Correlation Risk and the Term Structure of Interest Rates

Andrea Buraschi, Anna Cieślak and Fabio Trojani*

ABSTRACT

We develop a structural model of the term structure of interest rates, in which state variables evolve as a matrix-valued process of stochastically correlated factors. This setting grants a new element of flexibility in the simultaneous modeling of stochastic volatilities and correlations of factors, and in the formulation of the market price of risk. We demonstrate that the model provides a unified answer to several empirical regularities of the yield curve such as the predictability of excess bond returns, the persistence of conditional volatilities and correlations of yields, and the humped term structure of forward rate volatilities. At the same time, it remains very parsimonious and analytically tractable with closed-form solutions for zero-coupon bonds, and semi closed-form solutions for the prices of interest rate derivatives.

EFM classification: 310, 340, 440
JEL classification: D51, E43, G13, G12
Keywords: Affine term structure models, Wishart process, stochastic correlations, derivative pricing
First version: November 16, 2006
This version: January 15, 2007, preliminary and incomplete

*Andrea Buraschi is at Tanaka Business School, Imperial College London. Anna Cieślak and Fabio Trojani are at the University of St. Gallen. We thank Paolo Porchia and Paul Söderlind for useful comments. Fabio Trojani gratefully acknowledges the financial support of the Swiss National Science Foundation (NCCR FINRISK and grants 101312-103781/1 and 100012-105745/1). The usual disclaimer applies.

I. Introduction

THIS PAPER DEVELOPS A STRUCTURAL CONTINUOUS-TIME MODEL of the yield curve with stochastically correlated risk factors. Our approach extends the standard affine class of term structure models,[1] as it grants a new element of flexibility in the modeling of the first and second moments of yields. Consequently, we are able to explain, all at a time, the violations of the expectations hypothesis (e.g., the Campbell-Shiller slope coefficients), the persistence in the conditional volatility of yields, the hump-shaped forward rate volatility curve, and the time-varying correlations among different segments of the term structure.

A vast literature has explored the ability of term structure models to account for the time-series and cross-sectional properties of bond yields and interest rate derivatives. The search has focused on analytically tractable models that allow for a general equilibrium interpretation in terms of investor preferences, and ensure economically meaningful (e.g. positive) yields. This set of requirements poses a significant challenge and has turned out difficult to put in unison with the available modeling tools. In affine term structure models (ATSMs), for instance, the tractability in pricing and estimation—their key advantage—comes at the price of restrictive assumptions that guarantee admissibility of the underlying state processes and their econometric identification. Dai and Singleton (2000) emphasize that under the risk-neutral measure admissibility implies a trade off between factors' dependence and their stochastic volatilities. Moreover, Grasselli and Tebaldi (2004) show that continuous factor paths require the positive factors to be mutually independent for the closed-form solution for bond prices to exist.[2] In order to have both correlations and time-varying volatilities, positive square-root processes need to be melded together with the conditionally Gaussian ones. The inclusion of Gaussian dynamics allows for an unconstrained sign of factor correlations, yet at the expense of a loss in the ability to explain stochastic volatilities, and with a possibility to generate negative realizations of the short interest rate.

The empirical literature has highlighted at least five regularities that characterize the dynamics of the US Treasury yields.[3] First, yields are highly autocorrelated, and the extent of their persistence may differ in subsamples. Second, excess returns on bonds are on average close to zero, but vary systematically with the term structure. Accordingly, they can be predicted with several yield curve variables such as the slope, or the spot-forward spread (Fama and Bliss, 1987; Duffee, 2002; Cochrane and Piazzesi, 2005). As a corollary to this fact, one can also forecast the future changes of yields with the current slope of the yield curve (Campbell and Shiller, 1991). Third, in some periods the term structure of

[1] By "standard affine" we denote the state space $D = \mathbb{R}_+^m \times \mathbb{R}^{n-m}$ of a regular affine n-dimensional process (see Duffie, Filipović, and Schachermayer, 2003).

[2] Dai and Singleton (2000) (p. 1945) note, however, that the standard assumption of independent risk factors in the CIR-style models is not necessary either for admissibility of these models or for zero-coupon prices to be known in (essentially) closed-form.

[3] For a thorough review of the empirical features of yields see e.g., Piazzesi (2003); Dai and Singleton (2003); Singleton (2006).

unconditional volatilities peaks for maturities of 2–3 years, and is moderately downward sloping for longer yields (Piazzesi, 2001). Fourth, conditional volatilities of yields are time-varying and positively correlated with the yield level. Finally, although correlations among yields of different maturity are on average very high, they can change broadly over time.

In order to match the physical dynamics of the yield curve, reduced-form models have exploited rich specifications of the market price of risk. The early "completely affine" literature focussed on defining the market price of risk as a constant multiple of factor volatility.[4] Summarizing the empirical failure of this literature, Duffee (2002) proposed an "essentially affine" extension, in which the market prices of risk of unrestricted (conditionally Gaussian) factors are inversely proportional to factor volatility and have a switching sign, but they preserve the completely affine formulation for the square-root (volatility) factors. More recently, Cheridito, Filipović, and Kimmel (2005) suggested a new "extended affine" generalization, in which the market price of risk of *all* factors—both Gaussian and square-root—is inversely proportional to their volatility. Thus, it permits to further relax the link between the market price of risk and the factor volatility.

The levers obtained by augmenting the market price of risk are twofold. First, by incorporating Gaussian factors, the market price of risk can switch sign. Second, unconditional correlations between (some) factors can take both positive and negative values. This gain in flexibility appears crucial for matching the behavior of yields over time. Duffee (2002) and Dai and Singleton (2002) highlight the key role of correlated factors for the model's ability to forecast yield changes and excess bond returns. Notwithstanding a good fit to *some* features of the data, the structural limitations of the essentially[5] affine models—manifest in the trade-off between stochastic volatilities and correlations—allow to match the first *or* the second moments of yields, but not both at the same time (see e.g., Brandt and Chapman (2002); Dai and Singleton (2003)). Moreover, the extensions of the market price of risk are not innocuous from an economic perspective. In a general equilibrium setting, the market price of risk reflects investor risk attitudes. More complex formulations are therefore equivalent to increasingly complex investor preferences. Indeed, models with extended or essentially affine market price of risk may be difficult to justify with general equilibrium arguments.[6]

This paper takes a different approach. While we start with a representative agent economy in the tradition of Cox, Ingersoll, and Ross (1985a) (CIR), we diverge from the literature in making a non-

[4]Examples of completely affine models are Vasicek (1977), or Cox, Ingersoll, and Ross (1985b). The models have been systematically characterized by Dai and Singleton (2000).

[5]In the lack of extensive studies on the fit of extended affine models, we are restricted to discuss only the results available for the essentially affine specifications.

[6]Some equilibrium motivation for the essentially affine form of the market price of risk can be found in the term structure models with habit formation, like in Buraschi and Jiltsov (2006), and Dai (2003). The extended affine family of Cheridito, Filipović, and Kimmel (2005), instead, seems more difficult to reconcile with the standard expected utility maximization, because it entails that agents become more concerned about risk precisely when it goes away.

standard assumption about the production technology. In particular, the fundamental risk factors in our economy evolve as a continuous-time process of positive definite matrices, whose transition probability is Wishart, and by construction can be *stochastically* correlated. With the aim of exploring the yield curve implications of this assumption, we deliberately assign very simple log-preferences to the agent. We demonstrate that our setting grants a remarkable flexibility along several dimensions.

First, the market price of risk involves elements that take both positive and negative values, and hence it translates into excess bond returns that can switch sign. We document that the variation in the model-implied term premia is consistent in size and direction with the historical deviations from the expectations hypothesis.

Second, we are not bound to give up stochastic volatility of factors for their correlations, both of which are truly stochastic in our setting. By construction, the model accommodates conditional and unconditional as well as positive and negative correlations among state variables. The rich dependence of factors is preserved not only under the physical but also under the risk-neutral measure and, contrary to the previous affine literature, does not need to be induced through more complex preferences. We show that the model-generated yields bear the degree of mutual co-movement and volatility persistence which is compatible with the historical evidence.

Third, the model is solved in closed-form for zero-coupon bond prices, and in semi closed-form (Fourier inversion) for the prices of interest rate derivatives. Its tractability is especially noteworthy in light of the admissible factor correlations and their stochastic volatility. Furthermore, while the short rate and yields are linear combinations of the elements of the state matrix, some of which have unrestricted sign, their positivity is ensured with a straightforward restriction. This provides a contrast to the standard affine $A_m(N)$ case with $m < N$, in which the presence of correlated factors entails a probability of negative interest rates.

Fourth, and as a consequence of the properties just discussed, our setup promises good results for the pricing of interest rate derivatives. For one, we show that the model produces a hump-shaped term structure of forward interest rate volatilities, which is crucial for the pricing of caps. Moreover, since stochastic correlations arise by the definition of the state dynamics, there is no need to force them into the model with an additional independent state variable, as is typically done in the literature (see the related literature below). This feature shall have a bearing on the consistent pricing of caps (portfolios of options) relative to swaptions (options on a portfolio).

Finally, the calibration of our model reveals that the regularities found in the data can be successfully reproduced with the most parsimonious three-factor specification, which implies fewer parameters even than the completely affine three-factor CIR model.

Our findings contribute not only to the theoretical yield curve literature, but they also provide a response to the practical concerns of the financial industry. Consider a fairly typical problem of a trader or market maker who wants to hedge a $\$10Ml$ long position in a 10-year zero bond with a position in a 2-year bond, say. If the yield curve were to move in a parallel fashion, i.e. the correlation between different yields to maturity were equal to one, the hedging position would be simply proportional to the ratio of the present values of a basis point (PVBP) of the two bonds, $\frac{PVBP_{10}}{PVBP_2}$. If the correlation is less than perfect, traders often follow the textbook suggestion to adjust the hedge ratio with the predictive regression coefficient between the yields, i.e. $\frac{PVBP_{10}}{PVBP_2}\beta$. The lower the sensitivity of the 10-year yield to changes in the 2-year rate, the lower the hedge ratio. In practice, however, this strategy is far from perfect. For instance, in January 2000 a trader using ten years of data would compute a β equal to 0.83. Accordingly, he would short $\$24.1Ml$ face value of the 2-year bond.[7] Unfortunately, this adjustment factor fluctuates broadly over time: Since $\beta = \frac{\sigma_{10}}{\sigma_2}\rho_{10,2}$, the variation is due both to the variation in the correlation coefficient and in the ratio of their volatilities (see Figure 1). In the last week of January 2000, the conditional value of β is equal to 1.08, as estimated with the exponential moving average. The market maker who assumes a constant value of $\beta = 0.83$ would suffer major losses in a matter of just one week. Clearly, the effectiveness of the hedge depends crucially on the predictive power of the risk weight implied by a yield curve model. The purely Gaussian models are at odds with the observed time variation in the hedge ratio, as they entail constant conditional correlations and volatilities of yields. Similarly, an investor who follows the recommendation of a three-factor CIR model with independent factors would not be able to appropriately adjust the position, because the model-generated variation in the hedge ratio is only about a third of the historical number.[8] This paper addresses the issue of the stochastic volatilities and correlations of yields in a structural general equilibrium setting.

[7] Let $PVBP_{iY}$ denote the change in the dollar value of a i-maturity bond given one basis point decrease in the yield. $PVBP$ is quoted per 100 face value of the bond. Let F_{iY} be the face value of the position in the i-year bond, and β be the predictive slope coefficient of a regression on yield changes. The change in the portfolio value induced by one basis point increase in the 2-year yield can be computed as

$$\Delta V_p = -F_{10}\frac{PVBP_{10} \times \beta}{100} - F_2\frac{PVBP_2}{100},$$

Thus, the optimal hedging position in the 2-year bond is:

$$F_2 = -F_{10}\frac{PVBP_{10} \times \beta}{PVBP_2} = -10Ml \times 0.83 \times \frac{0.050814}{0.017569} = -24.1Ml.$$

[8] This conclusion is born out by the simulation of the three-factor CIR model at parameters estimated by Dai and Singleton (2002) for the Treasury yields and Jagannathan, Kaplin, and Sun (2003) for the swap rates. The conditional correlation of yield changes, and more so, the ratio of their conditional volatilities implied by the three-factor CIR model show a too low variation. As a result, the risk weight $\frac{\sigma_{10}}{\sigma_2}\rho_{10,2}$ has a very low volatility; its standard deviation is only about a third of the historical figure at both sets of parameter estimates.

Related literature

Apart from affine yield curve models, our work is related to the pricing and hedging of interest rate derivatives. Two features of the interest rate derivatives data have spurred extensive research. First, the observed term structure of implied cap volatilities is on average hump-shaped (e.g., De Jong, Driessen, and Pessler, 2004). This requires a model that generates a hump in the volatility curve of forward interest rates (e.g., Brigo and Mercurio, 2006). Second, prices of swaptions and caps are mutually inconsistent (Longstaff, Santa-Clara, and Schwartz, 2001). The cap-implied volatilities are typically higher and more volatile than those extracted from swaptions (Collin-Dufresne and Goldstein, 2001). Simple affine models such as the three-factor CIR fail to explain this feature. In consequence, Jagannathan, Kaplin, and Sun (2003) find the CIR factors backed out from swap and LIBOR rates to be negatively correlated, which contradicts the model itself. Taken together, these findings suggest that in order to successfully price interest rate derivatives, a model needs to accommodate time-varying correlations and volatilities, both driven by factors unrelated to bond prices themselves. In this vein, Collin-Dufresne and Goldstein (2001) and Han (2005) develop models that aim to bypass the trade-off between stochastic volatilities and correlations that pertains to the affine class. Their key insight lies in the notion of "unspanned" factors and boils down to untangling state variables that move yields from the ones that move their stochastic volatilities and correlations. The immediate presence of stochastic correlations in the Wishart factor model sets us apart from the approaches just mentioned. With this property, our framework opens up a new perspective on the discussion of unspanned risk factors, as we can investigate the relative importance of stochastic volatility versus correlation.

Our work draws upon earlier results on the properties of the Wishart process. The Wishart process has been introduced by Bru (1991), and incorporated into the finance domain by Gourieroux, Jasiak, and Sufana (2004) and Gourieroux and Sufana (2003, 2004). Recently, it has found applications to the solution of an intertemporal portfolio choice problem with a correlation hedging demand (Buraschi, Porchia, and Trojani, 2006), and to multivariate option pricing (da Fonseca, Grasselli, and Tebaldi, 2005, 2006). In the yield curve context, Gourieroux and Sufana (2003) have suggested the Wishart process as a convenient way of representing the yield factors, and thus of extending the state space of the standard affine processes, as characterized by Duffie and Kan (1996), to the space of positive definite matrices.[9] Their focus is on a statistical representation of the yield curve in a discrete-time model with an exogenous specification of the pricing kernel. We, by contrast, propose a continuous-time equilibrium framework, which enables us to investigate the economic properties of the term structure. We demonstrate that the model goes a long way toward resolving the empirical "puzzles" of the yield

[9] See the discussion in Gourieroux and Sufana (2003), Section 4.

curve. Moreover, to our knowledge, we are the first to study the interest rate derivatives in the Wishart factor setting.

The plan of the paper is as follows. Section II defines the economy and derives the equilibrium interest rate. Section III provides the solution for the term structure and discusses its asset pricing implications. Section IV investigates the properties of factors based on the model calibrated to the US Treasury yields. In Section V, the model is scrutinized for its consistency with the stylized facts of the yield curve literature. Section VI concludes. All proofs and figures are in Appendices.

II. The economy

We study a continuous time Cox, Ingersoll, and Ross (1985a) production economy with a utility maximizing representative agent. While our representative agent is endowed with simple preferences, we depart from the literature in the specification of the available real investment opportunities. Specifically, the fundamental risk factors in this economy are non-standard, and evolve as a continuous time process of symmetric positive definite matrices.

Assumption 1 (Preferences)**.** *The representative agent maximizes an infinite horizon utility function*

$$E_t\left[\int_t^{\infty} e^{-\rho s}\ln(C_s)ds\right], \tag{1}$$

where $E_t(\cdot)$ is the conditional expectations operator, ρ is the time discounting factor, and C_t is the consumption at time t.

Investor's objective is to select the optimal level of consumption C, and the optimal fractions of wealth to invest in the available assets. The opportunity set comprises: a locally riskless bond in zero net supply with return r, a technology Y producing a single physical good that can be either consumed or reinvested, and a vector of financial assets, with prices F, in zero net supply.

Assumption 2 (Production technology)**.** *The return to the production technology evolves as*

$$\frac{dY_t}{Y_t} = Tr\left(D\Sigma_t\right)dt + Tr\left(\sqrt{\Sigma_t}dB_t\right), \tag{2}$$

where dB_t is a $n \times n$ matrix of independent standard Brownian motions; Σ_t (as well as its square root) is a $n\times n$ symmetric positive definite matrix of state variables; D is a symmetric matrix of deterministic coefficients. Tr indicates the trace operator.

While the dynamics of returns to production Y are specified exogenously, the risk premia on F are an output of equilibrium. The agent chooses the level of consumption C, and the fractions of wealth to be invested in the production technology (real asset), v_Y, and in the financial assets, v_F. The agent's wealth evolves according to

$$\frac{dW_t}{W_t} = v_Y\frac{dY_t}{Y_t} + v_F' I_{F_t}^{-1} dF_t + (1 - v_Y - v_F'\mathbf{1})r_t dt - v_C dt,$$

where $v_C = C/W$ is the fraction of wealth, which is consumed, $\mathbf{1}$ is a vector of ones having the same dimension as F_t, and $I_{F_t} = diag(F_t)$.

Assumption 3 (Risk factors)**.** *The physical dynamics of the risk factors is governed by the Wishart process* Σ_t*, given by the matrix diffusion system*

$$d\Sigma_t = (\Omega\Omega' + M\Sigma_t + \Sigma_t M')\, dt + \sqrt{\Sigma_t} dB_t Q + Q' dB_t' \sqrt{\Sigma_t}, \tag{3}$$

where Ω, M, Q*,* Ω *invertible, are square* $n \times n$ *matrices. Throughout, we assume that* $\Omega\Omega' = kQQ'$ *with degrees of freedom* $k > n$*.*

The specification of the risk factors represents the key idiosyncracy of our—otherwise standard—economy. It also motivates the assumption of log-preferences: As we are primarily interested in exploring the term structure implications of the posited state dynamics, to be able to disentangle them from other effects, we select the simplest possible form of preferences, despite limitations this choice entails (e.g., myopic portfolio choice).

Before we move on to defining the equilibrium, we discuss the properties of the factor dynamics. The Wishart process is a multivariate extension of the well-known square-root (CIR) process. In a special case when k is an integer, it can be interpreted as a sum of several outer products of multidimensional Ornstein-Uhlenbeck (OU) processes. A number of qualities make the process particularly suitable for the modeling of multivariate sources of risk in finance (Gourieroux, Jasiak, and Sufana, 2004; Gourieroux, 2006).

First, the restriction $\Omega\Omega' \gg QQ'$ guarantees that Σ_t is positive definite, and hence can be used to represent a variance-covariance matrix. Thus, the diagonal elements of Σ_t (and $\sqrt{\Sigma_t}$) are always positive, but the out-of-diagonal elements can take negative values. Moreover, if $\Omega\Omega' = kQQ'$ for some $k > 1$, then Σ_t follows the Wishart distribution.

Second, the Wishart process admits time and portfolio aggregation, implying that the integrals as well as the weighted sums of the Wishart process are again Wishart. This property is in general not true for multivariate GARCH processes, which are not invariant under linear aggregation.

Third, the process is affine Markov in the sense that its first and second moments are affine in Σ_t. It follows that the conditional Laplace transform both of the Wishart process and of the integrated

Wishart process is an exponential affine function of Σ_t.[10] This gives rise to convenient closed-form solutions to prices of bonds or equity options, and facilitates the econometric inference based on the model.

Fourth, the coefficients of the dynamics (3) admit intuitive interpretation: The M matrix is responsible for the mean reversion of factors, and the Q matrix—for their conditional dependence. Typically, in order to ensure non-explosive features of the process, M is assumed negative definite.

Finally, the different elements of the Wishart matrix of factors provide for rich interdependencies, and their conditional and unconditional correlations are unrestricted in sign (see Section IV and Result 11 in Appendix A.1 for details). Thus, besides the convenient statistical properties, the dynamics (3) carries a deeper economic sense. Even though we do not attach names to the factors in this economy, we treat them as a dynamic system of dependent sources of risk.

II.A. The equilibrium short interest rate

The investor's objective is to select the optimal level of consumption C^*, and the optimal fractions of wealth invested in the production technology v_Y^*, and in the financial assets v_F^*. The maximization problem is given by the following value function

$$J(W, \Sigma, t) = \max_{v_Y, v_F, v_C} E_t \left[\int_t^{\infty} e^{-\rho s} \ln(C_s) ds \right]. \tag{4}$$

The definition of equilibrium is analogous to Cox, Ingersoll, and Ross (1985a).[11]

Definition 4 (Equilibrium)**.** *The equilibrium is represented as the set of processes (C_t^*, r_t, F_t) for the optimal consumption, the equilibrium interest rate and the prices of the financial assets, such that*

- *The maximization problem in equation (4) is solved;*
- *The optimal consumption is financed by a trading strategy according to which all wealth is invested in the technology Y $(v_Y^* = 1, v_F^* = 0)$.*

The market clearing condition at equilibrium implies that the representative investor selects the optimal consumption $(C^* = \rho W)$, and reinvests all unconsumed wealth in the production technology $(v_Y^* = 1)$.

[10] The conditional Laplace transform of the integrated Wishart process is of exponentially affine form

$$\varphi_{t,\tau}(\Gamma) = E_t[\exp Tr(\Gamma \int_t^{t+\tau} \Sigma_s ds)] = \exp\left[Tr\left[A(\tau,\Gamma)\Sigma_t\right] + B(\tau,\Gamma)\right],$$

where coefficients $A(\tau,\Gamma)$ and $B(\tau,\Gamma)$ solve an appropriate (matrix) ordinary differential equation (see e.g., Gourieroux (2006), p. 19).

[11] See the definition in Cox, Ingersoll, and Ross (1985a), p. 371.

Financial assets are in zero net supply ($v_F^* = 0$). From the first order conditions of the log-investor it is straightforward to obtain the short interest rate and the market price of risk supported in equilibrium.

By applying the Feynman-Kač argument to (4), J solves the Bellman equation of the form

$$\max_{v_Y, v_F, v_C} \left[\mathcal{L}_{W,\Sigma} J + U\right] + \frac{\partial J}{\partial t} - \rho J = 0, \tag{5}$$

where $\mathcal{L}_{W,\Sigma}$ is the infinitesimal generator of the couple (W, Σ); $U(C_t) = e^{-\rho t} \ln C_t$ represents the felicity function, and $\partial J / \partial t$ is the first derivative of J with respect to time. Since risk factors are represented by a matrix-valued process, the infinitesimal generator $\mathcal{L}_{W,\Sigma}$ is non-standard, and has the following form[12]

$$\mathcal{L}_{W,\Sigma} = \mu_W \frac{\partial}{\partial W} + \frac{1}{2}\sigma_W^2 \frac{\partial^2}{\partial W^2} + Tr\left[\left(\Omega\Omega' + M\Sigma + \Sigma M'\right)\mathcal{R} + 2\Sigma\mathcal{R}(Q'Q\mathcal{R})\right] + 2Tr\left[\Sigma Q \mathcal{R}\right]\frac{\partial}{\partial W}, \tag{6}$$

where by $\mathcal{R}$ we denote the matrix differential operator defined as[13]

$$\mathcal{R} := \left(\frac{\partial}{\partial \Sigma_{ij}}\right)_{n \times n} \quad \text{for } 1 \leq i, j \leq n. \tag{7}$$

The generator $\mathcal{L}_{W,\Sigma}$ comprises three components: (i) the standard infinitesimal generator for the wealth process $\mathcal{L}_W$ (first two terms), (ii) the generator for the Wishart process $\mathcal{L}_\Sigma$ (third term), and (iii) the cross generator involving the covariation terms between wealth and the Wishart process (fourth term). While the standard interpretation of the generator is preserved,[14] its novelty comes from the notation in terms of matrices.

The drift of the wealth process is given by

$$\mu_W = \left[v_Y\left[Tr(D\Sigma) - r\right] + v_F'(\mu_F - r\mathbf{1}) + (r - v_C)\right] W, \tag{8}$$

where μ_F is the drift vector with i-th element defined as $\mu_{F,i} := \frac{1}{dt} E\left(\frac{dF_i}{F_i}\right)$. We note that the quadratic variation of the technology in equation (2) is a linear combination of the diagonal elements of the state matrix, and can be written as $Tr(\Sigma_t)$.[15] The quadratic variation of the wealth process follows as

[12] See Section 2 in Bru (1991), and Propositions 1 and 2 in da Fonseca, Grasselli, and Tebaldi (2005).

[13] Note that in our case $\mathcal{R}$ is symmetric, since Σ_t is symmetric.

[14] Indeed, the $n \times n$ dimensional process Σ_t can be viewed as a vector of $\frac{n(n+1)}{2}$ different factors.

[15] Note that $Tr(\sqrt{\Sigma_t} dB_t) = \text{vec}(\sqrt{\Sigma_t})' \text{vec}(dB_t)$, where vec denotes stacking elements of a matrix in a vector. The quadratic variation is

$$\begin{aligned} \langle Tr(\sqrt{\Sigma} dB)\rangle_t &= \text{vec}(\sqrt{\Sigma})'\text{vec}(dB)\text{vec}(\sqrt{\Sigma})'\text{vec}(dB) = \text{vec}(\sqrt{\Sigma})'\text{vec}(dB)\text{vec}(dB)'\text{vec}(\sqrt{\Sigma}) \\ &= \text{vec}(\sqrt{\Sigma})'\text{vec}(\sqrt{\Sigma})dt = Tr(\Sigma)dt. \end{aligned}$$

$$\sigma_W^2 = \left[v_Y^2 Tr(\Sigma) + 2v_Y v_F' \sigma_{YF} + v_F' \sigma_{FF} v_F\right] W^2, \tag{9}$$

where σ_{YF} is the vector of covariances with i-th element $\sigma_{YF_i} := \frac{1}{dt} Cov\left(\frac{dY}{Y}, \frac{dF_i}{F_i}\right)$, and σ_{FF} is a covariance matrix with ij-th element $\sigma_{FF,ij} := \frac{1}{dt} Cov\left(\frac{dF_i}{F_i}, \frac{dF_j}{F_j}\right)$.

We now specialize the general expressions in (5) and (6) to the log-utility case. By the standard result of Merton (1971), for the log-utility investor the value function in equation (4) is separable and can be written as

$$J(W, \Sigma, t) = \frac{1}{\rho} e^{-\rho t} \ln(W) + G(\Sigma, t). \tag{10}$$

The log-utility assumption considerably simplifies the solution, as the investor's maximization problem reduces to

$$\psi = \max_{v_Y, v_F, v_C} \left[\mu_W J_W + \frac{1}{2}\sigma_W^2 J_{WW} + U\right],$$

where expressions for μ_W and σ_W^2 are provided in equations (8) and (9). At equilibrium ($v_Y^* = 1$, $v_F^* = 0$), the interest rate r follows directly from the first order condition $\psi_Y = 0$. The next proposition states the result.

Proposition 5 (The equilibrium interest rate and the market price of risk)**.** *Under Assumptions 1–3, the equilibrium interest rate is given by*

$$r_t = Tr\left[(D - I_n)\Sigma_t\right]. \tag{11}$$

The market price of risk equals the square root of the matrix of Wishart factors

$$\Lambda_t = \sqrt{\Sigma_t}. \tag{12}$$

Proof: Appendix A.2. ■

The instantaneous variance V_t of the changes in the equilibrium interest rate is obtained by applying Ito's Lemma to equation (11), and computing the quadratic variation of the process dr, i.e. $V_t = \langle dr \rangle_t$. Similarly, the instantaneous covariance CV_t between the changes in the interest rate and their variance can be obtained as the quadratic co-variation of dr and dV, i.e. $CV_t = \langle dr, dV \rangle_t$. The instantaneous variance of the interest rate changes is given by

$$\frac{1}{dt} V_t = 4Tr\left[\Sigma_t (D - I_n) Q' Q (D - I_n)\right]. \tag{13}$$

The covariance between the changes in level and the changes in volatility of the interest rate reads

$$\frac{1}{dt}CV_t = 4Tr\left[\Sigma_t(D - I_n)Q'Q(D - I_n)Q'Q(D - I_n)\right]. \tag{14}$$

Expressions (13) and (14) show that both quantities preserve the affine property in the elements of Σ_t. Appendix A.3 provides the derivation.

II.B. Discussion

The short interest rate. In the conventional (non-trace) notation, the equilibrium short interest rate in (11) can be written as

$$r_t = \sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}\Sigma_{ij,t},$$

where d_{ij} denotes the ij-th element of matrix $D - I_n$. Thus, the short rate is a linear combination of the Wishart factors, which are conditionally and unconditionally dependent, with their correlations being unrestricted in sign. The short rate comprises both positive factors on the diagonal of Σ_t and out-of-diagonal factors that can take both signs.

As soon as we abstract from the real economy, the positivity of the short rate becomes a concern. Notably, in our setup the positivity is imposed with a straightforward restriction requiring the $D - I_n$ matrix to be positive definite (see Result 12 in Appendix A.1). The simplicity of this condition is striking in the context of the traditional affine models, in which the coexistence of a positive short rate is at odds with unrestricted and correlated factors. According to the classification of Dai and Singleton (2000), $A_N(N)$ is the only subfamily of ATSMs that guarantees the positivity of the short rate. In $A_N(N)$ models, all three state variables determine the volatility structure of factors, and thus remain conditionally uncorrelated. At the same time, the restrictions imposed on the mean reversion matrix require the unconditional correlations among state variables to be non-negative.

Remark 6 (Link to the Longstaff and Schwartz (1992) model)**.** *By restricting D in equation (2) to be a diagonal 2×2 matrix, our expression for the equilibrium interest rate resembles the two-factor model of Longstaff and Schwartz (1992). Indeed, in this case the drift of the production technology and the interest rate become linear combinations of two positive processes Σ_{11}, Σ_{22}—the diagonal elements of Σ_t. Even in this special case, however, our model has a richer structure than in Longstaff and Schwartz (1992), as the variance of the changes in the interest rate is driven by all three factors $(\Sigma_{11}, \Sigma_{12}, \Sigma_{22})$, which are correlated.*

The market price of risk. In our setting, the market price of risk Λ_t is derived by the same equilibrium argument that underlies the completely affine models (Cox, Ingersoll, and Ross, 1985b). A well-recognized critique of this family is its inability to match the empirical properties of yields due to (i) sign restriction (positivity) on the market price of risk, and (ii) its one-to-one link with the volatility of factors. We emphasize that despite analogous derivation, the market price of risk in our model is very distinct from the standard completely affine specification, as it reflects not only volatilities but also co-volatilities of factors, and thus involves elements that can change sign. Positive factors on the diagonal of matrix Σ are endowed with a positive market price of risk, while the market price of risk of the remaining out-of-diagonal factors is unrestricted. It may occur that our specification shares some analogy to the essentially affine formulation of Duffee (2002). However, as we show in Section V, the theoretical and empirical consequences of our specification are very different from those of the essentially affine class.

Alternative representations of the model. It is sometimes useful to re-express the latent risk factors in Σ_t in terms of variables with an economic interpretation. The linearity in Wishart factors is a convenient feature for deriving alternative representations of the model. For instance, assuming a 2×2 dimension of Σ_t, the expressions (11), (13) and (14) imply a linear mapping between the equilibrium interest rate r, the variance V and the covariance CV, and the distinct elements Σ_{11}, Σ_{12}, and Σ_{22} of the state matrix. Thus, to retrieve the dynamics of the risk factors in Σ we can use the change of variable implied by this mapping. Similar technique is routinely exploited in the standard affine models to move away from latent factors to variables with an economic meaning (Longstaff and Schwartz, 1992; Dai and Singleton, 2000).

III. The term structure of interest rates

Given expression (11) for the equilibrium interest rate, the price at time t of a zero-coupon bond maturing at time T is

$$P(t,T) = E_t^* \left(e^{-\int_t^T r_s ds} \right) = E_t^* \left(e^{-Tr\left[(D-I_n)\int_t^T \Sigma_s ds\right]} \right), \tag{15}$$

where $E_t^*(\cdot)$ denotes the conditional expectation under the risk neutral measure. To move from the physical dynamics of the Wishart state in equation (3) to the risk neutral dynamics, we can apply the standard change of drift technique. Because of the separability of the value function, the risk neutral drift adjustment Φ_Σ is simple to compute as

$$\Phi_\Sigma = -\frac{J_{WW}}{J_W} Cov_t(dW, d\Sigma) = \Sigma Q + Q'\Sigma, \tag{16}$$

where $Cov_t(dW, d\Sigma)$ denotes the instantaneous covariance between changes in wealth W and changes in the state Σ, and is given as a symmetric matrix of dimension $n \times n$ (see Appendix A.4 for details). As in the scalar case, Φ_Σ can be interpreted as the expected excess return on a set of securities constructed so that it perfectly reflects the risk embedded in the elements of the state matrix Σ.

The partial differential equation (PDE) for pricing contingent claims follows from the application of the discounted Feynman-Kač formula to expectation in (15), where the change of drift of Σ_t from the physical to the risk-neutral dynamics is motivated by the equilibrium argument provided in equation (16).

Proposition 7 (The pricing PDE)**.** *Under Assumptions 1–3, the price at time t of a contingent claim F maturing at time $T > t$, whose value is independent of wealth, satisfies the partial differential equation*

$$Tr\left[\left(\Omega\Omega' + M\Sigma + \Sigma M' - \Phi_\Sigma\right)\mathcal{R}F\right] + 2Tr\left[\Sigma\mathcal{R}(Q'Q\mathcal{R}F)\right] + \frac{\partial F}{\partial t} - rF = 0, \tag{17}$$

with the boundary condition

$$F(\Sigma, T, T) = \Psi(\Sigma, T), \tag{18}$$

where r is the equilibrium interest rate, the risk neutral change of drift for the Wishart process is given by $\Phi_\Sigma = \Sigma Q + Q'\Sigma$, and the matrix differential operator is defined in equation (7).

Proof: See Appendix A.4. ■

In expression (15), we recognize the Laplace transform of the integrated Wishart process. By the affine property of the process, the solution to the PDE in equation (17) is an exponentially affine function in the risk factors. The next proposition states this result in terms of prices of zero-coupon bonds.

Proposition 8 (Bond prices)**.** *At time t, the price of a zero-coupon bond P with maturity $T > t$, under the model dynamics (2)–(3), is of the exponentially affine form*

$$P(\Sigma, t, T) = e^{b(t,T) + Tr[A(t,T)\Sigma]}, \tag{19}$$

for state independent scalar $b(t,T)$, and a symmetric matrix $A(t,T)$ solving the system of matrix Riccati equations[16]

$$\begin{aligned} -\frac{db}{dt} &= Tr\left(\Omega\Omega' A\right) && (20) \\ -\frac{dA}{dt} &= A(M - Q') + (M' - Q)A + 2AQ'QA - (D - I_n), && (21) \end{aligned}$$

with terminal conditions $A(T,T) = 0$ and $b(T,T) = 0$. The closed-form solution to this system is provided in Appendix A.5. ■

[16] Equivalently, we can express the coefficients $A(t,T)$ and $B(t,T)$ as functions of the time to maturity $\tau = T - t$. Under the change of variable $t \to \tau$ the left hand side of the above ODE has to be multiplied by factor -1.

Under the assumed factor dynamics, bond prices are given in closed form. In particular, the solution for the coefficient matrix $A(t, T)$ does not entail any numerical integration, which typically arises in ATSMs except for the purely Gaussian and the multiple independent factor CIR models.

Yields. From equation (19), we obtain the yield of a zero-bond maturing in $\tau = T - t$ periods

$$y_t^\tau = -\frac{1}{\tau}\left[b(\tau) + Tr\left(A(\tau)\Sigma_t\right)\right]. \tag{22}$$

Note that to ensure the positivity of yields, the matrix $A(\tau)$ needs to be negative definite (see Appendix, Result 12). Like in the case of the short interest rate, this restriction is satisfied if $D - I_n$ is positive definite.

The modeling of covariances between yields is a nontrivial issue in applications such as the bond portfolio selection. Therefore, it is important to understand their properties arising in the Wishart setting. The instantaneous covariance of the changes in yields with different (but fixed) time to maturity (τ_1, τ_2) becomes[17]

$$\frac{1}{dt}Cov_t\left[dy_t^{\tau_1}, dy_t^{\tau_2}\right] = \frac{4}{\tau_1\tau_2}Tr\left[A(\tau_1)\Sigma_t A(\tau_2)Q'Q\right].$$

Our model implies that yields of different maturities co-vary in a non-deterministic way, evident in the presence of Σ_t in the above equation. Moreover, given the indefiniteness of matrix $A(\tau_1)Q'QA(\tau_2)$, the correlations of yields can stochastically change sign over time. We investigate the empirical consequences of this fact in Section V.C.

III.A. Excess bond returns

The price dynamics of a zero-coupon bond follow from the application of Ito's Lemma to $P(\Sigma_t, \tau)$:

$$\frac{dP(\Sigma_t, \tau)}{P(\Sigma_t, \tau)} = (r_t + e_t^\tau)dt + Tr\left[\left(\sqrt{\Sigma_t}dB_tQ + Q'dB_t'\sqrt{\Sigma_t}\right)A(\tau)\right], \tag{23}$$

where e_t^τ is the term premium (instantaneous expected excess return) to holding a τ-period bond (see Appendix A.6). The functional form of the expected excess return can be inferred from the fundamental pricing PDE (17), and is represented by a linear combination of the Wishart factors

$$e_t^\tau = Tr\left[\left(A(\tau)Q' + QA(\tau)\right)\Sigma_t\right]. \tag{24}$$

[17] Note that in this expression we do not let bond maturity decrease as time passes, which is equivalent to looking at yields with a constant time to maturity, and is consistent with the way the yield data are actually quoted. Alternatively, as time goes on, one could let the time to maturity decrease. This would correspond to observing the evolution of a yield on one particular bond.

The instantaneous variance of bond returns can be written as

$$v_t^\tau = 4Tr\left[A(\tau)\Sigma_t A(\tau)Q'Q\right]. \tag{25}$$

A stylized empirical observation is that excess returns on bonds are on average close to zero, but vary broadly taking both positive and negative values.[18] This means that the ratio $\frac{e_t^\tau}{\sqrt{v_t^\tau}}$ is low for all maturities τ. The essentially and extended affine models of Duffee (2002) and Cheridito, Filipović, and Kimmel (2005) are able to replicate this empirical regularity by assigning to non-volatility factors a market price of risk that can change sign. Can our model fare equally well? The answer is "Yes": We can generate excess returns that have a switching sign if the symmetric matrix $A(\tau)Q' + QA(\tau)$ premultiplying Σ in equation (24) is indefinite, i.e. has at least one positive and one negative eigenvalue. The set of matrices (and model parameters) satisfying this condition is large, giving us the latitude to capture the combination of low excess returns on bonds with their high volatilities. Using calibration results, in Sections V.A and V.B we study the properties of model-implied excess bond returns, and explore whether they coincide with the model's ability to fit the observed shapes of the yield curve.

III.B. Forward interest rate

Let $f(t,T)$ be the instantaneous forward interest rate at time t for a contract beginning at time $T = t+\tau$. The instantaneous forward rate is defined as $f(t,T) := -\frac{\partial \ln P(t,T)}{\partial T}$. Taking the derivative of the log-bond price in equation (19), we have

$$f(t,T) = -\frac{\partial b}{\partial \tau} - Tr\left(\frac{\partial A}{\partial \tau}\Sigma_t\right),$$

where $\partial A/\partial \tau$ denotes the derivative with respect to the elements of matrix A given in equation (21). The dynamics of the forward rate are given by (see Appendix A.7)

$$df(t,T) = -\frac{\partial f(t,T)}{\partial \tau}dt - Tr\left(\frac{\partial A}{\partial \tau}d\Sigma_t\right). \tag{26}$$

De Jong, Driessen, and Pessler (2004), for instance, argue that a humped shape in the volatility term structure of the instantaneous forward rate leads to possibly large humps in the implied volatility curves of caplets and caps that are typically observed in the market. We examine the magnitude and sources of the hump in the model-implied volatility of the calibrated forward interest rate in Section V.D.

[18] See e.g. Figure 5 in Piazzesi (2003).

III.C. Pricing of interest rate derivatives

Our framework allows to derive convenient expressions for the prices of simple interest rate derivatives. The price of a call option with strike K and maturity S written on a zero bond maturing at $T \geq S$ is:

$$\begin{aligned} \text{ZBC}(t, \Sigma_t; S, T, K) &= E_t^* \left[e^{-\int_t^S r_u du} (P(S,T) - K)^+ \right] \\ &= P(t,T) Pr_t^T \{P(S,T) > K\} - K P(t,S) Pr_t^S \{P(S,T) > K\}, \end{aligned}$$

where $Pr_t^T\{X\}$ denotes the conditional probability of the event X (exercise of the option) based on the forward measure related to the T-maturity bond. We can take the logarithm to obtain

$$Pr_t^T\{P(S,T) > K\} = Pr_t^T\{b(S,T) + Tr\,(A(S,T)\Sigma_S) > \ln K\}.$$

To solve for the option price, we need to determine the conditional distribution of the log-bond price under the S- and T-forward measures. In our framework, the characteristic function of log-bond prices—due to the affine property in Σ—is available in closed form. Thus, the pricing of the option amounts to performing two *one-dimensional* Fourier inversions under the two forward measures. The next proposition provides the pricing formula for a call option on a zero bond.

Proposition 9 (Zero-coupon bond call option price)**.** *Under Assumptions 1–3, the time-t price of an option with strike K, expiring at time S, written on a zero-bond with maturity $T \geq S$ can be computed by Fourier inversion according to*

$$\begin{aligned} ZBC(t,S,T) &= P(t,T)\left\{\frac{1}{2} + \frac{1}{\pi}\int_0^\infty Re \frac{e^{-iz[\log K - b(S,T)]}\Psi_t^T(iz)}{iz} dz\right\} \\ &\quad - K P(t,S)\left\{\frac{1}{2} + \frac{1}{\pi}\int_0^\infty Re \frac{e^{-iz[\log K - b(S,T)]}\Psi_t^S(iz)}{iz} dz\right\}, \end{aligned}$$

where $\Psi_t^j(iz), j = S,T$, is the characteristic function of $Tr\,[A(S,T)\Sigma_S]$ under the S- and T-forward measure, respectively. Details and closed-form expressions for the characteristic function are provided in Appendix A.8. ■

With this result at hand, we can price interest rate caps and floors, which effectively are portfolios of put and call options on zero-bonds.[19]

[19] The pricing of caps and floors involves the put-call parity relation for the zero-bond option. At time t, the prices of a put and a call option with maturity S, strike K, written on a zero-bond maturing at $T > S$, satisfy the following put-call parity relation

$$\text{ZBP}(t,S,T,K) = \text{ZBC}(t,S,T,K) - P(t,T) + KP(t,S).$$

IV. The model mechanics

In the current and the following sections, we are guided by the criteria laid down by Dai and Singleton (2003) and study the correspondence between our theory and the historical behavior of the yield curve. The model is scrutinized for its ability to match: (*i*) the predictability of yields, (*ii*) the persistence of conditional volatilities of yields, (*iii*) the correlations between different segments of the yield curve, and (*iv*) the behavior of forward rates.

Taken together, these criteria present a challenge for any yield curve model. Essentially affine models that perform well on one front, by construction fail on another. The top performing models in forecasting—the conditionally Gaussian $A_0(3)$ subfamily—fail completely in generating the time variation in the volatility of yields, while the models that capture the stochastic volatility of yields—the $A_1(3)$ and $A_2(3)$ subfamily—disappoint in terms of prediction (Dai and Singleton, 2002). The success of ATSMs either in forecasting or in fitting the volatility crucially depends on the essentially affine specification of the market price of risk. With the completely affine formulation, instead, the superiority of either subfamily is less clear (e.g., Singleton, 2006). The recent specification of Cheridito, Filipović, and Kimmel (2005) is a potential step towards reconciling the forecasting of yields with their time-varying volatility, as the extended affine market price of risk improves the time-series properties especially in models with multiple restricted state variables, i.e. $A_2(3)$ and $A_3(3)$ subfamilies. However, relative to the essentially affine models, this improvement is achieved by adding several new parameters, and is typically accompanied by a deterioration in the cross-sectional fit to the shapes of the yield curve.

To illustrate the ability of our model to match the criteria (*i*)–(*iv*), we choose the most parsimonious framework, in which the technology is driven by a 2×2 matrix of factors, and the market price of risk is derived from equilibrium with log-investors. Effectively, we place ourselves in a three-factor setting, with two positive factors Σ_{11}, Σ_{22}, and one factor Σ_{12} that can change sign. We assume $\Omega\Omega' = kQQ', k = 3$, thus ensuring the positive definiteness of the process. The realistic values of parameter matrices M, Q and D are obtained by calibrating the model to the unconditional moments of yields.

In this way, we put our model to a rigorous test along two dimensions: First, using merely the unconditional information in yields, we require the model to *simultaneously* match both their conditional and unconditional properties. This allows us to assess the richness of the structure implicit in the Wishart framework. Second, due to the choice of a low dimension for the state matrix, the model is equipped with only 11 parameters to perform tasks (*i*)–(*iv*) mentioned above. This low number is hard to match by any (unrestricted) model in the affine class, in which already the completely affine CIR with three independent factors requires 12 parameters. The tight parametrization of the Wishart setting becomes even more salient in the context of essentially affine and extended affine models. For instance, the essentially affine Gaussian $A_0(3)$ model studied in Duffee (2002) involves the estimation of 21 parameters in the preferred—i.e., restricted—version, and 28 parameters in the unrestricted version, whereas the respective numbers for the essentially affine $A_1(3)$ model are 22 (preferred) and 29 (unrestricted). The extended affine specification introduced by Cheridito, Filipović, and Kimmel (2005) for the restricted $A_1(3)$ model requires 24 parameters. It is worth remarking that the number of parameters in the three-factor essentially and extended affine models is comparable to the 3×3 specification of the Wishart factor model, which implies six factors and 24 parameters (given the restriction $\Omega\Omega' = kQQ'$ with fixed degrees of freedom k). In what follows, we show that 11 parameters in our model are already enough to explain several puzzling features of the term structure data.

IV.A. The Calibration

We use monthly data on zero-coupon US Treasury bonds for the period from January 1952 through June 2005. The sample includes the following maturities: 3 and 6 months, 1, 2, 3, 5, 7 and 10 years. The yields for the period from January 1952 through December 1969 are from McCulloch and Kwon data set; the yields from January 1970 through December 1999 are from the Fama and Bliss CRSP tapes; finally for the last period from January 2000 thorough June 2005 we use the US Treasury constant maturity yields provided by Datastream.[20]

To obtain realistic values of parameters, the model is calibrated to the unconditional moments of yields with maturities 6 months, 2 years, and 10 years. The moments that provide a "static" description of the term structure comprise means, standard deviations and correlations of yields. The "intertemporal" information is introduced by augmenting the set of moment conditions with the Campbell-Shiller regression coefficients for the 2- and 10-year yields.[21] At this point the stationarity of yields in our sample is not a concern: We assume that the respective moments exist, and that the model is stationary. In order to reach the stationary distribution, we simulate a large number of yields, and sample them at monthly frequency. Additionally, we set a warm-up period of 500 months before the actual simulated sample starts. This procedure leaves us with 36000 monthly realizations of yields from our model. The calibration is based on minimizing the squared distances between the empirical and the model-implied moments. The parameter matrices D, M and Q, and the details on the calibration are provided in Appendix B.

Our model implies two restrictions on the parameter matrices: (i) the positive definiteness of matrix $D - I_n$, which ensures a positive short rate and well-defined yields,[22] and (ii) negative definiteness of matrix M, which guarantees the stationarity of factors. Apart from requiring the D matrix to be

[20] The sample is an extension of the one used in Duffee (2002). The properties of yields from Datastream are consistent with the Fama-Bliss data set for the overlapping part of both samples.
[21] For the computation of the Campbell-Shiller coefficients, see the discussion in Section V.B.
[22] The positivity of the short rate is important here, as we work with nominal yields.

symmetric, we do not impose any constraints on the parameters. Nevertheless, both above restrictions are fulfilled in the calibration itself. This result can be viewed as an *ad hoc* check of the meaningfulness of the yield curve dynamics arising in our model.

In the remaining part of the paper, we use the calibrated parameters to study the properties of the model. If not stated otherwise, the simulation evidence is based on 36000 realizations from the model. We emphasize that results provided below are based on the single set of parameters. This fact is important for understanding the ability of the model to explain joint properties of the term structure of interest rates.

By the positive definiteness of $D - I_n$ in equation (11), the short rate is guaranteed to remain positive. In terms of the level and the slowly decaying autocorrelations, the simulated short rate is consistent with the 3-month yield in the US data set. For instance, in the simulated 54 years of its monthly realizations (equal to the length of the US Treasury yield sample), the autocorrelation at the 1-month lag is 0.985, declines to 0.920 at the 5-month lag, and further down to 0.859 at the 10-month lag, while in the data the respective numbers are 0.984, 0.910, and 0.841.

IV.B. Properties of the risk factors

We argue thoroughout that the Wishart process gives much freedom in the modeling of the conditional dependence between positive factors, while allowing for their stochastic volatility. Thus, contrary to the classical affine models, we need not trade off one feature for another. Our calibration substantiates this claim.

Factor correlations. In the simulated sample, the unconditional correlations of factors are

$$Corr(\Sigma_{11}, \Sigma_{12}, \Sigma_{22}) = \begin{pmatrix} 1 & -0.75 & 0.47 \\ \cdot & 1 & -0.86 \\ \cdot & \cdot & 1 \end{pmatrix},$$

where Σ_{ij} is the ij-th element of matrix Σ. To demonstrate how the time variation in correlations arises in our setting, we consider an example of the instantaneous covariance between the positive elements of a 2×2 matrix of factors:

$$Corr_t(\Sigma_{11}, \Sigma_{22}) = \frac{\langle \Sigma_{11}, \Sigma_{22} \rangle_t}{\sqrt{\langle \Sigma_{11} \rangle_t}\sqrt{\langle \Sigma_{22} \rangle_t}}. \tag{27}$$

Instantaneous variances and covariance of the elements Σ_{11} and Σ_{22} are straightforward to compute (see Result 11 in Appendix A.1):

$$
\begin{aligned}
d\langle\Sigma_{11}\rangle_t &= 4\Sigma_{11}(Q_{11}^2 + Q_{21}^2)dt, \\
d\langle\Sigma_{22}\rangle_t &= 4\Sigma_{22}(Q_{12}^2 + Q_{22}^2)dt, \\
d\langle\Sigma_{11}, \Sigma_{22}\rangle_t &= 4\Sigma_{12}(Q_{11}Q_{12} + Q_{21}Q_{22})dt, \qquad (28)
\end{aligned}
$$

where Q_{ij} is the ij-th element of matrix Q.

The conditional second moments are linear in the elements of the factor matrix. The covariance between positive factors is determined by the out-of-diagonal element Σ_{12}, which can take both signs. As a result, the instantaneous correlation of Σ_{11} and Σ_{22} is time-varying, unrestricted in sign, and depends on the elements of Σ in a non-linear way.[23]

In Figure 2, panels *a* through *c*, we display instantaneous correlations among all factors. The calibrated parameters allow positive factors to be conditionally negatively correlated. To exclude the possibility that this pattern is just an instantaneous phenomenon, we use a window of 650 months (equal to the length of the US Treasury sample), and compute the rolling correlation between Σ_{11} and Σ_{22} (Figure 2, panel *d*). The result is preserved. The negative model-implied conditional correlation of positive factors is a peculiarity in the context of ATSMs, in which positive (volatility) factors can be at best unconditionally positively correlated. These are the negative correlations, instead, that seem to matter in fitting the observed yields. For instance, Dai and Singleton (2000) report that in a CIR setting with two independent factors, studied earlier in Duffie and Singleton (1997), the correlation between the state variables backed out from yields is approximately -0.5, instead of zero. Their estimation results for the completely affine $A_1(3)$ subfamily give further support to the importance of negative conditional correlations among (conditionally Gaussian) factors.[24]

Factor volatilities. The rich dependence of factors in our model does not handicap their stochastic volatilities. It is clear from the dynamics of the Wishart process in equation (3) that all three state variables feature stochastic volatility. This is also manifest in highly significant GARCH coefficient computed for the simulated sample of factors.[25] Ahn, Dittman, and Gallant (2002) note that the goodness-of-fit of the standard ATSMs may be weakened precisely in settings, in which state variables have pronounced conditional volatility and are simultaneously strongly negatively correlated. We are able to fully accommodate such situations. The ease at which correlations and stochastic volatilities coexist in our model is one of its most powerful characteristics, as we prove in Section V.

[23] When the elements of the Wishart matrix admit an interpretation as a variance-covariance matrix of multiple assets, Buraschi, Porchia, and Trojani (2006) show that the correlation diffusion process of $\rho = \Sigma_{12}/\sqrt{\Sigma_{11}\Sigma_{22}}$ is non-linear, with the instantaneous drift and the conditional variance being quadratic and cubic in ρ, respectively. The non-linearity of the correlation process arises despite the affine structure of the covariance process itself.

[24] See Dai and Singleton (2000), Table II and III.

[25] For the sake of brevity, we do not report the coefficients here, but just remark that for all state variables the GARCH(1,1) coefficient is above 0.85.

$A_m(N)$-type interpretation of factors. It is informative to look at our setting from the perspective of the $A_m(N)$ taxonomy developed by Dai and Singleton (2000). This, however, requires an additional qualifier. For instance, the 2×2 Wishart factor model is similar to: (*i*) the essentially affine $A_2(3)$ specification in terms of the number of unrestricted versus positive factors; (*ii*) the completely affine $A_3(3)$ specification in terms of the number of stochastic volatility factors; (*iii*) the $A_0(3)$ specification in terms of the unrestricted (positive and negative) correlations among factors; and it does not have a counterpart within the $A_m(3)$ family in terms of conditional (stochastic) correlations among factors. We find the out-of-diagonal element of the Wishart matrix, Σ_{12}, to be negative in more than 80 percent of the simulated sample. This result conforms with the affine literature, which evidences the superior performance of the $A_{m<N}(N)$ class, with some factors having unrestricted signs, over the multifactor CIR models.

IV.C. Factors manifest in yields

In order to verify the plausibility of the assumed state dynamics in our model, we study what yields can tell us about factors.

Shapes of the term structure. The flexibility of our setting is evident in the variety of shapes that the yield curve can accommodate for just one set of calibrated parameters, and the simplest 2×2 formulation of the state matrix. The term structure can be normal, inverted, humped up- or downwards, or flat (see Figure 3). The typical shapes occur in a way that is roughly consistent with the data. For instance, we observe that the simulated term structure can be (*i*) on average upward sloping when the short yield is low (up to around 4 percent), (*ii*) normal, inverted or flat when the short yield is in intermediate range (from 4 to 11 percent), or (*iii*) humped and downward sloping when the short yield is high (above 11 percent). Due to its parsimony, the model seems to have difficulties in generating sufficient concavity in the short segment of the yield curve, and in producing the right behavior of the term structure for extreme levels of yields. For an extremely high short yield (of above 15 percent), the term structure that prevails in the historical sample is downward sloping, possibly with a hump. While the model is able to generate both negative slopes and humps for this range of yields, even more frequently it displays an upward sloping term structure. This result may seem counterintuitive. However, it should be recognized that in sample—unlike in our simulation—extreme levels of interest rates are rare events, and may not be representative of the true tail distribution of yields.

Yield responses to factors. To get more insight into how the shapes of the term structure arise and change with the underlying factors, we examine the coefficients in the yield equation (22). The scaled elements of the $A(\tau)$ matrix, $-A_{11}(\tau)/\tau$, $-2A_{12}(\tau)/\tau$, and $-A_{22}(\tau)/\tau$, represent the loadings on Σ_{11}, Σ_{12} and Σ_{22}, respectively. The constant term $b(\tau)$ contributes to the on average positive slope of the yield curve,

Table I: Correlation of factors with the yield curve variables

a. Correlation in changes

	$\Delta\Sigma_{11}$	$\Delta\Sigma_{22}$	$\Delta\Sigma_{12}$
$\Delta\bar{y}_t$	0.83	0.10	-0.45
Δslope	0.10	-0.71	0.41
Δy_t^1	0.89	0.22	-0.56
Δy_t^5	0.78	0.00	-0.36
Δy_t^{10}	0.78	0.01	-0.37

b. Correlation in levels

	Σ_{11}	Σ_{22}	Σ_{12}
$\bar{y}_t$	0.51	0.04	0.03
slope	0.02	-0.85	0.47
y_t^1	0.49	0.18	-0.04
y_t^5	0.49	-0.06	0.09
y_t^{10}	0.53	-0.10	0.08

Note: The table shows correlations between latent factors and several term structure variables obtained in simulation. In panel *a* the correlations are for first differences, in panel *b*—for levels. $\overline{y}_t$ is computed by taking average of yields across maturities at time t. Slope is the difference between the longest and the shortest yield in our simulation. y_t^1, y_t^5 and y_t^{10} denote yields with one, 5 and 10 years to maturity, respectively.

but its overall impact is small. Figure 4 plots the respective coefficients. Judging by the form of the loadings as a function of maturity, our factors lend themselves to the classical interpretation in terms of the level, slope and curvature. Most clearly, the variation in Σ_{22} induces the changes in the slope of the yield curve, while the variation in Σ_{11} impacts upon the changes in the level of yields. Even though the loading on Σ_{12} is reminiscent of a curvature factor, its highly negative correlation with both other factors begs caution in this interpretation.

We gain more understanding about the roles of factors by looking at how they correlate with the term structure variables such as the average level of yields across maturities, or the slope of the yield curve. Table I displays the relevant correlation coefficients computed for the simulated sample. Our earlier intuition for the level and the slope factor is confirmed: The correlation of $\Delta\Sigma_{11}$ with the level changes is high, while the correlation of $\Delta\Sigma_{11}$ with the slope changes is low; the pattern reverts for $\Delta\Sigma_{22}$ (see panel *a* in Table I). Specifically, we expect the rise of Σ_{11} to shift the entire yield curve upwards, and the rise of Σ_{22} to reduce the slope. Qualitatively similar results arise when we consider correlations in levels of variables, though Σ_{11} seems to impact more changes in level of yields than the yield level itself (see panel *b* in Table I).

While our "level" and "slope" labels attached to Σ_{11} and Σ_{22} seem warranted, the picture for Σ_{12} is more complex. From Figure 2, it follows that most of the time Σ_{12} is negatively correlated with positive factors. Accordingly, $\Delta\Sigma_{12}$ has opposite effect on the level and the slope of the term structure than $\Delta\Sigma_{11}$ and $\Delta\Sigma_{22}$ do. By comparing the results in panels *a* and *b* of Table I, we deduce that Σ_{12}

Table II: Yield variation explained by j-th principal component

j	1	2	3	4	5
US data	97.29	2.51	0.18	0.02	0.00
Model	97.09	2.82	0.09	0.00	0.00

Note: The table displays the percentages of variation in yield levels explained by the j-th principal component. We use monthly US yields for eight maturities: 3, 6 months, 1, 2, 3, 5, 7, 10 years, and the sample period 1952:01–2005:06. The principal components implied by our model are computed from simulated monthly yields with the same maturities as the empirical ones. The length of the simulation is 36000 observations.

influences both the curvature and the level of the term structure. This double role of Σ_{12} is interesting for the interpretation of the number of factors driving the yield curve, and is studied below.

Factors in disguise of principal components. To verify if the historical yields could have been generated by the assumed factor dynamics, we apply the standard principal component analysis to the data and to the yields simulated from the model. It is well-documented that three principal components explain over 99 percent of the total variation in yields (Litterman and Scheinkman, 1991; Piazzesi, 2003). Since we use a 2×2 specification of the state matrix, the simulated yields are spanned by at most three factors. Table II shows that the portions of yield variation explained by the first two principal components in our model largely coincide with the empirical evidence. Moreover, the traditional factor labels are evident in Figure 5. While weights on the first and second principal components virtually overlap with those computed from the data, the discrepancy is somewhat larger for the third one. Indeed, the term structure produced by the model slightly understates the real curvature of the yield curve.

Shifting number of factors. Although the three-factor structure of the US yields seems robust across different data frequencies and types of interest rates, recent research points to a time variation in the number of common factors underlying the bond market. Pérignon and Villa (2006) reject the hypothesis that the covariance matrix of US yields is constant over time, and document that both factor weights and the percentage of variance explained by each factor change concurrently with switches in the monetary policy under subsequent FED chairmen.

The behavior of instantaneous correlations between Wishart factors in Figure 2 leads us to investigate whether a similar phenomenon is present in our simulated economy. We find that the model is in fact able to adopt the changing factor structure. To this end, we sort yields according to the level of instantaneous correlations between state variables, and for each group we obtain the principal components. A facts that stands out from this exercise is that the percentage of yield variation explained by the consecutive principal components changes considerably and in a systematic way across different correlation bins.

In Figure 6 we plot, as a function of instantaneous factor correlation, the portions of yield variation explained by each principal component.[26] The results for sorts based on $Corr_t(\Sigma_{11}, \Sigma_{22})$ show that even when Σ_{11} and Σ_{22} are highly positively correlated (bin 1 in the plots), three clearly distinguishable factors may be still at work in explaining the variance of yields. The key to understanding this lies in the role of Σ_{12}. Recall from Table I that Σ_{11} and Σ_{22} have distinct impacts on the level and the slope of the term structure, while the impact of Σ_{12} is ambiguous. We observe that in periods when $Corr_t(\Sigma_{12}, \Sigma_{22})$ is high and positive, which happens to coincide with strongly negative $Corr_t(\Sigma_{11}, \Sigma_{22})$, Σ_{12} influences both the level and the slope of the term structure right in the same way as Σ_{11} and Σ_{22} do. In such phases the number of state variables is seemingly reduced.[27] This provides a useful intuition for the role of Σ_{12}, which can be summarized as follows: For one, Σ_{12} co-determines the yield curve as a stand-alone factor. For the other, it effectively diminishes the dimension of the state matrix by mimicking the impact of other factors. Σ_{12} is thus the variable that induces the changes in the factor structure of our model. This interpretation of the intuitive role of Σ_{12} is consistent with expression (28) for the instantaneous covariance between the positive factors, being just a scaling of Σ_{12}. This evidence suggest that conditionally our model is a three-factor one, but unconditionally the stochastic correlation among state variables may work as an additional factor.

V. Yield curve puzzles revisited

In what follows we assess the model in terms of the goodness-of-fit criteria advertised in the introduction to Section IV. The analysis indicates that our simple framework is able to provide a consistent answer across several dimensions of the bond market.

V.A. Excess returns on bonds

The key implication of essentially affine models proposed by Duffee (2002) is that excess bond returns have unrestricted sign. This feature is crucial for matching their empirical properties. Owing to the properties of the Wishart process, our model "by default" grants the degree of flexibility that is (at least) comparable with the essentially affine class. To substantiate this claim, Figures 7 and 8 present excess bond returns obtained from the simulation at the calibrated parameters. The plots confirm the consistency of the model with the empirical evidence in several respects. First, we observe the switching

[26] Since the correlation between Σ_{11} and Σ_{22} is high precisely when the correlation between Σ_{12} and Σ_{22} is low, it is not surprising that plots in Figure 6 are mirror reflections of each other. We use instantaneous correlations both between Σ_{11} and Σ_{22}, and between Σ_{22} and Σ_{12} as conditioning information according to which yields are grouped. We do not present the results for instantaneous correlation between Σ_{11} and Σ_{12}, since due to its high persistence results for only some of the bins are available.

[27] Detailed results, not reported here, show that in periods of high $Corr_t(\Sigma_{22}, \Sigma_{12})$, Σ_{12} correlates strongly and positively (correlation of about 0.9) with the average level of the term structure, and negatively with the slope of the term structure (correlation of about -0.5). Thus, it works in the same direction as Σ_{11} and Σ_{22}.

sign of the risk premia, both instantaneous (Figure 7) and those computed from discrete realizations of the bond price process (Figure 8). Second, excess returns are highly volatile, with the ratio $e_t^\tau/\sqrt{v_t^\tau}$ being below 1. Third, the above properties hold true across bonds with different maturities.

Although we only use the unconditional moments of yields to calibrate the model, simulated data match the magnitudes and the distributional properties of the US bond return dynamics. Empirically, the expected excess returns on long bonds are on average higher and more volatile than on short bonds. Consistent with this evidence,[28] the model produces expected excess returns and volatilities that rise as a function of maturity. In the simulated sample, the expected excess return (the volatility of excess returns) increases from 0.025 percent (1 percent) for a 3-month bond to 5.7 percent (25 percent) for a 10-year bond. These features shall play an important role in the model's ability to replicate the failure of the expectations hypothesis, which we discuss next.

V.B. The failure of the expectations hypothesis

The expectation hypothesis states that yields are a constant plus expected values of the current and average future short rates. Thus, bond returns are unpredictable.

Campbell-Shiller regressions. This implication can be tested in a linear projection of the change in yields onto the (weighted) slope of the yield curve, known as the Campbell and Shiller (1991) regression:

$$y_{t+m}^{n-m} - y_t^n = \beta_0 + \beta_1 \frac{n}{n-m}\left(y_t^n - y_t^m\right) + \varepsilon_t, \tag{29}$$

where y_t^n is the yield at time t of a bond maturing in n periods, and n, m are given in months. While the expectations hypothesis implies the β_1 coefficient of unity for all maturities n,[29] a number of empirical studies point to its rejection. Moreover, there is a clear pattern to the way the expectations hypothesis is violated: In the data, β_1 is found to be negative and increasing in absolute value with maturity. This means that an increase in the slope of the term structure is associated with a decrease in the long term yields; or paraphrased in terms of returns, the expected excess returns on bonds are high when the slope of the yield curve is steeper than usual.

The findings of Section V.A attest the ability of our model to produce the time variation in expected returns. To study if expected returns vary in the "right" way with the term structure, we estimate the model-implied population[30] coefficients of Campbell-Shiller regressions, and compare them with their empirical counterparts (see panel a in Table III and in Figure 9). For the sake of comparison,

[28] See e.g. Table I in Duffee (2002).
[29] See Campbell and Shiller (1991) or Dai and Singleton (2002) for the derivation.
[30] In computing the population coefficients, we follow Dai and Singleton (2002) who claim that matching the population coefficients to the historical estimates is a much more demanding task than matching the fitted yields from an ATSM.

Table III: Regressions of the yield changes onto the slope of the term structure

a. Wishart factor model

Maturity	12	24	36	60	84	120
Data β_1 (1952–2005)	-0.174	<u>-0.615</u>	-0.852	-1.250	-1.660	<u>-2.244</u>
t-stat	-0.4	-1.1	-1.4	-1.8	-2.1	-2.4
Model β_1	-0.349	<u>-0.618</u>	-0.895	-1.454	-1.940	<u>-2.254</u>
t-stat	-6.4	-9.2	-10.8	-12.2	-12.5	-11.5

b. ATSMs

Maturity	12	24	36	60	84	120
Data β_1 (1952–1994)	-0.392	-0.696	-0.890	-1.291	-1.738	-2.451
t-stat	-0.8	-1.2	-1.4	-1.7	-2.0	-2.3
$A_0(3)$ (essentially)	-0.037	-0.401	-0.597	-0.986	-1.462	-2.248
t-stat	-0.8	-7.0	-9.0	-12.3	-15.4	-18.8
$A_1(3)$ (essentially)	0.522	0.445	0.545	0.653	0.620	0.472
t-stat	7.1	4.6	4.8	4.6	3.5	2.0
$A_2(3)$ (completely)	1.354	1.416	1.454	1.369	1.226	1.007
t-stat	18.6	14.2	12.6	10.3	8.3	5.6

Note: The table displays the parameters of the Campbell-Shiller regression in equation (29). The maturities n are quoted in months. The value of m is taken to be six months, for all n. Panel a, the first row presents historical coefficients based on US yields in the period 1952:01–2005:06. The third row shows the model-implied population coefficients. The underline indicates the coefficients used as moment conditions in our calibration. Panel b shows analogous results for the preferred affine specifications of Duffee (2002) at his parameter estimates. The historical coefficients for the period 1952:01–1994:12 concur with the sample used in estimation. All model-implied t-statistics are computed using Newey-West adjustment of the covariance matrix and represent the population values. Due to unobservability of yields with a half-year spacing of maturity, we follow Campbell and Shiller (1991) (their Table I, p. 502), and approximate y_{t+m}^{n-m} by y_{t+m}^{n}. This approximation is used consistently for the simulated and historical data.

we perform an analogous exercise for three preferred affine specifications of Duffee (2002), i.e. ones in which the insignificant parameters have been pruned (see panel b in Table III and in Figure 9). This gives rise to two essentially affine models $A_0(3)$ and $A_1(3)$ and one completely affine model $A_2(3)$.

The results indicate that our model can accommodate the predictability of the yield changes by the term structure slope. The failure of the expectations hypothesis in our setting is evident from the sheer time variation of expected returns. Recall, however, that in the calibration we virtually induce the model to violate the expectations hypothesis by including two Campbell-Shiller coefficients as moment conditions. Otherwise, the information content of the remaining unconditional moments of yields is insufficient to generate the "right" direction of this violation.[31] While the two coefficients used in the calibration are by construction matched almost perfectly (see the underline in Table III, panel a), the model turns out to do a good job also in fitting other parameters, not used in the calibration. Panel a of Figure 9 visualizes this match by showing that all model-implied coefficients lie within the 80 percent confidence bounds computed from the historical sample. At the same time, the results for

[31] In fact, it is already enough to include one Campbell-Shiller coefficient to obtain the predictability of yields consistent with the empirical evidence.

the ATSMs concur with the previous literature. The essentially affine Gaussian model $A_0(3)$ conforms with the empirical evidence, whereas both $A_1(3)$—notwithstanding its essentially affine market price of risk—and $A_2(3)$ models have counterfactual predictability implications.

To corroborate the predictability of yields present in our framework, we consider two additional regressions, which are detached from the calibration procedure, but reflect the same reasons for the failure of the expectations hypothesis as the Campbell-Shiller regressions.[32]

Prediction from the 5Y–3M spread. Following Duffee (2002), we first report the coefficients obtained from projecting the monthly excess bond returns onto the lagged slope of the term structure:

$$R_{t+1}^{\tau} - R_{t+1}^{3M} = \gamma_0 + \gamma_1 \left(y_t^{5Y} - y_t^{3M} \right) + \varepsilon_t, \tag{30}$$

where $R_{t+1}^{\tau} = \ln \left(P_{t+1}^{\tau} / P_t^{\tau} \right)$ is a monthly return on a bond with constant maturity τ. The excess return on the τ-maturity bond is obtained by subtracting the return on a 3-month bond; the slope is defined as the difference between the 5-year and 3-month yield. Even though this regression merely restates the information conveyed by Campbell-Shiller coefficients, it provides a robustness check to the previous results, as neither of the yields which construct the slope is directly used in calibrating the model. Moreover, the frequency of returns is now monthly, and thus it is free from the overlapping samples problem. Panel a in Table IV summarizes the regression output based on the simulated and historical yields. The predictability is confirmed: A steep slope of the term structure in our model forecasts high excess returns to bonds during the next month. Due to the noisiness of the historical estimates, we refrain from assessing the proximity between the model and the data.[33] We note, however, that the implications of our setting are much in line with the evidence collected in Duffee (2002). In both the model and the data the forecasting power of the term structure slope is most pronounced at the long end of the yield curve.

Predictability from the forward–spot spread. As a second check, in panel b of Table IV we report the model-implied coefficients of regressions proposed by Fama and Bliss (1987), who project the excess one-year *holding period* bond return onto the spot-forward spread:

$$hpr_{t+1}^{\tau \to \tau-1} - y_t^{1Y} = \delta_0 + \delta_1 \left(f_t^{\tau-1 \to \tau} - y_t^{1Y} \right) + \varepsilon_t, \tag{31}$$

where $hpr_{t+1}^{\tau \to \tau-1}$ is the return from holding the τ-maturity bond for one year, from t to $t+1$, $hpr_{t+1}^{\tau \to \tau-1} = \ln(P_{t+1}^{\tau-1}/P_t^{\tau})$. $f_t^{\tau-1 \to \tau}$ is the one-year forward rate at time t, $f_t^{\tau-1 \to \tau} = \ln(P_t^{\tau-1}/P_t^{\tau})$. y_t^{1Y} is the one-year yield. The basic intuition from Campbell-Shiller regressions carries over also to this equation.

[32]See Dai and Singleton (2002) for a discussion of relations between the different predictability regressions.
[33]For all maturities the coefficients from the model lie within the 80 percent confidence interval.

Table IV: Predictability regressions

a. Duffee (2002) regression

Maturity	1	2	3	5	7	10
Data γ_1 (1952–2005)	-0.007	0.021	0.060	0.153	0.248	0.386
t-stat	-0.30	0.45	0.93	1.65	2.15	2.64
Model γ_1	0.015	0.038	0.064	0.118	0.167	0.222
t-stat	9.40	10.47	10.85	10.85	10.61	10.27

b. Fama and Bliss (1987) regression

$hpr_t^{\tau\to\tau-1}$	2→1	3→2	4→3	5→4
CP δ_1 (2005)*	0.99	1.35	1.61	1.27
t-stat	3.00	3.29	3.35	1.98
R^2	0.16	0.17	0.18	0.09
Model δ_1	1.30	1.53	1.78	2.04
t-stat	42.28	37.72	33.50	30.07
R^2	0.37	0.31	0.26	0.21

Note: The table displays two predictability regressions. Panel *a* presents regression of the monthly excess return of a constant maturity bond onto the term structure slope, studied by Duffee (2002); see equation (30). We report the historical estimates and the model-implied population coefficients. Panel *b* shows regressions of holding period returns onto the forward-spot spread studied by Fama and Bliss (1987); see equation (31). The column header $2 \to 1$ indicates the return from holding a today's 2-year bond over the next year. t-statistics use the Newey-West adjustment of the covariance matrix; model-implied t-test is the population value.
*) The empirical coefficients, t-statistics and R^2's in the first three rows of panel *b* are from Cochrane and Piazzesi (2005), who compute them for the 1964:01–2003:12 sample of US monthly yields (see their Table 2, p. 145).

However, by looking at the holding period returns we abstract from the concept of constant maturity bond returns used earlier.

High R^2 values of these regressions is what became a stylized fact in the yield curve literature. Since our historical sample does not allow to compute all one-year forward rates, we benchmark the population estimates from the model against those of Cochrane and Piazzesi (2005), who update the results of Fama and Bliss (1987) to include more recent data. Table IV, panel *b*, and the scatter plots in Figure 10 confirm the correct qualitative behavior of the model. Indeed, the slope of the forward yield curve is a predictor of the holding period returns to bonds. While R^2 values are high, they tend to decline with maturity. It should be noted that high R^2's are unique only to Fama-Bliss regressions, both in the historical sample and in the model. In Campbell-Shiller regressions the maximal R^2 is 2.6 percent for the model and 3.8 percent for the data. The respective numbers for Duffee's regressions are even lower.

The above discussion provides further insight into the relevance of the factor structure imposed by the Wishart process. Within the essentially affine $A_m(3)$ family, these are the Gaussian models that dominate other subfamilies in terms of prediction, to the extent that they allow for correlated factors as well as the changing sign of the market price of risk (Dai and Singleton, 2002). This fact appears to convey the intuition behind the success of our model, which by construction is equipped with both correlated factors and a flexible market price of risk. Moreover, and in contrast to affine Gaussian models, to achieve an acceptable forecasting performance in our setting, it is not necessary to sacrifice the stochastic volatilities. This is the aspect we address next.

V.C. Second moments of yields

Two issues that occupy the term structure research agenda are (*i*) the time variation and persistence of the conditional second moments of yields, (*ii*) the humped term structure of unconditional yield volatilities. A less studied feature concerns the dynamic structure of correlations between yields of different maturities.

Conditional volatility of yields. In our model, stochastic volatilities of factors are an immediate consequence of the definition of the Wishart process. We now explore how they translate into the conditional second moments of yields. Is the degree of the time variation and of the persistence in the yield volatility commensurate with the historical evidence? To answer this question, we follow Dai and Singleton (2003), and estimate a GARCH(1,1) model for the 5-year yield (see Table V). The choice of the 5-year yield is motivated by the fact that this maturity is not involved in our calibration. Therefore, its conditional and unconditional properties can be traced back to the intrinsic structure of the model. In panel *a* of Table V, we report the coefficients for our model and compare them with the historical estimates. To be able to infer the relative significance of the two sets of parameters, we compute the median GARCH estimate based on 1000 simulated samples with 54 years of monthly observations each. We take the same approach to assess the volatility implications of the preferred $A_1(3)$ and $A_2(3)$ models of Duffee (2002) (see Table V, panel *b*). Due to its constant volatility assumption, the $A_0(3)$ model is not considered here. Figure 11 contrasts the finite sample GARCH coefficients implied by the ATSMs and by our model.

The results confirm that the degree of the volatility persistence implied by the Wishart factor model is well aligned with the historical figures. For instance, the median model-implied GARCH coefficient is 0.829 against 0.820 found empirically. Furthermore, our model is able to reproduce the dynamic link between yield levels and their conditional volatilities: We find the correlation between the level of the 5-year yield and its GARCH(1,1) conditional volatility to be 0.72 in the simulated sample. In combination with the discussion of Section V.B, these results lend support to our claim that the model can solve at the same time the predictability and the volatility puzzle.

The volatility persistence in the affine models, instead, shows a larger discrepancy with the data, and is typically too low. Similar to Dai and Singleton (2003), we document that the essentially affine $A_1(3)$ specification exhibits conditional volatility that is roughly in line with the historical evidence. Still, as

Table V: GARCH(1,1) parameters for the simulated and historical 5-year yield

a. Wishart factor model

	α	β	$\bar{\sigma}$
Data (1952–2005)	0.180	0.820	0.000
t-stat	7.6	39.0	3.3
Model (popul.)*	0.098	0.894	0.000
Model (648 obs.)†	0.141	0.829	0.000
t-stat	4.7	26.1	2.7

b. ATSMs

	α	β	$\bar{\sigma}$
Data (1952–1994)	0.243	0.757	0.003
t-stat	6.8	23.3	3.7
$A_1(3)$ (popul.)*	0.257	0.670	0.000
$A_1(3)$ (516 obs.)†	0.153	0.707	0.000
t-stat	3.0	8.5	2.7
$A_2(3)$ (popul.)*	0.409	0.590	0.000
$A_2(3)$ (516 obs.)†	0.370	0.547	0.000
t-stat	5.2	9.7	4.4

Note: The table displays the estimates of a GARCH(1,1) model: $\sigma_t^2 = \bar{\sigma} + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$, where ε_t is the innovation from the AR(1) representation of the level of the 5-year yield. Panel *a* shows the ML estimates for the Wishart factor model, and compares them to the historical coefficients based on the sample period 1952:01–2005:06. Panel *b* displays estimates for the preferred affine models of Duffee (2002): the essentially affine $A_1(3)$ and the completely affine $A_2(3)$ and compares them to the historical coefficients. Accordingly, the simulation of the ATSMs uses the estimates from Duffee (2002) for the sample period 1952:01–1994:12.
*) The simulated sample comprises 36000 observations.
†) The coefficients and t-statistics are the median of 1000 estimates based on the simulated sample of 648 and 516 months, respectively. The simulated path reflects the length of the sample used to calibrate/estimate the different models.

shown is Section V.B, it also largely fails at explaining the conditional first moments of yields. Although the $A_2(3)$ specification allows for two CIR-type factors, the volatility persistence it implies is even lower than in the $A_1(3)$ case. Recall, however, that the preferred $A_2(3)$ model of Duffee (2002) is equivalent to the completely affine formulation, because the essentially affine market price of risk parameters turn out insignificant in estimation. This outcome reinforces the claim of Dai and Singleton (2003) that the essentially affine market price of risk is the key to modeling the persistence in the conditional second moments of yields. Finally, Figure 11 conveys useful intuition about the proximity between the different models and the true process driving the volatility of yields. Out of the three models considered, the median GARCH coefficient in the Wishart factor setting is not only the closest to the historical number, but it is also the least dispersed one.

Humped term structure of unconditional volatilities. The term structure of unconditional volatilities of yields (and yield changes) is another recurring aspect in the yield curve debate. Its shape varies across different subsamples, and it has aroused increased interest with the appearance of a hump at around 2-year maturity during the Greenspan era (1987–2006). In this period, the volatility curve is high for very short maturities (up to 3 months), declines quickly afterwards only to hump at maturities of 2–3 years, and finally it stabilizes for longer yields. Piazzesi (2001) calls this behavior a "volatility snake." In the model calibration, we notice an interplay between matching the unconditional second moments, and predicting yields by the term structure slope. In particular, the inclusion of the Campbell-Shiller coefficients in the set of moment conditions tends to worsen the model's fit to the term structure of unconditional volatilities. The largest calibration error of 14 percent occurs for the unconditional volatility of the 5-year yield. In spite of this tension, our model leads to the changing shape of the volatility curve, and in some simulated subsamples is able to produce the snake-like pattern in the volatility curve of yields and yield changes. Figure 12 presents two possible term structures of the volatilities of yield changes.

Dai and Singleton (2000) examine the volatility curves that arise in different affine models. They conclude that the key to modeling the hump lies either in correlations between the state variables or in the respective factor loadings in the yield equation. In contrast to $A_m(3), m = 1, 2$, subfamilies, models with independent factors such as multifactor CIR cannot induce a hump. To discern the mechanism that leads to the non-monotonicity in the volatility curve in our model, we decompose the unconditional variance of yields into contributions of factor variances and covariances scaled by the respective loadings (not reported). The decomposition reveals that the hump in the volatility curve is driven by the (weighted) unconditional variances of Σ_{11} and Σ_{12} and their covariance. This result fits nicely within the interpretation of Dai and Singleton (2000), as Σ_{12} also determines the correlation between the positive factors.

Correlation structure of yields. Finally, we provide a brief account of the dependence structure among yields in our model. Although correlations between yields of different maturities are on average very high, they may also change quite noticeably over time, and even become negative. Modeling directly the stochastic correlations allows us to match high correlations of yields, as well as their considerable time-variability. The maximum calibration error for unconditional correlations between the yield levels across all maturities is 0.73 percent.[34] Furthermore, the model-implied correlations between different segments of the yield curve have different characteristics. For instance, the unconditional correlation between the 3Y–1Y yield spread and 5Y–3Y yield spread is 0.90 in the simulated sample, and 0.85 in the data, while correlation between the 6M–3M spread and 10Y–7Y spread is −0.29 in simulation, and −0.16 in the data. In Figure 13, we are able to mimic the empirical time variation of conditional correlations between the yield levels. Even though the variation of the model-implied correlations of yield changes (not reported in any figure) is well pronounced, its extent is somewhat smaller than in the data.

[34]The calibration error is understood as the difference between the model-implied and the historical moment per unit of the historical moment.

V.D. Second moments of forward rates

Forward interest rate volatility is the second element, besides stochastic correlations, shown to play an important role in derivative pricing. Similar to unconditional second moments of yields, the term structure of forward rate volatilities tends to be hump-shaped for shorter maturities, as evidenced in e.g. Amin and Morton (1994) and Moraleda and Vorst (1997). This pattern, in turn, translates into the hump-shaped Black (1976) implied volatility of caps and caplets, by which these instruments are typically quoted in the market. The "transmission" mechanism from forward rates to caplets can be understood by noticing that the volatility of a caplet is the integrated instantaneous volatility of the forward rate (see e.g. Brigo and Mercurio, 2006). It follows that models which can display a hump in the instantaneous volatility of the forward rate should also perform well in the pricing of caplets.

In our model, the instantaneous forward rate is given in closed form in expression (26). We can readily obtain the whole term structure of the instantaneous forward rate volatilities. Figure 14 plots instantaneous volatility curves for several dates in our simulation. Since the hump is also an unconditional feature of the data, we compute the standard deviation of the one-year forward rates in our simulated sample, and plot them against maturities in Figure 15. Additionally, we put the results into perspective with the three affine models used before. At the calibrated parameters, both the instantaneous and the unconditional term structure of forward rate volatilities in our model exhibit a pronounced hump. The decomposition of the unconditional forward rate variance reveals that the non-monotonicity is introduced by three elements: the two variances and the covariance of Σ_{12} and Σ_{11}, scaled by the respective products of the elements of matrix $A(\tau)$.[35] Thus, we identify that the source of a hump in volatility curves is the same both for yields and for forward rates. Figure 15 shows that, in contrast to the Wishart factor setting, the hump is absent from the standard affine specifications. In the $A_0(3)$ model, the forward volatility curve is monotonically decreasing; the mixed models $A_1(3)$ and $A_2(3)$, on the other hand, imply its increase for longer maturities—an implication which is not valid empirically.

The research into the pricing of caps and swaptions points to a link between correlations of different yields/forward rates and the humped term structure of their volatilities (Collin-Dufresne and Goldstein, 2001; Han, 2005). Our discussion of the sources of the volatility hump indicates that this link can be retrieved in the Wishart setting from the role played by the factor Σ_{12}. For one, Σ_{12} determines the stochastic dependence structure between different elements of the state matrix, and thus between yields. For the other, it introduces the non-monotonicity of the volatility curves. We can therefore state that the shape of volatility curves in our model is just a different manifestation of the same underlying force: the stochastic relationship between factors.

[35] The decomposition refers to the variance of the forward rate $f(t; S, T)$, where $f(t; S, T) = \ln P(t, S) - \ln P(t, T) = b(t, S) - b(t, T) + Tr\left[(A(t, S) - A(t, T))\Sigma_t\right]$.

Bond portfolio management and interest rate derivative pricing are two obvious examples, in which the modeling of stochastic relations between factors and yields is essential for successful applications. In the context of interest rate derivatives, the restrictions on the correlation structure imposed by the affine class have turned out to be a major obstacle in reliable and consistent pricing of caps and swaptions (see e.g. Jagannathan, Kaplin, and Sun (2003)). Since recent research highlights the need for stochastic correlations among the interest rates, our model provides a simple and tractable framework to price interest rate derivatives. Moreover, by increasing the dimension of the Wishart state matrix (e.g. to a 3×3 case), we can imagine incorporating additional factors that are detached from the spot yields, but are able to span the derivative market.

VI. Conclusions

In this article, we have explored the theoretical and empirical implications of a term structure model, in which the risk factors are stochastically correlated. Our contribution is twofold. First, we propose a general equilibrium continuous time economy, in which the production technology is driven by a matrix-valued process of dependent factors. We solve the model in closed form for the bond prices and in semi-closed form for the prices of interest rate derivatives. Second, we systematically investigate the empirical implications of the proposed setting, and document that it helps explain several term structure puzzles at the same time.

Our model is endowed with three important characteristics: (*i*) the market price of risk can take both positive and negative values, (*ii*) the correlation structure of factors is stochastic and unrestricted in sign, and (*iii*) all factors display stochastic volatility. While the first feature is shared with some other essentially and extended affine models, the latter two in combination with each other represent the peculiarity of our approach. With these levers, we provide answers to the following issues:

First, we replicate the distributional properties and the dynamic behavior of expected bond returns. The predictability of returns in our calibrated economy violates the expectations hypothesis in line with the historical evidence. The model-implied population coefficients in Campbell and Shiller (1991) regressions are negative and increase in absolute value with the time to maturity. Similarly, a steeper slope of the term structure and a larger spot-forward spread forecast higher excess bond returns in the future.

Second, the extent of persistence of the model-implied conditional yield volatilities matches what is found in the data by estimating a GARCH model. Moreover, the model is able to accommodate the changes in the term structure of unconditional yield volatilities across different subsamples.

Third, the model generates correlations between levels of different yields and between different yield spreads that change broadly but persistently over time, and can switch signs. We observe that the pattern and the extent of this time variation is largely consistent with the data.

Fourth, we find that the term structure of the forward rate volatilities implied by the model is marked by a hump around the 2-year maturity. The result is preserved both instantaneously, and for the unconditional volatilities of discrete observations simulated from the model. Since there is a functional relationship between the caplet and the forward rate volatilities, we expect our model to explain the prices of caplets observed in the market. Preempting our future research, we tentatively claim that since the model is able to produce the humped forward rate volatility and stochastic correlations of factors, it has also the latitude to price caps and swaptions in a consistent way.

Several facts strengthen the findings just described. First, to illustrate its explanatory power, we use the most parsimonious formulation of the model. The choice of the 2×2 state matrix puts us in a three-factor framework, with two positive and one unrestricted factor, while the assumption of log-investors leads us to the simplest possible market price of risk. In this form, the model is endowed with only 11 parameters to successfully perform all tasks listed above, while many of the best performing three-factor essentially or extended affine parameterizations comprise at least twice this number.

Second, using just a *single* set of calibrated parameters, we are largely able to reconcile the first and second moments of model-implied yields with their historical counterparts. Even though the calibration is performed with the unconditional moments of yields, the intrinsic structure of the model permits to reproduce the conditional features of the data.

Third, to appreciate the full extent of our framework, it is important to recognize its analogies with the standard ATSMs. For an arbitrary dimension of the Wishart state matrix, we maintain a tractability which is comparable with the multifactor CIR model. Likewise, the equilibrium argument we apply to derive the market price of risk, is closest to the one that stands behind the completely affine class. In spite of the apparent similarities, the explanatory content together with the parsimony of our framework make the model very attractive.

Finally, the generality of the presented approach allows for easy extensions into higher dimensions of the state matrix. One conceivable enlargement of the model lies in incorporating additional factors that drive prices of interest rate derivatives, but at the same time are detached from the spot interest rates. With the increased dimension of the state space to the 3×3 case, the model has six factors, three of which are unrestricted, and only 24 parameters. The additional factors enhance the flexibility of the model without impairing its analytical tractability. Measuring by the number of parameters, the 3×3

setting is comparable to the essentially or extended affine three-factor specifications, and therefore it shall be readily available for empirical applications.

References

Ahn, D.-H., R. F. Dittman, and A. R. Gallant (2002): "Quadratic Term Structure Models: Theory and Evidence," *Review of Financial Studies*, 15, 243–288.

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

Amin, K. I., and A. J. Morton (1994): "Implied Volatility Functions in Arbitrage-free Term Structure Models," *Journal of Financial Economics*, 35, 141–180.

Ang, A., and M. Piazzesi (2003): "A No-Arbitrage Vector Autoregression of Term Structure with Macroeconomic and Latent Variables," *Journal of Monetary Economics*, 50, 745–787.

Audrino, F., G. Barone-Adesi, and A. Mira (2005): "The Stability of Factor Models of Interest Rates," *Journal of Financial Econometrics*, 3, 422–441.

Black, F. (1976): "The Pricing of Commodity Contracts," *Journal of Financial Economics*, 3, 167–179.

Brandt, M. W., and D. A. Chapman (2002): "Comparing Multifactor Models of the Term Structure," Working paper, University of Pennsylvania and University of Texas.

Brigo, D., and F. Mercurio (2006): *Interest Rate Models: Theory and Practice*. Springer, Berlin, Heidelberg.

Brown, S. J., and P. H. Dybvig (1986): "The Empirical Implications of the Cox, Ingersoll, Ross Theory of the Term Structure of Interest Rates," *Journal of Finance*, 41, 617–630.

Bru, M.-F. (1991): "Wishart Processes," *Journal of Theoretical Probability*, 4, 725–751.

Buraschi, A., and A. Jiltsov (2006): "Term Structure of Interest Rates Implications of Habit Persistence," *Journal of Finance*, forthcoming.

Buraschi, A., P. Porchia, and F. Trojani (2006): "Correlation Hedging," Working Paper, Imperial College, London and University of St. Gallen.

Campbell, J. Y., and R. J. Shiller (1991): "Yield Spreads and Interest Rate Movements: A Bird's Eye View," *Review of Economic Studies*, 58, 495–514.

Carr, P., and D. B. Madan (1999): "Option Valuation Using the Fast Fourier Transform," *Journal of Computational Finance*, 2, 61–73.

Cheridito, P., D. Filipović, and R. L. Kimmel (2005): "Market Price of Risk Specifications for Affine Models: Theory and Evidence," Working Paper, Princeton University, University of Munich, *Journal of Financial Economics*, forthcoming.

Cochrane, J. H., and M. Piazzesi (2005): "Bond Risk Premia," *American Economic Review*, 95, 138–160.

Collin-Dufresne, P., and R. S. Goldstein (2001): "Stochastic Correlations and the Relative Pricing of Caps and Swaptions in a Generalized-Affine Framework," Working Paper, Carnegie Mellon University and Washington University.

——— (2002): "Do Bonds Span the Fixed Income Markets? Theory and Evidence for the Unspanned Stochastic Volatility," *Journal of Finance*, 58(4), 1685–1730.

Conley, T. G., L. P. Hansen, E. G. J. Luttmer, and J. A. Scheinkman (1997): "Short-Term Intrest Rates as Subordinated Diffusions," *Review of Financial Studies*, 10, 525–577.

Cox, J. C., J. E. Ingersoll, and S. A. Ross (1985a): "An Intertemporal General Equilibrium Model of Asset Prices," *Econometrica*, 53, 363–384.

——— (1985b): "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 373–384.

da Fonseca, J., M. Grasselli, and C. Tebaldi (2005): "Wishart Multi-Dimensional Stochastic Volatility," Working Paper, ESILV, University of Padova and University of Verona.

——— (2006): "Option Pricing when Correlations Are Stochastic: An Analytical Framework," Working Paper, ESILV, University of Padova and University of Verona.

Dai, Q. (2003): "Term Structure Dynamics in a Model with Stochastic Habit," Working Paper, New York University.

Dai, Q., and K. Singleton (2000): "Specification Analysis of Affine Term Structure Models," *Journal of Finance*, 55, 1943–1978.

——— (2002): "Expectationns Puzzles, Time-Varying Risk Premia, and Affine Models of the Term Structure," *Journal of Financial Economics*, 63, 415–441.

——— (2003): "Term Structure Dynamics in Theory and Reality," *Review of Financial Studies*, 16(3), 631–678.

De Jong, F., J. Driessen, and A. Pessler (2004): "On the Information in the Interest Rate Term Structure and Option Prices," *Review of Derivatives Research*, 7, 99–127.

Driessen, J., P. Klaassen, and B. Melenberg (2002): "The Performance of Multi-Factor Term Structure Models for Pricing and Hedging Caps and Swaptions," Working Paper, Tilburg Unversity.

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

Duffie, D., D. Filipović, and W. Schachermayer (2003): "Affine Processes and Applications in Finance," *Annals of Applied Probability*, 13, 984–1053.

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

Duffie, D., J. Pan, and K. Singleton (2000): "Transform Analysis and Asset Pricing for Affine Jump-Diffusions," *Econometrica*, 68, 1343–1376.

Duffie, D., and K. Singleton (1997): "An Econometric Model of the Term Structure of Interest-Rate Swap Yields," *Journal of Finance*, 52, 1287–1321.

——— (1999): "Modeling Term Structure of Defaultable Bonds," *Review of Financial Studies*, 12(4), 687–720.

Engle, R. (2002): "Dynamic Conditional Correlation – A Simple Class of Multivariate GARCH Models," *Journal of Business and Economic Statistics*, 20, 339–350.

Fama, E. F. (1984): "The Information in the Term Structure," *Journal of Financial Economics*, 13, 509–528.

Fama, E. F., and R. R. Bliss (1987): "The Information in Long-Maturity Forward Rates," *American Economic Review*, 77, 680–692.

Fisher, M., and C. Gilles (1996a): "Estimating Exponential-Affine Models of the Term Structure," Working Paper, Federal Reserve Bank of Atlanta.

Fisher, M., and C. Gilles (1996b): "Term Premia in Exponential-Affine Models of the Term Structure," Working Paper, Federal Reserve Bank of Atlanta.

Gibbons, M. R., and K. Ramaswamy (1993): "A Test of the Cox, Ingersoll, and Ross Model of the Term Structure," *Review of Financial Studies*, 6, 619–658.

Gourieroux, C. (2006): "Continuous Time Wishart Process for Stochastic Risk," *Econometric Reviews*, 25, 177–217.

Gourieroux, C., J. Jasiak, and R. Sufana (2004): "The Wishart Autoregressive Process of Multivariate Stochastic Volatility," Working Paper, CREST, CEPREMAP, and University of Toronto.

Gourieroux, C., and R. Sufana (2003): "Wishart Quadratic Term Structure Models," Working Paper, CREST, CEPREMAP, and University of Toronto.

——— (2004): "Derivative Pricing with Wishart Multivariate Stochastic Volatility: Application to Credit Risk," Working Paper, CREST, CEPREMAP, and University of Toronto.

Grasselli, M., and C. Tebaldi (2004): "Solvable Affine Term Structure Models," Working Paper, University of Verona, *Mathematical Finance*, forthcoming.

Han, B. (2005): "Stochastic Volatilities and Correlations of Bond Yields," Working Paper, Ohio State University, *Journal of Finance*, forthcoming.

Heston, S. L., and S. Nandi (1999): "A Discrete-Time Two-Factor Model for Pricing Bonds and Interest Rate Derivatives under Random Volatility," Working Paper, Federal Reserve Bank of Atlanta.

Jagannathan, R., A. Kaplin, and S. Sun (2003): "An Evaluation of Multi-Factor CIR Models Using LIBOR, Swap Rates, and Cap and Swaption Prices," *Journal of Econometrics*, 116, 113–146.

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

Kimmel, R. L. (2004): "Modeling the Term Structure of Interest Rates: A New Approach," *Journal of Econometrics*, 72, 143–183.

Lamoureux, C. G., and H. D. Witte (2002): "Empirical Analysis of the Yield Curve: The Information in the Data Viewed through the Window of Cox, Ingersoll, and Ross," *Journal of Finance*, 58, 1479–1520.

Litterman, R., and J. Scheinkman (1991): "Common Factors Affecting Bond Returns," *Journal of Fixed Income*, 1, 54–61.

Longstaff, F. A., P. Santa-Clara, and E. S. Schwartz (2001): "The Relative Valuation of Caps and Swaptions: Theory and Empirical Evidence," *Journal of Finance*, 56, 2067–2109.

Longstaff, F. A., and E. S. Schwartz (1992): "Interest Rate Volatility and the Term Structure: A Two-Factor General Equilibrium Model," *Journal of Finance*, 47, 1259–1282.

Merton, R. C. (1969): "Lifetime Portfolio Selection Under Uncertainty: The Continuous-Time Case," *Review of Economics and Statistics*, 51, 247–257.

——— (1971): "Optimal Consumption and Portfolio Rules in a Continuous-Time Model," *Journal of Economic Theory*, 3, 373–413.

Moraleda, J. M., and T. C. Vorst (1997): "Pricing Americal Interest Rate Claims with Humped Volatility Models," *Journal of Banking and Finance*, 21, 1131–1157.

Musiela, M., and M. Rutkowski (2005): *Martingale Methods in Financial Modeling*. Springer, Berlin, Heidelberg.

Pérignon, C., and C. Villa (2006): "Sources of Time Variation in the Covariance Matrix of Interest Rates," *Journal of Business*, 79, 1535–1549.

Piazzesi, M. (2001): "An Econometric Model of the Yield Curve with Macroeconomic Jump Effects," Working Paper, UCLA and NBER.

——— (2003): "Affine Term Structure Models," Working Paper, University of Chicago.

Roduner, C. (1994): "Die Riccati-Gleichung," Manuscript, ETH Zürich.

Sangvinatsos, A., and J. A. Wachter (2005): "Does the Failure of the Expectations Hypothesis Matter for Long-Term Investors?," *Journal of Finance*, 60, 179–230.

Singleton, K., and L. Umantsev (2002): "Pricing Coupon-Bond Options and Swaptions in Affine Term Structure Models," *Mathematical Finance*, 12, 427–446.

Singleton, K. J. (2006): *Empirical Dynamic Asset Pricing*. Princeton University Press, Princeton and Oxford.

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

A. Proofs

A.1. Useful results for the Wishart process

Result 10. *The following result facilitates the computation of the second order moments of (quadratic forms of) the Wishart process (see Gourieroux (2006)):*

$$\begin{aligned}
& Cov_t\left(\alpha' d\Sigma_t \alpha, \beta' d\Sigma_t \beta\right) \\
= \;& Cov_t\left[\alpha'\left(\sqrt{\Sigma_t} dB_t Q + Q' dB_t' \sqrt{\Sigma_t}\right)\alpha, \beta'\left(\sqrt{\Sigma_t} dB_t Q + Q' dB_t' \sqrt{\Sigma_t}\right)\beta\right] \\
= \;& E_t\left[\left(\alpha'\sqrt{\Sigma_t} dB_t Q\alpha + \alpha' Q' dB_t'\sqrt{\Sigma_t}\alpha\right)\left(\beta'\sqrt{\Sigma_t} dB_t Q\beta + \beta' Q' dB_t'\sqrt{\Sigma_t}\beta\right)\right] \\
= \;& 4\left(\alpha'\Sigma_t\beta\alpha' Q' Q\beta\right) dt,
\end{aligned}$$

where we use the fact that for any n-dimensional vectors u and v it holds

$$\begin{aligned}
E_t\left(dB_t uv' dB_t\right) &= E_t\left(dB_t' uv' dB_t'\right) = vu' dt \\
E_t\left(dB_t uv' dB_t'\right) &= E_t\left(dB_t' uv' dB_t\right) = v' u I_n dt
\end{aligned}$$

Result 11 (Covariances between quadratic forms associated with the Wishart matrix)**.** *Given $n \times n$ Wishart matrix $d\Sigma$ and arbitrary n-dimensional vectors a, b, c, f it follows*

$$Cov_t\left[a' d\Sigma_t b, c' d\Sigma_t f\right] = \left[a' Q' Q f b' \Sigma_t c + a' Q' Q c b' \Sigma_t f + b' Q' Q f a' \Sigma_t c + b' Q' Q c a' \Sigma_t f\right] dt$$

Covariances between arbitrary quadratic forms of $d\Sigma$ are linear combinations of quadratic forms of Σ. In particular, both drift and instantaneous covariances of the single components of the matrix process Σ are affine functions of Σ itself. Using this formula, it is straightforward to compute

$$\begin{aligned}
d\langle\Sigma_{11}\rangle_t &= 4\Sigma_{11}\left(Q_{11}^2 + Q_{21}^2\right) dt \\
d\langle\Sigma_{22}\rangle_t &= 4\Sigma_{22}\left(Q_{22}^2 + Q_{12}^2\right) dt \\
d\langle\Sigma_{12}\rangle_t &= \left[\Sigma_{11}\left(Q_{12}^2 + Q_{22}^2\right) + \Sigma_{22}\left(Q_{11}^2 + Q_{21}^2\right) + 2\Sigma_{12}\left(Q_{11}Q_{12} + Q_{21}Q_{22}\right)\right] dt \\
d\langle\Sigma_{11}, \Sigma_{22}\rangle_t &= 4\Sigma_{12}\left(Q_{11}Q_{12} + Q_{21}Q_{22}\right) dt \\
d\langle\Sigma_{11}, \Sigma_{12}\rangle_t &= \left[2\Sigma_{11}\left(Q_{11}Q_{12} + Q_{21}Q_{22}\right) + 2\Sigma_{12}\left(Q_{11}^2 + Q_{21}^2\right)\right] dt \\
d\langle\Sigma_{22}, \Sigma_{12}\rangle_t &= \left[2\Sigma_{22}\left(Q_{11}Q_{12} + Q_{21}Q_{22}\right) + 2\Sigma_{12}\left(Q_{22}^2 + Q_{12}^2\right)\right] dt,
\end{aligned}$$

where Σ_{ij} and Q_{ij} denote the ij-th element of matrix Σ and Q, respectively.

Result 12. *If Σ_t is a Wishart process and C is a positive definite matrix, then the scalar process $Tr(C\Sigma_t)$ is positive (see Gourieroux (2006)).*

By the singular value decomposition, a symmetric (positive or negative) definite matrix D can be written as $D = \sum_{j=1}^n \lambda_j m_j m_j'$, where λ_j and m_j are the eigenvalues and eigenvectors of D, respectively. Let D be positive definite, and we get:

$$\begin{aligned}
Tr(D\Sigma) &= Tr\left(\sum_{j=1}^n \lambda_j m_j m_j' \Sigma\right) = \sum_{j=1}^n \lambda_j Tr(m_j m_j' \Sigma) \\
&= \sum_{j=1}^n \lambda_j Tr(m_j' \Sigma m_j) = \sum_{j=1}^n \lambda_j m_j' \Sigma m_j > 0,
\end{aligned}$$

where we use the facts that: (i) we can commute within the trace operator, (ii) $\lambda_j > 0$ for all j, and (iii) Σ is positive definite. Note that for a positive definite matrix $D = \sum_{j=1}^n a_j a_j'$, where $a_j = \sqrt{\lambda_j} m_j$.

Result 13. *For any symmetric square matrices A and C, it follows*

$$Cov_t\left[Tr(Ad\Sigma_t), Tr(Cd\Sigma_t)\right] = 4Tr[A\Sigma_t CQ'Q].$$

To derive the result we can directly consider the product of the two traces

$$\begin{aligned}
&Cov_t\left[Tr(Ad\Sigma_t), Tr(Cd\Sigma_t)\right] = \\
&Tr\left(A\sqrt{\Sigma_t}dB_tQ + AQ'dB_t'\sqrt{\Sigma_t}\right)Tr\left(C\sqrt{\Sigma_t}dB_tQ + CQ'dB_t'\sqrt{\Sigma_t}\right).
\end{aligned}$$

Apply several times the fact that

$$Tr\left(QA\sqrt{\Sigma_t}dB_t\right) = \mathrm{vec}(\sqrt{\Sigma_t}AQ')'\mathrm{vec}(dB_t)$$

and note that

$$\mathrm{vec}(dB_t)\mathrm{vec}(dB_t)' = I_{n^2}dt,$$

where vec *denotes the operation of stacking elements of a matrix in a vector.*

A.2. Proof of Proposition 5: Equilibrium interest rate and market price of risk

Let $\psi \equiv U + \mathcal{L}_{W,\Sigma}J$, where J is given in (10), and U is the logarithmic utility function. The log-utility simplifies the investor's maximization problem since due to separability of the value function the mixed partial derivatives (with respect to wealth and the state) are zero and ψ becomes

$$\begin{aligned}
\psi &= U + \left[\upsilon_Y\left(Tr(D\Sigma) - r\right) + \upsilon_F'(\mu_F - r\mathbf{1}) + (r - \upsilon_C)\right]WJ_W \\
&\quad + \frac{1}{2}\left[\upsilon_Y^2 Tr(\Sigma) + 2\upsilon_Y\upsilon_F'\sigma_{YF} + \upsilon_F'\sigma_{FF}\upsilon_F\right]W^2J_{WW} \\
&\quad + Tr\left[\left(\Omega\Omega' + M\Sigma + \Sigma M'\right)\mathcal{R}J + 2\Sigma\mathcal{R}(QQ'\mathcal{R}J)\right],
\end{aligned}$$

where the μ_F is the vector of drifts with i-th element $\mu_{F_i} = \frac{1}{dt}E\left(\frac{dF_i}{F_i}\right)$, σ_{YF} is the vector of covariances between the returns to production and financial assets with i-th element $\sigma_{YF_i} = \frac{1}{dt}Cov\left(\frac{dY}{Y}, \frac{dF_i}{F_i}\right)$, and σ_{FF} is the covariance of returns to financial assets with ij-th element $\sigma_{FF,ij} = \frac{1}{dt}Cov\left(\frac{dF_i}{F_i}, \frac{dF_j}{F_j}\right)$.

The first order conditions for maximization are

$$\psi_C = U_C - J_W = 0, \tag{32}$$

$$\psi_Y = \left[Tr(D\Sigma) - r\right]WJ_W + \upsilon_Y Tr(\Sigma_t)W^2J_{WW} = 0, \tag{33}$$

$$\psi_{F_i} = \left(\mu_{F_i} - r\right)WJ_W + \left(\upsilon_Y\sigma_{YF_i} + \upsilon_{F_i}\sigma_{FF,ii}\right)W^2J_{WW} = 0. \tag{34}$$

The standard envelope condition in (32) implies the optimal consumption: $C^* = \rho W$. In equilibrium, the agent is fully invested in the production technology Y, thus $\upsilon_Y^* = 1$ and $\upsilon_F^* = 0$. From (33), we obtain the equilibrium interest rate

$$r = Tr(D\Sigma) - \left(-\frac{WJ_{WW}}{J_W}Tr(\Sigma)\right),$$

which for the case of log-utility investor has a particularly simple form

$$r_t = Tr\left[(D - I_n)\Sigma_t\right].$$

From the definition of the market price of risk, we have

$$\mu_Y - r = \sigma_Y' \lambda, \tag{35}$$

where $\mu_Y := Tr(D\Sigma)$, $\sigma_Y := \text{vec}(\sqrt{\Sigma})$, and $\lambda := \text{vec}(\Lambda)$.

$$\begin{aligned} \mu_Y - r &= \text{vec}(\sqrt{\Sigma})'\text{vec}(\Lambda) = Tr(\sqrt{\Sigma}\Lambda) \text{ (by definition)} \\ &= Tr(\Sigma) \text{ (from equilibrium).} \end{aligned}$$

It thus follows $\Lambda_t = \sqrt{\Sigma_t}$. ■

A.3. Second moments of the equilibrium short interest rate

Let $D - I_n = C$, a symmetric matrix. We assume the matrix C to be positive definite. Therefore, it can be written as $C = \sum_{i=1}^{n} c_i c_i'$. In a first step, we derive the expression for the instantaneous variance of the interest rate. Applying Ito's Lemma to the interest rate, and using Result 12 in Appendix A.1 we have

$$dr = Tr\left(C d\Sigma\right) = \sum_{i=1}^{n} c_i' d\Sigma c_i.$$

By Result 10, we obtain the expression (13):

$$\begin{aligned} Var_t\left(\sum_{i=1}^{n} c_i' d\Sigma c_i\right) &= Cov_t\left(\sum_{i=1}^{n} c_i' d\Sigma c_i, \sum_{j=1}^{n} c_j' d\Sigma c_j\right) \\ &= 4\sum_{i=1}^{n}\sum_{j=1}^{n}\left(c_i'\Sigma c_j \; c_i' Q' Q c_j\right) \\ &= 4\sum_{j=1}^{n} c_j' \Sigma C Q' Q c_j dt \\ &= 4Tr\left(C\Sigma C Q' Q\right) dt \\ &= 4Tr\left[(D - I_n)\Sigma(D - I_n)Q'Q\right] dt \\ &= 4Tr\left[\Sigma(D - I_n)QQ'(D - I_n)\right] dt. \end{aligned}$$

By similar arguments the expression for the covariance between the changes in the level and the variance of interest rate follows. Let $(D - I_n)Q'Q(D - I_n) = P$, a symmetric matrix, and apply Ito's Lemma to V_t.

$$
\begin{aligned}
Cov_t\left(dr, dV\right) &= Cov_t\left[Tr\left(Cd\Sigma\right), Tr\left(Pd\Sigma\right)\right] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} c_i' d\Sigma c_i \; p_j' d\Sigma p_j \\
&= 4\sum_{i=1}^{n}\sum_{j=1}^{n} c_i'\Sigma_t p_j c_i' Q' Q p_j dt \\
&= 4\sum_{i=1}^{n}\sum_{j=1}^{n} p_j'\Sigma c_i c_i' Q' Q p_j dt \\
&= 4\sum_{j=1}^{n} p_j'\Sigma C Q' Q p_j dt \\
&= 4Tr\left[P\Sigma C Q' Q\right] dt \\
&= 4Tr\left[\Sigma\left(D - I_n\right) Q' Q\left(D - I_n\right) Q' Q\left(D - I_n\right)\right] dt.
\end{aligned}
$$

Note that the multiplier of Σ, i.e. $H = \left(D - I_n\right) Q'Q\left(D - I_n\right) Q'Q\left(D - Id_n\right)$, is again a symmetric matrix. ■

A.4. Proof of Proposition 7: Pricing PDE

The only term that requires clarification is the drift adjustment Φ_Σ. By Theorem 2 in CIR (1985a, p. 374), the risk adjustment (excess return) for the factor Σ_{kl} is given by

$$
\Phi_{\Sigma_{kl}} = \frac{-J_{WW}}{J_W} Cov_t(dW, d\Sigma_{kl}) + \sum_{i=1}^{n}\sum_{j=1}^{n}\left(-\frac{J_{W\Sigma_{kl}}}{J_W} Cov_t(d\Sigma_{kl}, d\Sigma_{ij})\right).
$$

For the separable log-utility case, the drift adjustment has a particularly simple form since all cross derivatives are zero, and it can be comfortably written in a matrix form as

$$
\Phi_\Sigma = \frac{-J_{WW}}{J_W} Cov_t(dW, d\Sigma) = \frac{1}{W} d\langle W, \Sigma\rangle, \tag{36}
$$

where the covariance is represented by a $n \times n$ symmetric matrix with a typical element $d\langle W, \Sigma_{ij}\rangle$, e.g. in the 2×2 case

$$
d\langle W, \Sigma\rangle = \begin{pmatrix} d\left\langle W, \Sigma_{11}\right\rangle & d\left\langle W, \Sigma_{12}\right\rangle \\ d\left\langle W, \Sigma_{12}\right\rangle & d\left\langle W, \Sigma_{22}\right\rangle \end{pmatrix}.
$$

Therefore, it is enough to show that for an arbitrary dimension $n \times n$ of matrices Σ, Q and dB, it holds:

$$
\Phi_\Sigma = Cov_t\left[Tr(\sqrt{\Sigma} dB), \sqrt{\Sigma} dB Q + Q' dB' \sqrt{\Sigma}\right] = \Sigma Q + Q'\Sigma.
$$

Let us define $\sigma_{ki}, 1 \le k, i \le n$ as a typical element $\sqrt{\Sigma}$. First, we note that

$$
Tr\left(\sqrt{\Sigma} dB\right) = \sum_{k=1}^{n}\sum_{i=1}^{n} \sigma_{ki} dB_{ik}, \tag{37}
$$

and that the element (l,m) of the matrix $\sqrt{\Sigma}dB$ is

$$(\sqrt{\Sigma}dB)^{l,m} = \sum_{w=1}^{n} \sigma_{lw} dB_{wm} = \sum_{w=1}^{n} \sigma_{wl} dB_{wm}.$$

Thus, we obtain the following covariance

$$\begin{aligned} \frac{1}{dt}\langle Tr(\sqrt{\Sigma}dB), (\sqrt{\Sigma}dB)^{l,m}\rangle &= \frac{1}{dt}\sum_{k=1}^{n}\sum_{i=1}^{n}\sigma_{ki}dB_{ik}\sum_{w=1}^{n}\sigma_{lw}dB_{wm} \\ &= \sum_{i=1}^{n}\sigma_{mi}\sigma_{li} = \Sigma_{ml} = \Sigma_{lm} \end{aligned}$$

The covariation matrix follows as

$$\frac{1}{dt}\langle Tr(\sqrt{\Sigma}dB), \sqrt{\Sigma}dBQ\rangle = \Sigma Q.$$

In the similar manner, we obtain the second summand $Q'\Sigma$ in (37).

Note that we arrive at the same drift adjustment by using the Girsanov's theorem, where the Girsanov's kernel is $-\Lambda_t$, and Λ_t is the market price of risk $\Lambda_t = \sqrt{\Sigma_t}$. Define the Radon-Nikodym derivative for the transformation from the physical measure $\tilde{\mathbb{Q}}$ to the risk neutral measure $\mathbb{Q}^*$:

$$\frac{d\mathbb{Q}^*}{d\tilde{\mathbb{Q}}}|_{\mathcal{F}_t} := e^{Tr\left[-\int_0^t \Lambda_u' dB_u - \frac{1}{2}\int_0^t \Lambda_u'\Lambda_u du\right]}. \tag{38}$$

Under the risk neutral measure $\mathbb{Q}^*$ the process B_t^* defined by

$$B_t^* = B_t + \int_0^t \Lambda_u du \tag{39}$$

is a $n \times n$ matrix of standard Brownian motions. The risk neutral dynamics of the Wishart process is given by:

$$d\Sigma_t = (\Omega\Omega' + (M - Q')\Sigma_t + \Sigma_t(M' - Q))\,dt + \sqrt{\Sigma_t}dB_t^*Q + Q'dB_t^{*\prime}\sqrt{\Sigma_t}. \tag{40}$$

■

A.5. Proof of Proposition 8: Solution to the matrix Riccati equation

The coefficients $A(t,T)$ and $b(t,T)$ in the bond price expression are identified by inserting function (19) into the pricing PDE (17) and solving the resulting matrix Riccati equation. Note that $\mathcal{R}P = A(t,T)P$, and

$$\frac{\partial P}{\partial t} = P\left[\frac{d}{dt}b(t,T) + Tr\left(\frac{d}{dt}A(t,T)\Sigma\right)\right], \tag{41}$$

where for brevity with P we denote the time t price of a bond maturing at time T.

The fundamental PDE. The pricing PDE (17) can be expressed as

$$Tr\left[(\Omega\Omega' + (M-Q')\Sigma + \Sigma(M'-Q))A + 2\Sigma AQ'QA\right] + \frac{db}{dt} + Tr\left(\frac{dA}{dt}\Sigma\right) - Tr\left[(D-I_n)\Sigma\right] = 0.$$

Matrix Riccati equation. The above equation holds for all t, T and Σ. After identifying coefficients of Σ, we get a system of matrix Riccati equations in A and b

$$-\frac{db}{dt} = Tr\left(\Omega\Omega' A\right) \tag{42}$$

$$-\frac{dA}{dt} = A(M-Q') + (M'-Q)A + 2AQ'QA - (D-I_n). \tag{43}$$

with the respective terminal conditions $b(T,T) = 0$ and $A(T,T) = 0$. For convenience, we consider $A(\cdot)$ and $b(\cdot)$ as parametrized by the time to maturity $\tau = T - t$. Clearly, this reparametrization merely requires the LHS of the above system to be multiplied by -1

$$\frac{db}{d\tau} = Tr\left(\Omega\Omega' A\right) \tag{44}$$

$$\frac{dA}{d\tau} = A(M-Q') + (M'-Q)A + 2AQ'QA - (D-I_n), \tag{45}$$

with boundary conditions $A(0) = 0$ and $b(0) = 0$. We remark that the instantaneous interest rate is

$$r_t = \lim_{\tau\to 0} -\frac{1}{\tau}\log P(t,\tau) = -\frac{db(0)}{d\tau} - Tr\left(\frac{dA(0)}{d\tau}\Sigma_t\right) = Tr\left[(D-I_n)\Sigma_t\right].$$

Matrix Riccati linearization. To solve the equation (45), we use the matrix Riccati linearization proposed in da Fonseca, Grasselli, and Tebaldi (2005). We express $A(\tau)$ as

$$A(\tau) = H(\tau)^{-1}G(\tau), \tag{46}$$

for $H(\tau)$ invertible and $G(\tau)$ being a square matrix. Differentiating (46), we have

$$\frac{d}{d\tau}\left[H(\tau)A(\tau)\right] = \frac{dG(\tau)}{d\tau}$$

$$\frac{d}{d\tau}\left[H(\tau)A(\tau)\right] = \frac{dH(\tau)}{d\tau}A(\tau) + H(\tau)\frac{dA(d\tau)}{d\tau}.$$

Premultiplying (45) by $H(\tau)$ gives

$$H\frac{dA}{d\tau} = HA(M-Q') + H(M'-Q)A + 2HAQ'QA - H(D-I_n).$$

This is equivalent to

$$\frac{dG}{d\tau} - \frac{dH}{d\tau}A = G(M-Q') + H(M'-Q)A + 2GQ'QA - H(D-I_n),$$

where for brevity we suppress the argument τ of $A(\cdot), H(\cdot)$ and $G(\cdot)$. After collecting coefficients of A in the last equation, we obtain the following matricial system of ODEs

$$\begin{aligned}\frac{dG(\tau)}{d\tau} &= G(M-Q') - H(D-I_n)\\ \frac{dH(\tau)}{d\tau} &= -2GQ'Q - H(M'-Q),\end{aligned}$$

or written compactly

$$\frac{d}{d\tau}\begin{pmatrix} G(\tau) & H(\tau)\end{pmatrix} = \begin{pmatrix} G(\tau) & H(\tau)\end{pmatrix}\begin{pmatrix} M-Q' & -2Q'Q \\ -(D-I_n) & -(M'-Q)\end{pmatrix}.$$

The solution to the above ODE is obtained by exponentiation

$$\begin{aligned}\begin{pmatrix} G(\tau) & H(\tau)\end{pmatrix} &= \begin{pmatrix} G(0) & H(0)\end{pmatrix}\exp\left[\tau\begin{pmatrix} M-Q' & -2Q'Q \\ -(D-I_n) & -(M'-Q)\end{pmatrix}\right]\\ &= \begin{pmatrix} A(0) & I_n\end{pmatrix}\exp\left[\tau\begin{pmatrix} M-Q' & -2Q'Q \\ -(D-I_n) & -(M'-Q)\end{pmatrix}\right]\\ &= \begin{pmatrix} A(0)C_{11}(\tau)+C_{21}(\tau) & A(0)C_{12}(\tau)+C_{22}(\tau)\end{pmatrix}\\ &= \begin{pmatrix} C_{21}(\tau) & C_{22}(\tau)\end{pmatrix},\end{aligned}$$

where we use the fact that $A(0)=0$, and

$$\begin{pmatrix} C_{11}(\tau) & C_{12}(\tau) \\ C_{21}(\tau) & C_{22}(\tau)\end{pmatrix} := \exp\left[\tau\begin{pmatrix} M-Q' & -2Q'Q \\ -(D-I_n) & -(M'-Q)\end{pmatrix}\right].$$

From equation (46), the closed-form solution to (45) is given by

$$A(\tau) = C_{22}(\tau)^{-1}C_{21}(\tau).$$

Given the solution for A, the coefficient b is obtained directly by integration

$$b(\tau) = Tr\left(\Omega\Omega'\int_0^\tau A(s)\right)ds.$$

■

A.6. Bond returns

By Ito's Lemma, for a smooth function $\phi(\Sigma,t)$ we have

$$d\phi = \left(\frac{\partial\phi}{\partial t} + \mathcal{L}_\Sigma\phi\right)dt + Tr\left[(\sqrt{\Sigma}dBQ + Q'dB'\sqrt{\Sigma})\mathcal{R}\phi\right], \tag{47}$$

where $\mathcal{L}_\Sigma$ denotes the infinitesimal generator of the Wishart process. Using this result, the drift of the bond price $P(\Sigma,t,T)$ can be written as:

$$\begin{aligned}\frac{1}{dt}E_t\left(dP\right) &= \frac{\partial P}{\partial t}+\mathcal{L}_\Sigma P \\ &= \frac{\partial P}{\partial t}+Tr\left[\left(\Omega\Omega'+M\Sigma+\Sigma' M\right)\mathcal{R}P+2\Sigma\mathcal{R}(Q'Q\mathcal{R}P)\right].\end{aligned}$$

From the fundamental PDE (17), we note that at equilibrium the drift must satisfy

$$\frac{1}{dt}E_t\left(dP\right)-Tr(\Phi_\Sigma\mathcal{R}P)=rP.$$

By taking derivatives of the bond price with respect to the Wishart matrix, $\mathcal{R}P=A(\tau)P$, it follows that the expected excess bond return (over the short rate) is given by

$$e_t^\tau=Tr\left[\left(A(\tau)Q'+QA(\tau)\right)\Sigma_t\right]. \tag{48}$$

For completeness, we also provide the expression for the instantaneous variance of the bond return. From equation (47), the diffusion part of the bond dynamics dP is given by $Tr\left[\left(\sqrt{\Sigma}dBQ+Q'dB'\sqrt{\Sigma}\right)A(\tau)P\right]$. Since $A(\tau)$ is a symmetric negative definite matrix, we can write

$$Tr(Ad\Sigma)=\sum_{i=1}^n Tr\left(\lambda_i a_i a_i' d\Sigma\right)=\sum_{i=1}^n \lambda_i a_i' d\Sigma a_i,$$

where λ_i is the i-th eigenvalue of $A(\tau)$ and a_i is its i-th eigenvector. Using Result 10, the instantaneous variance of the bond returns is

$$\begin{aligned}Var_t\left(\frac{dP}{P}\right) &= Var_t\left[Tr\left(\left(\sqrt{\Sigma}dBQ+Q'dB'\sqrt{\Sigma}\right)A\right)\right] \\ &= 4\sum_{i=1}^n\sum_{j=1}^n \lambda_i\lambda_j a_i'\Sigma a_j a_i' Q' Q a_j dt \\ &= 4\sum_{i=1}^n\sum_{j=1}^n \lambda_i\lambda_j a_j'\Sigma a_i a_i' Q' Q a_j dt \\ &= 4\sum_{j=1}^n \lambda_j a_j'\Sigma A Q' Q a_j dt \\ &= 4Tr\left(A\Sigma A Q' Q\right)dt.\end{aligned}$$

■

A.7. Dynamics of the forward rate

The dynamics of the instantaneous forward rate is

$$df(t,\tau)=-\frac{\partial d\log P(t,\tau)}{\partial\tau}.$$

By Ito's Lemma, we first obtain the dynamics of the logarithm of the bond price in (19):

$$d\log P = \left[\frac{\partial b}{\partial t} + Tr\left(\frac{\partial A}{\partial t}\Sigma\right) + Tr\left[(\Omega\Omega' + M\Sigma + \Sigma M')A\right]\right]dt + Tr\left[(\sqrt{\Sigma}dBQ + Q'dB'\sqrt{\Sigma})A\right].$$

By noting that $f(t,T) = \frac{\partial b}{\partial t} + Tr\left(\frac{\partial A}{\partial t}\Sigma_t\right)$, we arrive at the instantaneous forward rate dynamics

$$df(t,\tau) = -\left(\frac{\partial f}{\partial \tau} + Tr\left[(\Omega\Omega' + M\Sigma + \Sigma M')\frac{\partial A}{\partial \tau}\right]\right)dt - Tr\left[\left(\sqrt{\Sigma}dBQ + Q'dB'\sqrt{\Sigma}\right)\frac{\partial A}{\partial \tau}\right]. \blacksquare$$

A.8. Pricing of zero-bond options

A.8.1. Change of drift for the Wishart factors

Let ZBC$(t,\Sigma_t;S,T,K)$ denote the price of a European option with expiry date S and exercise price K, written on a zero-bond maturing at time $T \geq S$

$$\text{ZBC}(t,\Sigma_t;S,T,K) = P(t,T)Pr_t^T\{P(S,T) > K\} - KP(t,S)Pr_t^S\{P(S,T) > K\}.$$

To evaluate the two probabilities Pr_t^T and Pr_t^S in this expression, we need to obtain the dynamics of the Wishart process under the two forward measures associated with bonds maturing at time S and T, respectively. The risk-neutral dynamics of a S-maturity zero-bond $P(t,S)$ are

$$\frac{dP(t,S)}{P(t,S)} = r_t dt + Tr(\Theta'(t,S)dB_t^*) + Tr(\Theta(t,S)dB_t^{*\prime}), \tag{49}$$

where $\Theta(t,S) = \sqrt{\Sigma_t}A(t,S)Q'$, $A(t,S)$ and $\sqrt{\Sigma_t}$ symmetric, and $A(t,S)$ solves the matrix Riccati equation (43). The transformation from the risk neutral measure $\mathbb{Q}^*$ to the forward measure $\mathbb{Q}^S$ is given by

$$\frac{d\mathbb{Q}^S}{d\mathbb{Q}^*}|_{\mathcal{F}_S} = e^{Tr\left[\int_0^S \Theta'(u,S)dB_u^* - \frac{1}{2}\int_0^S \Theta'(u,S)\Theta(u,S)du\right]},$$

where we use the fact that $Tr(\Theta'(t,S)dB_t^*) = \text{vec}(\Theta(t,S))'\text{vec}(dB_t^*)$. By Girsanov's theorem it follows

$$dB_t^* = dB_t^S + \sqrt{\Sigma_t}A(t,S)Q'dt, \tag{50}$$

where dB_t^S is a $n \times n$ matrix of standard Brownian motions under $\mathbb{Q}^S$. By a similar argument, we have

$$dB_t^* = dB_t^T + \sqrt{\Sigma_t}A(t,T)Q'dt, \tag{51}$$

where dB_t^T is a $n \times n$ matrix of standard Brownian motions under $\mathbb{Q}^T$.

Remark 14. *The measure transformations presented here are standard, but for the matrix-trace notation. Equivalently, we could use the vector notation for the T-maturity bond dynamics*

$$\frac{dP_t}{P_t} = r_t dt + \text{vec}(\Theta)'\text{vec}(dB_t^*) + \text{vec}(\Theta')'\text{vec}(dB_t^{*\prime}) \tag{52}$$

Then,

$$\text{vec}(dB_t^*) = \text{vec}(\Theta)dt + \text{vec}(dB_t^T) = \text{vec}(\Theta dt + dB_t^T).$$

By operation reverse to vectorizing, we obtain

$$dB_t^* = dB_t^T + \Theta dt.$$

Recall that the risk-neutral dynamics of the Wishart process is given by

$$d\Sigma_t = (\Omega\Omega' + (M - Q')\Sigma_t + \Sigma_t(M' - Q))\, dt + \sqrt{\Sigma_t} dB_t^* Q + Q' dB_t^{*\prime} \sqrt{\Sigma_t}. \tag{53}$$

We are now ready to express the dynamics of the process under the S-forward measure

$$d\Sigma_t \quad = \quad \{\Omega\Omega' + [M - Q'(I_n - QA)]\Sigma_t + \Sigma_t[M' - (I_n - AQ')]Q\}dt + \sqrt{\Sigma_t} dB_t^S Q + Q' dB_t^{S\prime} \sqrt{\Sigma_t},$$

where for brevity we write A for $A(t, S)$. The dynamics under the T-forward measure $\mathbb{Q}^T$ differ only in A standing for $A(t, T)$.

A.8.2. Pricing of zero-bond option by Fourier inversion

Due to the affine property of the Wishart process, the conditional characteristic function of log-bond prices is available in closed form. Thus, the pricing of bond options amounts to performing two *one-dimensional* Fourier inversions under the two forward measures (see e.g., Duffie, Pan, and Singleton (2000)). We note that

$$Pr_t^j\{P(S,T) > K\} = Pr_t^j\{b(S,T) + Tr\,[A(S,T)\Sigma_S] > \ln K\}, \text{ where } j = \{S, T\}.$$

To evaluate this probability by Fourier inversion, we find the characteristic function of the random variable $Tr[A(S,T)\Sigma_S]$ under the S- and T-forward measures. Let $\tau = S - t$, then the conditional characteristic function is

$$\Psi_t^S(iz) = E_t^S\left(e^{izTr[A(t+\tau,T)\Sigma_{t+\tau}]}\right), \tag{54}$$

where E_t^S denotes the conditional expectation under the S-forward measure, $i = \sqrt{-1}$, and $z \in \mathbb{R}$. In the sequel, we show the argument for the S-forward measure, the argument for the T-forward measure being analogous. By the affine property of Σ_t, the relevant characteristic function is itself of the exponentially affine form in Σ_t

$$\Psi_t^S(iz) = e^{Tr[\hat{A}(\tau)\Sigma_t] + \hat{b}(\tau)}, \tag{55}$$

where $\hat{A}(\tau)$ and $\hat{b}(\tau)$ are, respectively, a symmetric matrix and a scalar with possibly complex coefficients, which solve the system of matrix Riccati equations (57)–(58) detailed below. With the characteristic functions of $Tr[A(S,T)\Sigma_S]$ for the S- and T-forward measure at hand, we can express the bond option price by the Fourier inversion as

$$\begin{aligned}\text{ZBC}(t,S,T) &= P(t,T)\left\{\frac{1}{2}+\frac{1}{\pi}\int_0^\infty \text{Re}\frac{e^{-iz[\log K-b(S,T)]}\Psi_t^T(iz)}{iz}dz\right\}\\ &\quad -KP(t,S)\left\{\frac{1}{2}+\frac{1}{\pi}\int_0^\infty \text{Re}\frac{e^{-iz[\log K-b(S,T)]}\Psi_t^S(iz)}{iz}dz\right\},\end{aligned}$$

in which the integral can be evaluated by numerical methods.

The coefficients $\hat{A}(\tau)$ and $\hat{b}(\tau)$ in (55) are derived by the same logic as in Appendix A.5. By the Feynman-Kač argument applied to (54), Ψ_t^S solves the following PDE

$$\frac{\partial \Psi_t^S}{\partial \tau} = \mathcal{L}_\Sigma \Psi_t^S. \tag{56}$$

Then, plugging for Ψ_t^S the expression (55), and collecting terms, gives the system of ordinary differential equations

$$\frac{\partial \hat{b}(\tau)}{\partial \tau} = Tr[\Omega\Omega'\hat{A}(\tau)] \tag{57}$$

$$\frac{\partial \hat{A}(\tau)}{\partial \tau} = \hat{A}(\tau)M^S + M^{S'}\hat{A}(\tau) + 2\hat{A}(\tau)Q'Q\hat{A}(\tau), \tag{58}$$

where $M^S = M - Q'[I_n - QA(t,S)]$ results from the drift adjustment under the S-forward measure (see Appendix A.8.1). The boundary conditions at $\tau = 0$ are

$$\begin{aligned}\hat{b}(0) &= 0\\ \hat{A}(0) &= ziA(S,T).\end{aligned}$$

By matrix Riccati linearization, the solution for $\hat{A}(\tau)$ reads

$$\hat{A}(\tau) = (ziA(S,T)\hat{C}_{12} + \hat{C}_{22})^{-1}(ziA(S,T)\hat{C}_{11} + \hat{C}_{21}), \tag{59}$$

with

$$\begin{pmatrix} \hat{C}_{11}(\tau) & \hat{C}_{12}(\tau) \\ \hat{C}_{21}(\tau) & \hat{C}_{22}(\tau) \end{pmatrix} := \exp\left[\tau \begin{pmatrix} M^S & -2Q'Q \\ 0 & -M^{S'} \end{pmatrix}\right].$$

The coefficient $\hat{b}(\tau)$ is obtained by integration

$$\hat{b}(\tau) = \int_0^\tau Tr[\Omega\Omega'\hat{A}(u)]du. \tag{60}$$

■

B. Details on the calibration

Table VI summarizes the calibration errors for the respective moment conditions and yields used in calibration.

Table VI: Calibration errors

	6M	2Y	10Y
Means	7	-22	7
	(1.3)	(-3.8)	(1.1)
Volatilities	-12	12	14
	(-4.1)	(4.3)	(5.2)
Correlations*	17	-9	14
	(0.2)	(-0.1)	(0.1)
CS coeff.**	n.a.	0	0
	n.a.	(0.0)	(0.0)

Note: The table displays absolute calibration errors for the moments of the 6-month, 2-year and 10-year yields. The percentage calibration errors are given in parentheses. The absolute error is computed as the difference between the model-implied and the empirical value of a given moment, scaled by 10000; thus errors for means and volatilities are in basis points. The percentage error is the absolute error per unit of the empirical value of the respective moment.
*) The columns are $Corr(y^{6M}, y^{2Y}), Corr(y^{6M}, y^{10Y})$ and $Corr(y^{2Y}, y^{10Y})$.
**) CS coeff. denotes the Campbell-Shiller regression coefficients. Both the 2- and 10-year value of the coefficient is computed using the 6-month yield as the shorter maturity; see Section V.B, expression (29) for the relevant regression equation.

The calibrated parameters are:

$$D = \begin{pmatrix} 1.3844 & 0.5426 \\ 0.5426 & 2.0704 \end{pmatrix},$$

$$M = \begin{pmatrix} -0.1173 & 0.0625 \\ 0.0444 & -0.0256 \end{pmatrix},$$

$$Q = \begin{pmatrix} 0.0338 & -0.1147 \\ -0.0183 & 0.0499 \end{pmatrix}.$$

C. Figures

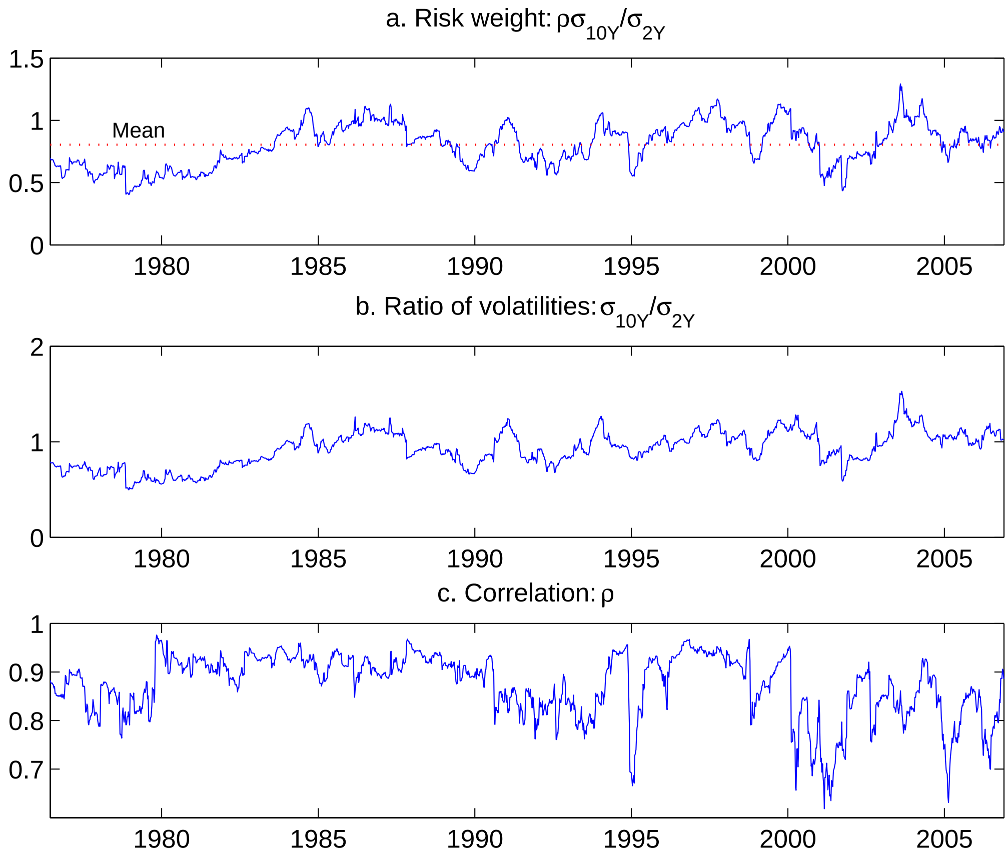

Figure 1: **Time-variation in the hedge ratio.** Panel *a* shows the "optimal" hedge ratio $\beta = \frac{\sigma_{10}}{\sigma_2}\rho_{10,2}$ for hedging a 10-year zero bond with a 2-year bond. Panel *b* displays the ratio of the volatilities of the 10-year and the 2-year yield changes. Panel *c* plots the correlation of the yield changes. The sample comprises weekly changes in the Treasury constant maturity rates from FRED in the period 1976:06–2006:11. All plots are obtained with the exponential moving average (EMA) estimator of the covariance matrix of the yield changes: $V_t = (1-\lambda)u_{t-1}u'_{t-1} + \lambda V_{t-1}$, where u_t is the 2×1 vector of demeaned yield changes, and λ is assumed to equal 0.94.

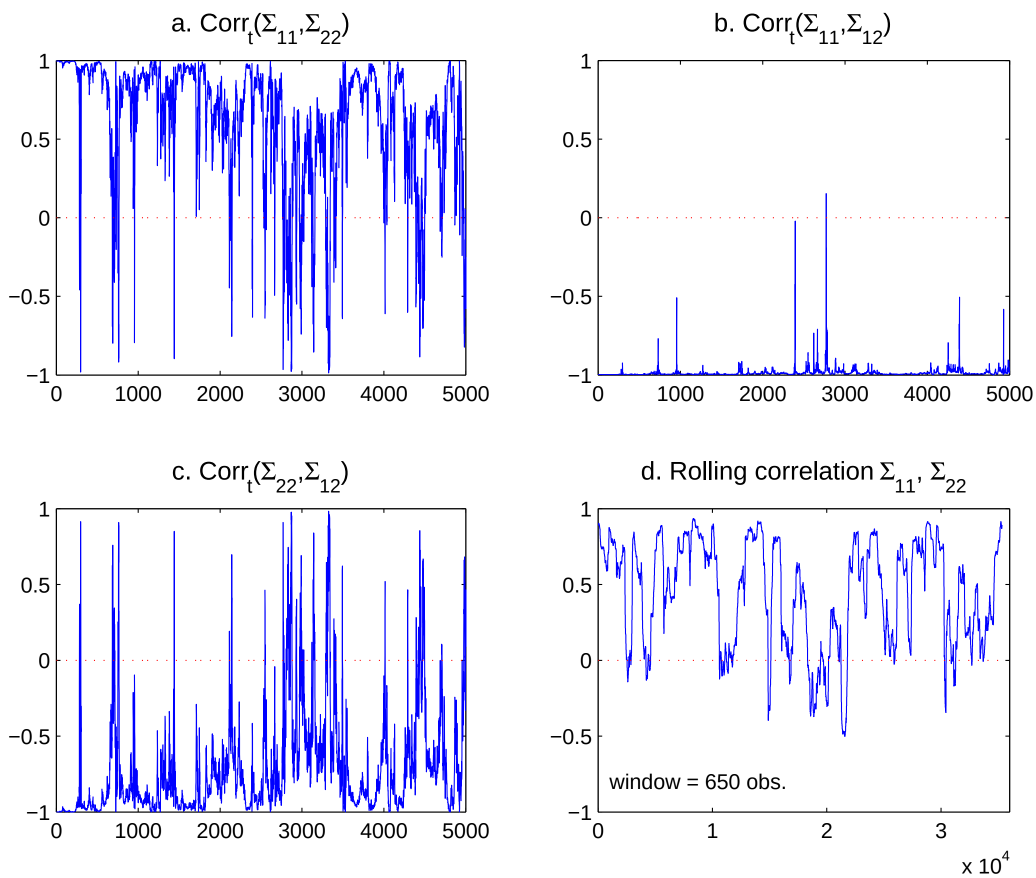

Figure 2: **Factor correlation.** Panels *a*–*c* show instantaneous correlations of factors. For readability, we only show the first 5000 realizations. Panel *d* displays the rolling correlation between the positive factors Σ_{11}, Σ_{22} for the whole simulated sample. The rolling window is 650 months, which is the length of our US yield sample.

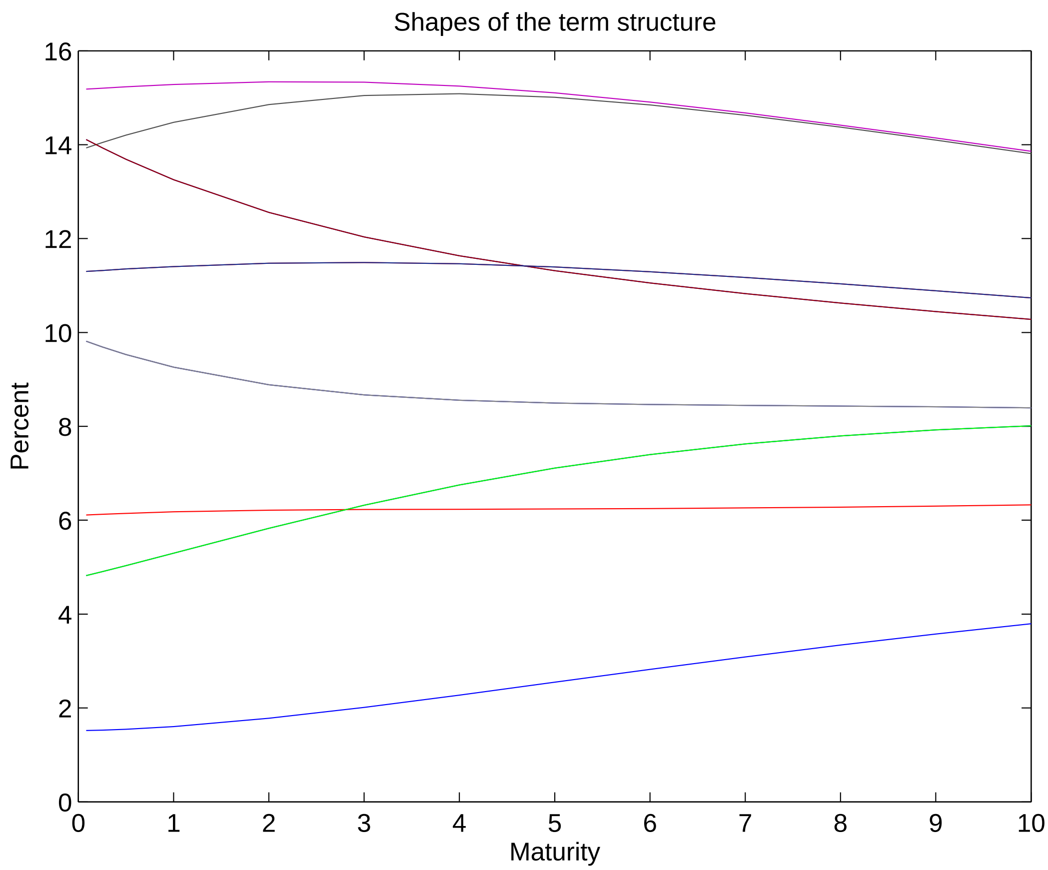

Figure 3: **Possible model-implied shapes of the term structure.** The figure displays a variety of shapes of the term structure arising at different dates of the simulated sample. The maturities are given in years.

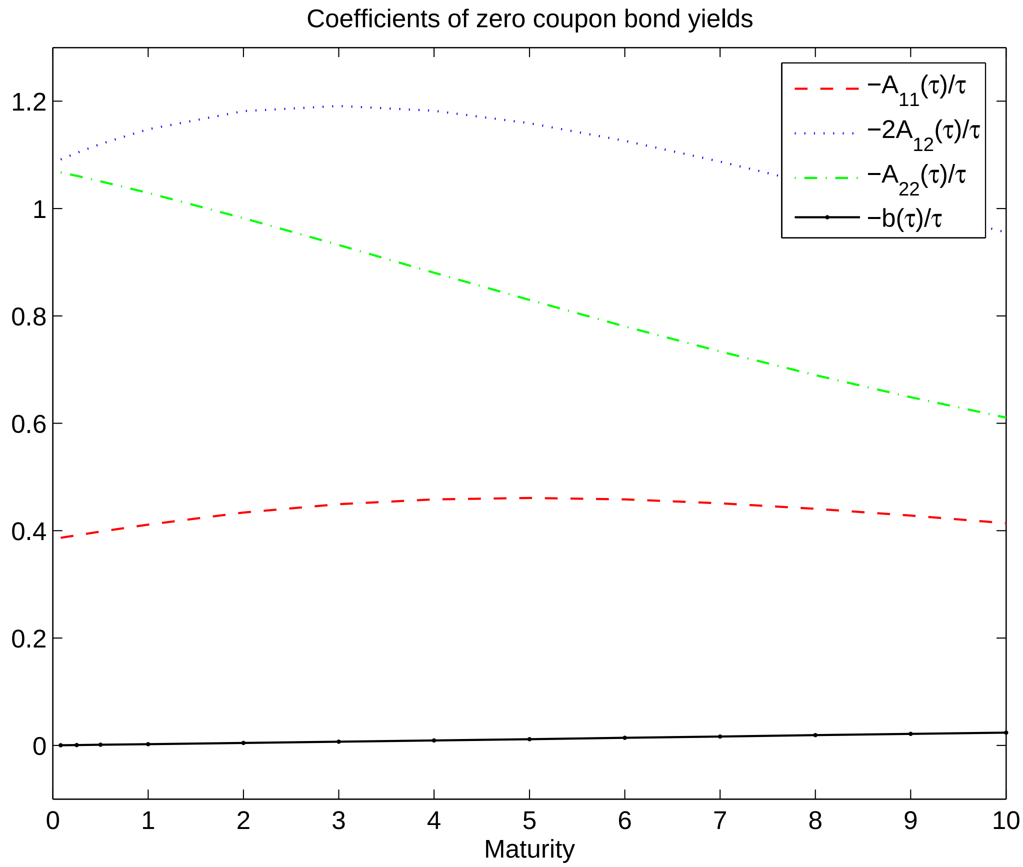

Figure 4: **Factor loadings.** The figure displays the scaled elements of the $A(\tau)$ matrix in equation (22), $-A_{11}(\tau)/\tau$, $-2A_{12}(\tau)/\tau$, and $-A_{22}(\tau)/\tau$, which represent the loadings on Σ_{11}, Σ_{12} and Σ_{22}, respectively; $b(\tau)$ is a constant term.

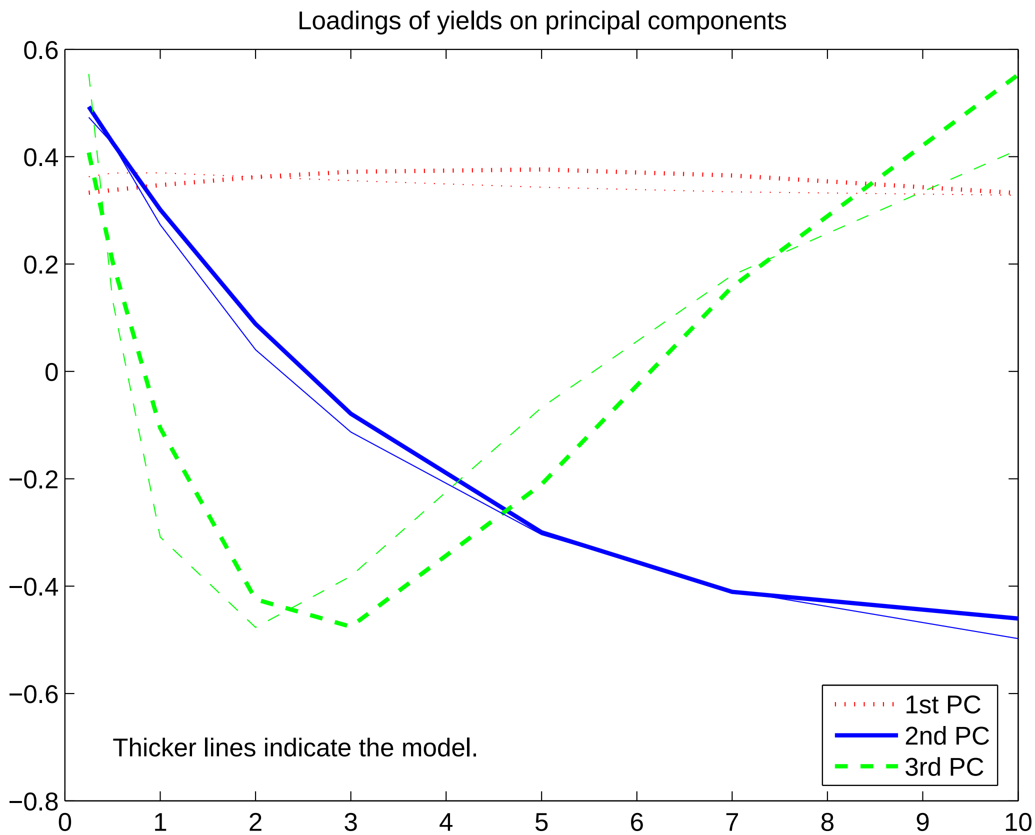

Figure 5: **Loadings of yields on principal components: model versus data.** The covariance matrix of yields is decomposed as $U\Lambda U'$, where U is the matrix of eigenvectors normalized to have unit lengths, and Λ is the diagonal matrix of associated eigenvalues. The figure shows columns (factor loadings) of U associated with the three largest eigenvalues. Thicker lines indicate loadings of yields obtained from the model; finer lines are loadings obtained from the sample of US yields, 1952:01–2005:06.

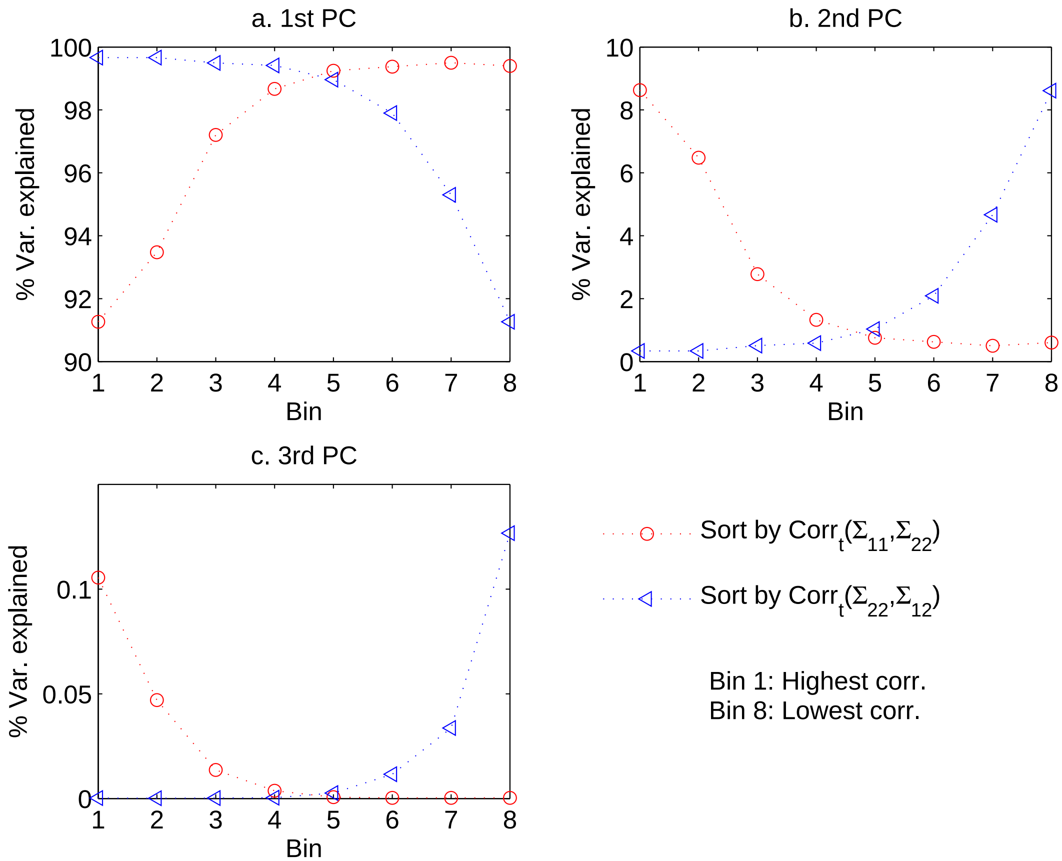

Figure 6: **Variance explained by principal components, conditional on factor correlation.** The figure shows the portions of yield variance explained by each principal component. The principal components are computed for yields grouped by the level of instantaneous correlation of factors. We form eight correlation bins, and number them from 1 to 8. The bins are in descending order: $(1, .9)$, $(.9, .8)$, $(.8, .5)$, $(.5, 0)$, $(0, -.5)$, $(-.5, -.8)$, $(-.8, -.9)$, $(-.9, -1)$. Two sort criteria are used: circles denote sorts by the level of $Corr_t(\Sigma_{11}, \Sigma_{22})$, triangles indicate sorts by the level of $Corr_t(\Sigma_{22}, \Sigma_{12})$.

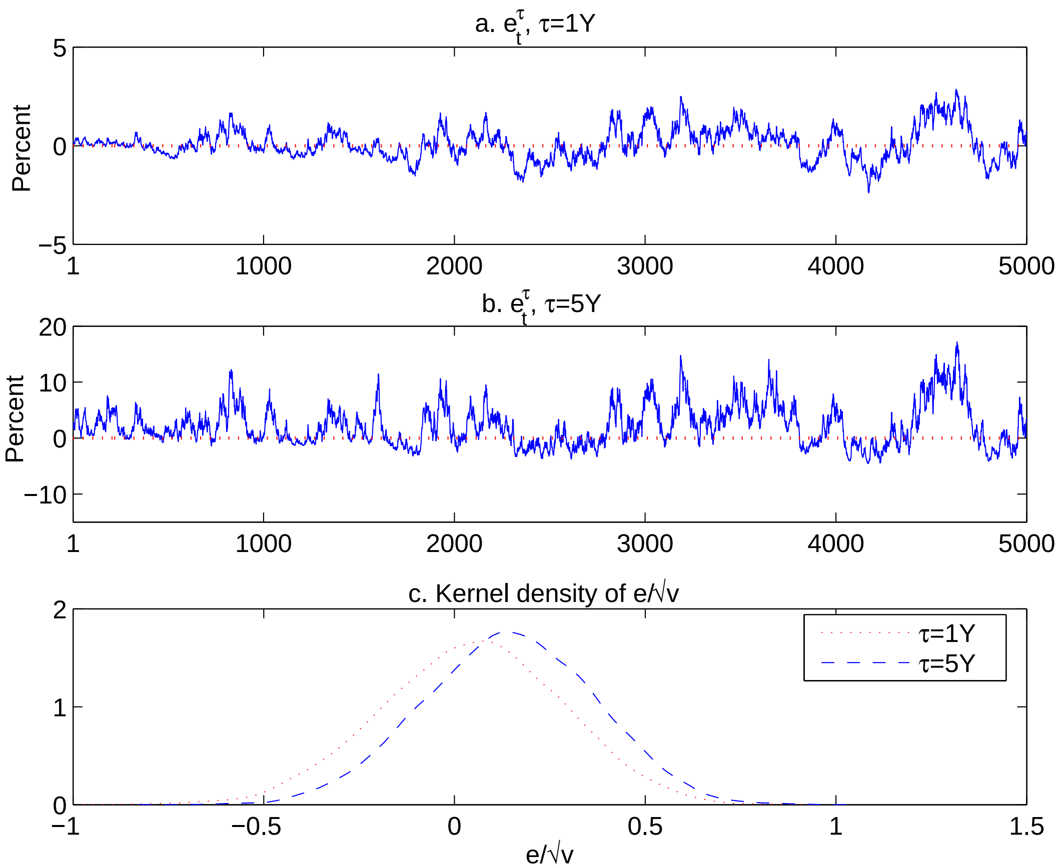

Figure 7: **Properties of expected excess bond returns.** Panels a and b display the instantaneous excess returns on a one-year and 5-year bond, respectively. The excess returns are computed as $e_t^\tau = Tr\left[\left(A(\tau)Q' + QA(\tau)\right)\Sigma_t\right]$. Panel c plots kernel density of the ratio of the instantaneous expected excess returns to their volatility, $e^\tau/\sqrt{v^\tau}$. For readability, we only present the initial 5000 realizations from the simulation. The kernel densities are obtained using the whole simulated sample of 36000 observations.

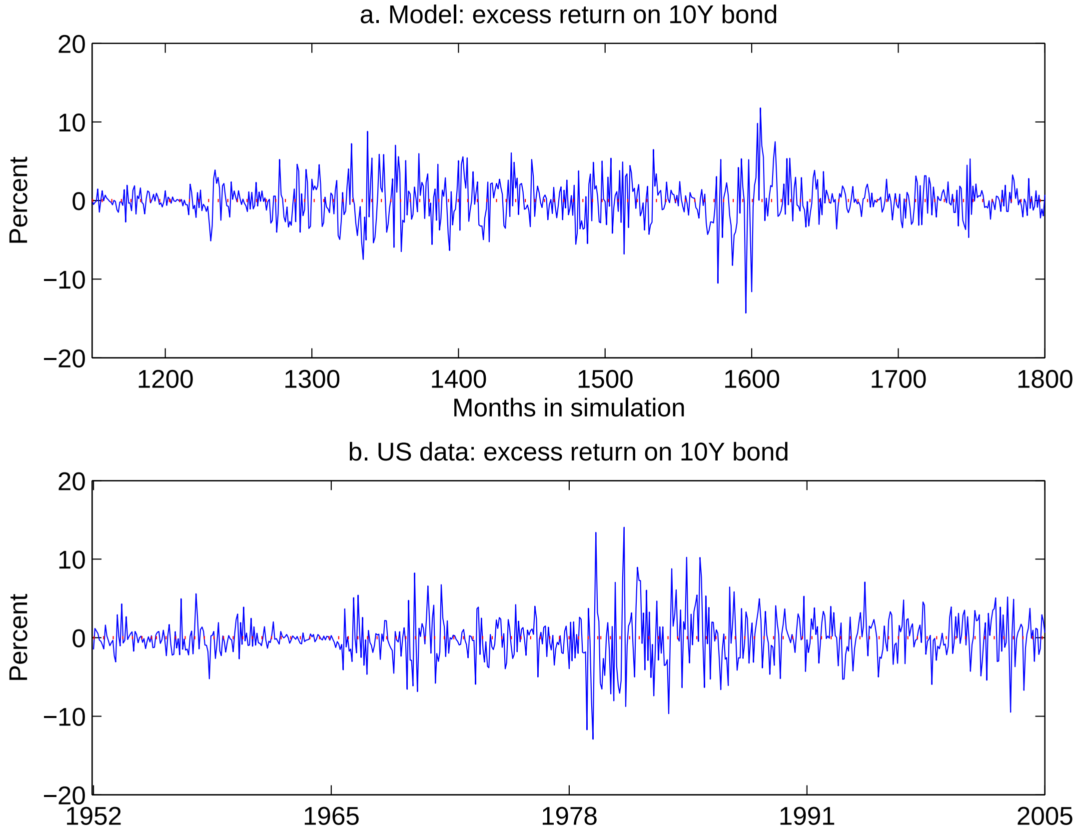

Figure 8: **Properties of realized excess bond returns.** Panel *a* displays 650 realized monthly excess returns on a 10-year bond in a subsample of simulated data. Panel *b* displays the realized monthly excess returns on the 10-year US Treasury bond. In both panels, the realized excess return is computed as the return on the 10-year bond over the 3-month bond.

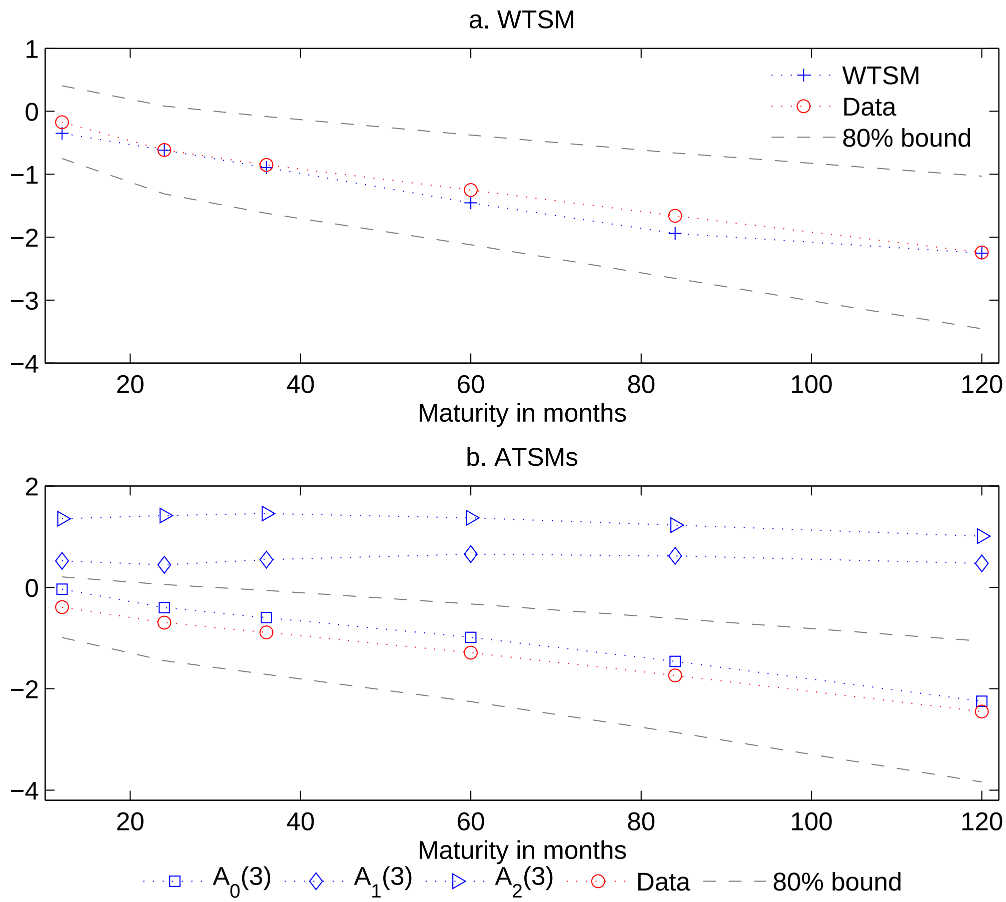

Figure 9: **Campbell-Shiller regression coefficients.** The figure plots as a function of maturity the parameters of Campbell and Shiller (1991) regression in equation (29). Panel a displays the coefficients obtained from the US yields in the sample period 1952:01–2005:06 and the population coefficient implied by the Wishart factor model. Panel b performs the same exercise for the preferred affine models estimated by Duffee (2002), and compares them to the empirical coefficients for the relevant sample period 1952:01–1994:12. The dashed lines plot the 80 percent confidence bounds for the historical estimates based on the Newey-West covariance matrix. Further remarks from Table III apply.

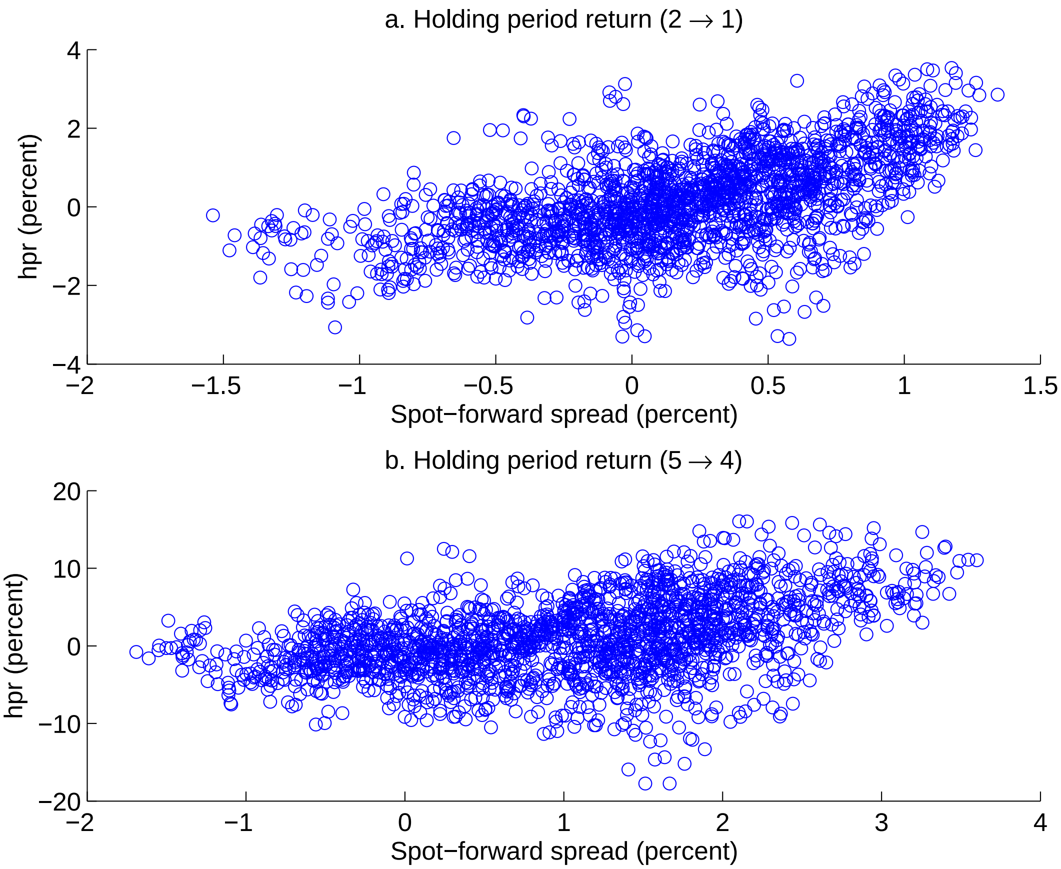

Figure 10: **Predictability of excess bond returns.** The figure presents scatter plots of excess bond returns against the spot-forward spread; see regression equation (31). The forward-spot spread in panel a is $(f_t^{2\rightarrow 1} - y_t^{1Y}) \times 100$, and in panel b: $(f_t^{5\rightarrow 4} - y_t^{1Y}) \times 100$. The plots use 2000 data points simulated from the model.

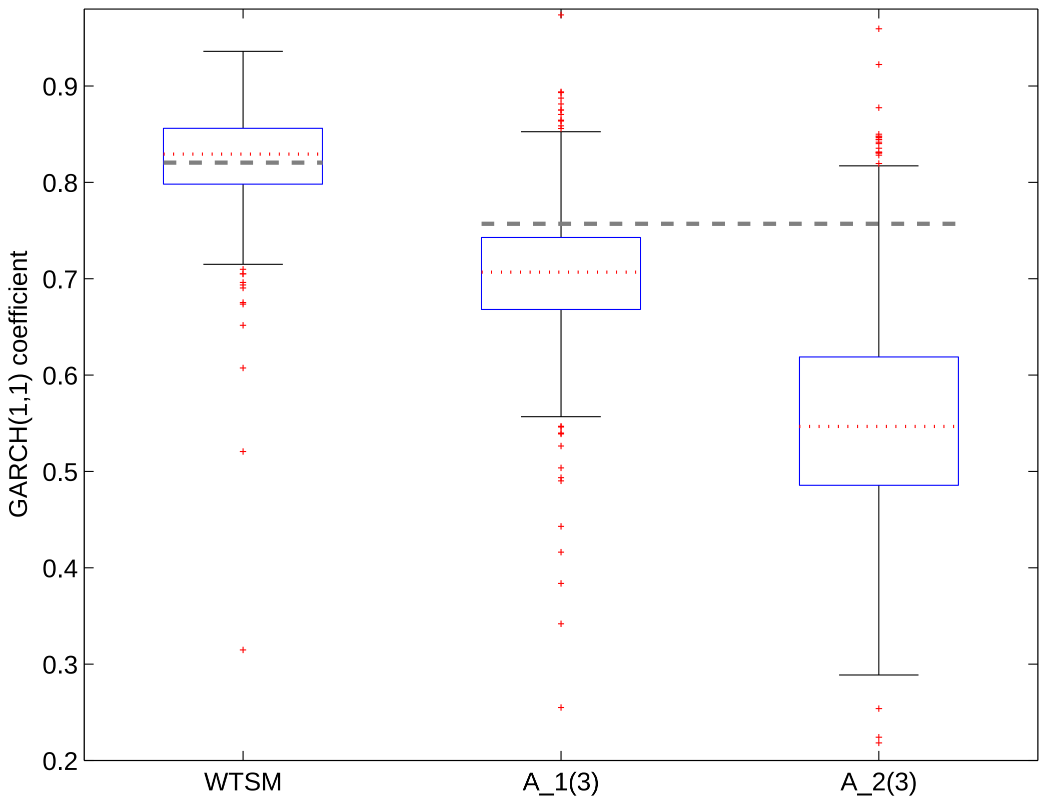

Figure 11: **GARCH(1,1) coefficients.** The figure displays box plots of 1000 GARCH coefficients for the 5-year yield. The yield is simulated from three term structure models: the Wishart factor model (WTSM), and the affine $A_1(3)$ and $A_2(3)$ models. The simulation of ATSMs is based on the estimates from Duffee (2002). The dotted lines are the median estimates; the dashed lines are the historical GARCH estimates in two different sample periods: 1952:01–2005:06 for the Wishart factor model, and 1952:01–1994:12 for the affine models. The letter sample period coincides with the sample used by Duffee (2002) for estimation. Consequently, the length of each simulated path is 54 and 43 years of monthly observations respectively for the Wishart factor model and both ATSMs.

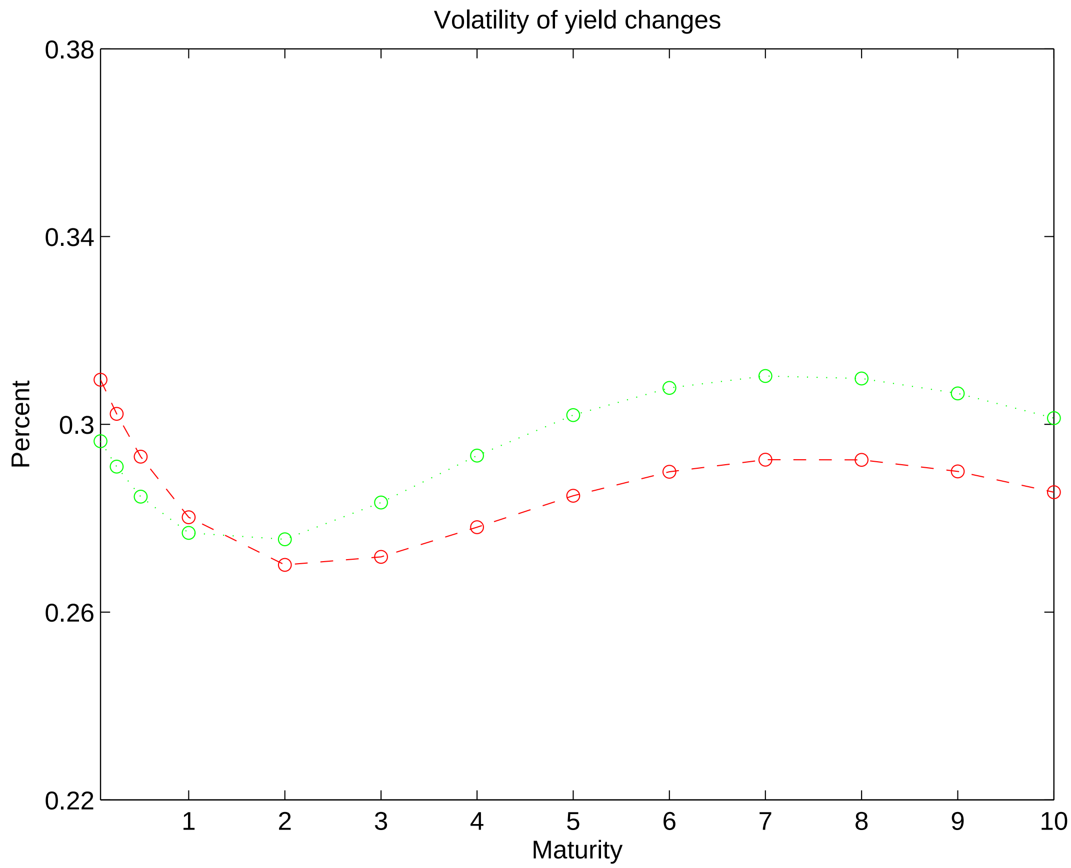

Figure 12: **Unconditional second moments of yields.** The figure plots the term structure of unconditional standard deviations of yield changes computed in two simulated subsamples, including 650 realizations each.

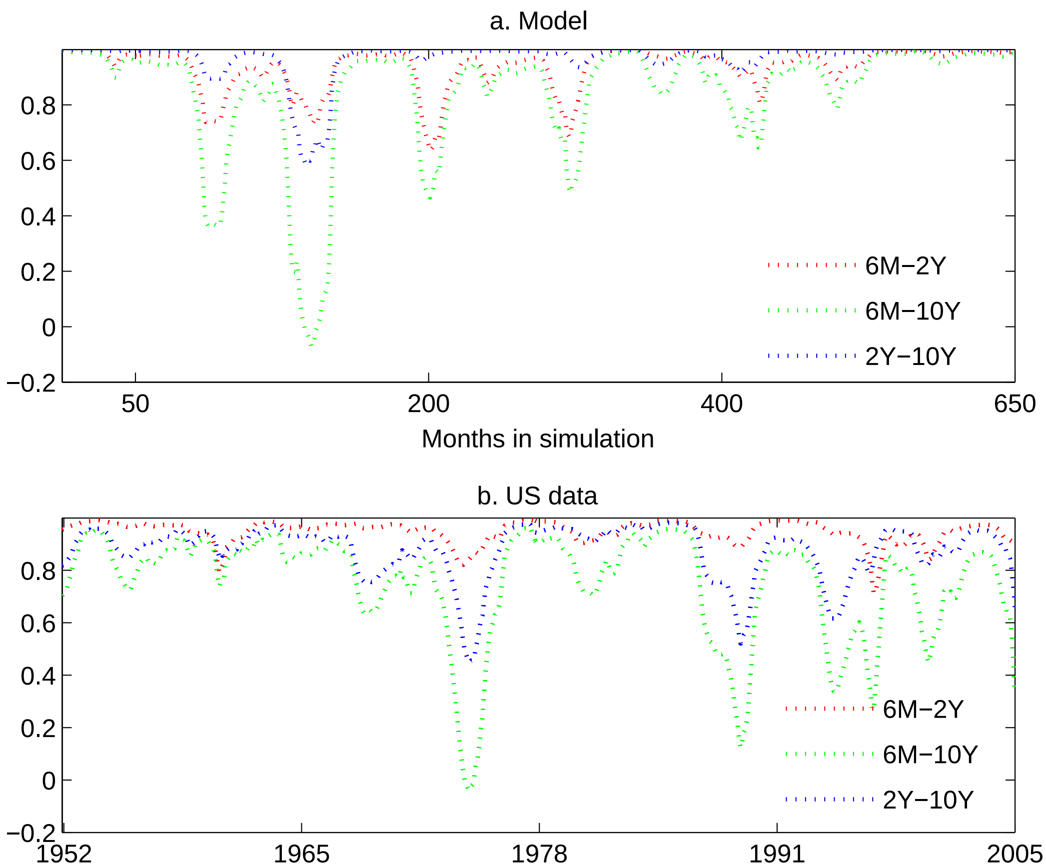

Figure 13: **Correlations of yields.** The figure displays the rolling window correlations between the levels of yields with different maturity. We use the rolling window of 48 monthly observations, and present three correlation pairs. Panel a presents the values from the calibrated model; panel b depicts the behavior of the data.

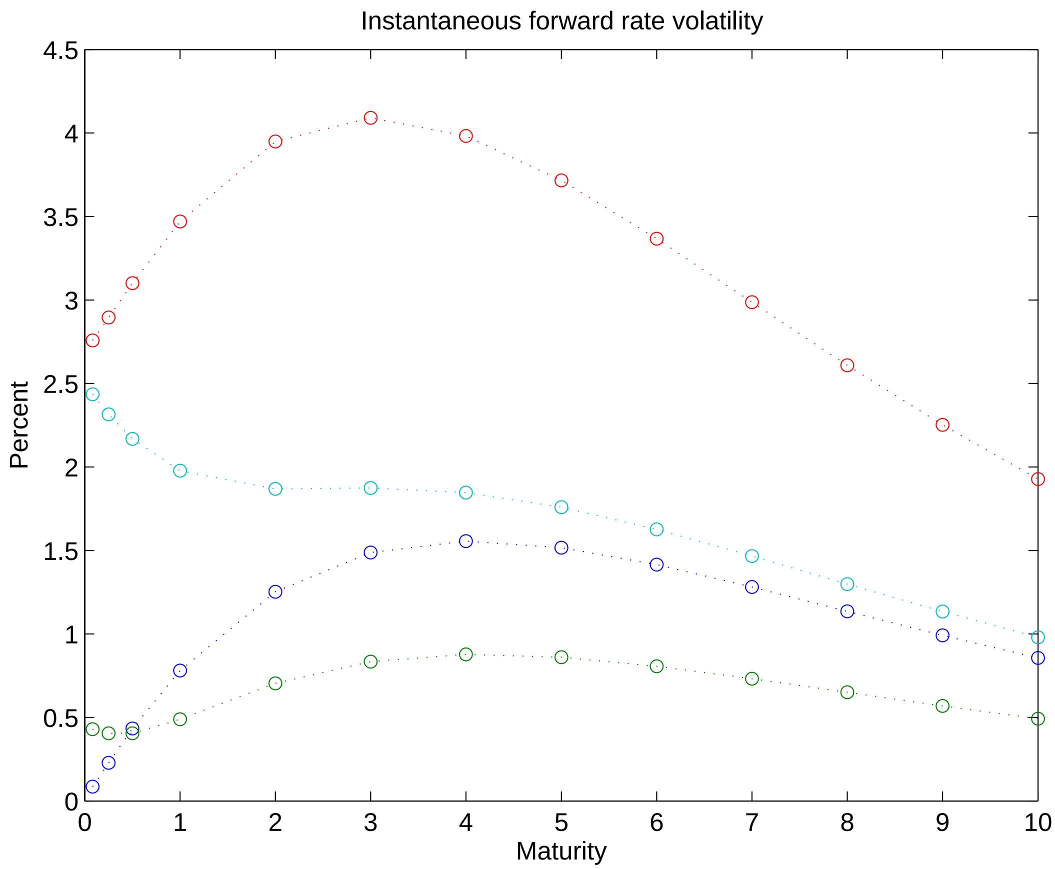

Figure 14: **Forward interest rate volatilities in the Wishart factor model.** The figure presents the term structure of the instantaneous volatility of the (instantaneous) forward rate given in equation (26). The instantaneous volatility is computed as $v^f(t,\tau) = 4Tr[\frac{dA(\tau)}{d\tau}\Sigma_t \frac{dA(\tau)}{d\tau} Q'Q]$, where $\frac{dA(\tau)}{d\tau}$ is given in closed form in equation (21).

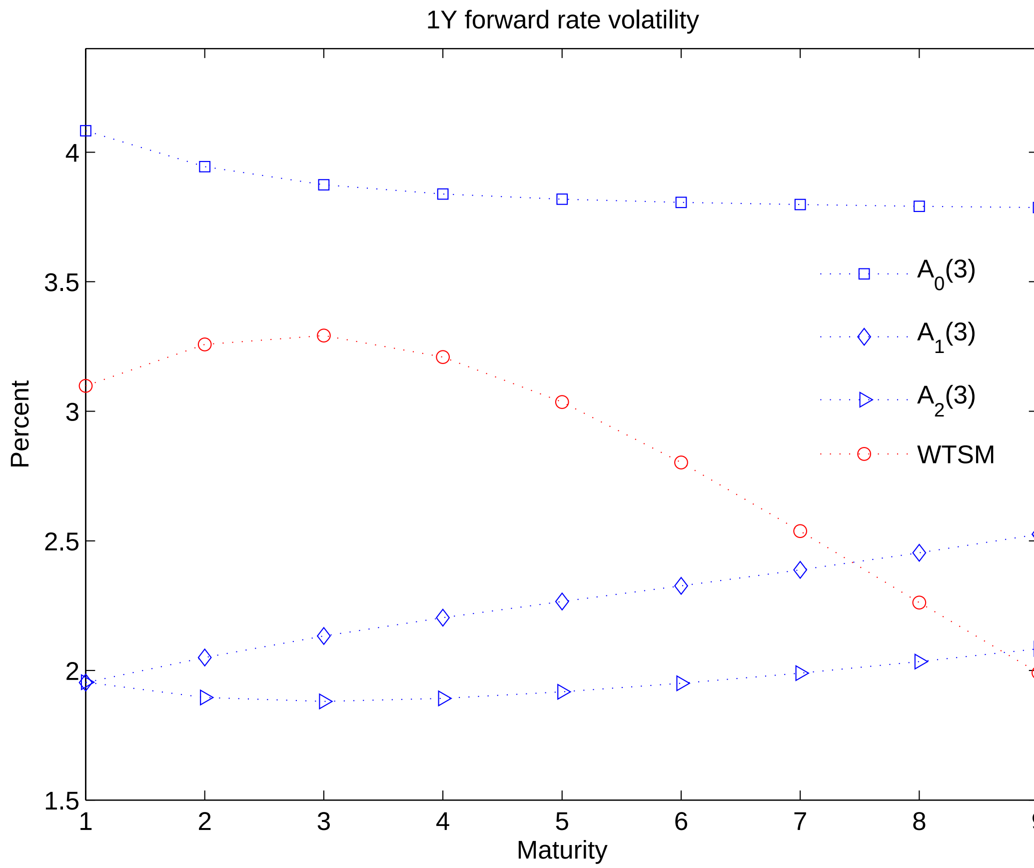

Figure 15: **Comparison of forward interest rate volatilities implied by the Wishart factor model and ATSMs.** The figure shows the unconditional volatility of the one-year forward rate obtained from the simulated sample of 36000 monthly observations. The models considered are the Wishart factor model (WTSM), and three preferred affine models of Duffee (2002): $A_0(3)$, $A_1(3)$ and $A_2(3)$. The maturity indicates the expiry of the forward contract, e.g. at maturity τ the forward rate is $f_t^{\tau \to \tau+1}$.