Correlation Risk and the Term Structure of Interest Rates

ANDREA BURASCHI, ANNA CIESLAK and FABIO TROJANI*

ABSTRACT

We study a completely affine term structure model with matrix-valued state variables, which grants a new element of flexibility in the joint modeling of stochastic volatilities and correlations of factors, and in the specification of the market prices of risk. The most parsimonious three-factor setup provides a unified answer across a set of empirical regularities such as (*i*) the predictability of excess bond returns, (*ii*) the persistence of conditional volatilities and correlations of yields, or (*iii*) the hump in the term structure of forward rate volatilities and implied volatilities of caps. By a straightforward enlargement of the state space, the model can also successfully accommodate more complex features of the bond market such as the Cochrane-Piazzesi (2005) return-forecasting factor, or the unspanned dynamics of interest rate derivatives.

THIS PAPER STUDIES A COMPLETELY AFFINE CONTINUOUS-TIME yield curve model, in which matrix-valued state variables generate stochastic correlations of risk factors. By providing additional flexibility in modeling first and second moments of yields, this setting extends the standard affine class of term structure models.[1] We document its success in reconciling a range of empirical patterns such as (i) the violations of the expectations hypothesis (e.g., the Campbell and Shiller (1991) and the Cochrane and Piazzesi (2005) slope coefficients), (ii) the persistence in the conditional volatility of yields, (iii) the hump-shaped forward rate and cap implied volatility curves, (iv) the time-varying correlations among different segments of the term structure, and (v) the appearance of unspanned factors in the interest rate derivatives market.

The complexity of modeling jointly the first and second moments of yields is well illustrated by the recent bond market behavior itself. One commonly cited example comes from a comparison of two monetary tightenings by the US Fed which occurred a decade apart, 1994/95 and 2004/05. In both periods, the reaction at the long end of the curve provided for a surprise, yet for opposing reasons: The first interest rate hike puzzled many observers because long term yields rose, the latter did so because they fell. This wedge in comovements between short and long yields in the two periods has undermined the fit of some otherwise successful models, and spurred considerable research in recent years. In search for an economic explanation of these facts, various studies have argued about the differences in risk compensation and in the amount of risk, manifest in shifting bond market volatility.[2] Figure 1 summarizes this contrasting bond market behavior. The consensus view suggests that while premia on long bonds were high and accompanied by an increased yield volatility in the early episode, they turned low or even negative in the more recent and calmer period. This brief account summarizes quite well a more general mechanism present in term structure dynamics. It also reflects the key demand for a yield curve model to combine flexible specifications of bond excess returns with a rich time-variation in volatilities and correlations of risk factors (and thus yields).

More generally, the empirical literature has highlighted a number of 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, the spot-forward spread, or linear combination of forward rates (Fama and Bliss, 1987; Campbell and Shiller, 1991; Cochrane and Piazzesi, 2005). Third, in some periods the term structure of unconditional volatilities peaks for maturities of 2–3 years, and is moderately downward sloping for longer yields (Piazzesi, 2001). Finally, the term structure of conditional second moments of yields is highly time-varying, and exhibits a multi-factor structure (e.g. Andersen and Benzoni, 2008).

A vast literature has explored the ability of term structure models to account for the time-series and cross-sectional properties of bond market dynamics. The research has focused on analytically tractable models that

---

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

[2] See Campbell (1995), Backus and Wright (2007) and Cochrane and Piazzesi (2008) for an interesting account of these tightening events. Rudebusch, Swanson, and Wu (2006) document the failure of two established macro-finance Gaussian models in fitting the 2004/05 "conundrum" period. They ascribe a large portion of the conundrum to declines in long-term bond volatility, and contrary to common argument, find little or no role for foreign official purchases of US Treasuries in explaining the puzzle.

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

ensure economically meaningful behavior of yields and bond returns. This combination of theoretical and empirical requirements poses a significant challenge. 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. In order to have both correlations and time-varying volatilities, positive square-root processes need to be melded together with conditionally Gaussian ones. The inclusion of Gaussian dynamics allows for an unconstrained sign of factor correlations, but instead sacrifices some ability to accommodate stochastic yield volatilities.

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 assumed the market price of risk to be a constant multiple of factor volatility.[4] Summarizing the empirical failure of this specification, Duffee (2002) proposed an "essentially affine" extension, in which the market prices of risk are inversely proportional to factor volatility in the case of unrestricted (conditionally Gaussian) factors and have a switching sign, but preserve the completely affine form for the square-root (volatility) factors. More recently, Cheridito, Filipovic, and Kimmel (2007) suggested a new "extended affine" generalization making the market price of risk of all factors—both Gaussian and square-root—inversely proportional to their volatilities.

The levers obtained by augmenting the market price of risk are twofold. First, by incorporating Gaussian factors, the expected excess bond returns can switch sign. Second, 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) stress the key role of correlated factors for the model's ability to forecast yield changes and excess bond returns. Yet, despite a good fit to some features of the data, the essentially affine models face limitations evident in the trade-off between stochastic volatilities and correlations of factors, and thus allow to match the first or the second conditional moments of yields, but not both at a time.[5] While the extended affine market price of risk helps mitigate some of these tensions, recent research has exposed its restrictions in terms of matching higher order moments of yields (Feldhütter, 2007). However useful for fitting the data, the extensions of the market price of risk are not innocuous from the perspective of both theory and applications. In an equilibrium setting, the market price of risk reflects investor risk attitudes. More complex formulations are therefore equivalent to increasingly complex investor preferences which can be difficult to justify by standard arguments.[6] More importantly, gains in empirical

---

[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]See e.g., Brandt and Chapman (2002); Bansal and Zhou (2002); Dai and Singleton (2003).

[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 Dai (2003), and Buraschi and Jiltsov (2006). The extended affine family of Cheridito, Filipovic, and Kimmel (2007), instead, seems difficult to reconcile with the standard expected utility maximization, because it entails that agents become more concerned about risk precisely when it goes away. While potentially inconsistent with standard preferences, such behavior can arise quite naturally in an economy with ambiguity aversion as demonstrated by Gagliardini, Porchia, and Trojani (2007). In their non-affine term structure model, the contribution of the ambiguity premium to excess bond returns dominates the standard risk premia precisely when the aggregate risk in the economy is low.

performance implied by a richer market price of risk come about by adding new parameters to already large models, many of which turn out difficult to identify from yield curve data alone.

This paper takes a different approach. Rather than focussing on market prices of risk in the first place, we start with a single non-standard assumption about the state space. Specifically, we assume risk factors in the economy to follow a continuous-time affine process of positive definite matrices whose transition probability is Wishart. By construction, such factors allow correlations and volatilities both to be stochastic. This approach builds on the work of Gourieroux and Sufana (2003) who propose the Wishart process as a convenient theoretical framework to represent yield factors. We start from their insight and take the first attempt to investigate the properties of a continuous-time Wishart yield curve model. By exploiting the flexibility of the posited state space, we do not need to introduce a highly parametrized market price of risk, and thus are able to resurrect the parsimonious completely affine class. Studying the properties of the term structure, bond returns and standard interest rate derivatives, we document the ability of this setting to match features of the data along several important dimensions.

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

Second, the model does not face the volatility-correlation trade-off pertaining to ATSMs. Both under the physical and risk-neutral measures, it accommodates switching sign of conditional and unconditional correlations among state variables. 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 framework allows for a new aspect of flexibility also from the derivatives pricing perspective. We find that it produces a hump-shaped term structure of forward interest rate and cap implied volatilities. The direct presence of stochastic correlations in the state dynamics, both under the risk-neutral and physical measure, does not require to impose them with an additional independent variable, as is typically done in the literature. These features are key for the consistent pricing of caps relative to swaptions.

Finally, even though many regularities found in the data can be already reproduced with the most parsimonious three-factor specification, the Wishart setting affords a direct extension to a larger dimension at no harm to analytical solutions. With the state space enlarged to six variables, the framework additionally supports the single return-forecasting factor of Cochrane and Piazzesi (2005), is able to incorporate realistic dynamic correlations of yield changes, and the comovement of yield volatilities. Parallel to providing a comprehensive description of the spot yield curve, the enlarged model also allows for the appearance of unspanned factors in interest rate derivatives. The flexibility gained with extra factors does not seriously impair the parsimony. In terms of the number of parameters, the six-factor setting still looks more favorable than the standard essentially or extended affine three-factor specifications.

In our analysis, we exploit several results on the non-central Wishart distribution. Even though the statistical properties of this distribution have been extensively covered in the early multivariate statistics literature as summarized for instance by Muirhead (1982), the extension involving time dependence has only been explored of late. The Wishart process was first proposed by Bru (1991), and more recently studied by Donati-Martin, Doumerc, Matsumoto, and Yor (2004) and Gourieroux (2006) in continuous time and Gourieroux, Jasiak, and Sufana (2004) in discrete time. This work has provided a foundation for such applications to finance as derivatives pricing or portfolio choice with stochastically correlated assets (see e.g., Gourieroux and Sufana, 2004; Buraschi, Porchia, and Trojani, 2006; da Fonseca, Grasselli, and Tebaldi, 2006).

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 (e.g., 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 these features. Consequentially, 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 multi-factor stochastic structure of second moments, likely driven by factors unrelated to bond prices themselves. In this vein, Collin-Dufresne and Goldstein (2001) and Han (2007), among others, develop HJM-type of models that allow to bypass the trade-off between stochastic volatilities and correlations in the affine class.

The plan of the paper is as follows. Section I introduces our completely affine market price of risk specification, derives the short interest rate, provides the solution for the term structure and discusses its asset pricing implications. Section II presents the empirical approach, derives the moments of the Wishart process and yields, and investigates the properties of factors based on the model fitted to the US Treasury zero curve. In Section III, the most simple three-factor specification is scrutinized for its consistency with the stylized facts of yield curve literature. Section IV extends the discussion to a six-factor framework and highlights the payoffs from the enlarged model. Section V concludes. All proofs and figures are in Appendices.

I. The Economy

In analogy to the standard completely affine models, we motivate a parsimonious form of the market price of risk within the general Cox, Ingersoll, and Ross (1985a) framework.

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

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

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

We depart from the affine literature only in the specification of risk factors driving the production technology dynamics. These are assumed to follow an affine continuous-time process of symmetric positive definite matrices.

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 is a $n \times n$ symmetric positive definite matrix of state variables, and $\sqrt{\cdot}$ denotes the square root in the matrix sense; D is a symmetric $n \times n$ matrix of deterministic coefficients. Tr indicates the trace operator.

In the above diffusion, the drift $\mu_Y = Tr(D\Sigma_t)$ and the quadratic variation $\sigma_Y^2 = Tr(\Sigma_t)$ are of affine form. Therefore, the univariate process of returns $\frac{dY_t}{Y_t}$ is affine in $\frac{n(n+1)}{2}$ distinct elements of the symmetric Σ_t matrix.

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

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

where Ω, M, Q, Ω invertible, are square $n \times n$ matrices. Throughout, we assume that $\Omega\Omega' = kQ'Q$ with integer degrees of freedom $k > n - 1$, ensuring the Σ_t matrix is of full rank.

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 outer products of k independent copies of an Ornstein-Uhlenbeck process, each with dimension n. A number of qualities make the process particularly suitable for modeling multivariate sources of risk in finance (Gourieroux, Jasiak, and Sufana, 2004; Gourieroux, 2006). First, the conditional Laplace of the Wishart and the integrated Wishart process are exponential affine in Σ_t. As such, these processes are affine in the sense of Duffie, Filipovic, and Schachermayer (2003). This feature gives rise to convenient closed-form solutions to prices of bonds or options, and simplifies the econometric inference. Second, the Wishart process lives in the space of positive definite matrices. Thus, it is distinct from the affine processes defined on $\mathbb{R}_+^m \times \mathbb{R}^{n-m}$. The restriction $\Omega\Omega' \gg Q'Q$ guarantees that Σ_t is positive semi-definite. By additionally assuming that $k > n - 1$, it is ensured that the state matrix is of full rank. Thus, the diagonal elements of Σ_t (and $\sqrt{\Sigma_t}$) are always positive, but the out-of-diagonal elements can take on negative values. Moreover, if $\Omega\Omega' = kQ'Q$ for some $k > n - 1$ (not necessarily an integer), then Σ_t follows the Wishart distribution (Muirhead, 1982, p. 443). Third, the elements of the Wishart matrix feature rich dependence structure, and their conditional and unconditional correlations are unrestricted in sign (see Section II and Result 1 in Appendix D for details). This feature implies a term structure setting with weak restrictions on the stochastic cross-sectional dependence between yields. Finally, 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.

I.A. The Short Interest Rate and the Market Price of Risk

Since we are primarily interested in exploring the term structure implications of the state dynamics (3), we maintain the simple completely affine form of the market price of risk implied by our assumptions. Given Assumption 1, the optimal consumption plan of the representative agent is $C_t^* = \rho Y_t$, and the short interest rate can be computed by the equation $r_t = -\mathbb{E}_t \left[\frac{du'(C_t^*)}{u'(C_t^*)} \right]$, with $u(C) = \ln(C)$. The $n \times n$ matrix of market prices of risk Λ_t follows as the unique solution of the equation $\mu_Y - r_t = Tr \left[\sqrt{\Sigma_t} \Lambda_t \right]$. A standard application of Itô's Lemma implies the following expressions for r_t and Λ_t.

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

$$r_t = Tr\left[(D - I_n) \Sigma_t \right], \tag{4}$$

where I_n is an $n \times n$ identity matrix. The market price of risk equals the square root of the matrix of Wishart factors:

$$\Lambda_t = \sqrt{\Sigma_t}. \tag{5}$$

The short interest rate in (4) can be equivalently 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, r_t is a linear combination of the Wishart factors, which are conditionally and unconditionally dependent, with correlations being unrestricted in sign.[7] While the short rate comprises both positive factors on the diagonal of Σ_t and out-of-diagonal factors that can take both signs, its positivity is ensured with a straightforward restriction requiring the $D - I_n$ matrix to be positive definite (see Result 4 in Appendix D). In Section I.D below, we prove that the same condition is necessary and sufficient to impose the positivity on the whole term structure of interest rates. The instantaneous variance V_t of changes in the short rate is obtained by applying Ito's Lemma to equation (4), 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 short rate and their variance can be found 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:

$$\frac{1}{dt} V_t = 4 Tr \left[\Sigma_t (D - I_n) Q' Q (D - I_n) \right]. \tag{6}$$

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

---

[7]By restricting D in equation (2) to be a diagonal 2×2 matrix, this expression for the short interest rate resembles the two-factor model of **Longstaff and Schwartz** (1992). Even in this special case, however, the model has a richer structure, as the variance of the changes in the interest rate is driven by all three factors ($\Sigma_{11}, \Sigma_{12}, \Sigma_{22}$), which are pairwise correlated.

$$\frac{1}{dt}CV_t = 4Tr\left[\Sigma_t(D - I_n)Q'Q(D - I_n)Q'Q(D - I_n)\right]. \tag{7}$$

Expressions (6) and (7) show that both quantities preserve the affine form in the elements of Σ_t. Appendix A.1 provides the derivation.

The completely affine form of the market price of risk (5) fosters parsimony and tractability under the risk neutral and the physical measure. In order to fit the data, the Wishart model exploits the flexibility of the underlying risk factors, rather than a rich parametrization of Λ_t. Given the difficulties in estimating the market price of risk parameters frequently reported in the affine literature, the plain specification adopted here appears to be an asset for empirical applications.

A well-recognized critique of the completely affine models is their inability to match the empirical properties of yields due to (i) the sign restriction on the market price of risk, and (ii) its one-to-one link with the volatility of factors. Despite analogous derivation, the market price of risk in (5) is 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. Intuitively, the signs of the elements of Λ_t reflect different perceptions of volatility and co-volatility risks by investors.

Remark 1 (Extended Wishart specification of the market price of risk)**.** In a reduced-form Wishart factor model, one could easily construct a richer market price of risk à la Cheridito, Filipovic, and Kimmel (2007):

$$\Lambda_t = \Sigma_t^{-1/2}\Lambda_0 + \Sigma_t^{1/2}\Lambda_1,$$

for some $n \times n$ constant matrices Λ_0 and Λ_1. This specification preserves the affine property of the process (3) under the risk neutral measure, but modifies both the constant and the mean reversion terms in its drift. In order to maintain the Wishart distribution under the risk neutral measure, one can set $\Lambda_0 = vQ'$ for $v \in \mathbb{R}_+$ such that $k - 2v > n - 1$.[8] While the extended affine market price of risk adds significant flexibility to the standard affine class, we show such an extension is not required to match the data, given the flexibility of the Wishart process itself. $\square$

I.B. The Term Structure of Interest Rates

Given expression (4) for the short rate, the price at time t of a zero-coupon bond maturing at time T is:

$$P(t,T) = E_t^*\left(e^{-\int_t^T r_s ds}\right) = E_t^*\left(e^{-Tr\left[(D - I_n)\int_t^T \Sigma_s ds\right]}\right), \tag{8}$$

where $E_t^*(\cdot)$ denotes the conditional expectation under the risk neutral measure. To move from physical dynamics of the Wishart state in equation (3) to risk neutral dynamics, we can apply the standard change

---

[8]For practical applications, this form of Λ_t requires a careful analysis of conditions ensuring no arbitrage. In the classic affine case, Cheridito, Filipovic, and Kimmel (2007) show that the extended affine market price of risk does not admit arbitrage provided that under both measures the state variables cannot achieve their boundary values. In the Wishart factor setting, this condition holds true if the Q matrix is invertible, and $k - 2v > 1$.

of drift technique. Due to the completely affine market price of risk specification, the risk neutral drift adjustment of $d\Sigma_t$ takes on a very simple form: $\Phi_\Sigma = \Sigma Q + Q'\Sigma$.[9]

Remark 2. This form of the drift adjustment is straightforward to justify in a reduced form setting by defining the Radon-Nikodym derivative for the transformation from the physical measure $\tilde{\mathbb{Q}}$ to the risk neutral measure $\mathbb{Q}^*$:

$$\frac{d\mathbb{Q}^*}{d\tilde{\mathbb{Q}}}\big|_{\mathcal{F}_t} := e^{Tr\left[-\int_0^t \Lambda'_u dB_u - \frac{1}{2}\int_0^t \Lambda'_u \Lambda_u du\right]}, \tag{9}$$

where $\Lambda_t = \sqrt{\Sigma_t}$ is our completely affine market price of risk. It follows from the Girsanov's theorem that the process $B_t^* = B_t + \int_0^t \Lambda_u du$ is an $n \times n$ matrix of standard Brownian motions under $\mathbb{Q}^*$. Therefore, the SDE

$$d\Sigma_t = \left(\Omega\Omega' + (M - Q')\Sigma_t + \Sigma_t(M' - Q)\right)dt + \sqrt{\Sigma_t}dB_t^* Q + Q' dB_t^{*\prime}\sqrt{\Sigma_t}$$

represents the risk neutral dynamics of the Wishart process. $\square$

Given the change of drift Φ_Σ, the term structure pricing equation follows by applying the discounted Feynman-Kač formula to expectation (8).

Proposition 2 (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:[10]*

$$Tr\{[\Omega\Omega' + (M - Q')\Sigma + \Sigma(M' - Q)]\mathcal{R}F\} + 2Tr[\Sigma\mathcal{R}(Q'Q\mathcal{R}F)] + \frac{\partial F}{\partial t} - Tr[(D - I_n)\Sigma]F = 0, \tag{10}$$

with the boundary condition:

$$F(\Sigma, T, T) = \Psi(\Sigma, T), \tag{11}$$

where $\mathcal{R}$ is a matrix differential operators with $ij-th$ component equal to $\frac{\partial}{\partial\Sigma_{ij}}$.

In expression (8), we recognize the Laplace transform of the integrated Wishart process. By the affine property of the process, the solution to the PDE in equation (10) is an exponentially affine function of risk factors. The next proposition states this result in terms of prices of zero-coupon bonds.

Proposition 3 (Bond prices)**.** *If there exists a unique continuous solution to the Wishart equation, then under the model dynamics (2)–(3), the price at time t of a zero-coupon bond P with maturity $T > t$, is of the exponentially affine form:*

$$P(\Sigma, t, T) = e^{b(t,T) + Tr[A(t,T)\Sigma]}, \tag{12}$$

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

$$-\frac{db(t, T)}{dt} = Tr[\Omega\Omega' A(t, T)] \tag{13}$$

$$-\frac{dA(t, T)}{dt} = A(t, T)(M - Q') + (M' - Q)A(t, T) + 2A(t, T)Q'QA(t, T) - (D - I_n), \tag{14}$$

---

[9]As in the scalar case, the elements of Φ_Σ can be interpreted as the expected excess returns on securities constructed so that they perfectly reflect the risk embedded in the corresponding elements of the state matrix Σ.

[10]See Bru (1991) equation (5.12) for the infinitesimal generator of the Wishart process.

with terminal conditions $A(T,T) = 0$ and $b(T,T) = 0$. Letting $\tau = T - t$, the closed-form solution to (14) is given by:

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

where $C_{12}(\tau)$ and $C_{22}(\tau)$ are $n \times n$ blocks of the following matrix exponential:

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

Given the solution for $A(\tau)$, the coefficient $b(\tau)$ is obtained directly by integration:

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

Proof: See Appendix A.2.1. ∎

Under the assumed factor dynamics, bond prices are given in closed form. The solution for $A(\tau)$ implied by the matrix Riccati ODE (14) has been known at least since the work of Radon in 1929. The general (non-symmetric) case has been discussed by Levin (1959). The matrix form of the coefficients facilitates the characterization of the definiteness and monotonicity of the solution, given in the next corollary.

Corollary 1 (Definiteness and monotonicity of the solution). *If matrix $A(\tau)$ is the solution to the matrix Riccati equation (14), then $A(\tau)$ is negative definite and monotonically decreasing for all $\tau \in [0, T]$, i.e. $A(\tau) < 0$ and $A(\tau_2) < A(\tau_1)$ for $\tau_2 > \tau_1$, if and only if $D - I_n$ is positive definite, $D - I_n > 0$.*

Proof: See Appendix A.2.2. ∎

I.C. Yields.

From equation (12), we obtain the yield of a zero-bond maturing in $\tau = T - t$ periods:

$$y_t^\tau = -\frac{1}{\tau}\left[b(\tau) + Tr\left(A(\tau)\Sigma_t \right) \right]. \tag{15}$$

The affine property of yields in the elements of Σ_t allows the relation (15) to be uniquely inverted for the factors. In particular, the symmetric $n \times n$ state matrix can be identified from $\frac{n(n+1)}{2}$ yields (see Appendix B.1 for details on the inversion). Moreover, to ensure positive yields, the $A(\tau)$ matrix needs to be negative definite, for which it is both necessary and sufficient that $D - I_n > 0$. The coexistence of this simple positivity condition together and unrestricted factor correlations is striking in the context of traditional ATSMs, in which a positive short rate is at odds with unrestricted and correlated Gaussian factors. In the taxonomy of Dai and Singleton (2000), $A_N(N)$ is the only subfamily of standard ATSMs that guarantees the positivity of yields. In $A_N(N)$ models, all 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.[11]

---

[11] Feldhütter (2007) shows that the negativity of yields in ATSMs can indeed become a concern. For instance, in the Gaussian $A_0(3)$ model, which offers a maximal flexibility in modeling correlations of factors, the probability of negative 1-year (5-year)

Finally, the modeling of covariances between yields is a nontrivial issue in applications such as 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:

$$\frac{1}{dt}Cov_t\left[dy_t^{\tau_1}, dy_t^{\tau_2}\right] = \frac{4}{\tau_1\tau_2}Tr\left[A(\tau_1)\Sigma_t A(\tau_2)Q'Q\right]. \tag{16}$$

Our model implies that yields of different maturities co-vary in a non-deterministic and multivariate way, evident in the presence of Σ_t in the above equation. Given the indefiniteness of matrix $A(\tau_1)Q'QA(\tau_2)$, the correlations of yields can stochastically change sign over time. The secular decline in long yields during 2004/05 tightening period (see Figure 1) provides one example of an interest rate environment, in which such feature can be important. While new empirical evidence emphasizes the presence of multiple factors in second moments of yields (Andersen and Benzoni, 2008), this property cannot be captured by the univariate volatility $A_1(N)$ specification preferred in several recent applications (e.g., Joslin, 2007; Thompson, 2008). We investigate the consequences of the multifactor volatility structure in the Wishart setting in Section IV.B.

Remark 3 (Relationship to quadratic term structure models). As pointed out by Gourieroux and Sufana (2003), in a special case the yield curve in equation (15) collapses to an n-factor quadratic term structure model (QTSM) in the spirit of Ahn, Dittmar, and Gallant (2002) and Leippold and Wu (2002). With one degree of freedom ($k = 1$) in the factor dynamics (3), the state becomes a singular matrix of rank one, $\Sigma_t = X_t^1 X_t^{1\prime}$, where X_t^1 is an n-dimensional Ornstein-Uhlenbeck (OU) process (see Appendix B for details). Using the fact that $Tr\left[A(\tau)X_t^1 X_t^{1\prime}\right] = X_t^{1\prime}A(\tau)X_t^1$, a purely quadratic expression for yields emerges:

$$y_t^\tau = -\frac{1}{\tau}\left[b(\tau) + X_t^{1\prime}A(\tau)X_t^1\right].$$

Apart from the state degeneracy, we show that this special case places important limitations on the form of yield correlations. Specifically when $k = 1$, correlations between the diagonal elements $\Sigma_{ii,t}, \Sigma_{jj,t}$, $i \neq j$, turn out to be piecewise constant and take plus/minus the same value (see Result 2 in Appendix D). To see this, note that:

$$Corr_t\left(\Sigma_{ii}, \Sigma_{jj}\right) = \frac{\langle\Sigma_{ii}, \Sigma_{jj}\rangle_t}{\sqrt{\langle\Sigma_{ii}\rangle_t}\sqrt{\langle\Sigma_{jj}\rangle_t}} = \text{sgn}(X_{i,t}^1 X_{j,t}^1)\frac{Q^{i\prime}Q^j}{\sqrt{Q^{i\prime}Q^i}\sqrt{Q^{j\prime}Q^j}},$$

where $X_{i,t}^1, X_{j,t}^1$ are scalar OU processes given by the elements of vector X_t^1, Q^i, Q^j denote the i-th and j-th column of the Q matrix, respectively, and sgn is the sign function. By studying the more general non-degenerate case we achieve two main goals.[12] First, the out-of-diagonal elements of the state matrix become non-trivial stand-alone factors with unrestricted sign. This feature permits us to conveniently capture the predictability of bond returns. Second, since factor correlations are allowed to take any value between -1 and 1, they provide additional flexibility in modeling the stochastic second moments of yields. The relevance of this case extends beyond the basic zero yield curve modeling, into areas such as the pricing

---

yields amounts to non-negligible 5.98 (3.91) percent. The history of US nominal yields makes this probability look high in comparison. Indeed, until the 2008 credit crisis negative nominal yields in the US remained very much a theoretical concept.

[12]In fact, when we estimate the model with $k = 1$ its overall fit deteriorates considerably compared to the non-degenerate case.

of correlation-sensitive interest rate derivatives, or term structure models with stochastically correlated observable macro-factors. □

I.D. Excess Bond Returns

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

$$\frac{dP(\Sigma_t, \tau)}{P(\Sigma_t, \tau)} = (r_t + e_t^\tau)dt + Tr\left[\left(\sqrt{\Sigma_t}dB_tQ + Q'dB_t'\sqrt{\Sigma_t}\right)A(\tau)\right], \tag{17}$$

where e_t^τ is the term premium (instantaneous expected excess return) to holding a τ-period bond (see Appendix A.3). The functional form of the expected excess return can be inferred from the fundamental pricing PDE (10), and is represented by a linear combination of the Wishart factors:

$$e_t^\tau = Tr\left[(A(\tau)Q' + QA(\tau))\Sigma\right] = 2Tr\left[\Sigma_t QA(\tau)\right]. \tag{18}$$

The instantaneous variance of bond returns can be written as:

$$v_t^\tau = 4Tr\left[A(\tau)\Sigma_t A(\tau)Q'Q\right]. \tag{19}$$

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.[13] This means that the mean 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, Filipovic, and Kimmel (2007) 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 (18) is indefinite. The set of matrices (and model parameters) satisfying this condition is large, giving us the latitude to capture the combination of low expected excess returns on bonds with their high volatilities. Using estimation results, in Sections III.A and III.B we study the properties of model-implied excess bond returns, and establish the model's ability to fit their historical behavior.

I.E. 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 (12), we have:

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

where $\partial A(\tau)/\partial \tau$ denotes the derivative with respect to the elements of the matrix $A(\tau)$ given in equation (14). The dynamics of the forward rate are given by (see Appendix A.4):

---

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

$$df(t,T) = -\frac{\partial f(t,T)}{\partial \tau}dt - Tr\left(\frac{\partial A(\tau)}{\partial \tau}d\Sigma_t\right). \tag{20}$$

De Jong, Driessen, and Pessler (2004), for example, argue that a humped shape in the volatility term structure of the instantaneous forward rate leads to possible 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 forward interest rate in Section III.D.

I.F. Interest Rate Derivatives

Our framework allows us 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:

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

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:

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

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 the log bond price—due to the affine property in Σ—is available in closed form. Thus, we can readily apply the techniques developed in Heston (1993), Duffie, Pan, and Singleton (2000), and Chacko and Das (2002). The pricing of the option amounts to performing two one-dimensional Fourier inversions under the two forward measures.

Proposition 4 (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:*

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

where $\Psi_t^j(iz)$, $j = S,T$, are characteristic functions of $Tr\left[A(S,T)\Sigma_S\right]$ under the S- and T-forward measure, respectively. Details and closed-form expressions for the characteristic function are provided in Appendix A.5. ∎

The price of the corresponding put bond option can be obtained by the following put call parity relation:

$$ZBP(t,S,T,K) - ZBC(t,S,T,K) = KP(t,S) - P(t,T).$$

With these results at hand, we can price interest rate caps and floors, which are respectively portfolios of put and call options on zero-bonds.

II. The Model-Implied Factor Dynamics

In this and the following sections, we are guided by the criteria laid down by Dai and Singleton (2003) and study how the completely affine Wishart yield curve model corresponds to the historical behavior of the term structure of interest rates. 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 interest rate derivatives.

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 generalization of Cheridito, Filipovic, and Kimmel (2007) is a potential step towards reconciling the forecasting of yields with their time-varying volatility. The extended affine market price of risk improves the time-series fit especially in models with multiple restricted state variables, i.e. $A_2(3)$ and $A_3(3)$ subfamilies. However, in comparison to essentially affine models, this improvement is achieved by adding new parameters, and is typically accompanied by a deterioration in the cross-sectional fit to the yield curve (Feldhütter, 2007).

To study the ability of the Wishart model to match the criteria (i)–(iv), we begin with the most simple 2×2 specification $(k = 3)$ of the state matrix (3) and the completely affine market price of risk (5). 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. Section III shows that this simple model suffices to match the first and second conditional moments of yields, and to replicate the hump term structure of cap implied volatilities. In order to study more complex features of the bond market such as the Cochrane-Piazzesi factor or unspanned factors in derivatives, we subsequently extend the dimension of the state matrix to 3×3 $(k = 3)$. The results are discussed in Section IV.

We put the model to the test in two ways: First, we require it to simultaneously match both conditional and unconditional properties of yields. This allows us to assess the richness of the structure implicit in the Wishart framework. Second, due to the choice of a low dimensional state matrix and a parsimonious market price of risk, after imposing the identification restrictions, the estimated 2×2 model is equipped with only 9 parameters to perform tasks (i)–(iv) mentioned above. This low number is hard to match by models in the affine class, in which the completely affine CIR with three independent factors already 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. The respective numbers for the essentially affine $A_1(3)$ model are 22 (preferred) and 29 (unrestricted). The extended affine specification requires 24 parameters in the restricted $A_1(3)$ model. With 18 identified parameters, even the completely affine 3×3 Wishart six-factor model compares favorably with the three-factor essentially and extended affine counterparts.

Before proving the model's ability to explain the yield curve puzzles, we outline our econometric approach based on closed-form moments of yields, discuss the model identification, and review the features of the underlying state space implied by the estimated parameters. Proofs are delegated to Appendix C.

II.A. Econometric approach

We use end-of-month 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. Yields for the period from January 1952 through December 1969 are from McCulloch and Kwon data set; yields from January 1970 through December 1999 are from the Fama and Bliss CRSP tapes. For the last period from January 2000 through June 2005, we use yields compiled by Gurkaynak, Sack, and Wright (2006).[14]

II.A.1. Moments of yields

The estimation of the parameter vector θ comprising the elements of matrices M, Q and D can be posed as the method of moments by setting:

$$\hat{\theta}_T = \arg\min_\theta \|\hat{\mu}_T - \mu(\theta)\|,$$

where $\hat{\mu}_T$ represents a vector-valued function of empirical moments based on the historical yields with different maturities, and $\mu(\theta)$ is its theoretical counterpart obtained from the model. The function μ involves the first and second (cross)-moments of yields with different maturities. We estimate the parameter vector θ by using moment conditions that provide both stationary and dynamic information about the term structure. The moments that provide a stationary description of the term structure comprise means, standard deviations and correlations of yields. The conditional information is introduced by augmenting the set of moment conditions with Campbell-Shiller regression coefficients. More specifically, in the 2×2 case we use the unconditional moments of yields with maturities 6 months, 2 years, and 10 years, and the Campbell-Shiller regression coefficients for the 2- and 10-year yields. In estimating the 3×3 specification, we further expand this set with correlations of yield changes and forward rate volatilities, and by adding the 5-year yield. This leaves us with 11 and 20 moment restrictions for the 2×2 and 3×3 model, respectively.

Affine expressions for the term structure facilitate the computation of the theoretical moments of yields as a function of the moments of the Wishart state variable. For brevity, we only provide the unconditional mean and covariance:

---

[14]The sample is an extension of the one used in Duffee (2002). The three sources use different filtering procedures, thus yields they report for the overlapping period do not match exactly. However, the descriptive statistics for yields of Gurkaynak, Sack, and Wright (2006) are consistent with the Fama-Bliss data set for the overlapping part of both samples.

$$E(y_t^\tau) = -\frac{1}{\tau} \{b_\tau + Tr\left[A_\tau E\left(\Sigma_t\right)\right]\}$$

$$Cov(y_{t+s}^{\tau_1}, y_t^{\tau_2}) = \frac{1}{\tau_1 \tau_2} vec\left(\Phi_s' A_{\tau_1} \Phi_s\right) \left[Cov(vec\Sigma_t)\right] vec(A_{\tau_2}),$$

where $Cov(vec\Sigma_t) = E\left[vec\Sigma_t(vec\Sigma_t)'\right] - vecE\Sigma_t\left(vecE\Sigma_t\right)'$.

In that the moments of the Wishart process, stated in the next lemma, are particularly simple and available in fully closed form, the computation of the moments of the term structure becomes a straightforward task.

Lemma 2 (Conditional moments of the Wishart process). *Given the Wishart process* (3) *of dimension* n *with* k *degrees of freedom, the first and the second conditional moments of* $\Sigma_{t+\tau}|\Sigma_t$ *are of the form, respectively:*

$$E(\Sigma_{t+\tau}|\Sigma_t) = \Phi_\tau \Sigma_t \Phi_\tau' + kV_\tau, \tag{21}$$

and

$$E[vec\Sigma_{t+\tau}\left(vec\Sigma_{t+\tau}\right)'|\Sigma_t] = vec(\Phi_\tau \Sigma_t \Phi_\tau' + kV_\tau)\left[vec(\Phi_\tau \Sigma_t \Phi_\tau' + kV_\tau)\right]'$$
$$+ \left(I_{n^2} + K_{n,n}\right)\left[\Phi_\tau \Sigma_t \Phi_\tau' \otimes V_\tau + k(V_\tau \otimes V_\tau) + V_\tau \otimes \Phi_\tau \Sigma_t \Phi_\tau'\right], \tag{22}$$

where I_{n^2} *is the* $n^2 \times n^2$ *identity matrix,* $K_{n,n}$ *denotes the* $n^2 \times n^2$ *commutation matrix,* $\otimes$ *is the Kronecker product, and* vec *denotes vectorization. Matrices* Φ_τ *and* V_τ *are given as:*

$$\Phi_\tau = e^{M\tau}$$

$$V_\tau = \int_0^\tau \Phi_s Q'Q\Phi_s' ds. \tag{23}$$

Proof and the closed-form expression for the integral (23) *are detailed in Appendix C.* ∎

The stationarity of the state variables requires the M matrix to be negative definite, in which case the unconditional moments of Σ_t are readily available:[15]

$$E\left(\Sigma_t\right) = kV_\infty$$

$$E\left[vec\Sigma_t\left(vec\Sigma_t\right)'\right] = k^2 vecV_\infty\left(vecV_\infty\right)' + k(I_{n^2} + K_{n,n})\left(V_\infty \otimes V_\infty\right),$$

where V_∞ can be efficiently computed as:

$$vecV_\infty = vec\left(\lim_{\tau\to\infty}\int_0^\tau \Phi(s)Q'Q\Phi'(s)ds\right) = -\left[(I_n \otimes M) + (M \otimes I_n)\right]^{-1} vec(Q'Q)$$

by exploiting the link between the matrix integral and the Lyapunov equation $MX + XM' = Q'Q$ (see Appendix C.1.3).

---

[15]Clearly, the negative definiteness of M matrix ensures that $\lim_{\tau\to\infty}\Phi(\tau) = 0$.

The unobservability of the state spells out the possibility that two distinct models lead to distributionally equivalent yields. Therefore, the identification of the parameters becomes a concern for empirical applications. A simple way to characterize the identification restrictions in our setting is to exploit its link to QTSMs. In the general case of an integer $k > 1$, the Wishart factor model shows analogy to an $(n \cdot k)$-factor "super-quadratic" model without a linear term and with repeated parameters. To see this, stack k n-dimensional OU processes X and the corresponding Brownian motions W as $Z_t = (X_t^{1\prime}, X_t^{2\prime}, \ldots X_t^{k\prime})'$ and $W_t = (W_t^{1\prime}, W_t^{2\prime}, \ldots W_t^{k\prime})'$. Then, the factors, the short rate and the market price of risk can be recast in a vector form as:

$$
\begin{aligned}
dZ_t &= (I_k \otimes M)Z_t dt + (I_k \otimes Q')dW_t \\
r_t &= Z_t'\left[I_k \otimes (D - I_n)\right]Z_t \\
\Lambda_t &= Z_t,
\end{aligned}
$$

where $\otimes$ denotes the Kronecker product. This model implies a risk neutral dynamics for the state variables of the form $dZ_t = \left[I_k \otimes (M - Q')\right]Z_t dt + (I_k \otimes Q')dW_t$. For instance, the 3×3 Wishart model with $k = 3$ degrees of freedom has a super-quadratic nine-factor interpretation.

With this equivalence, the parameter identification is ensured under conditions similar to those of Ahn, Dittmar, and Gallant (2002) or Leippold and Wu (2002). Since our completely affine market price of risk does not contain additional parameters, we can identify $I_k \otimes (M - Q')$, $I_k \otimes M$ and $I_k \otimes Q'$ from the Gaussian distribution of the state variables, both under physical and risk neutral probabilities. To pin down the parameters with respect to invertible linear transformations, it is enough to require matrices $I_k \otimes (M - Q')$ and $I_k \otimes M$ to be lower triangular. This condition is equivalent to imposing that matrices M and Q' are both lower triangular.

In addition to identification restrictions, the model implies two mild conditions on the parameter matrices: (i) negative definiteness of matrix M, which guarantees the stationarity of factors, and (ii) the invertibility of matrix Q, which ensures that Σ_t is reflected towards the domain of positive definite matrices when the boundary of the state space is reached. We can additionally require the $D - I_n$ matrix to be positive definite, thus ensuring positive yields. Appendix E provides details on the estimation procedure along with the estimated parameter values.

II.B. Properties of Risk Factors

The Wishart process gives much freedom in modeling 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. The current section substantiates this claim using the parameters of the estimated 2×2 model.

Factor correlations. To demonstrate how the time variation in correlations comes up in our setting, we consider an example of the instantaneous covariance between the positive elements of a 2×2 matrix of factors:

$$Corr_t(\Sigma_{11}, \Sigma_{22}) = \frac{\langle \Sigma_{11}, \Sigma_{22} \rangle_t}{\sqrt{\langle \Sigma_{11} \rangle_t} \sqrt{\langle \Sigma_{22} \rangle_t}}. \tag{24}$$

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

$$
\begin{aligned}
d\langle \Sigma_{11} \rangle_t &= 4\Sigma_{11}(Q_{11}^2 + Q_{21}^2)dt, \\
d\langle \Sigma_{22} \rangle_t &= 4\Sigma_{22}(Q_{12}^2 + Q_{22}^2)dt, \\
d\langle \Sigma_{11}, \Sigma_{22} \rangle_t &= 4\Sigma_{12}(Q_{11}Q_{12} + Q_{21}Q_{22})dt,
\end{aligned}
\tag{25}
$$

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 is either positive or negative. 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.[16] 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. In fitting the observed yields, however, the possibility of negative correlations plays a crucial role. 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.[17]

The degrees of freedom parameter k. The properties of the conditional factor correlations in our model are controlled by the degrees of freedom parameter k. Recall that k integer fixes the number of OU processes used to construct the state dynamics in equation (3) (see also Remark 3 and Appendix B for a related discussion). As such, it drives the non-singularity of Σ_t. By going beyond unitary degrees of freedom, but keeping them integer, we reach several goals all at once. First, we introduce a non-trivial time-variation in conditional correlations of positive factors, and thus break the link between our model and the n-factor QTSMs. Second, since the diagonal factors are non-central $\chi^2(k)$ distributed, k influences different moments of yields. To illustrate this, we plot the instantaneous correlations of diagonal factors (Figure 2) as well as the distribution of the 5-year yield implied by the estimated 2×2 model for different k's (Figure 3). In the special case where $k = 1$, the conditional correlations of positive factors are piecewise constant. Moreover, the restrictive $\chi^2(1)$ factor distribution translates into a by far too high skewness of yields, as compared

---

[16] 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.

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

to the historical distribution. A higher k tends to mitigate the misfit to the higher moments of yields by making the distribution closer (but, clearly, not isomorphic) to the Gaussian. To compare the model's performance with the affine class, Figure 4 superposes the 5-year US yield against the densities implied by our 2×2 model (panel a) and three benchmark ATSMs estimated by Duffee (2002) (panel b). The mixed $A_1(3)$, $A_2(3)$ ATSMs face severe problems in matching the higher order moments of yields, and the purely Gaussian $A_0(3)$ implies a non-negligible probability of negative yields. At the same time, the Wishart model $(k = 3, 7)$ is free from both shortcomings.[18] Intuitively, the multiple roles played by the degrees of freedom parameter explain why the flexibility of the model does not come at a loss to its parsimony.

[Insert Figure 2, 3 and 4 here]

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 manifested in the highly significant GARCH coefficient computed for the simulated sample of factors.[19] Ahn, Dittmar, 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 accommodate such situations. The ease at which correlations and stochastic volatilities coexist in the Wishart model is one of its powerful characteristics, as we prove in Section III.C and IV.B.

$A_m(N)$-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 framework nests under one name distinctive and so far also mutually exclusive features of previous models: (i) the number of unrestricted versus positive factors of the essentially affine $A_2(3)$ specification; (ii) the number of stochastic volatility factors of the completely affine $A_3(3)$ specification; (iii) the unrestricted (positive and negative) correlations among factors of the $A_0(3)$ specification; and it does not have a counterpart within the $A_m(3)$ family in terms of stochastic correlations among factors. We find the out-of-diagonal element of the Wishart matrix, Σ_{12}, to be negative in more than 90 percent of the simulated sample.[20] This result conforms with the affine literature, which provides evidence for the superior performance of the $A_{m<N}(N)$ class, with some factors having unrestricted signs, over the multifactor CIR models.

II.C. Factors Manifest in Yields

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

---

[18]Feldhütter (2007) provides extensive evidence of the different abilities of essentially affine, extended affine, and semi-affine models to match the higher-order moments of yields. The general conclusion from his work about the poor performance of the essentially affine family is confirmed in our Figure 4.

[19]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.

[20]The simulated sample comprises 72000 monthly observations from the model. Whenever subsequently we refer to simulation results, we always use this sample length, unless otherwise stated.

Factors in disguise of principal components. As a basic check whether 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 with 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 model-generated yields are spanned by at most three factors. We find that the portions of yield variation explained by the first two principal components in the model largely coincide with the empirical evidence. Moreover, the traditional factor labels are evident in Figure 5 with weights on the first three principal components virtually overlapping with those computed from the data.

[Insert Figure 5 here]

Shifting number of factors. Although the three-factor structure of 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 changes in monetary policy under various FED chairmen.

The behavior of instantaneous correlations between the Wishart factors in Figure 2 leads us to investigate whether the model can help explain this seemingly changing risk structure. We sort yields according to the level of instantaneous correlations between state variables, and for each group we compute the principal components. An immediate fact stands out from this exercise: the percentage of yield variation explained by the consecutive principal components changes considerably and in a systematic way across different subsamples. For instance, the loadings on the first and second principal components can vary from over 99 to 95 percent and from 5 to almost zero percent, respectively, depending on the level of instantaneous correlations. Such variability is consistent with the decompositions of the US yield levels in different subsamples. In Figure 6 we plot, as a function of instantaneous factor correlation, the portions of yield variation explained by each principal component.

[Insert Figure 6 here]

The closed-form expression for the conditional covariance of yields (16) allows us to perform a dynamic principal component analysis (Figure 7). This decomposition has an intriguing gist, as it leads to a much higher variation in factors explaining yields than what could be expected from the previous decomposition based on crudely defined subsamples. This discrepancy indicates that an empirical finding of some changeability in the yield factor structure across subsamples may significantly understate the true conditional variability. It stands to reason that, depending on the state of the economy, the relative impact of different macroeconomic variables (e.g., inflation expectations, real activity) onto the yield curve can vary considerably over time. This intuition finds support in the historical behavior of US yields, which seemed to be dominated by the variation of inflation expectations in the 1970s, and by the variation of real rates in the 1990s. Motivated

by this evidence, Kim (2007) highlights the relevance of structural instabilities and changing conditional correlations of macro quantities for explaining the term structure variation in the last 40 years.

[Insert Figure 7 here]

III. Yield Curve Puzzles Revisited

We assess the model in terms of the goodness-of-fit criteria discussed in the introduction to Section II. The analysis based on the 2×2 specification and a single set of estimated parameters (given in Appendix E.1) indicates that already this simple framework is able to provide a consistent answer across several dimensions of the spot and derivative bond markets. This coherence turns out to be a powerful tool to differentiate between our model and several preferred ATSMs which we choose as a benchmark.

III.A. Excess Returns on Bonds

The key implication of the essentially affine market price of risk proposed by Duffee (2002) is that excess bond returns have unrestricted sign. This feature is crucial for matching their empirical properties. Due to the properties of the Wishart process, our model "by default" grants a degree of flexibility that is (at least) comparable with the essentially affine class. To substantiate this claim, Figures 8 and 9 present excess bond returns obtained from the 2×2 Wishart term structure model (WTSM). The plots confirm the consistency of the model with the empirical evidence in several respects. First, we observe a switching sign of the risk premia, both instantaneous (Figure 8) and those computed from discrete realizations of the bond price process (Figure 9). Second, excess returns are highly volatile, with the conditional ratio $e_t^\tau / \sqrt{v_t^\tau}$ having a large probability mass between ± 1. Third, the above properties hold true across bonds with different maturities.

[Insert Figure 8 and 9 here]

The model-implied returns 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 (see Table I in Duffee (2002)). In line with this evidence, the model produces expected excess returns and volatilities that rise as a function of maturity. Moreover, the ratio of mean excess returns to their volatility is well below one (0.17 on average) across all maturities. These features play an important role in the model's ability to replicate the failure of the expectations hypothesis, which we discuss next.

III.B. The Failure of the Expectations Hypothesis

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

$$y_{t+m}^{n-m} - y_t^n = \beta_0 + \beta_1 \frac{m}{n-m}(y_t^n - y_t^m) + \varepsilon_t, \tag{26}$$

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[21] for all maturities n, 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. Rephrased in terms of returns, the expected excess returns on bonds are high when the slope of the yield curve is steeper than usually.

The findings of Section III.A confirm the ability of the model to produce the time variation in expected returns. To study whether expected returns vary in the "right" way with the term structure, we compute the model-implied theoretical coefficients of Campbell-Shiller regressions,[22] and benchmark them against their empirical counterparts (see panels a in Table I and in Figure 10). For the sake of comparison, we perform an analogous exercise for three preferred affine specifications of Duffee (2002) at his parameter estimates (see panels b in Table I and in Figure 10). In Duffee's convention, preferred models drop the parameters that contribute little to the models' QML values. 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. While the two Campbell-Shiller coefficients used as moment conditions are matched almost perfectly (see the shading in Table I, panel a), the model turns out to do a good job also in fitting other parameters, not used in the estimation. Panel a of Figure 10 visualizes this match by showing that all model-implied coefficients lie within the 80 percent confidence bounds computed from the historical sample.[23] At the same time, the results for 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.

[Insert Figure 10 here]

---

[21] To see this, note that m-month return on n-maturity bond is $r_{t,t+m}^n = \ln P_{t+m}^{n-m}/P_t^n = -(n-m)y_{t+m}^{n-m} + ny_t^n$. Then, the monthly excess return over the risk free return y_t^m is:

$$rx_{t,t+m}^n = \frac{1}{m}r_{t,t+m}^n - y_t^m = -\frac{m}{n-m}\left(y_{t+m}^{n-m} - y_t^n\right) + (y_t^n - y_t^m).$$

Reformulating and taking expectation yields:

$$E_t\left(y_{t+m}^{n-m} - y_t^n\right) = -\frac{m}{n-m}E_t\left(rx_{t+m}^n\right) + \frac{m}{n-m}(y_t^n - y_t^m).$$

Under the expectations hypothesis, the first term on the RHS is a constant, and the slope coefficient in a regression based on the above equation is unity.

[22] In computing the theoretical 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 coefficients implied by yields fitted to an ATSM.

[23] Note that the 80 percent bound—however lax for the data—is a more rigid gauge for the model's performance than a wider (e.g. 90 percent) bound. In fact, the coefficients obtained from $A_1(3)$ and $A_2(3)$ still fall outside the 95 percent bound.

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

The table presents the parameters of the Campbell-Shiller regression in equation (26). 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 theoretical coefficients along with population t-statistics below. The shading indicates the coefficients used as moment conditions in estimation. The fifth row shows small sample results obtained from the model by Monte Carlo. 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. 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 all model-implied and historical data.

a. Wishart 2×2 factor model

Maturity (n months)	12	24	36	60	84	120
Data β_1 (1952–2005)	−0.174	−0.615	−0.852	−1.250	−1.660	−2.244
t-stat	−0.4	−1.1	−1.4	−1.8	−2.1	−2.4
Model β_1 (popul.)	−0.070	−0.614	−1.070	−1.713	−2.120	−2.244
t-stat	−6.4	−9.2	−10.8	−12.2	−12.5	−11.5
Model β_1 (648 obs.)*	0.008	−0.521	−0.960	−1.568	−1.922	−2.065
t-stat	0.03	−1.5	−2.2	−2.4	−2.3	−1.9

b. ATSMs

Maturity (n months)	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

*) The coefficients and t-statistics are the median of 1000 estimates based on the simulated sample of 648 observations. The simulated path reflects the length of the sample used to estimate the WTSM. To conserve space, for ATSMs we only provide the population results.

As an additional test, we have also considered two additional regressions, which are detached from the estimation procedure, but reflect the same reasons for the failure of the expectations hypothesis as the Campbell-Shiller regressions. First, following Duffee (2002) we study projections of the monthly excess constant maturity bond returns on the lagged slope of the term structure defined as the difference between the 5-year and the 3-month yield. Even though this regression merely restates the information conveyed by Campbell-Shiller coefficients, it provides a robustness check of the previous results, because the yields which construct the slope are not used directly in estimating the model. As a second check, we replicate the regressions of Fama and Bliss (1987) projecting the excess one-year holding period bond return on the spot-forward spread. For brevity, we only state the main results without reporting the details. The predictability is confirmed. Consistent with the empirical evidence, the model-implied coefficients in both regressions increase with the maturity of the bond used as the dependent variable. A steep slope of the term structure

forecasts high excess returns during the next month. Similarly, a positive spot-forward spread is a predictor of higher holding period returns.

The above discussion provides further insight into the relevance of the Wishart factor structure in our completely affine setting. Within the essentially affine $A_m(3)$ family, these are the Gaussian models that dominate other subfamilies in terms of prediction because 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 aspect is addressed next.

III.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.

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 time variation and persistence in yield volatility commensurate with 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 II). The choice of the 5-year yield is motivated by the fact that this maturity is not involved in estimation of the 2×2 model. Therefore, its conditional and unconditional properties can be traced back to the intrinsic structure of the model. In panel a of Table II, 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 II, panel b). Due to its constant conditional volatility assumption, the Gaussian $A_0(3)$ model is not taken into consideration.

The results confirm that the degree of volatility persistence implied by the Wishart factor model aligns well with the historical figures. For example, the median model-implied GARCH coefficient is 0.847 as compared with 0.820, which is found empirically. Furthermore, our model is able to reproduce the positive 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 III.B, these results lend support to our claim that the model can simultaneously solve the predictability and the volatility puzzle.

The volatility persistence in the benchmark affine models, instead, shows a larger discrepancy to the data, and is typically too low. As in 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. And yet, as shown is Section III.B, it also largely fails in explaining the conditional first moments of yields. Although

Table II: GARCH(1,1) parameters for the model-implied and historical 5-year yield

The table presents 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 population values are based on 72000 observations.

a. Wishart 2×2 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.)	0.116	0.870	0.000
Model (648 obs.)*	0.123	0.847 [0.71, 0.96]	0.000
t-stat	4.4	27.6	2.4

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 [0.50, 0.86]	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 [0.33, 0.84]	0.000
t-stat	5.2	9.7	4.4

*) 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 estimate the different models. The numbers in square brackets show the small sample 99 percent confidence intervals based on Monte Carlo.

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. Note, however, that the preferred $A_2(3)$ model of Duffee (2002) is equivalent to the completely affine formulation, because the parameters of the essentially affine market price of risk turn out to be 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 in ATSMs. Finally, the small sample confidence intervals for the GARCH estimates convey useful intuition about the proximity of the different models and the true process driving the volatility of yields. Out of the three models considered, the GARCH coefficient in the Wishart factor setting is not only on average 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, which varies across different subsamples, has aroused increased interest with the appearance of a hump at around 2-year maturity during the Greenspan era (Piazzesi, 2001). Dai and Singleton (2000) conclude that the key to modeling the hump in ATSMs lies either in correlations between the state variables or in the respective factor loadings in the yield equation. Consistent with this interpretation, our model allows for non-monotonic behavior of yield

volatilities. To uncover the mechanism that leads to this non-monotonicity, 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 predominantly driven by the (weighted) variance of the Σ_{12} factor. This result fits nicely within the interpretation of Dai and Singleton (2000) in that Σ_{12} also determines the correlation between the positive factors. The forward yield volatilities and cap implied volatilities are discussed in greater detail next.

III.D. Aspects in Derivative Pricing

There is an intimate link between factor correlations, forward rate volatilities, and the prices of interest rate derivatives.

The term structure of forward rate volatilities. Similar to the unconditional second moments of yields, the term structure of forward rate volatilities tends to be hump-shaped for shorter maturities, as reported in Amin and Morton (1994) and Moraleda and Vorst (1997), among others. In WTSM, the instantaneous forward rate is given in closed form in expression (20). Thus, we can readily obtain the whole term structure of instantaneous forward rate volatilities. Consistent with the empirical evidence, at the estimated parameters a pronounced hump becomes evident. The decomposition of the model-implied forward rate variance reveals that the non-monotonicity is due to two elements: the variances of Σ_{12} and Σ_{11}, scaled by the respective functions of the elements of matrix $A(\tau)$.[24] Thus, yields and forward rates share the same source of a hump in volatility curves. Figure 11 (panel a) plots the instantaneous volatility curves for several dates in the simulated WTSM. In that in reality the hump emerges for discretely spaced (in contrast to instantaneous) forward rates, we also compute the theoretical standard deviations of one-year forward rates, and plot them against maturities in Figure 11 (panel b). To put the results into perspective with ATSMs, the hump is absent from the affine specifications at the parameters estimated by Duffee (2002). In the $A_0(3)$ model, the forward volatility curve is monotonically decreasing. The mixed models $A_1(3)$ and $A_2(3)$, instead, imply its increase for longer maturities—an implication which is not valid empirically.

[Insert Figure 11 here]

Implied volatilities of interest rate caps. The empirical properties of the conditional second moments of yields can be inferred from the implied volatility quotes for the interest rate derivatives, such as caps. Also in this case, the evidence of a hump is ubiquitous (e.g., Leippold and Wu (2003); De Jong, Driessen, and Pessler (2004)). In a yield curve model, the hump of cap volatilities can be induced via forward rates. The "transmission" mechanism follows from the fact that the volatility of a caplet is the integrated instantaneous volatility of the forward rate (see e.g., Brigo and Mercurio, 2006). Thus, those models able to display a hump in the instantaneous forward rate volatility should also perform well in the pricing of interest rate caps.

To verify this statement, we use the model-implied prices and the corresponding Black (1976) volatilities for caps with maturities from one to 15 years. A cap struck at rate $\overline{C}$ starting at T_0 and making equidistant

---

[24]The decomposition refers to the variance of the forward rate $f(t; S, T)$ prevailing at time t for the expiry at time $S > t$, and maturity $T > S$, defined as $f(t; S, T) = \frac{1}{T-S} \ln \frac{P(t,S)}{P(t,T)} = \frac{1}{T-S} \{b(t, S) - b(t, T) + Tr\left[(A(t, S) - A(t, T))\Sigma_t\right]\}$.

25

payments at times T_i, $i = 1, \ldots, n$, based on the simply compounded floating rate $L(T_{i-1}, T_i)$, can be priced according to:

$$
\begin{aligned}
\text{Cap}_t &= \sum_{i=2}^{n} E_t^* \left[e^{-\int_t^{T_i} r_s ds} \delta \left(L(T_{i-1}, T_i) - \overline{C} \right)^+ \right] \\
&= (1 + \delta\overline{C}) \sum_{i=2}^{n} E_t^{T_{i-1}} \left[\left(\frac{1}{1 + \delta\overline{C}} - P(T_{i-1}, T_i) \right)^+ \right],
\end{aligned}
\tag{27}
$$

where $T_i - T_{i-1} = \delta$. To get the last expression note that $\delta L(T_{i-1}, T_i) = \frac{1}{P(T_{i-1}, T_i)} - 1$. The last payment date T_n determines the maturity of the cap. $E_t^{T_{i-1}}$ specifies the conditional expectation under the forward measure induced by the zero bond maturing in $T_{i-1} - t$ periods. Thus, using Proposition 4, we can value caps as portfolios of put options on zero bonds. By market convention, we focus on at-the-money (ATM) contracts, for which the strike rate $\overline{C}$ of the T_n-maturity cap is set equal to the current T_n-year swap rate:

$$
\overline{C} = \frac{1}{\delta} \frac{P(0, T_1) - P(0, T_n)}{\sum_{i=2}^{n} P(0, T_i)}.
\tag{28}
$$

The mapping from cap prices to Black volatilities assumes an identical volatility for each caplet constituting the contract. Figure 12 presents the results as a function of the contract's maturity for several dates in the simulation. Indeed, our 2×2 factor specification is able to adapt an empirically plausible behavior of implied volatilities, which are lower for short maturity caps (one-year), increase in the intermediate range and decline smoothly for longer maturities.

[Insert Figure 12 here]

The implied cap volatilities carry new information about the conditional features of the model. By using an empirically relevant specification of the cap contract, we fend off concerns that a hump is just an instantaneous feature of the model. Moreover, and in contrast to the preceding discussion, the investigation of derivatives helps us assess the *risk-adjusted* properties of the state dynamics. The appearance of a hump in volatilities of forward rates *and* caps reflects the flexibility of the model under both physical and risk-adjusted measures.

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, 2007). Rich interrelations of factors also lie at the origins of the conditional hump in the quadratic term structure models (Leippold and Wu, 2003). A similar link can be retrieved in the Wishart setting from the role played by the out-of-diagonal factors. In the 2×2 example, Σ_{12} determines the stochastic dependence among different elements of the state matrix, and contributes to the non-monotonicity of the volatility curves. Therefore, the shape of volatility curves is a manifestation of the underlying stochastic correlations of factors, that are equally present under the physical and risk-adjusted measures. This puts the model aside from the standard affine class, in which factor correlations—while possibly restricted under the risk-adjusted measure—are passed to the model through the market price of risk. Even though this permits a better fit to the time series of yields, the constrains on the risk-neutral dynamics still impede the pricing in

the cross section, as suggested by the inconsistent valuations of caps and swaptions in the multifactor affine settings (see e.g. Jagannathan, Kaplin, and Sun (2003)).

IV. Extensions

By construction, any three-factor model comes under strain at least in three important respects. First, with the low dimension of the state space, the correlation of factors plays multiple (possibly conflicting) roles, steering simultaneously the time-series dynamics of the yield curve (e.g. predictability) along with its cross-sectional characteristics (e.g. humped term structure of yield volatilities). We note, for instance, that the inclusion of Campbell-Shiller coefficients in the set of moment conditions tends to somewhat worsen the model's fit to the unconditional volatility curve. Second, the number of risk factors restricts the number of non-collinear variables (e.g. forward rates or yields) that can be used in a forecasting regression of excess returns. Finally, with three factors necessary to explain the spot interest rates, there is no room for the so-called unspanned factors manifested in the prices of interest rate derivatives. An obvious cure to these tensions lies in the enlargement of the state space to a higher dimension. In this section, we estimate the 3×3 model along the lines of Section II.A (see Appendix E.2 for parameters). This extension takes us to a six-factor setting with three positive and three unrestricted factors and 18 parameters. It is worth pointing out that six-factor models with stochastic volatility are very difficult to come by in the standard affine setup due to the large number of parameters they involve. We show that the completely affine six-factor WTSM is able to accommodate further characteristics of fixed income markets in addition to those already captured in the 2×2 case.

IV.A. Cochrane-Piazzesi Single Forecasting Factor

Cochrane and Piazzesi (2005) strengthen the case against the expectations hypothesis by showing that excess returns across bonds of different maturities can be predicted with a single factor—a linear combination of five forward rates. Notably, the coefficients of a projection of one-year holding period bond returns on a constant and five one-year forward rates exhibit a systematic (tent-like) pattern. The implications of different models for the Cochrane-Piazzesi-type predictability have been studied by Bansal, Tauchen, and Zhou (2003) and Dai, Singleton, and Yang (2004). These works have focussed on the restricted case with at most three factors. However, to replicate the evidence in full (in an affine framework), at least five factors are required to avoid collinearity. We work with six. In our model, log bond prices are linear in the Wishart factors; hence the predictability due to a single linear combination of forward rates is equivalent to the predictability due to a single linear combination of the elements of Σ. We scrutinize the 3×3 model for the presence and the form of the single return-forecasting factor.

Single return-forecasting factor. While the single common factor appears to be an established feature of the data, its specific shape could be an artifact of a smoothing method used in constructing the zero-curve. Dai, Singleton, and Yang (2004) document that the pattern can turn into wave-like if, instead of the *unsmoothed* Fama-Bliss (UFB) yields used by Cochrane and Piazzesi (2005), the *smoothed* Fama-Bliss (SFB) data set is

employed.[25] Some caveats are in order: First, yields generated from the model naturally lead to a "smooth" zero curve. Second, in our model-generated term structure, as in reality, four factors (at a maximum) are sufficient to explain 100 percent of the variance of yields. Thus, in the absence of any cross-sectional measurement error, the collinearity of regressors becomes a concern; the oscillating shape of the forecasting factor found by Dai, Singleton, and Yang (2004) in the smoothed data is a likely signal thereof.[26] Figure 13 plots the slope coefficients in regressions of individual one-year excess bond returns on the set of five one-year forward rates. The pattern of factor loadings in the model (Figure 13, panel a) resembles closely the one found in the SFB data (Figure 13, panel b).

[Insert Figure 13 here]

Judging by the regularity of the slope coefficients, the key intuition for the single factor seems to be supported by the model. Moreover, by including only three forward rates $(f_t^{0 \to 1}, f_t^{2 \to 3}, f_t^{4 \to 5})$ in the regression, the upward pointed triangular shape of the loadings (a "restricted tent") becomes apparent (not reported). The pertinent question, however, is whether our 3×3 model can recover the high degree of predictability due to the *single* forecasting factor, rather than whether it can recover the particular shape. In Table III, we report the univariate second stage regressions of excess returns on the linear combination of forward rates, in Cochrane-Piazzesi notation:

$$\frac{1}{4} \sum_{\tau=2}^{5} rx_{t+1}^{(\tau)} = \gamma' \mathbf{f_t} + \bar{\varepsilon}_{t+1} \qquad \text{(1st stage)} \tag{29}$$

$$rx_{t+1}^{(\tau)} = b^{(\tau)} \left(\gamma' \mathbf{f_t} \right) + \varepsilon_{t+1}^{(\tau)} \qquad \text{(2nd stage)} \tag{30}$$

where $rx_{t+1}^{(\tau)} = hpr_{t+1}^{\tau \to \tau-1} - y_t^{1Y}$ is the return on holding the τ-maturity bond in excess of one-year yield, and $\mathbf{f_t} = \left[1, f_t^{0 \to 1} \text{ (spot)}, \ldots, f_t^{4 \to 5} \right]$ is the vector of forward rates. With γ coefficients in equation (30) fixed at the values from the first stage regression, $\gamma' \mathbf{f_t}$ represents the return-forecasting factor.

Note that the moments used to estimate the model do not provide direct information about the above regressions. Nevertheless, the model can replicate the empirical evidence. We study its population as well as small sample implications and juxtapose them with the UFB and SFB data sets. The main conclusion of Cochrane and Piazzesi (2005)—that a single factor accounts for a large portion of time-variation in excess returns—is confirmed by the high R^2 values in Table III. The model tends to generate R^2's that are very much in line with the empirical figures in panel a. In support of the single factor hypothesis, the $b^{(\tau)}$

---

[25] Both data sets are constructed from the same underlying coupon bond prices. The method used to extract the UFB yields assumes the forward rate to be a piecewise linear (step) function of maturity, whereas the SFB data is computed by smoothing the UFB rates with a Nelson-Siegel exponential spline. See Dai, Singleton, and Yang (2004) and Singleton (2006) for a discussion.

[26] Bansal, Tauchen, and Zhou (2003) suggest that in the presence of three latent factors driving the historical yields, the use of five regressors creates near perfect co-linearity problem, up to cross-sectional measurement errors that mask the singularity. As a standard remedy to collinear regressors, we add a very small amount of i.i.d. noise to model-generated variables. This has virtually no impact on the first four moments of the interest rates distribution. The noise is generated from $N(0, 4.5 \times 10^{-6})$; a different distribution, e.g. t_3-student, does not change the results. Especially, the level of the R^2 statistics remains largely unchanged. This finding is in line with the argument of Cochrane and Piazzesi (2005) that the predictability is not driven by measurement errors. We note that running the regression without noise gives qualitatively identical results, but leads to unreasonably high coefficients, which is again a diagnostic of the collinearity problem.

Table III: Cochrane-Piazzesi single forecasting factor regressions

The table reports the coefficients $b^{(\tau)}$, R^2, and t-statistics for the restricted Cochrane-Piazzesi regressions in equation (30). τ in the first row refers to the maturity of the bond whose excess return is forecasted. Panel a displays the results for two data sets: smoothed Fama-Bliss (SFB) and unsmoothed Fama-Bliss (UFB) yields. The yields are monthly and span the period 1970:01–2000:12. Panel b presents the model-implied estimates in the population (72000 observations), and in a small sample of length 360 months. All reported t-statistics use the Newey-West adjustment of the covariance matrix with 15 lags.

τ	2	3	4	5		2	3	4	5
				a. Data (1970-2000)					
		SFB data					UFB data		
$b^{(\tau)}$	0.46	0.85	1.19	1.50		0.46	0.87	1.22	1.45
t-stat	5.63	5.44	5.18	4.95		7.99	7.60	7.54	7.05
R^2	0.30	0.31	0.31	0.32		0.36	0.37	0.39	0.36
				b. Wishart 3×3 factor model					
		Population					Small sample* (360 obs.)		
$b^{(\tau)}$	0.46	0.86	1.20	1.48		0.46	0.86	1.20	1.48
t-stat	59.48	60.99	58.63	55.42		6.46	6.52	6.32	6.05
R^2	0.31	0.33	0.31	0.29		0.41	0.43	0.41	0.40

*) The coefficients and t-statistics are the median of 1000 estimates, each based on the sample of 360 realizations from the model.

coefficients are all significant and increase smoothly with bond's maturity τ. Such behavior persists across all data sets, irrespective of the underlying pattern of the γ's in the first stage regressions.

Information content of $\gamma'\mathbf{f_t}$. A salient empirical property of the return-forecasting factor is that it carries information beyond what is captured in the level, slope (typically considered to be the return predicting variable) and curvature. The evidence in Cochrane and Piazzesi (2005) suggests that $\gamma'\mathbf{f_t}$ is related to the fourth principal component of yields, which in turn has only a weak impact on the yields themselves. In our setting, this outcome has two different shades. Despite the great stability of the model-implied finite sample estimates in Table III, the relationship between $\gamma'\mathbf{f_t}$ and the yield factors turns out highly susceptible to the small sample biases. The Monte Carlo analysis (based on 360 observations from the model) indicates that the portions of $\gamma'\mathbf{f_t}$ variance explained by the respective principal components are highly uncertain quantities: They can range from 4 to 43 percent for the slope factor, and from 2 to 28 percent for the fourth factor (as measured by the upper and lower decile). This changes as the length of the sample becomes large. Then, the slope accounts for as little as 12 percent of the $\gamma'\mathbf{f_t}$ variance, while the total contribution of the fourth and fifth factor approaches 40 percent.

The historical behavior of term premia highlights the necessity of going beyond low dimensional models to provide a comprehensive description of the bond market dynamics. The completely affine WTSM appears as a convenient model to withstand this challenge.

Table IV: Properties of the conditional hedge ratio

The table reports the statistics for the conditional hedge ratio HR_t, the ratio of conditional volatilities of yield changes $\sigma_{10}(t)/\sigma_2(t)$, and the conditional correlation of yield changes $\rho_{2,10}(t)$. The conditional volatilities are estimated with a GARCH(1,1) model, the conditional correlations—with the DCC model of Engle (2002). We report the means and volatilities of the estimated conditional quantities. For the model, we provide the population statistics along with their small sample 99 percent confidence intervals (in brackets underneath) based on Monte Carlo with 1000 repetitions of 54 years of monthly data each.

a. Data (1952–2005)

	HR_t	$\sigma_{10}(t)/\sigma_2(t)$	$\rho_{2,10}(t)$
Mean	0.61	0.76	0.80
Volatility	0.16	0.17	0.06
$\mathrm{Corr}(\sigma_2(t), \sigma_{10}(t))$	0.86		

b. Wishart 3×3 factor model

	HR_t	$\sigma_{10}(t)/\sigma_2(t)$	$\rho_{2,10}(t)$
Mean	0.51	0.61	0.83
conf. bound	[0.41, 0.64]	[0.50, 0.74]	[0.80, 0.87]
Volatility	0.13	0.14	0.03
conf. bound	[0.05, 0.15]	[0.06, 0.17]	[0.01, 0.06]
$\mathrm{Corr}(\sigma_2(t), \sigma_{10}(t))$	0.76		
conf. bound	[0.40, 0.91]		

IV.B. Conditional Hedge Ratio

The evidence in the literature suggests that low dimensional affine models have a hard time capturing the right dynamics of volatilities and correlations of different segments of the yield curve (Bansal, Tauchen, and Zhou, 2003; Dai and Singleton, 2003). It comes as no surprise that the 2×2 model—while providing a good explanation for several features of the data—cannot fully nail down the cross-sectional dynamics of yields. Six factors appear to be the remedy.

To add insight on the (co-)volatilities of yields in the 3×3 setting, we focus on the conditional hedge ratio between the 10- and 2-year bonds, i.e. $HR(t) = \frac{\sigma_{10}(t)}{\sigma_2(t)}\rho_{2,10}(t)$. $\sigma_2(t)$ and $\sigma_{10}(t)$ are the conditional volatilities of yield changes obtained with a univariate GARCH(1,1); $\rho_{2,10}(t)$ is the conditional correlation of the yield changes estimated with the dynamic conditional correlation (DCC) model of Engle (2002). To match the characteristics of HR_t, the model needs to accommodate not only the time-varying correlations but also the variation in the relative riskiness of the two bonds, i.e. the comovement of volatilities. In this context, Gaussian ATSMs, for example, turn out completely untenable. Our model, in contrast, replicates the general properties of HR_t relatively well (see Table IV). The model-implied volatility of both the conditional correlation process $\rho_{2,10}(t)$ and of the volatility ratio $\sigma_{10}(t)/\sigma_2(t)$ is close to the empirical one. Overall, the implications for the conditional hedge ratio seem realistic as measured by the small sample confidence intervals. We reach a similar conclusion for the correlation of volatilities of the two bonds.

The conditional behavior of second moments of yields has attracted considerable attention in the latest term structure literature (e.g., Collin-Dufresne, Goldstein, and Jones, 2006; Joslin, 2007). This research

tends to agree in that low dimensional affine models face severe difficulties capturing the conditional yield volatility across maturities. Not surprisingly, those models in which more factors have stochastic volatility seem to perform better. Based on the examination of $A_m(3)$ models, Jacobs and Karoui (2006), for instance, tentatively suggest the $A_3(4)$ or $A_4(4)$ class as the best potential candidate for modeling volatility. At the same time, they recognize that its heavy parametrization may frustrate the estimation effort. The completely affine Wishart setting seems to provide a convenient way to combine a low number of parameters with a maximal flexibility in the stochastic (co-)volatilities of factors.

IV.C. Unspanned Factors

With six state variables at hand, but only three factors needed to explain the variation of yields, our model naturally lends itself to exploring the presence of factors unspanned by the spot market.

In the empirical literature, the discovery of unspanned factors follows from the poor performance of bond portfolios in hedging interest rate derivatives. Recent research reports a considerable variation in cap and swaption prices that appears to be weakly related to the underlying bonds. For instance, Heidari and Wu (2003) document that three common factors in Libor and swap rates can explain little over 50 percent in swaption implied volatility. Li and Zhao (2006) arrive at a similar conclusion for at-the-money (ATM) difference caps. Even though the yield factors can explain around 90 percent of the variation in the short maturity cap returns, their explanatory power deteriorates dramatically at longer maturities approaching just 30 percent for the 10-year maturity.

Two observations lead us to expect a similar phenomenon to arise in our estimated model. First, while the Σ_t matrix includes six autonomous factors,[27] the principal component decomposition of the model-generated yields reveals that three factors already explain nearly their total variation. Second, the estimated loading matrix $A(\tau)$ in the yield equation (15) is almost reduced rank across all maturities (see Table V, panel a), which indicates the possibility that the dimension of the state space generated by the yields is strictly smaller than the dimension of the state space generated by the Wishart factors.

In the absence of a theory for unspanning in the Wishart setting, we take an empirical backing out approach to test the presence of unspanned volatility features in our estimated model. We obtain the model prices of ATM caps for maturities from one to 10 years, and perform three different checks. First, we regress caps against three factors (level, slope and curvature) that span the spot yield curve in the model. Interestingly, even without imposing any structural restrictions, we do find evidence of bond market incompleteness, with an average R^2 of just 53 percent (see Table V, panel b). By using just three factors we prune some information, which—while inconsequential for spot yields—could possibly contribute to the variation of caps. To check this possibility, we subsequently include the remaining principal components in the regressions. Although the R^2 increases, the general conclusion persists, and is roughly consistent with the findings of Li and Zhao (2006) for ATM difference caps or Heidari and Wu (2003) for swaptions. Second, we decompose the covariance matrix of residuals from the first set of regressions (panel c). The decomposition exposes at least

---

[27]This can be seen from Σ_t being full rank. Equivalently, a principal component exercise performed on the unconditional covariance matrix of the state variables (i.e. covariance between the elements of Σ_t) also detects six factors.

31

Table V: Factors in the derivatives market

The table reports the level of spanning of cap prices by the spot yields in the 3×3 model. Panel a reports the eigenvalues of the $A(\tau)$ coefficients in the yield equation (15). We consider cap contracts with maturities of one to 10 years. The first line in panel b contains the R^2 values in regressions of caps on the first three principal components of yields (level, slope and curvature). The second line presents the R^2 values in regressions of cap prices on the (maximal set of) six principal components. Panel c provides the principal component decomposition of the covariance matrix of residuals from the first regressions of different maturity caps on the three yield PC's. Panel d reports the R^2 values in regressions of spot yields on six principal components retrieved from the covariance matrix of cap prices.

a. Eigenvalues of $A(\tau)$ coefficients										
τ	1	2	3	4	5	6	7	8	9	10
λ_1	-0.12	-0.23	-0.36	-0.50	-0.65	-0.81	-0.98	-1.15	-1.32	-1.48
λ_2	-0.04	-0.06	-0.07	-0.08	-0.08	-0.08	-0.09	-0.09	-0.09	-0.09
λ_3	2.7e-6	-2.3e-5	-3.6e-5	-4.4e-5	-4.8e-5	-5.2e-5	-5.5e-5	-5.7e-5	-5.8e-5	-6.0e-5

b. Regressions of ATM caps on yield PC's											
Cap maturity	1	2	3	4	5	6	7	8	9	10	Mean
3 PC's*, R^2	55.5	73.0	74.6	70.8	64.8	57.4	48.7	39.0	29.1	19.6	53.2
6 PC's, R^2	91.5	88.9	84.2	79.0	73.2	66.6	59.0	50.5	41.6	32.9	66.7

* Note that regressions of each of the yields on the first three PC's give the R^2 of 1.

c. Decomposition of the covariance matrix of residuals										
Eigenvector	1	2	3	4	5	6	7	8	9	10
% Explained	97.94	1.81	0.22	0.03	0.00	0.00	0.00	0.00	0.00	0.00

d. Regressions of yields on cap PC's											
Yield maturity	1	2	3	4	5	6	7	8	9	10	Mean
6 cap PC's, R^2	93.4	93.8	94.0	94.1	94.2	94.2	94.3	94.3	94.4	94.4	94.1

three additional factors influencing cap prices. These factors reflect the residual degrees of freedom which—while present in the model—have not been exploited in its estimation with the spot yield data. Finally, to fend off the criticism that the results are just a spurious effect of applying a linear regression to a nonlinear problem, we reverse the first exercise, and regress the yields on factors obtained from caps. The ability of caps to hedge the interest rate risk is evident in the high R^2 values (on average 94 percent) reported in panel d.

In ATSMs, the theoretical conditions for the existence of unspanned factors boil down to restricting the coefficient loadings in the yield equation to be linearly dependent for all maturities, as first noted by Collin-Dufresne and Goldstein (2002). Under such constraints, not all state variables are revealed through the yield curve dynamics alone. This pioneering analysis has been recently expanded by Joslin (2006) who provides general conditions for the incomplete bond markets in affine $\mathbb{R}_+^m \times \mathbb{R}^{n-m}$ models. His results, however, do not carry over to the Wishart factor framework due to the different form of the state space. In a companion paper,

Joslin (2007) shows that a four-factor ATSM with an unspanned volatility restriction is soundly rejected by the data. An important—and so far unexplored—issue is whether such a conclusion would also persist in a model of larger dimension and maximally flexible specification of factor volatilities.[28] Our findings seem to indicate that the completely affine enlarged WTSM provides a leeway to reconcile unspanned factors with the remaining stylized facts of the yield curve. The characterization of the theoretical conditions for bond market incompleteness in the Wishart yield curve setting is an interesting topic for future research.

V. Conclusions

In this article, we study the implications of a term structure model with stochastically correlated risk factors driven by a matrix-valued Wishart process. Under this new class, we resurrect the completely affine market price of risk specification and document that this setting provides a consistent explanation for several term structure puzzles.

The 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 leading to multivariate dynamics in second moments of yields. While the first feature is shared with some essentially and extended affine models, the combination of the last two represents the strength of our approach. With these levers, we provide answers to the following issues:

First, we are able to replicate the distributional properties of yields and the dynamic behavior of expected bond returns. The predictability of returns in the Wishart framework violates the expectations hypothesis in line with historical evidence. The model-implied population coefficients in the Campbell and Shiller (1991) regressions are negative and increase in absolute value with 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 model-implied conditional yield volatilities match the data in terms of GARCH estimates. We document the superior performance of the model in generating the empirically relevant degree of volatility persistence as compared to the preferred affine specifications estimated by Duffee (2002).

Third, we find that the term structure of the forward rate volatilities in the model is marked by a hump around the two-year maturity. The result is preserved both instantaneously, and for the unconditional volatilities of discretely spaced one-year forward rates generated by the model. The conditional hump is further confirmed in Black implied volatilities of caps. Implicit in the volatility curves are the correlations between the state variables, which unlike in the standard affine class are equally present under physical and risk-neutral measures.

Several facts strengthen the findings just described. First, to illustrate its basic properties, we use the most parsimonious formulation of the model. The choice of a 2×2 state matrix puts us in a three-factor framework, with two positive and one unrestricted factor, and the simplest (completely affine) market price of risk specification. In this form, the model is endowed with only 9 parameters to successfully perform the tasks

---

[28]Heidari and Wu (2003), for instance, point to a necessity of a Gaussian 3+3 factor model to explain the joint behavior of yields and interest rate derivatives.

listed above, while many of the best performing three-factor essentially or extended affine parameterizations require at least twice this number. Second, using just a single set of parameters, we are largely able to reconcile the first and second moments of model-implied yields with their historical counterparts. The factor structure of the model permits us to reproduce both the unconditional and conditional features of the data, as is evident in the persistent conditional volatilities and humped term structure of cap implied volatilities.

To appreciate this setting, it is also useful to recognize its analogies with the standard ATSMs. The argument we apply to derive the market price of risk is the one that stands behind the completely affine class. Likewise, for an arbitrary dimension of the Wishart state matrix, we benefit from the analytical tractability comparable with a multifactor CIR model. In spite of these similarities, the theoretical properties of the state space and its empirical coherence set this setting apart from several benchmark ATSMs.

The generality of the presented framework allows for a flexible extension beyond three factors. With the 3×3 dimension of the state space, the model has six factors, three of which are restricted in sign. While the gain in flexibility is non-trivial, in terms of the number of parameters (18) the model still compares favorably to the essentially or extended affine three-factor specifications. The enlarged model not only performs very well across all dimensions mentioned above, but also has the right scope to tackle more complex dynamics of the fixed income market. As a consequence, we are able to address several open issues exposed by the latest yield curve literature: First, in the enlarged framework, the predictability of excess bond returns is supported by the single forecasting factor of Cochrane and Piazzesi (2005). Second, the dynamic correlations of yields and the comovement in their volatilities result in realistic properties of conditional hedge ratios between bonds. Last but not least, some state variables—while loading weakly on yields—have an economically significant impact on the prices of interest rate caps, thus evoking the notion of unspanned factors.

References

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

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

AÏT-SAHALIA, Y., AND J. YU (2006): "Saddlepoint Approximations for Continuous-Time Markov Processes," *Journal of Econometrics*, 134, 507–551.

AMIN, K. I., AND A. J. MORTON (1994): "Implied Volatility Functions in Arbitrage-Free Term Structure Models," *Journal of Financial Economics*, 35, 141–180.

ANDERSEN, T. G., AND L. BENZONI (2008): "Do Bonds Span Volatility Risk in the US Treasury Market? A Specification Test for Affine Term Structure Models," FRB of Chicago Working Paper No. 2006-15.

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.

BACKUS, D. K., AND J. H. WRIGHT (2007): "Cracking the Conundrum," Working Paper, Stern School of Business and Board of Governors of the Federal Reserve System.

BANSAL, R., G. TAUCHEN, AND H. ZHOU (2003): "Regime-Shifts, Risk Premiums in the Term Structure, and the Business Cycle," Duke University and Federal Reserve Board.

BANSAL, R., AND H. ZHOU (2002): "Term Structure of Interest Rates with Regime Shifts," *Journal of Finance*, 57, 1997–2038.

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.

BROCKETT, R. (1970): *Finite Dimensional Linear Systems.* Wiley, New York.

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): "Habit Formation and Macroeconomic Models of the Term Structure of Interest Rates," *Journal of Finance*, 62, 3009–3063.

BURASCHI, A., P. PORCHIA, AND F. TROJANI (2006): "Correlation Risk and Optimal Portfolio Choice," Working Paper, Imperial College, London and University of St. Gallen.

CAMPBELL, J. (1995): "Some Lessons from the Yield Curve," *Journal of Economic Perspectives*, 9, 129–152.

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.

CASASSUS, J., P. COLLIN-DUFRESNE, AND B. GOLDSTEIN (2005): "Unspanned Stochastic Volatility and Fixed Income Derivatives Pricing," *Journal of Banking and Finance*, 29, 2723–2749.

CHACKO, G., AND S. DAS (2002): "Pricing Interest Rate Derivatives: A General Approach," *Review of Financial Studies*, 15, 195–241.

CHENG, P., AND O. SCAILLET (2007): "Linear-Quadratic Jump-Diffusion Modeling," *Mathematical Finance*, 17, 575–598.

CHERIDITO, P., D. FILIPOVIĆ, AND R. L. KIMMEL (2006): "A Note on the Dai-Singleton Canonical Representation of Affine Term Structure Models," Working Paper, Princeton University, Unversity of Munich, and The Ohio State University.

CHERIDITO, P., D. FILIPOVIC, AND R. L. KIMMEL (2007): "Market Price of Risk Specifications for Affine Models: Theory and Evidence," *Journal of Financial Economics*, 83, 123–170.

COCHRANE, J. H., AND M. PIAZZESI (2005): "Bond Risk Premia," *American Economic Review*, 95, 138–160.

———— (2008): "Decomposing the Yield Curve," Working Paper, University of Chicago.

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.

COLLIN-DUFRESNE, P., R. S. GOLDSTEIN, AND C. S. JONES (2006): "Can Interest Rate Volatility Be Extracted from the Cross Section of Bond Yields? An Investigation of Unspanned Stochastic Volatility," Working Paper, University of California Berkeley, University of Minnesota, and University of Southern California.

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

DAI, Q. (2003): "Term Structure Dynamics in a Model with Stochastic Internal 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): "Expectation 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, 631–678.

DAI, Q., K. J. SINGLETON, AND W. YANG (2004): "Predictability of Bond Risk Premia and Affine Term Structure Models," Working Paper, New York University and Stanford University.

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.

DONATI-MARTIN, C., Y. DOUMERC, H. MATSUMOTO, AND M. YOR (2004): "Some Properties of the Wishart Process and a Matrix Extension of the Hartman-Watson Laws," *Publications of the Research Institute for Mathematical Sciences*, 40, 1385–1412, Working Paper, University of P. and M. Curie, University of P. Sabatier, and Nagoya University.

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.

DUARTE, J. (2004): "Evaluating an Alternative Risk Preference in Affine Term Structure Models," *Review of Financial Studies*, 17, 379–404.

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

DUFFIE, D., D. FILIPOVIC, 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.

FELDHÜTTER, P. (2007): "Can Affine Models Match the Moments in Bond Yields?," Working Paper, Copenhagen Business School.

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.

GAGLIARDINI, P., P. PORCHIA, AND F. TROJANI (2007): "Ambiguity Aversion and the Term Structure of Interest Rates," *Review of Financial Studies*, forthcoming.

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, CEPREMAT, and University of Toronto.

GOURIEROUX, C., AND R. SUFANA (2003): "Wishart Quadratic Term Structure Models," Working Paper, CREST, CEPREMAT, and University of Toronto.

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

GURKAYNAK, R. S., B. SACK, AND J. H. WRIGHT (2006): "The U.S. Treasury Yield Curve: 1961 to the Present," Finance and Economics Discussion Series, Federal Reserve Board.

HAN, B. (2007): "Stochastic Volatilities and Correlations of Bond Yields," *Journal of Finance*, 62, 1491–1524.

HEIDARI, M., AND L. WU (2003): "Are Interest Rate Derivatives Spanned by the Term Structure of Interest Rates?," *Journal of Fixed Income*, 13, 75–86.

HEIDARI, M., AND L. WU (2005): "Term Structure of Interest Rates, Yield Curve Residuals, and Consistent Pricing of Interest Rate Derivatives," Caspian Capital Management and Zicklin School of Business, Baruch College.

HESTON, S. (1993): "A Closed-Form Solution for Options with Stochastic Volatility and Applications to Bond and Currency Options," *Review of Financial Studies*, 6, 327–343.

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.

JACOBS, K., AND L. KAROUI (2006): "Affine Term Structure Models, Volatility and the Segmentation Hypothesis," Working Paper, McGill University.

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.

JOSLIN, S. (2006): "Can Unspanned Stochastic Volatility Models Explain the Cross Section of Bond Volatilities?," Working Paper, Stanford Graduate School of Business.

——— (2007): "Pricing and Hedging Volatility Risk in Fixed Income Markets," Working Paper, Stanford Graduate School of Business.

KIM, D. H. (2007): "Challenges in Macro-Finance Modelling," BIS Working Paper, No. 240.

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.

LAUB, A. J. (2005): *Matrix Analysis for Scientists and Engineers*. SIAM, Davis, California.

LEIPPOLD, M., AND L. WU (2002): "Asset Pricing under the Quadratic Class," *Journal of Financial and Quantitative Analysis*, 37, 271–295.

——— (2003): "Estimation and Design of Quadratic Term Structure Models," *Review of Finance*, 7, 47–73.

LEVIN, J. (1959): "On the Matrix Riccati Equation," *Proceedings of the American Mathematical Society*, 10, 519–524.

LI, H., AND F. ZHAO (2006): "Unspanned Stochastic Volatility: Evidence from Hedging Interest Rate Derivatives," *Journal of Finance*, 61, 341–378.

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.

MAGNUS, J. R., AND H. NEUDECKER (1979): "The Commutation Matrix: Some Properties and Applications," *Annals of Statistics*, 7, 381–394.

——— (1988): *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley Series in Probability and Statistics, Chichester.

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 American Interest Rate Claims with Humped Volatility Models," *Journal of Banking and Finance*, 21, 1131–1157.

MUIRHEAD, R. J. (1982): *Aspects of Multivariate Statistical Theory*. Wiley Series in Probability and Mathematical Statistics.

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.

RUDEBUSCH, G. D., E. T. SWANSON, AND T. WU (2006): "The Bond Yield "Conundrum" from a Macro Finance Perspective," *Monetary and Economic Studies (Special Edition)*, 24 (S-1), 83–109.

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.

THOMPSON, S. (2008): "Identifying Term Structure Volatility from the LIBOR-Swap Curve," *Review of Financial Studies*, 21, 819–854.

VAN LOAN, C. F. (1978): "Computing Integrals Involving Matrix Exponential," *IEEE Transactions on Automatic Control*, 23, 395–404.

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

Appendices

This section has the following structure. Appendix A derives and characterizes the equilibrium term structure solution. Appendix B discusses the link between the QTSMs and the WTSM. Appendix C gives the closed-form expressions for the two first moments of the state matrix and yields. Appendix D collects several useful results for the Wishart process, which are applied throughout the remaining appendices. Appendices E and F provide the estimation details and figures, respectively.

Throughout this Appendix the notation $A > B$, for two conformable matrices A and B, should be understood as their difference $A - B > 0$ being a positive definite matrix.

A. Proofs: Term Structure of Interest Rates

A.1. Second Moments of the Short Interest Rate

Let $D - I_n = C$, a symmetric matrix. In a first step, we derive the expression for the instantaneous variance of the interest rate. Applying Ito's Lemma to the interest rate, we have that $dr = Tr\,(Cd\Sigma)$. By Result 3 in Appendix D, we obtain the expression (6):

$$
\begin{aligned}
Var_t\,[Tr(Cd\Sigma)] &= 4Tr\,(C\Sigma CQ'Q)\,dt = 4Tr\,[(D - I_n)\Sigma(D - I_n)Q'Q]\,dt \\
&= 4Tr\,[(D - I_n)Q'Q(D - I_n)\Sigma]\,dt > 0.
\end{aligned}
$$

With 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:

$$
\begin{aligned}
Cov_t\,(dr, dV) &= Cov_t\,[Tr\,(Cd\Sigma)\,,Tr\,(Pd\Sigma)] = 4Tr\,[P\Sigma CQ'Q]\,dt \\
&= 4Tr\,[(D - I_n)\,Q'Q\,(D - I_n)\,Q'Q\,(D - I_n)\,\Sigma]\,dt.
\end{aligned}
$$

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

A.2. Proof of Proposition 3: Solution for the Term Structure of Interest Rates

The coefficients $A(t, T)$ and $b(t, T)$ in the bond price expression are identified by inserting function (12) into the pricing PDE (10) and solving the resulting matrix Riccati equation. Note that $\mathcal{R}P = A(t, T)P$, and

$$
\frac{\partial P}{\partial t} = P\left[\frac{d}{dt}b(t, T) + Tr\left(\frac{d}{dt}A(t, T)\Sigma\right)\right], \tag{A-1}
$$

where for brevity P denotes the price at time t of a bond maturing at time T. The pricing PDE (10) can be expressed as:

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

Matrix Riccati equation. The above equation holds for all t, T and Σ. By the matching principle, we get a system of ODEs in A and b:

$$
-\frac{db}{dt} = Tr\,(\Omega\Omega'A) \tag{A-2}
$$

$$
-\frac{dA}{dt} = A(M - Q') + (M' - Q)A + 2AQ'QA - (D - I_n). \tag{A-3}
$$

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:

$$
\frac{db}{d\tau} = Tr\,(\Omega\Omega'A) \tag{A-4}
$$

$$
\frac{dA}{d\tau} = A(M - Q') + (M' - Q)A + 2AQ'QA - (D - I_n), \tag{A-5}
$$

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

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

A.2.1. Closed-form Solution to the Matrix Riccati Equation

The closed-form solution to the matrix Riccati equation (A-5) is obtained, via Radon's lemma, by linearizing the flow of the differential equation. For completeness, we give it in this Appendix. We express $A(\tau)$ as:

$$A(\tau) = H(\tau)^{-1}G(\tau), \tag{A-6}$$

for $H(\tau)$ invertible and $G(\tau)$ being a square matrix. Differentiating (A-6), we have:

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

Premultiplying (A-5) 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 matrix-valued 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:

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

The solution to the above ODE is obtained by exponentiation:

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

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

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

From equation (A-6), the closed-form solution to (A-5) is given by:

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

whenever it exists. Given the solution for $A(\tau)$, the coefficient $b(\tau)$ is obtained directly by integration, and admits the following closed form (da Fonseca, Grasselli, and Tebaldi, 2006):

$$b(\tau) = Tr\left(\Omega\Omega'\int_0^\tau A(s)\right)ds = -\frac{k}{2}Tr\left[\ln C_{22}(\tau) + \tau(M' - Q)\right].$$

■

41

A.2.2. Characterization of the Solution to the Riccati Equation

In this section, we discuss the definiteness and monotonic properties of the solution to the matrix Riccati equation (14) in Proposition 3. For convenience, we stick to the notation in equation (A-5) of the Appendix. Since both the equation (A-5) and the terminal condition $A(0)$ are real and symmetric, the solution must be real and symmetric on the whole interval $[0, \tau]$.

Negative definiteness of the solution. To discuss the definiteness of the solution, let us rewrite the equation:

$$\frac{dA(\tau)}{d\tau} = A(\tau)\widetilde{M} + \widetilde{M}'A(\tau) + 2A(\tau)Q'QA(\tau) + C, \tag{A-7}$$

$$A(\tau_0) = 0, \tag{A-8}$$

where $\tau_0 = 0$ and for brevity $C = -(D - I_n)$ and $\widetilde{M} = M - Q'$, as the time-varying *linear* ODE of the form:

$$\frac{dA(\tau)}{d\tau} = W'(\tau)A(\tau) + A(\tau)W(\tau) + C,$$

where $W(\tau) = \widetilde{M} + Q'QA(\tau)$. A well-known result from the control theory (see e.g. Brockett, 1970, p.59, p.162) allows us to state the solution to this equation as:[29]

$$A(\tau) = \Phi(\tau, \tau_0)A(\tau_0)\Phi'(\tau, \tau_0) + \int_{\tau_0}^{\tau} \Phi(s, \tau_0)C\Phi'(s, \tau_0)ds = \int_{\tau_0}^{\tau} \Phi(s, \tau_0)C\Phi'(s, \tau_0)ds, \tag{A-9}$$

where $\Phi(\cdot, \cdot)$ is the state transition matrix of the system matrix $W(\tau)$ solving $\dot{\Phi}(\tau, \tau_0) = W'(\tau)\Phi(\tau, \tau_0)$, $\Phi(\tau_0, \tau_0) = I_n$. The RHS of equation (A-9) represents a congruent transformation of matrix C. Congruent transformations may change the eigenvalues of a matrix but they cannot change the signs of the eigenvalues (Sylvester's law of inertia). Thus, given $A(\tau_0) = 0$, the necessary and sufficient condition for $A(\tau)$ to be negative definite is that $C < 0$, therefore $D - I_n > 0$.

Monotonic properties of the solution. To prove the monotonicity of the solution, let us differentiate the equation (A-7) with respect to time to maturity, τ:

$$\ddot{A}(\tau) = \dot{A}(\tau)\widetilde{M} + \widetilde{M}'\dot{A}(\tau) + 2\dot{A}(\tau)Q'QA(\tau) + 2A(\tau)Q'Q\dot{A}(\tau)$$
$$= V'(\tau)\dot{A}(\tau) + \dot{A}(\tau)V(\tau),$$

where $V(\tau) = \widetilde{M} + 2Q'QA(\tau)$, and for convenience we use dot and double-dot notation for the first and second derivative, respectively. The solution to this equation is given as:

$$\dot{A}(\tau) = \Phi(\tau, \tau_0)\dot{A}(\tau_0)\Phi'(\tau, \tau_0),$$

with the state transition matrix $\Phi(\tau, \tau_0)$ of the system matrix $V(\tau)$ solving $\dot{\Phi}(\tau, \tau_0) = V'(\tau)\Phi(\tau, \tau_0)$, $\Phi(\tau_0, \tau_0) = I_n$. By plugging the terminal condition $A(\tau_0) = 0$ into equation (A-7), we have:

$$\dot{A}(\tau_0) = C = -(D - I_n).$$

Therefore,

$$\dot{A}(\tau) = \Phi(\tau, \tau_0)C\Phi'(\tau, \tau_0),$$

which is negative definite if $C < 0$. By integrating the above expression on the interval $(s, t), s < t$ for some $s, t \in [\tau_0, \tau_1]$, we have:

$$A(t) - A(s) = \int_s^t \dot{A}(u)du < 0.$$

Hence, $A(\tau)$ declines with the time to maturity τ of the bond. ∎

A.3. Bond Returns

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

---

[29]To be exact, the statement (A-9) should express the solution in terms of some $\Pi(\tau)$ (rather than $A(\tau)$), which solves the matrix Riccati ODE (A-7). However, to keep the notation simple, with some abuse of notation, we state the solution in terms of $A(\tau)$. This has no impact on the argument which ensues.

$$d\phi = \left(\frac{\partial \phi}{\partial t} + \mathcal{L}_\Sigma \phi\right) dt + Tr\left[(\sqrt{\Sigma}dBQ + Q'dB'\sqrt{\Sigma})\mathcal{R}\phi\right], \tag{A-10}$$

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:

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

From the fundamental PDE (10), 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:

$$\begin{aligned}
e_t^\tau &= Tr\left[(A(\tau)Q' + QA(\tau))\Sigma_t\right] \tag{A-11} \\
&= 2Tr\left[QA(\tau)\Sigma_t\right].
\end{aligned}$$

For completeness, we also provide the expression for the instantaneous variance of the bond return. From equation (A-10), the diffusion part of the bond dynamics dP is given by $Tr\left[\left(\sqrt{\Sigma}dBQ + Q'dB'\sqrt{\Sigma}\right)A(\tau)P\right]$. Using Result 3 in Appendix D, the instantaneous variance of the bond returns is:

$$\begin{aligned}
Var_t\left(\frac{dP}{P}\right) &= Var_t\left[Tr\left(\left(\sqrt{\Sigma}dBQ + Q'dB'\sqrt{\Sigma}\right)A(\tau)\right)\right] \\
&= 4Tr\left[A(\tau)\Sigma A(\tau)Q'Q\right]dt.
\end{aligned}$$

∎

A.4. Dynamics of the Forward Rate

From the expression for the instantaneous forward rate:

$$f(t,\tau) = -\frac{\partial \log P(t, t+\tau)}{\partial \tau} = -\frac{\partial b(\tau)}{\partial \tau} - Tr\left[\frac{\partial A(\tau)}{\partial \tau}\Sigma_t\right],$$

the dynamics of $f(t,\tau)$ can be computed from:

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

By Ito's Lemma, we first obtain the dynamics of the logarithm of the bond price from equation (12):

$$d\log P = \left[-\frac{\partial b}{\partial \tau} - Tr\left(\frac{\partial A}{\partial \tau}\Sigma\right) + Tr\left[(\Omega\Omega' + M\Sigma + \Sigma M')A\right]\right]dt + Tr\left[(\sqrt{\Sigma}dBQ + Q'dB'\sqrt{\Sigma})A\right].$$

By noting that the first two terms in the drift of $d\log P$ equal the forward rate $f(t,\tau)$, we arrive at the instantaneous forward rate dynamics:

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

∎

A.5. Pricing of Zero-Bond Options

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 $T \geq S$:

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

A.5.1. Change of Drift for the Wishart Factors: The Forward Measure

To evaluate the two probabilities $\Pr_t^T$ and $\Pr_t^S$ in the above 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:

$$\frac{dP(t,S)}{P(t,S)} = r_t dt + Tr(\Theta'(t,S)dB_t^*) + Tr(\Theta(t,S)dB_t^{*'}), \tag{A-12}$$

where $\Theta(t, S) = \sqrt{\Sigma_t} A(t, S) Q'$, $A(t, S)$ and $\sqrt{\Sigma_t}$ are symmetric, and $A(t, S)$ solves the matrix Riccati equation (A-3). 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:

$$dB_t^* = dB_t^S + \sqrt{\Sigma_t}A(t,S)Q'dt, \tag{A-13}$$

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

$$dB_t^* = dB_t^T + \sqrt{\Sigma_t}A(t,T)Q'dt, \tag{A-14}$$

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

Remark 4. 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:

$$\frac{dP_t}{P_t} = r_t dt + \text{vec}(\Theta)'\text{vec}(dB_t^*) + \text{vec}(\Theta')'\text{vec}(dB_t^{*'}) \tag{A-15}$$

Then:

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

Reversing the vec operation, we obtain a matrix of Brownian motions: $dB_t^* = dB_t^T + \Theta dt$. $\square$

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

$$d\Sigma_t = \left(\Omega\Omega' + (M - Q')\Sigma_t + \Sigma_t(M' - Q)\right)dt + \sqrt{\Sigma_t}dB_t^*Q + Q'dB_t^{*'}\sqrt{\Sigma_t}. \tag{A-16}$$

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

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

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

A.5.2. 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:

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

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:

$$\Psi_t^S(iz; \tau) = E_t^S\left(e^{izTr[A(t+\tau,T)\Sigma_{t+\tau}]}\right), \tag{A-17}$$

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 characteristic function is itself of the exponentially affine form in Σ_t:

$$\Psi_t^S(iz; \tau) = e^{Tr[\hat{A}(z,\tau)\Sigma_t] + \hat{b}(z,\tau)}, \tag{A-18}$$

where $\hat{A}(z,\tau)$ and $\hat{b}(z,\tau)$ are, respectively, a symmetric matrix and a scalar with possibly complex coefficients, which solve the system of matrix Riccati equations (A-20)–(A-21) 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:

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

in which the integral can be evaluated by numerical methods.

The coefficients $\hat{A}(z,\tau)$ and $\hat{b}(z,\tau)$ in (A-18) are derived by the same logic as in Appendix A.2. By the Feynman-Kač argument applied to (A-17), Ψ_t^S solves the following PDE:

$$\frac{\partial\Psi_t^S}{\partial\tau} = \mathcal{L}_\Sigma\Psi_t^S. \tag{A-19}$$

Then, plugging for Ψ_t^S the expression (A-18), and collecting terms, gives the system of ordinary differential equations:

$$\frac{\partial\hat{b}(z,\tau)}{\partial\tau} = Tr[\Omega\Omega'\hat{A}(z,\tau)] \tag{A-20}$$

$$\frac{\partial\hat{A}(z,\tau)}{\partial\tau} = \hat{A}(z,\tau)M^S + M^{S'}\hat{A}(z,\tau) + 2\hat{A}(z,\tau)Q'Q\hat{A}(z,\tau), \tag{A-21}$$

where $M^S = M - Q'[I_n - QA(t,S)]$ results from the drift adjustment under the S-forward measure (given in Appendix A.5.1). The boundary conditions at $\tau = 0$ are:

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

By Radon's lemma, the solution for $\hat{A}(\tau)$ reads:

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

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 $\hat{b}(z,\tau)$ is obtained by integration:

$$\hat{b}(z,\tau) = \int_0^\tau Tr[\Omega\Omega'\hat{A}(u)]du = -\frac{k}{2}Tr\left[\log(ziA(S,T)\hat{C}_{12}(\tau) + \hat{C}_{22}(\tau)) + \tau M^{S'}\right].$$

Remark 5. Let $\mathbf{A} = \tau\begin{pmatrix} M & -2Q'Q \\ 0 & -M' \end{pmatrix}$. Since $\exp(\mathbf{A}) = \sum_{i=0}^\infty \frac{\mathbf{A}^i}{i!}$, then using the rules for the product of block matrices, it is easily seen that the blocks of the matrix $\exp(\mathbf{A})$ are of the simple form:

$$
\begin{aligned}
\hat{C}_{11}(\tau) &= e^{\tau M} \\
\hat{C}_{12}(\tau) &= 2\sum_{d=1}^\infty \frac{1}{d!}\tau^d\sum_{j=1}^d (-1)^j M^{d-j}Q'Q\left(M'\right)^{j-1} \\
\hat{C}_{21}(\tau) &= 0_{n\times n} \\
\hat{C}_{22}(\tau) &= e^{-\tau M'}.
\end{aligned}
$$

$\square$

B. Relation to the Quadratic Term Structure Models (QTSMs)

For an integer degree of freedom k, the $n\times n$ state matrix Σ_t can be represented as the sum of k outer products of n-dimensional Ornstein-Uhlenbeck (OU) processes: $\Sigma_t = \sum_{i=1}^k X_t^i X_t^{i'}$. The OU dynamics is given by $dX_t^i =$

$MX_t^i dt + Q' dW_t^i$, where dW_t^i is a n-vector of independent Brownian motions, and dW_t^i, dW_t^j are independent for $i \neq j$. First, we show the equivalence of the dynamics $d\left(\sum_{i=1}^{k} X_t^i X_t^{i\prime}\right)$ and $d\Sigma_t$. Then, we discuss the link to QTSMs.

By the independence of the OU processes, we can write $d(\sum_{i=1}^{k} X_t^i X_t^{i\prime}) = \sum_{i=1}^{k} d(X_t^i X_t^{i\prime})$, where:

$$
\begin{aligned}
d(X_t^i X_t^{i\prime}) &= X_t^i dX_t^{i\prime} + dX_t^i X_t^{i\prime} + dX_t^i dX_t^{i\prime} \\
&= X_t^i (MX_t^i dt + Q' dW_t^i)' + (MX_t^i dt + Q' dW_t^i) X_t^{i\prime} + (MX_t^i dt + Q' dW_t^i)(MX_t^i dt + Q' dW_t^i)' \\
&= (Q'Q + MX_t^i X_t^{i\prime} + X_t^i X_t^{i\prime} M')dt + Q' dW_t X_t^{i\prime} + X_t^i dW_t^{i\prime} Q.
\end{aligned}
$$

By summing over $i = 1, \ldots, k$, we get:

$$
d(\sum_{i=1}^{k} X_t^i X_t^{i\prime}) = (kQ'Q + M\sum_{i=1}^{k} X_t^i X_t^{i\prime} + \sum_{i=1}^{k} X_t^i X_t^{i\prime} M')dt + Q'\sum_{i=1}^{k} dW_t^i X_t^{i\prime} + \sum_{i=1}^{k} X_t^i dW_t^{i\prime} Q. \tag{B-23}
$$

Clearly, the drift in expression (B-23) is identical to the drift of $d\Sigma_t$ in equation (3). Next, we show the distributional equivalence between the diffusion parts in (3) and (B-23). It suffices to consider the instantaneous covariance between the following matrix forms (a, b, c, f are n-vectors):

$$
\begin{aligned}
Cov_t[a' d(X_t^i X_t^{i\prime})b, c' d(X_t^i X_t^{i\prime})f] &= E[a'(X_t^i dW_t^{i\prime} Q + Q' dW_t^i X_t^{i\prime})b \ c'(X_t^i dW_t^{i\prime} Q + Q' dW_t^i X_t^{i\prime})f] \\
&= (b'Q'Qf \, a'X_t^i X_t^{i\prime} c + b'Q'Qc \, a'X_t^i X_t^{i\prime} f + a'Q'Qf \, b'X_t^i X_t^{i\prime} c + a'Q'Qc \, b'X_t^i X_t^{i\prime} f)dt. \tag{B-24}
\end{aligned}
$$

Note that when a, b, c, f are different unit vectors in $\mathbb{R}^n$, expression (B-24) characterizes all second moments of $d(X_t^i X_t^{i\prime})$. By independence of $X_t^i, X_t^j, i \neq j$, the result easily extends to $\sum_{i=1}^{k} X_t^i X_t^{i\prime}$:

$$
Cov_t[a' d(\sum_{i=1}^{k} X_t^i X_t^{i\prime})b, c' d(\sum_{i=1}^{k} X_t^i X_t^{i\prime})f]
$$

$$
= (b'Q'Qf \, a'\sum_{i=1}^{k} X_t^i X_t^{i\prime} c + b'Q'Qc \, a'\sum_{i=1}^{k} X_t^i X_t^{i\prime} f + a'Q'Qf \, b'\sum_{i=1}^{k} X_t^i X_t^{i\prime} c + a'Q'Qcb'\sum_{i=1}^{k} X_t^i X_t^{i\prime} f)dt. \tag{B-25}
$$

The expression (B-25) is equivalent to the covariation between different elements of Σ_t in Result 1 of Appendix D. Thus, the diffusion parts of $d(\sum_{i=1}^{k} X_t^i X_t^{i\prime})$ and $d\Sigma_t$ are distributionally equivalent.

When $k = 1$, Σ_t becomes singular. We can recast the model in terms of the single OU vector process as:

$$
d(X_t^1 X_t^{1\prime}) = (Q'Q + MX_t^1 X_t^{1\prime} + X_t^1 X_t^{1\prime} M')dt + Q' dW_t X_t^{1\prime} + X_t^1 dW_t^{1\prime} Q, \tag{B-26}
$$

where $dX_t^1 = MX_t^1 dt + Q' dW_t^1$. In this special case, our model has a direct analogy to a $n-$factor quadratic term structure model (QTSM) of Ahn, Dittmar, and Gallant (2002) and Leippold and Wu (2002), with the underlying OU dynamics of X_t^1 that has no intercept. Table VI compares the corresponding elements in the two settings. We preserve the notation used by Ahn, Dittmar, and Gallant (2002) for the QTSMs, and map it into the notation used throughout our paper.

B.1. Unique invertibility of the state

The affine property of yields in the elements of Σ_t represents an advantage of our setting over the QTSMs. When $k \geq n$, the state variables in Σ_t can be *uniquely* backed out from the observed yields. In particular, an $n \times n$ state matrix Σ_t can be identified from $\bar{n} = \frac{n(n+1)}{2}$ yields. Let us stack the yields in a vector using the fact that $Tr[A(\tau)\Sigma_t] = [\text{vec}A(\tau)]' \text{vec}\Sigma_t$:

$$
\begin{pmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_t(\tau_{\bar{n}}) \end{pmatrix} = -\begin{pmatrix} \frac{1}{\tau_1}b(\tau_1) \\ \frac{1}{\tau_2}b(\tau_2) \\ \vdots \\ \frac{1}{\tau_n}b(\tau_{\bar{n}}) \end{pmatrix} - \begin{pmatrix} \frac{1}{\tau_1}\text{vec}A(\tau_1)' \\ \frac{1}{\tau_2}\text{vec}A(\tau_2)' \\ \vdots \\ \frac{1}{\tau_n}\text{vec}A(\tau_{\bar{n}})' \end{pmatrix} \text{vec}(\Sigma_t),
$$

or in a short-hand vector-matrix notation:

$$
\vec{y}_t = -\vec{b} - \mathbf{A} \, \text{vec}(\Sigma_t).
$$

To be able to invert the last expression for the unique elements of Σ_t, we prune the non-unique elements of $\mathbf{A}$ and $\text{vec}(\Sigma_t)$ by using the half-vectorization:

$$
\vec{y}_t = -\vec{b} - \mathbf{A}S_n \, \text{vech}(\Sigma_t),
$$

Table VI: Quadratic versus Wishart factor model

The table presents a mapping between the QTSM of Ahn, Dittmar, and Gallant (2002) and the WTSM. For readability, we preserve the respective notations.

QTSM(n)	WTSM($n \times n$) $k = 1$
State variables	
$dY_t = (\mu + \xi Y_t)dt + \Sigma dW_t$	$dX_t = MX_t dt + Q'dW_t$
	$\mu = 0, \ \xi = M, \ \Sigma = Q'$
Short rate	
$r_t = \alpha + \beta'Y_t + Y_t'\Psi Y_t$	$r_t = X_t'(D - I_n)X_t$
$\beta = 0$ (identification)	$\alpha = 0, \ \beta = 0$
	$\Psi = D - I_n$
Market price of risk	
$\Lambda_t = \delta_0 + \delta_1 Y_t$	$\Lambda_t = X_t$
	$\delta_0 = 0, \ \delta_1 = I_n$

where S_n is a duplication matrix of dimension $n^2 \times \frac{n(n+1)}{2}$ such that $S_n \text{vech}(\Sigma_t) = \text{vec}(\Sigma_t)$. It follows that the state is identified from yields as:

$$\text{vech}(\Sigma_t) = -(\mathbf{A}S_n)^{-1}(\vec{y}_t + \vec{b}).$$

C. Moments of the factors and yields

To provide a general formulation for the moments of the Wishart process, we proceed via the conditional Laplace transform. The derivation, which starts from the Laplace transform of the discrete time process, holds true also for non-integer degrees of freedom k, and thus does not require the restrictive interpretation of Σ_t as the sum of outer products of OU processes. The assumption of an integer k is implicit only via the mapping between the discrete and continuous time parameters of the process, which we discuss next.

C.1. Moments of the discrete time Wishart process

The Wishart process allows for an exact discretization, i.e. there exists an explicit mapping between the discrete and continuous time parameters of the process:

$$\Phi_\Delta = e^{M\Delta} \tag{C-27}$$

$$V_\Delta = \int_0^\Delta \Phi_s Q'Q\Phi_s' ds, \tag{C-28}$$

where Δ denotes the discretization horizon. These expressions are the well-known conditional moments of the underlying multivariate OU process and hence are stated without proof.[30] For the tractability of subsequent derivations, we use the above mapping along with the conditional Laplace transform of the *discrete* time process to compute the moments of the *continuous* time process.

C.1.1. Laplace transform of the Wishart process

Let $\Sigma_\Delta|\Sigma_0 \sim Wis(k, \Phi, V)$. The conditional Laplace transform of $\Sigma_\Delta|\Sigma_0 = \Sigma$ is given by (see e.g. Muirhead, 1982, p. 442):

$$\Psi_\Delta(\Theta) := E_0[\exp(Tr(\Gamma\Sigma_\Delta))|\Sigma_0 = \Sigma]$$

$$= \exp\left\{ Tr[\Phi_\Delta'\Gamma(I_n - 2V_\Delta\Gamma)^{-1}\Phi_\Delta\Sigma] - \frac{k}{2}\log\det(I_n - 2V_\Delta\Gamma) \right\},$$

where $\Gamma := \Gamma(\Theta)$ and $\Gamma = (\gamma_{ij})$, $i, j = 1, ..., n$ with $\gamma_{ij} = \frac{1}{2}(1 + \delta_{ij})\theta_{ij}$, where δ_{ij} is the Kronecker delta:

---

[30]See e.g. Fisher and Gilles (1996b) for a detailed derivation of the conditional moments of a general affine process.

$$\delta_{ij} = \left\{ \begin{array}{l} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j. \end{array} \right.$$

The cumulant generating function is:

$$\mathcal{K}_\Delta(\Theta) := \log \Psi_\Delta(\Theta) = \left\{ Tr[\Phi'_\Delta \Gamma(I_n - 2V_\Delta\Gamma)^{-1}\Phi_\Delta\Sigma] - \frac{k}{2}\log\det(I_n - 2V_\Delta\Gamma) \right\}.$$

In the sequel, we use the shorthand notation Γ to be understood as $\Gamma(\Theta)$. For brevity, the subscript Δ at Φ and V denoting the discretization horizon is neglected.

C.1.2. Moments of the Wishart process

The moments of the process are obtained by evaluating the derivatives of the Laplace transform (cumulant generating function) at $\Gamma = 0$.

We apply the following definition of the derivative of some function F (possibly matrix-valued) with respect to the matrix argument (Magnus and Neudecker, 1988, p. 173):

$$\mathcal{D}F(\Theta) := \frac{d\,\text{vec}F(\Theta)}{d(\text{vec}\Theta)'}.$$

Lemma 3. *The closed-form expression for the first order derivative of the conditional cumulant generating function $\mathcal{K}(\Theta)$:*

$$\mathcal{D}\mathcal{K}(\Theta) = \frac{d\,\mathcal{K}(\Theta)}{d\,(\text{vec}\Theta)'} = P_1(\Theta) + P_2(\Theta) \tag{C-29}$$

where

$$P_1(\Theta) = \text{vec}[(I_n - 2V\Gamma)^{-1}\Phi\Sigma\Phi'(I_n - 2\Gamma V)^{-1}]'$$
$$P_2(\Theta) = k\,\text{vec}[V(I_n - 2\Gamma V)^{-1}]'.$$

Proof. The proof is an application of matrix calculus rules to the cumulant generating function:

$$\mathcal{D}K(\Theta) = \frac{d\,K(\Theta)}{d\,(\text{vec}\Theta)'}$$
$$= \underbrace{\frac{d\,Tr[\Phi'\Gamma(I_n - 2V\Gamma)^{-1}\Phi\Sigma]}{d\,(\text{vec}\Theta)'}}_{P_1(\Theta)} - \underbrace{\frac{k}{2}\frac{d\,(\log\det(I_n - 2V\Gamma))}{d\,(\text{vec}\Theta)'}}_{P_2(\Theta)}$$

$\square$

Corollary 4 (First moment of the Wishart process)**.** *From the results in Lemma 3, the expression for the first conditional moment of $\Sigma_\Delta | \Sigma_0 = \Sigma$ follows immediately:*

$$E((\text{vec}\Sigma_\Delta)'|\Sigma_0 = \Sigma) = \frac{dK(0)}{d(\text{vec}\Theta)'} = P_1(0) + P_2(0) = \big(\text{vec}(\Phi\Sigma\Phi' + kV)\big)'.$$

Equivalently, in matrix notation:

$$E(\Sigma_\Delta | \Sigma_0 = \Sigma) = \Phi\Sigma\Phi' + kV.$$

$\square$

Remark 6. The first moment of the Wishart process is straightforward to obtain for the *integer* degree of freedom k. Let x_t^i denote the OU process with the discrete time dynamics:

$$x_t^i = \Phi x_{t-\Delta}^i + \epsilon_t^i, \quad \text{where } \epsilon_t^i \sim N(0, V).$$

For an integer k, the Wishart is constructed as $\Sigma_t = \sum_i x_t^i x_t^{i\prime}$, and we have:

$$\sum_i x_t^i x_t^{i\prime} = \sum_i (\Phi x_{t-\Delta}^i + \epsilon_t)(\Phi x_{t-\Delta}^i + \epsilon_t)'$$
$$E_{t-\Delta}(\sum_i x_t^i x_t^{i\prime}) = \Phi\Sigma_t\Phi' + E_{t-\Delta}(\sum_i \epsilon_t^i \epsilon_t^{i\prime}) = \Phi\Sigma_t\Phi' + kV.$$

48

In contrast, our derivation via the Laplace transform is more generic as it does not rely (at least in the discrete time case) on the link between the Wishart and the OU process. Thus, the assumption of integer degrees of freedom is not required. Moreover, if we relax the restriction that $\Omega\Omega' = kQ'Q$ in the drift of the continuous time Wishart, then the conditional first moment can be further generalized to:

$$E_t\left(\Sigma_{t+\Delta}\right) = e^{\Delta M}\Sigma_t e^{\Delta M'} + \int_0^\Delta e^{sM}\Omega\Omega' e^{sM'}ds,$$

where the last expression follows from the Laplace transform of the continuous time process. $\qquad\square$

The second derivative of $\mathcal{K}(\Theta)$ is defined as (see Magnus and Neudecker, 1988, p. 188):

$$\mathcal{HK}(\Theta) = \mathcal{D}\left(\mathcal{DK}(\Theta)\right)' = \frac{d\,\mathrm{vec}\left[\mathcal{DK}(\Theta)\right]'}{d\left(\mathrm{vec}\Theta\right)'}. \tag{C-30}$$

Lemma 5. *The closed-form expression for the second order derivative (C-30) is defined as:*

$$\mathcal{HK}(\Theta) = \left(I_{n^2} + K_n\right)\left\{R_1(\Theta) \otimes R_2(\Theta) + K\left[R_2(\Theta) \otimes R_2(\Theta)\right] + R_2(\Theta) \otimes R_1(\Theta)\right\},$$

where

$$R_1(\Theta) = \left(I_n - 2V\Gamma\right)^{-1}\Phi\Sigma\Phi'\left(I_n - 2\Gamma V\right)^{-1}$$
$$R_2(\Theta) = \left(I_n - 2V\Gamma\right)^{-1}V,$$

where I_{n^2} is an $n^2 \times n^2$ identity matrix, and $K_{n,n}$ is the commutation matrix defined as: $\mathrm{vec}S' = K_{n,n}\mathrm{vec}S$ for some square matrix S.

Proof. The above expression follows from taking the following derivatives:

$$\frac{d\,\mathrm{vec}\left[\mathcal{DK}(\Theta)\right]'}{d\left(\mathrm{vec}\Theta\right)'} = \frac{d}{d(\mathrm{vec}\Theta)'}\left[\frac{d\,\mathcal{K}(\Theta)}{d\left(\mathrm{vec}\Theta\right)'}\right]' = \frac{d\,P_1'(\Theta)}{d(\mathrm{vec}\Theta)'} + \frac{d\,P_2'(\Theta)}{d(\mathrm{vec}\Theta)'}.$$

$\qquad\square$

Corollary 6 (Second moment of the Wishart process). *With the results in Lemma 5, the second conditional moments of $\Sigma_\Delta|\Sigma_0 = \Sigma$ follow:*

$$E[\mathrm{vec}(\Sigma_\Delta)\mathrm{vec}(\Sigma_\Delta)'|\Sigma_0 = \Sigma] = \mathcal{H}\Psi(0) = \left[\mathcal{DK}(0)\right]'\left[\mathcal{DK}(0)\right] + \mathcal{HK}(0).$$

Therefore:

$$\begin{aligned}E[\mathrm{vec}(\Sigma_\Delta)\mathrm{vec}(\Sigma_\Delta)'|\Sigma_0 = \Sigma] = {}& \mathrm{vec}(\Phi\Sigma\Phi' + kV)\mathrm{vec}(\Phi\Sigma\Phi' + kV)' \\ & + (I_{n^2} + K_{n,n})[\Phi\Sigma\Phi' \otimes V + k(V \otimes V) + V \otimes \Phi\Sigma\Phi'].\end{aligned}$$

Importantly, Result 4 and 6 cover the general case of non-integer degrees of freedom $k > n-1$. The integer degrees of freedom is implicit only through the mapping between the continuous and discrete time parameters in (C-27)–(C-28).

C.1.3. Closed-form expressions for the matrix integrals

The conditional moments of the Wishart involve the evaluation of:

$$V_\tau = \int_0^\tau \Phi_s Q'Q\Phi_s'ds. \tag{C-31}$$

The closed-form expression for this integral is given as:

$$\int_0^\tau e^{Ms}Q'Q e^{M's}ds = -\frac{1}{2}\hat{C}_{12}(\tau)\hat{C}_{11}'(\tau), \tag{C-32}$$

where $\hat{C}_{11}(\tau)$ and $\hat{C}_{12}(\tau)$ are blocks of the matrix exponential associated with the coefficients of the Laplace transform of the continuous time Wishart process:

$$\exp\left[\tau\begin{pmatrix} M & -2Q'Q \\ 0 & -M' \end{pmatrix}\right] = \begin{pmatrix} \hat{C}_{11}(\tau) & \hat{C}_{12}(\tau) \\ \hat{C}_{21}(\tau) & \hat{C}_{22}(\tau) \end{pmatrix}. \tag{C-33}$$

In applications, the expression (C-31) turns out numerically stable (for finite maturities) and computationally very efficient.

Proof. The elements of the matrix exponential can be expressed as (see Van Loan, 1978, Thm. 1):

$$\hat{C}_{11}(\tau) = e^{M\tau}$$
$$\hat{C}_{12}(\tau) = \int_0^\tau e^{M(\tau-s)}(-2Q'Q)e^{-M's}ds$$
$$\hat{C}_{21}(\tau) = 0_{n\times n}$$
$$\hat{C}_{22}(\tau) = e^{-M'\tau}$$

Postmultiplying the expression for $\hat{C}_{12}(\tau)$ by $\hat{C}'_{11}(\tau) = e^{M'\tau}$ and applying the change of variable $u = \tau - s$, yields:

$$\hat{C}_{12}(\tau)\hat{C}'_{11}(\tau) = -\int_0^\tau e^{Mu}(2Q'Q)e^{M'u}du.$$

After reformulating, the result follows:

$$\int_0^\tau e^{Mu}Q'Qe^{M'u}du = -\frac{1}{2}\hat{C}_{12}(\tau)\hat{C}'_{11}(\tau).$$

$\square$

A similarly tractable and computationally efficient expression is readily available for the limit of the integral (C-31), which occurs in the unconditional moments of the Wishart process. In vectorized form, we have:

$$\text{vec}V_\infty = \text{vec}\left(\lim_{\tau\to\infty}\int_0^\tau \Phi_s Q'Q\Phi'_s ds\right)$$
$$= -[(I_n \otimes M) + (M \otimes I_n)]^{-1}\text{vec}(Q'Q). \tag{C-34}$$

Proof. We exploit the relationship between the integral (C-34) and the solution to the following Lyapunov equation:

$$MX + XM' = Q'Q,$$

which can be written as (see e.g. Laub, 2005, p. 145):

$$X = -\int_0^\infty e^{Ms}Q'Qe^{M's}ds.$$

At the same time, the solution for X can be expressed in closed-form using the relationship between the vec operator and the Kronecker product, $\text{vec}(IXM) = (M' \otimes I)\text{vec}X$. This results in:

$$\text{vec}X = [(I_n \otimes M) + (M \otimes I_n)]^{-1}\text{vec}(Q'Q).$$

Thus, the integral can be efficiently computed as:

$$\text{vec}\left(\int_0^\infty e^{Ms}Q'Qe^{M's}ds\right) = -[(I_n \otimes M) + (M \otimes I_n)]^{-1}\text{vec}(Q'Q).$$

$\square$

C.2. Moments of yields

For the unconditional moments of yields to exist, we require $M < 0$. Since the unconditional first moment is straightforward to obtain, we only focus on the second moment. Its general specification comprises a cross moments of yields, one of which possibly lagged:

$$Cov(y_{t+s}^{\tau_1}, y_t^{\tau_2}) = \frac{1}{\tau_1 \tau_2} \{ E\left[Tr(A_{\tau_1}\Sigma_{t+s})Tr\left(A_{\tau_2}\Sigma_t\right)\right] - E\left[Tr(A_{\tau_1}\Sigma_{t+s})\right] E\left[Tr\left(A_{\tau_2}\Sigma_t\right)\right] \}$$

$$= \frac{1}{\tau_1 \tau_2} \{ E\left[(\text{vec}A_{\tau_1})' \, \text{vec}\Sigma_{t+s} \, (\text{vec}\Sigma_t)' \, \text{vec}A_{\tau_2} \right] - (\text{vec}A_{\tau_1})' \, \text{vec}E\Sigma_{t+s} \, (\text{vec}E\Sigma_t)' \, \text{vec}A_{\tau_2} \}$$

$$= \frac{1}{\tau_1 \tau_2} \{ E\left[(\text{vec}A_{\tau_1})' \left[\text{vec}\left(\Phi_s \Sigma_t \Phi_s'\right) \right] (\text{vec}\Sigma_t)' \, (\text{vec}A_{\tau_2}) \right] - (\text{vec}A_{\tau_1})' \left[\text{vec}\left(\Phi_s E\Sigma_t \Phi_s'\right) \right] (\text{vec}E\Sigma_t)' \, \text{vec}A_{\tau_2} \}$$

$$= \frac{1}{\tau_1 \tau_2} (\text{vec}A_{\tau_1})' \left(\Phi_s \otimes \Phi_s\right) \left[E(\text{vec}\Sigma_t(\text{vec}\Sigma_t)') - (\text{vec}E\Sigma_t)\,(\text{vec}E\Sigma_t)' \right] \text{vec}A_{\tau_2}$$

$$= \frac{1}{\tau_1 \tau_2} \text{vec}\left(\Phi_s A_{\tau_1} \Phi_s'\right)' \left[Cov(\text{vec}\Sigma_t)\right] \text{vec}A_{\tau_2},$$

where when moving from the second to the third line, we have used the law of iterated expectations:

$$E\left[\text{vec}\Sigma_{t+s}\,(\text{vec}\Sigma_t)'\right] = E\left[E_t\left(\text{vec}\Sigma_{t+s}\,(\text{vec}\Sigma_t)'\right)\right]$$
$$= E\left[\text{vec}\left(\Phi_s \Sigma_t \Phi_s'\right)(\text{vec}\Sigma_t)' + k\,\text{vec}V_s\,(\text{vec}\Sigma_t)'\right].$$

To obtain a contemporaneous covariance, note that $\Phi_{s=0} = e^0 = I_n$. Therefore:

$$Cov(y_t^{\tau_1}, y_t^{\tau_2}) = \frac{1}{\tau_1 \tau_2} (\text{vec}A_{\tau_1})' \left[Cov(\text{vec}\Sigma_t)\right] \text{vec}A_{\tau_2}.$$

D. Useful Results for the Wishart Process

Result 1. The following result facilitates the computation of the second moments of the Wishart process. Given $n \times n$ Wishart SDE $d\Sigma$ in equation (3) and arbitrary n-dimensional vectors a, b, c, f it follows:

$$Cov_t\left(a'd\Sigma_t b, c'd\Sigma_t f\right) = \left(a'Q'Qfb'\Sigma_t c + a'Q'Qcb'\Sigma_t f + b'Q'Qfa'\Sigma_t c + b'Q'Qca'\Sigma_t f\right) dt.$$

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 themselves affine functions of Σ.

Using the above results, it is straightforward to compute the (cross-)second moments of factors in the 2×2 case:

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

where Σ_{ij} and Q_{ij} denote the ij-th element of matrix Σ and Q, respectively. More generally, for an arbitrary dimension n of the state matrix, we obtain:

$$d\langle\Sigma_{ii}\rangle_t = 4\Sigma_{ii}Q^{i'}Q^i dt$$
$$d\langle\Sigma_{ii}, \Sigma_{jj}\rangle_t = 4\Sigma_{ij}Q^{i'}Q^j dt,$$

where Q^i, Q^j denote the i-th and j-th column of the Q matrix, respectively.

Result 2. The special case of the above result has been given by Gourieroux (2006):

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

where for any n-dimensional vectors u and v it holds that:

$$E_t\left(dB_t uv'dB_t\right) = E_t\left(dB_t'uv'dB_t'\right) = vu'dt$$
$$E_t\left(dB_t uv'dB_t'\right) = E_t\left(dB_t'uv'dB_t\right) = v'uI_n dt.$$

Result 3. Given square matrices A and C, we have:

$$Cov_t\left[Tr(Ad\Sigma_t), Tr(Cd\Sigma_t)\right] = Tr\left[(A+A')\Sigma_t(C+C')Q'Q\right] dt. \tag{D-35}$$

Moreover, for a square matrix A it holds that:

$$Var_t\left[Tr(Ad\Sigma_t)\right] = Tr\left[(A+A')\Sigma_t(A+A')Q'Q\right] dt > 0 \quad \text{iff} \quad \Sigma_t > 0. \tag{D-36}$$

In contrast to Gourieroux and Sufana (2003), in obtaining these results we do not impose definiteness restrictions on A and C.

Proof. To derive the result, we can directly consider the expectation of the product of the two traces:

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

The above expression can be split into the sum of four expectations. For brevity, we only provide the derivation for one of the terms:

$$E_t\left[Tr\left(A\sqrt{\Sigma_t}dB_tQ\right)\ Tr\left(CQ'dB_t'\sqrt{\Sigma_t}\right)\right] =$$

$$= E_t\left[\left(\text{vec}\left(QA\sqrt{\Sigma_t}\right)'\right)'(\text{vec } dB_t)\left(\text{vec}\left(\sqrt{\Sigma_t}CQ'\right)'\right)'(\text{vec } dB_t')\right]$$

$$= E_t\left[\left(\text{vec}\left(QA\sqrt{\Sigma_t}\right)'\right)'(\text{vec } dB_t)(\text{vec } dB_t)' K_{n,n}\left(\text{vec}\left(\sqrt{\Sigma_t}CQ'\right)'\right)\right]$$

$$= \left(\text{vec}\left(QA\sqrt{\Sigma_t}\right)'\right)'\left(\text{vec}\left(\sqrt{\Sigma_t}CQ'\right)\right) dt = Tr\left[CQ'QA\Sigma\right] dt.$$

where $K_{n,n}$ is the commutation matrix (thus $K_{n,n} = K'_{n,n}$), and we apply the following facts:

$$E_t\left[(\text{vec } dB_t)(\text{vec } dB_t)'\right] = E_t\left[(\text{vec } dB_t')(\text{vec } dB_t')'\right] = I_{n^2}dt$$

and

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

To prove the positivity of $Var_t\left[Tr(Ad\Sigma_t)\right]$ in equation (D-36), note that:

$$Tr\left[(A+A')\Sigma_t(A+A')Q'Q\right] = Tr\left[Q(A+A')\Sigma_t(A+A')Q'\right].$$

The expression within the trace on the RHS is a congruent transformation of the Wishart matrix Σ, which (by Sylvester's law) can change the values but not the signs of matrix eigenvalues. Thus, provided that $\Sigma_t > 0$, it follows that $Q(A+A')\Sigma_t(A+A')Q' > 0$. Combining this result with the properties of the trace we have:

$$Tr\left[Q(A+A')\Sigma_t(A+A')Q'\right] = \sum_{i=1}^{n}\lambda_i > 0,$$

where λ_i denotes the eigenvalue of $Q(A+A')\Sigma_t(A+A')Q'$. $\qquad\square$

Result 4. 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)).

Proof. By the singular value decomposition, a symmetric (positive or negative) definite $n \times n$ 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 $n \times n$ matrix $D = \sum_{j=1}^{n}a_j a_j'$, where $a_j = \sqrt{\lambda_j}m_j$. $\qquad\square$

E. Details on the Estimation Approach

E.1. The 2×2 Model

The 2×2 model comprises 9 parameters: the elements of matrices M, Q' (both lower triangular) and D (symmetric), plus an integer value of the degrees of freedom parameter k. The estimation is based on 11 moment conditions (see Table VII). The following steps describe our optimization technique:

Step 1. For M, Q and D, generate $I_{max} = 200$ from the uniform distribution under the condition that: (*i*) the diagonal elements of M are negative, (*ii*) the eigenvalues of $D - I_n$ are positive.

Step 2. Select the possible degrees of freedom k on a grid of integers from 1 to 9.

Step 3. Run 9×200 optimizations (for each k and each I_{max}) in order to select 10 parameter sets with the lowest value of the loss function.

Step 4. To determine the final parameter values, improve on the selected parameter sets using a gradient-based optimization routine (Matlab `lsqnonlin`).

This optimization procedure gives rise to $k = 3$ and to the parameters:

$$D = \begin{pmatrix} 1.0281 & 0.0047 \\ 0.0047 & 1.0018 \end{pmatrix},$$

$$M = \begin{pmatrix} -0.1263 & 0 \\ 0.0747 & -0.6289 \end{pmatrix},$$

$$Q = \begin{pmatrix} 0.4326 & -2.9238 \\ 0 & -0.5719 \end{pmatrix}.$$

The matrices satisfy the theoretical requirements: (*i*) the M matrix is negative definite, i.e. Σ_t does not explode; (*ii*) the Q matrix is invertible, i.e. Σ_t is reflected towards positivity whenever the boundary of the state space is reached; and (*iii*) the $D - I_2$ matrix is positive definite with eigenvalues (0.0010, 0.0289), i.e. the positivity of yields is ensured. Table VII summarizes the percentage errors for the respective moment conditions and yields used in estimation.

Table VII: Fitting errors for the 2×2 model

The table presents percentage fitting errors for the moments of the 6-month, 2-year and 10-year yields. The percentage error is the difference between the model-implied and the empirical value of a given moment per unit of its empirical value. All errors are in percent.

	Maturities used		Error (%)	
Average y_t^τ	6M, 2Y, 10Y	-0.27	-0.59	0.86
Volatility y_t^τ	6M, 2Y, 10Y	0.35	-0.19	-0.78
Corr. $y^{\tau 1}, y^{\tau 2}$	(6M,2Y), (6M,10Y), (2Y,10Y)	-1.77	1.63	2.44
CS coeff.	$n = 2$Y, 10Y; $m = 6$M	-0.15	0.00	–

E.2. The 3×3 Model

Our 3×3 framework is described by 18 parameters in M, Q and D matrices, plus the degrees of freedom parameter k. The estimation involves 24 moment conditions summarized in Table VIII. The optimization technique follows the same steps as for the 2×2 model. The final set of parameters has $k = 3$ degrees of freedom, and the following M, Q and D matrices:

$$D = \begin{pmatrix} 1.0396 & 0.0224 & -0.0094 \\ 0.0224 & 1.1412 & 0.0117 \\ -0.0094 & 0.0117 & 1.0043 \end{pmatrix},$$

$$M = \begin{pmatrix} -0.8506 & 0 & 0 \\ 0.2249 & -0.0787 & 0 \\ -2.2800 & -2.4125 & -0.9121 \end{pmatrix},$$

$$Q = \begin{pmatrix} 1.3276 & -0.2950 & 5.2410 \\ 0 & -0.0453 & 0.0667 \\ 0 & 0 & -0.6443 \end{pmatrix}.$$

Note from the diagonal elements, the M matrix is negative definite, and the Q matrix is invertible. The matrix $D - I_3$, however, is not positive definite with eigenvalues $(-0.0001, 0.0387, 0.1465)$. Even though there is a theoretical probability that yields become negative, the empirical frequency of such occurrences is zero across all maturities (based on the simulated sample of 72000 monthly observations). The violation of positive definiteness is thus inconsequential.

Table VIII: Fitting errors for the 3×3 model

The table presents percentage fitting errors for the moments of the 6-month, 2-year, 5-year and 10-year yields. The percentage error is the difference between the model-implied and the empirical value of a given moment per unit of its empirical value. All errors are in percent.

Moment	Maturities used	Error (%)			
Average y_t^τ	6M, 2Y, 5Y, 10Y	0.22	-0.40	0.16	-0.03
Volatility y_t^τ	6M, 2Y, 5Y, 10Y	0.85	-2.45	2.06	-0.23
Corr. $y^{\tau 1}, y^{\tau 2}$	(6M,2Y), (6M,5Y), (6M,10Y), (2Y,10Y)	-0.39	0.17	0.16	-0.23
Corr. $\Delta y^{\tau 1}, \Delta y^{\tau 2}$	(6M,2Y), (6M,10Y), (2Y,10Y)	-0.01	-0.03	0.02	–
CS coeff.	$n = 2Y, 10Y;\ m = 6M$	-0.00	-0.00	–	–
Forward rate volatility	6M→2Y, 2Y→5Y, 5Y→10Y	-3.38	4.99	-2.15	–

a. Realized 10Y yield volatility (left axis) vs. 10Y–3M spread (right axis)

b. 1st episode: Jan94 - Feb95

c. 2nd episode: Jun04 - Dec05
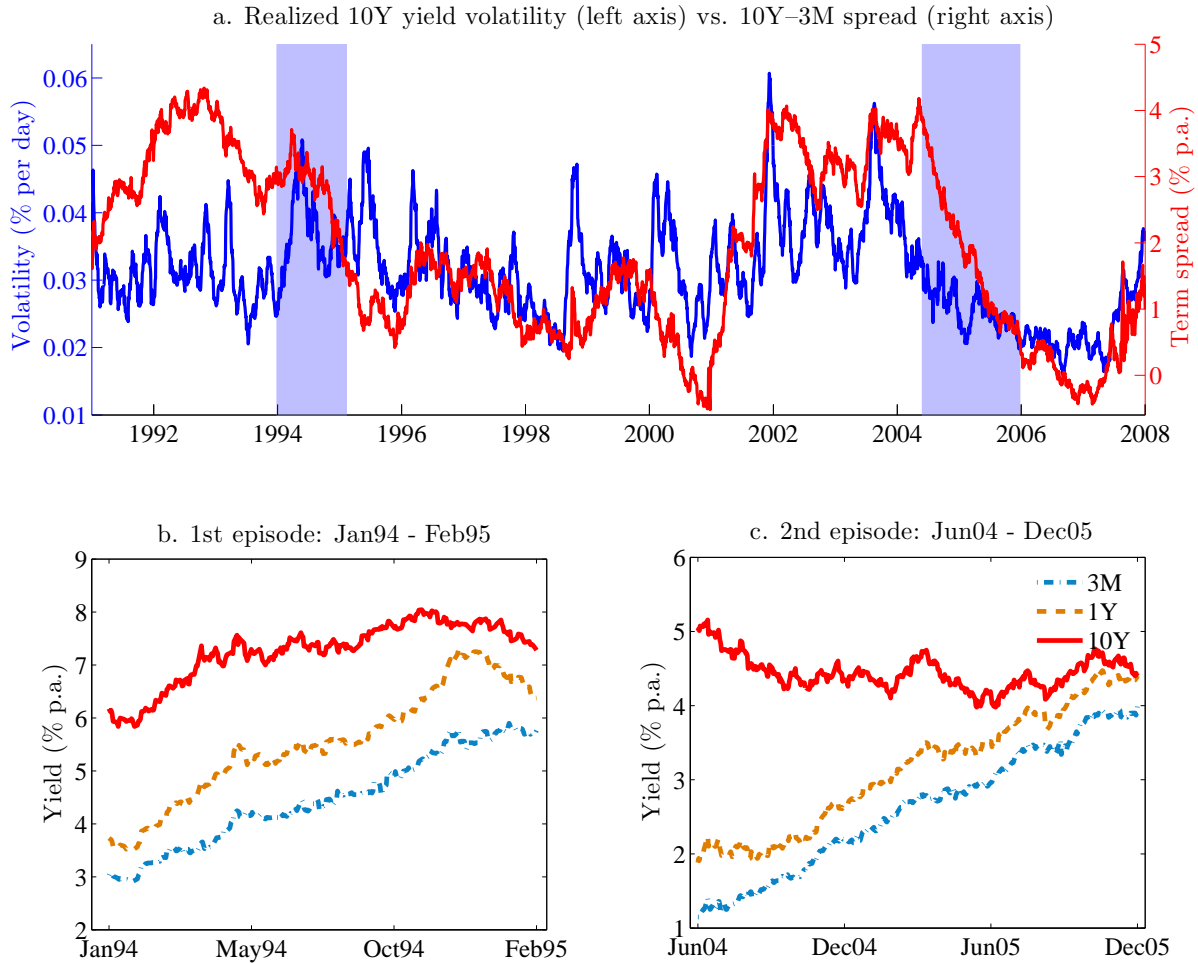
Figure 1: Term spreads and yield volatilities during 1994/95 and 2004/05 tightenings

Panel *a* plots interest rate volatilities (left axis) and yield spread (right axis) covering the period from 1991:01 to 2007:12. The shadings mark the 1994/95 and 2004/05 tightenings. The yield volatility is the 22-day moving average of realized daily volatilities ($v_\tau(t+h), h = 1$ day) obtained from high-frequency returns on 10-year Treasury note futures, where $v_\tau^2(t+h) = \frac{1}{\tau^2}\sum_{i=1}^n r^2(t + \frac{ih}{n}), n = 40$, and $r(t + \frac{ih}{n})$ is the 10-minute log return$\times$100 on the futures contract. The spread is computed as the difference between the 10-year yield minus the 3-month T-bill rate. Panels *b* and *c* give a zoomed view on the dynamics of the 3-month, 1-year and 10-year yields during 1994/95 and 2004/05 episodes, respectively. Data sources: Tick-by-tick interest futures prices are from TickData.com, zero yields—from Gurkaynak, Sack, and Wright (2006) files, and 3-month T-bill rate—from Fed's H.15 files.
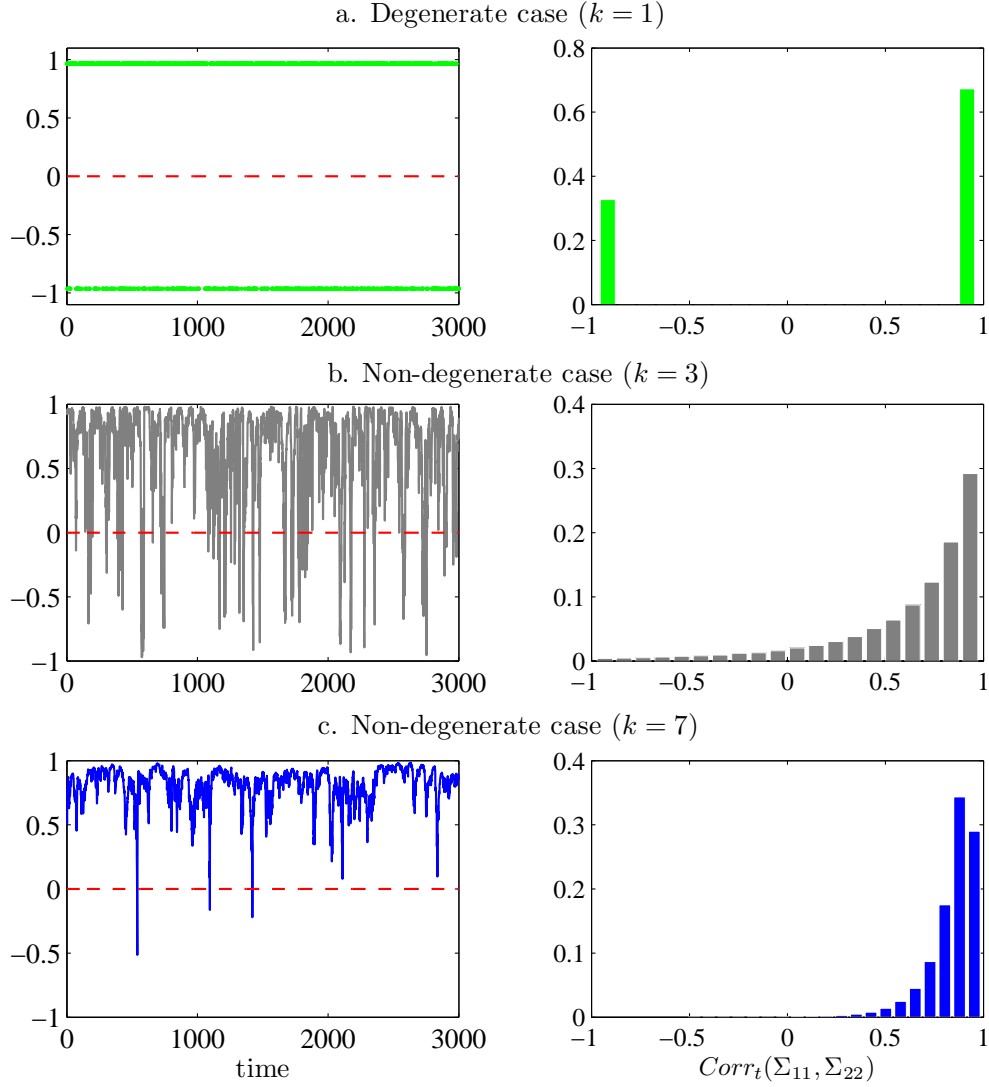
Figure 2: Instantaneous correlations of positive factors for different k's in the 2×2 WTSM

Panels a, b, and c plot the instantaneous correlations between the diagonal factors Σ_{11} and Σ_{22} in the 2×2 WTSM with $k = 1, 3$ and 7 degrees of freedom, respectively. The panels on the left show the consecutive realizations of the instantaneous correlations calculated according to equation (24), while the panels on the right display their respective histograms. In the degenerate case of $k = 1$, the WTSM narrows down to a QTSM.
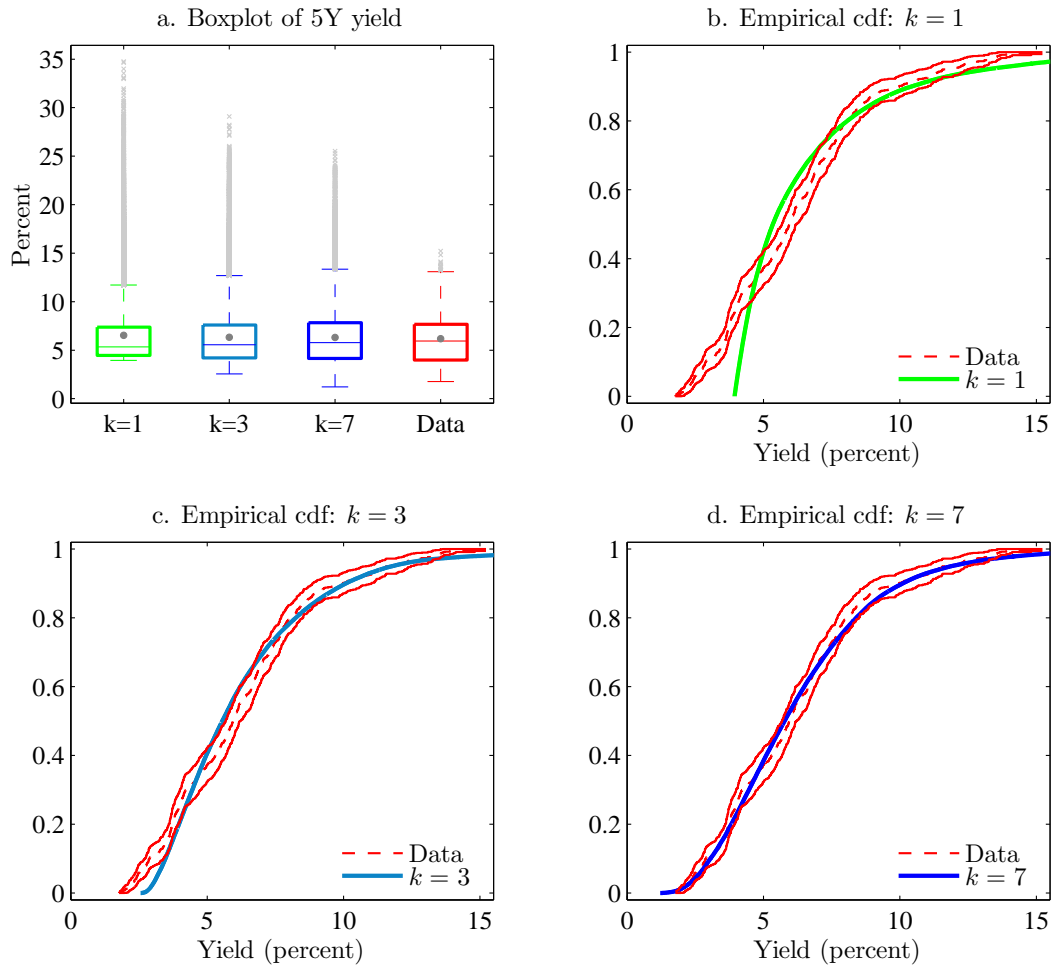
Figure 3: Unconditional distribution of the 5-year yield for different k in the 2×2 WTSM

Panel a presents boxplots of the 5-year yield in the 2×2 model with $k = 1, 3$ and 7 degrees of freedom, respectively. The results are juxtaposed with the US 5-year yield (1952:01–2005:06). Panels b, c and d illustrate the related empirical cumulative distribution functions both for the model (thick line) and for the data (dashed thin line). The solid thin lines mark the 99% upper and lower confidence bounds computed from the data with the standard Greenwood's formula.
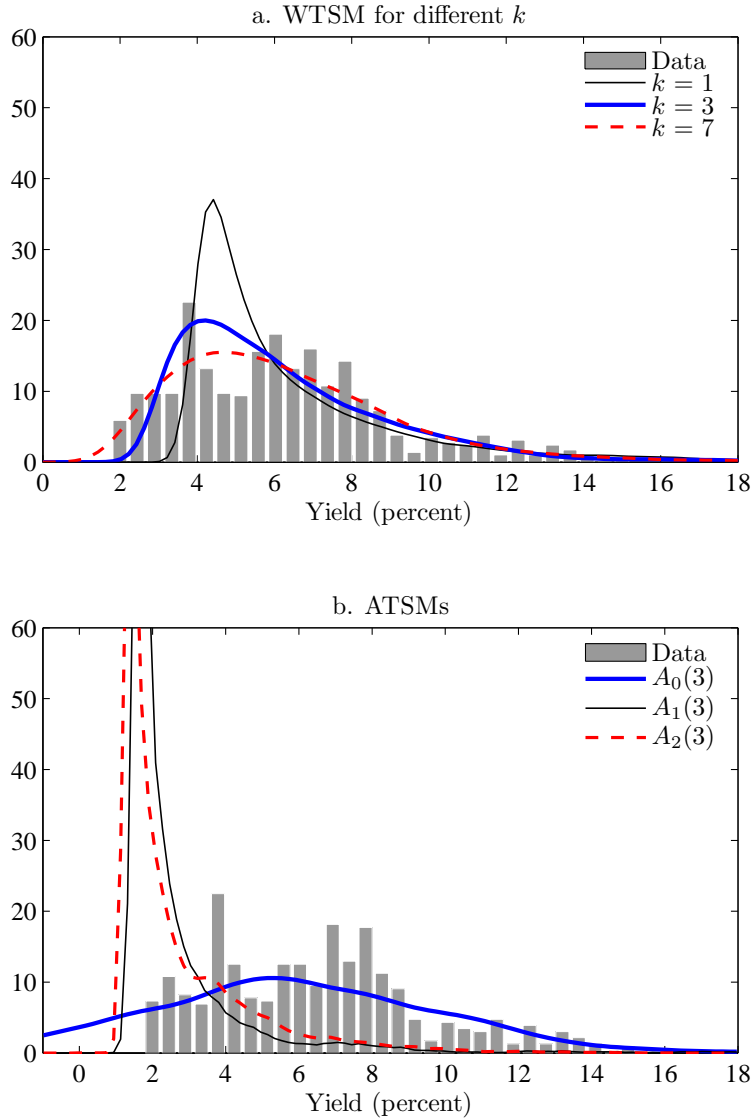
Figure 4: Actual and model-implied unconditional distributions of the 5-year yield

Panel *a* presents the densities of the 5-year yield obtained from the 2×2 WTSM with different degrees of freedom k. The results are superposed with the histogram of the 5-year US yield (1952:01–2005:06). Panel *b* shows the preferred ATSMs estimated by Duffee (2002): Gaussian $A_0(3)$, mixed models $A_1(3)$ and $A_2(3)$, and compares them with the histogram of the 5-year US yields (Duffee's sample 1952:01–1994:12).
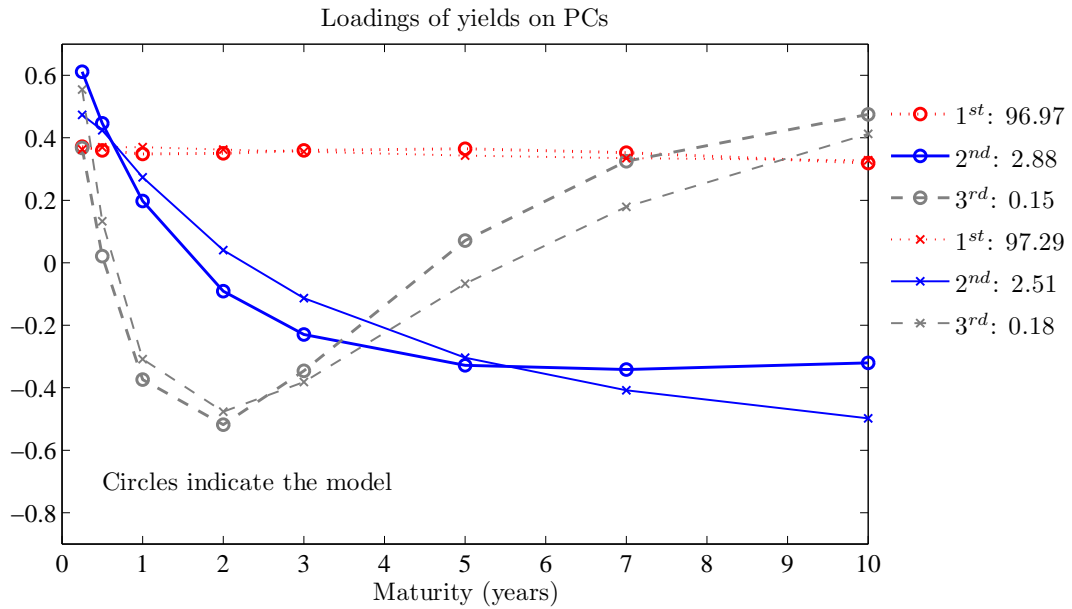
Figure 5: Loadings of yields on principal components: 2×2 WTSM vs. data

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 circled lines indicate loadings of yields obtained from the 2×2 WTSM ($k = 3$). Finer lines are loadings obtained from the sample of US yields, 1952:01–2005:06. The legend gives the corresponding percentages of yield variance explained by the first three principal components.
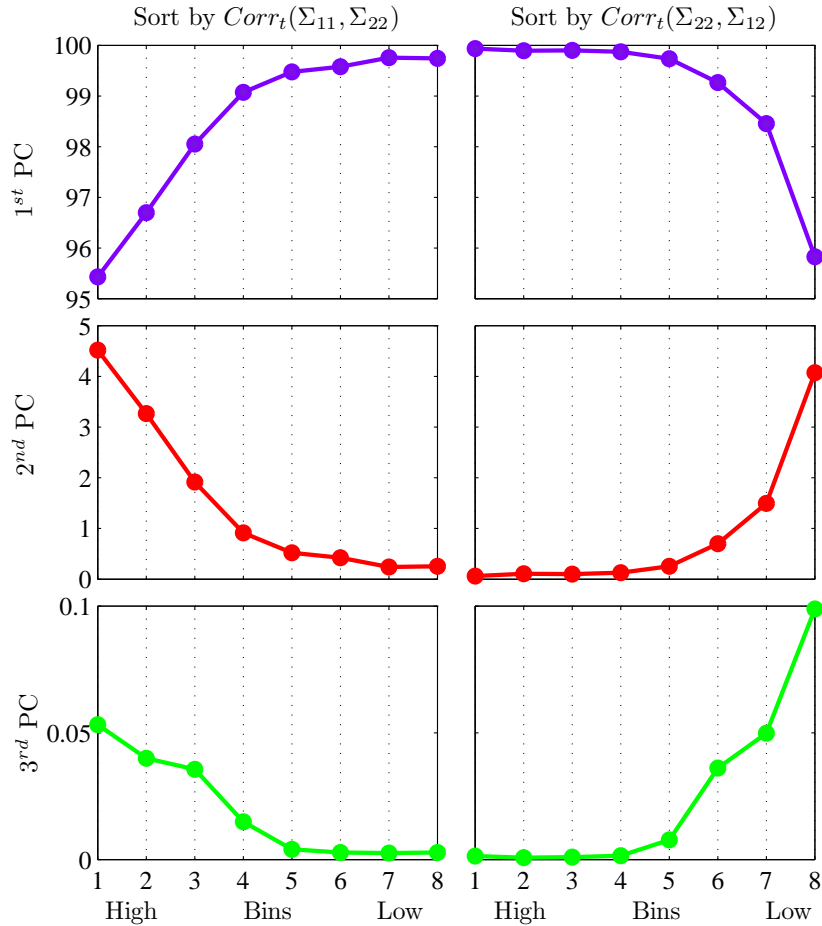
Figure 6: Variance explained by principal components, conditional on factor correlation

The figure shows the portions of yield variance explained by each principal component. The principal components of yields are computed conditional on the level of instantaneous correlation of factors in the 2×2 WTSM ($k = 3$). We form eight correlation bins, and number them from 1 (highest) to 8 (lowest). 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: the left-hand panel sorts by the level of $Corr_t(\Sigma_{11}, \Sigma_{22})$, the right-hand panel—by the level of $Corr_t(\Sigma_{22}, \Sigma_{12})$.
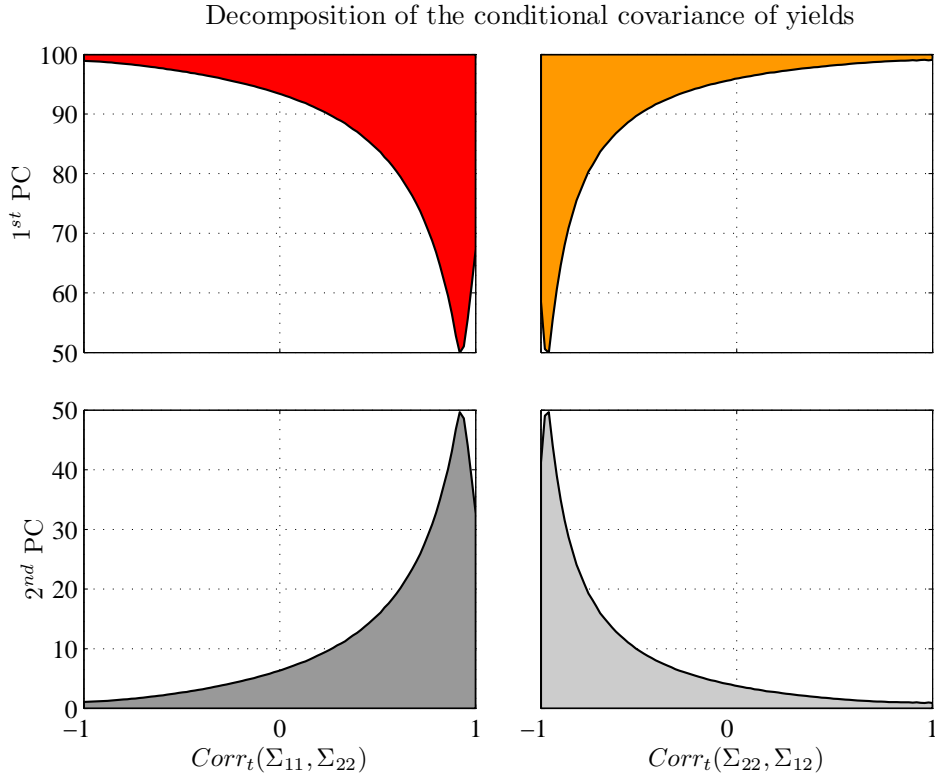
Figure 7: Conditional principal components of yields in the 2×2 WTSM

The figure displays the principal component decomposition of the instantaneous covariance of yields conditional on the level of the instantaneous correlation of factors in the 2×2 WTSM ($k = 3$). The shaded areas show the percentage of the conditional variance explained by the first principal component (upper panel) and the second principal component (bottom panel). The impact of the third factor ranges from 0 to 0.5 percent, and thus is not presented. The instantaneous conditional covariance of yields is given in expression (16). We consider 13 model-implied yields with maturities from 3 months to 10 years. The shaded areas are obtained as contours of scatter plots based on 72000 observations simulated from the model.
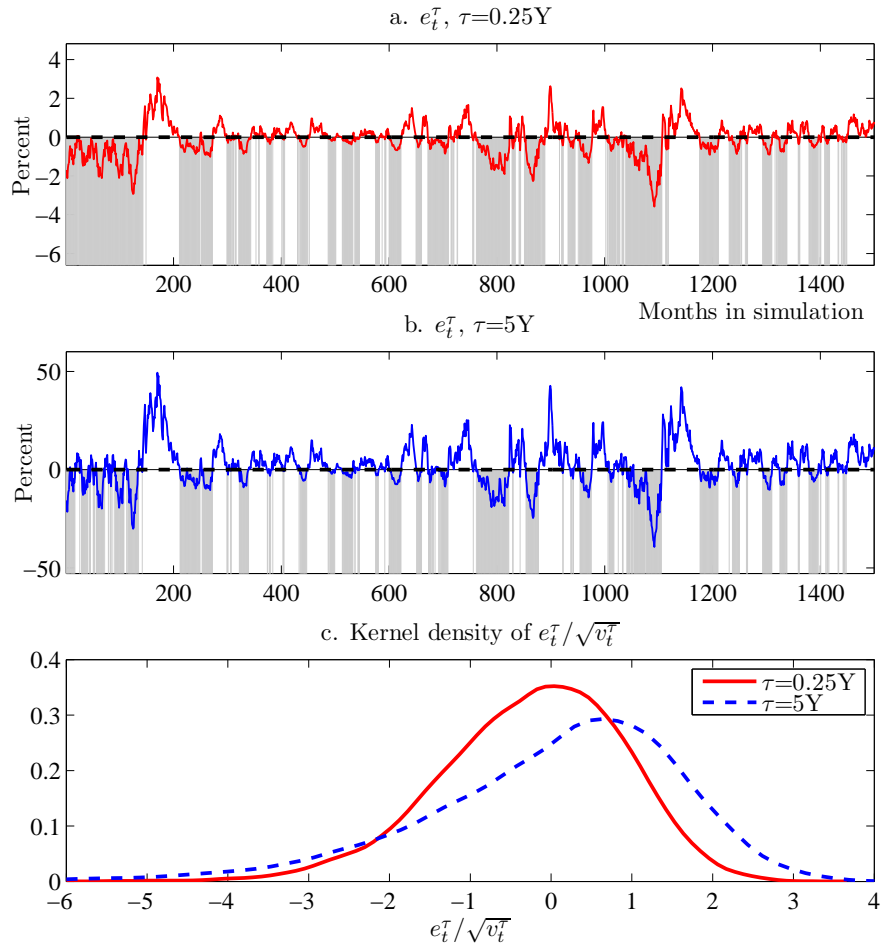
Figure 8: Properties of expected excess bond returns in the 2×2 WTSM

Panels a and b display the instantaneous expected excess returns on a 3-month and 5-year bond, respectively, as implied by the 2×2 Wishart factor model ($k = 3$). The expected excess returns are computed according to equation (18). Panel c plots the kernel density of the ratio of the instantaneous expected excess returns to their instantaneous volatility, $e_t^\tau / \sqrt{v_t^\tau}$, obtained by simulating 72000 observations from the model.
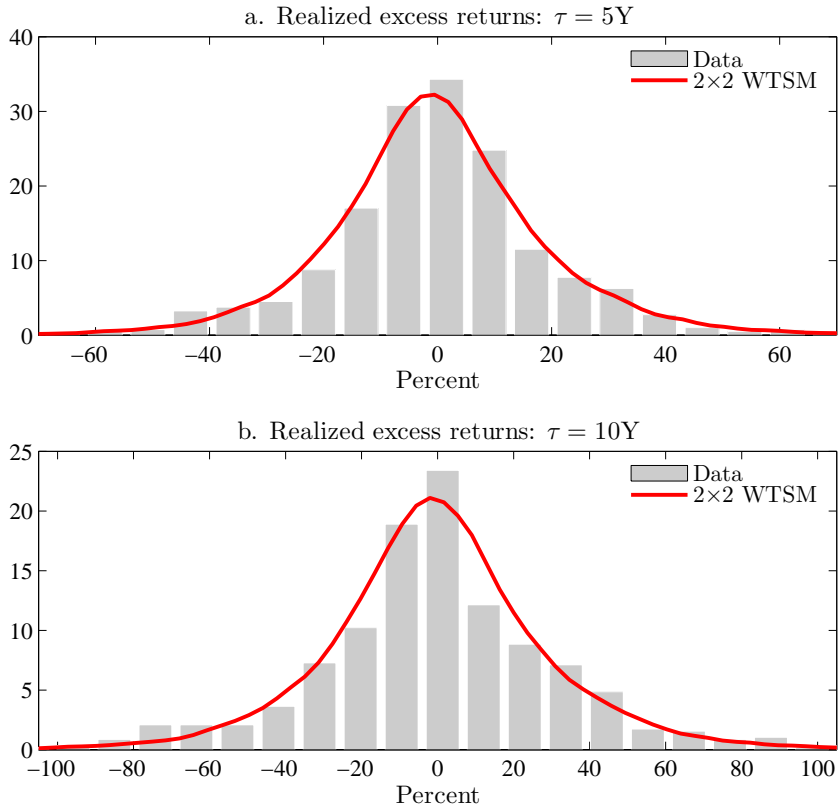
Figure 9: Properties of realized excess bond returns

Panels *a* and *b* display the distribution of realized monthly excess returns on 5-year and 10-year bonds, respectively. The realized excess returns implied by the 2×2 Wishart factor model ($k = 3$) are superposed with the histograms of realized excess returns on the corresponding US zero bonds (1952:01–2005:06). In both panels, the realized excess return is computed as the return on the long bond over the 3-month bond. All returns are annualized by multiplying with a factor 1200.
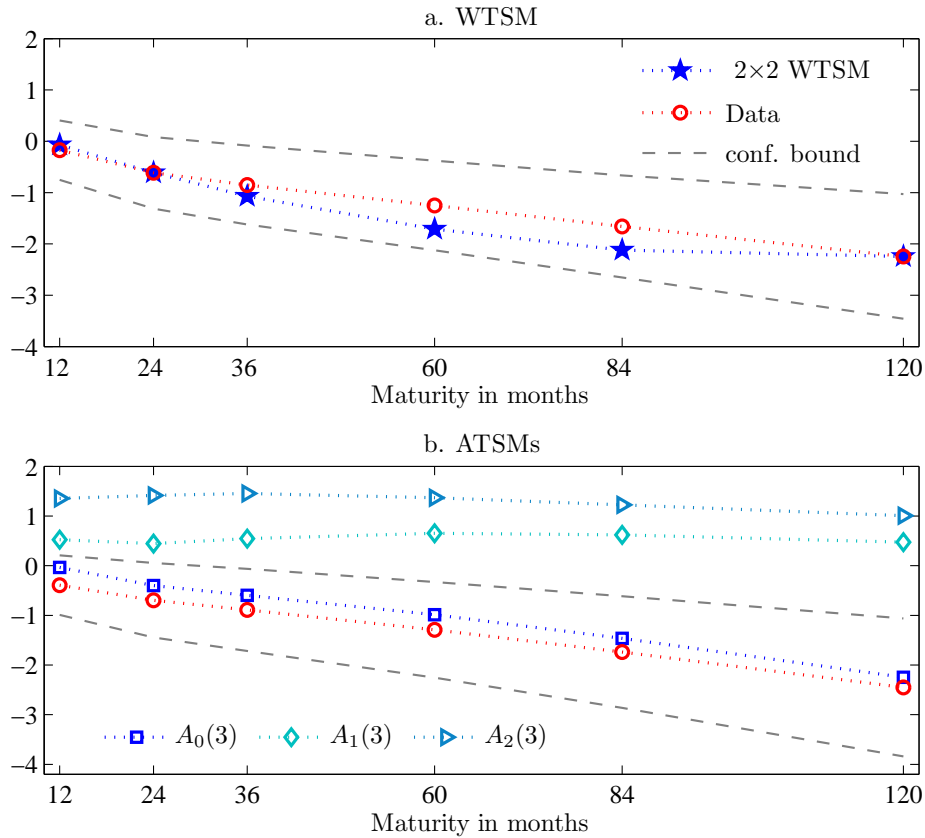
Figure 10: Campbell-Shiller regression coefficients

The figure plots—as a function of maturity—the parameters of Campbell and Shiller (1991) regression in equation (26). Panel *a* displays the coefficients obtained from the US yields in the sample period 1952:01–2005:06 and the theoretical coefficients implied by the 2×2 Wishart factor model ($k = 3$). 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. The 80 percent bound is lax for the data, but rigid for the model: Clearly, a more conservative choice (e.g. 90 percent) results in a still broader bound for the data, and thus is easier to match for the model. Further remarks from Table I apply.
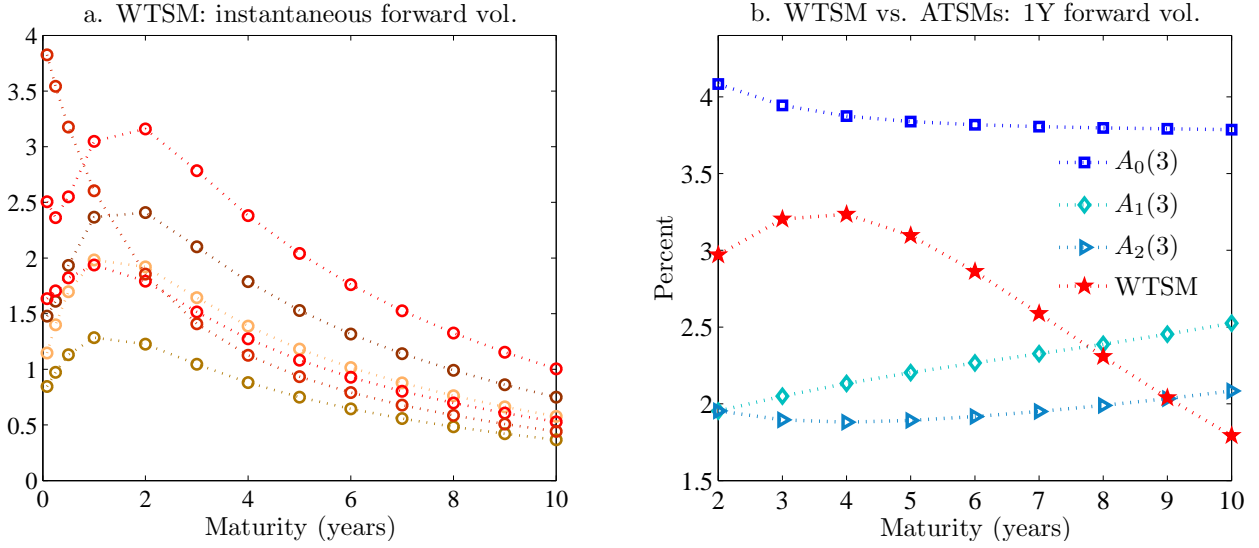
Figure 11: Term structure of forward interest rate volatilities

Panel a presents the term structure of the instantaneous volatility of the (instantaneous) forward rate given in equation (20), as implied by the 2×2 Wishart setting ($k = 3$). 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 (14). Panel b shows the theoretical unconditional volatility of the one-year forward rate from the same WTSM, and compares it with the preferred affine models of Duffee (2002): $A_0(3)$, $A_1(3)$ and $A_2(3)$. The x axis gives the maturity τ of the forward rate, $f_t^{\tau-1\to\tau} = \ln(P_t^{\tau-1}/P_t^\tau)$.
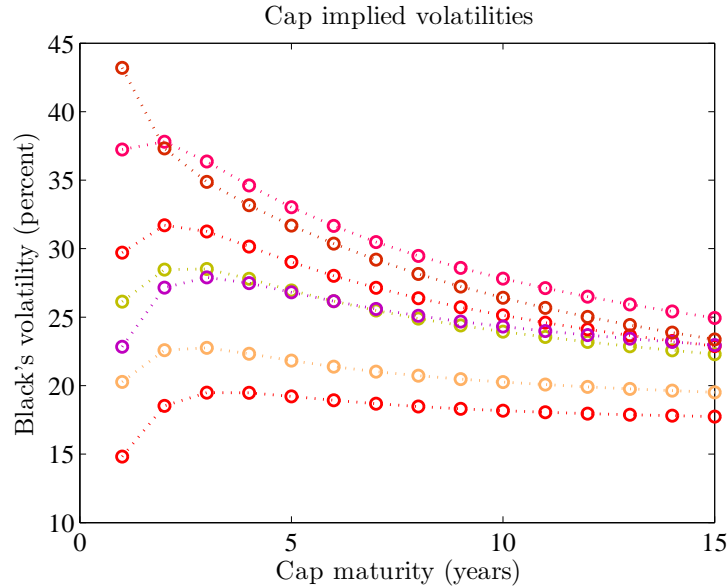

Figure 12: Term structure of cap implied volatilities

The figure exhibits several term structures of cap implied volatilities in the 2×2 Wishart model ($k = 3$), conditional on different values of the state matrix. The 3-month interest rate is the basis for each cap.
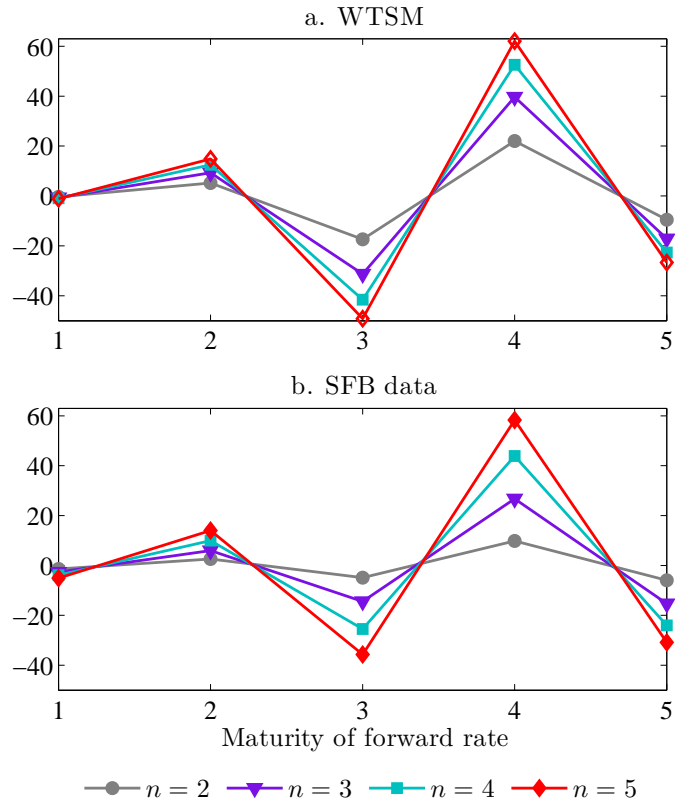
Figure 13: Cochrane-Piazzesi projection coefficients

The figure presents the slope coefficients in regressions of individual one-year excess bond returns on the set of one-year forward rates $f^{\tau-1\to\tau}, \tau = 1, 2, 3, 4, 5$, as a function of maturity τ. The legend $n = 2, 3, 4, 5$ refers to the maturity of the bond whose excess return is forecast. Panel a displays coefficients in the simulated 3×3 Wishart economy ($k = 3$). The regressions are run on a sample of 72000 monthly observations from the model. Panel b shows the loadings in the smooth Fama-Bliss (SBF) data set used by Dai, Singleton, and Yang (2004). The sample is monthly and spans the period 1970:01–2000:12.