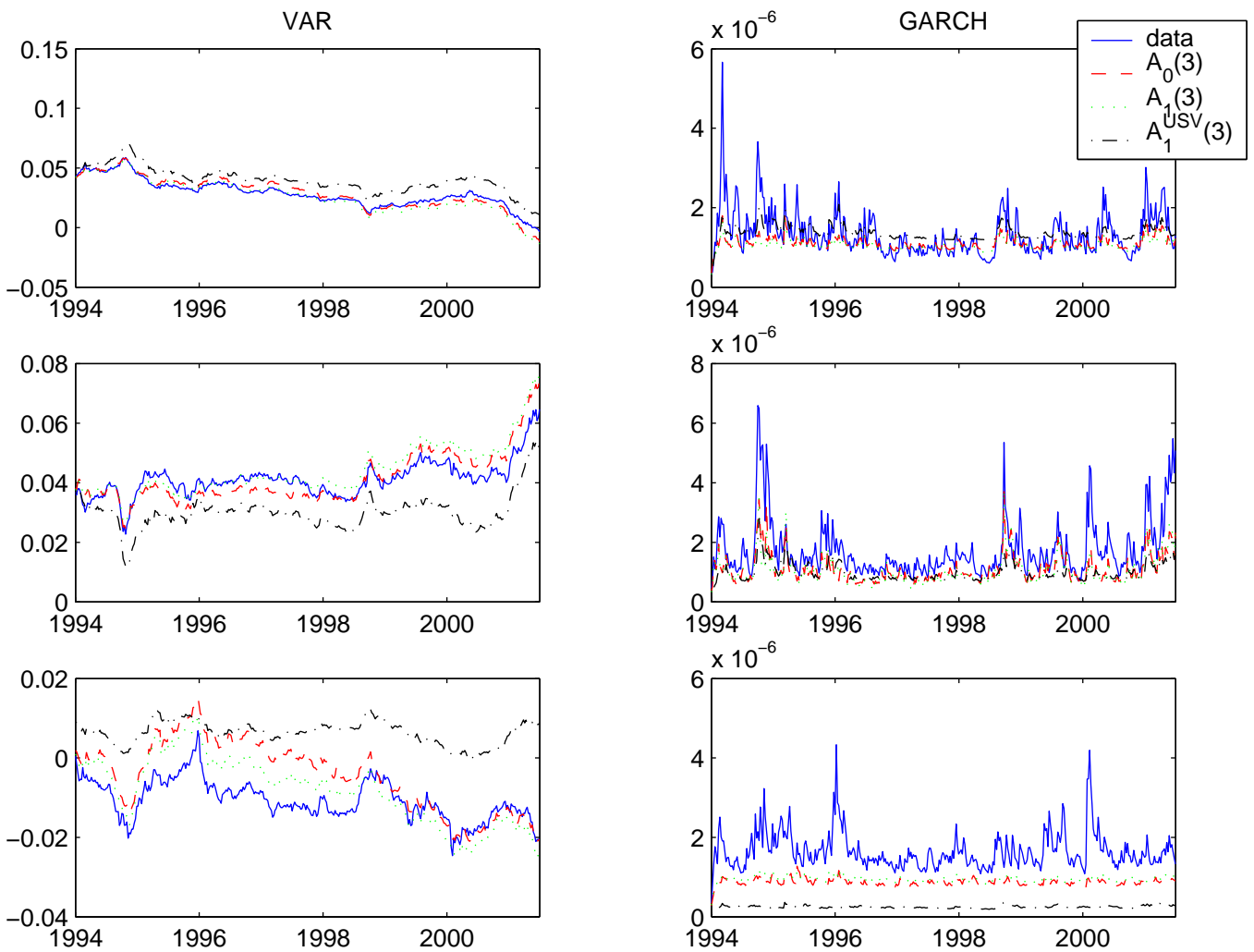
**Figure 2. Term Structure of unconditional and implied volatilities.**

We plot 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 average implied volatility computed from our sample.



**Figure 3. Conditional means and variances of the principal components: futures and options-based estimation.**

We plot conditional means and variances from the VAR(1)-GARCH(1,1) model estimated on the three principal components constructed from the futures from the data and the models. The first row reports the results for the level,  $y^f(0.5)$ , the second row reports for the slope,  $y^f(10) - y^f(0.5)$ , and finally the last row reports for the butterfly spread  $y^f(0.5) + y^f(10) - 2y^f(2)$ .



**Figure 4. At-the-money implied volatility as a function of the state variables.**

We plot how at-the-money implied volatilities is affected by the state variables in various models. The thinner lines correspond to the futures-only estimation.

