STOCHASTIC CORRELATION AND RISK PREMIA IN TERM STRUCTURE MODELS

CARL CHIARELLA[+], CHIH-YING HSIAO[‡], AND THUY-DUONG TÔ*

[+,‡]*School of Finance and Economics*
University of Technology, Sydney
PO Box 123, Broadway NSW2007
Australia
[+]carl.chiarella@uts.edu.au
[‡]chih-ying.hsiao@uts.edu.au

**School of Banking and Finance*
University of New South Wales
Sydney, NSW2052
Australia
td.to@unsw.edu.au

ABSTRACT. This paper proposes and analyses a term structure model that allows for both stochastic correlation between underlying factors and an extended market price of risk specification. The issues of invariant transformation and different normalization are then considered so that a comparison between different restrictions can be made. We show that significant improvement in bond fitting and portfolio hedging performance is obtained by the model. However, the restriction on market prices of risk and the restriction on correlated factors has different impacts for different uses of the model. The first restriction has a much more negative impact on bond fitting and forecasting, whereas the second restriction has a much more negative impact on hedging performance. Overall, the stochastic correlation is priced significantly by market participants, and contributes to the predictive power of the model for discrete bond risk premia. Once our model factors are taken into account, other predictive factors that have been considered in the literature become insignificant.

Key words: Term structure; Stochastic correlation, Risk premium; Wishart; Affine; Extended affine; Multidimensional CIR.

JEL classifications: G12, E43, C58

Date: September 14, 2011

* Corresponding author. Tel (61) (2) 9385 5865. Fax (61) (2) 9385 6730.

1. Introduction

Much effort has been exerted on term structure modelling to deliver models that can capture both time series and cross-sectional features of yield curves, and at the same time offer some analytical tractability. The affine term structure models (ATSMs) offer a tractable family of models that can deliver economically meaningful behaviour of bond yields. The completely affine models have long been of interest, from the early models of Vasicek (1977) and Cox, Ingersoll & Ross (1985b), to the reduced-form model of Duffie & Kan (1996) and then systematically characterized by Dai & Singleton (2000). Although the reduced-form models do not rely on specific modelling of investor behaviour, they do give much more scope to better match real data. It has been proved that a richer specification of the market price of risk is needed to capture behaviour of bond returns and the premium, such as the switching sign of both the market price of risk and unconditional correlations of (some) underlying factors. Developments in this direction include the essentially affine model of Duffee (2002), the semi-affine square root model of Duarte (2004), and the extended affine model of Cheridito et al. (2007). The latter authors show that the extended affine model provides a much better fit, especially in terms of time-series fit, and has strong statistical significance.

The developments in market price of risk modelling have delivered a much better model fit. However, a draw-back of these models is the restrictions they impose on the correlation structure of the state variables. As an example take the two fundamental factors long term yield and yield spread. These two factors are usually found to be the important ones influencing the term structure of interest rates (eg. Duffee (1999), Duan & Simonato (1999)). Figure 1 illustrates the realized correlations between a 30-year yield and the spread of 3-month and 30-year yields, and shows that they are in the negative range for most of the observation period and are highly volatile. However, the Duffie & Kan (1996) framework only allows for positive correlation between positive factors, and consequently ignores the stochastic nature of the correlation. Even though, as far as we are aware, there has not been a study focusing on the impact of this stochastic correlation on interest rate markets, evidence from other markets are available. Buraschi et al. (2010) show that the stochastic correlation has a significant impact on optimal stock portfolio formation as well as hedging strategy. Driessen et al. (2009) show that the correlation risk is priced in the option market, where the implied volatility smile from individual stock options is flatter than that from index options. It is therefore of particular importance to model the stochastic correlation between the underlying factors of the term structure of interest rates if bond portfolios are of interest.

[Figure 1 around here]

Recently, financial market researchers have explored the use of the Wishart distribution to model the dynamic correlation structure of the state variables. The risk factors are assumed to follow a time affine process of positive definite matrices, whose transition probability is a Wishart distribution. Application can be found in multivariate stochastic volatility model of Gourieroux et al. (2009), credit risk model of Gourieroux & Sufana (2006), option pricing model of Da Fonseca et al. (2007), and term structure models of Gourieroux (2006), Grasselli & Tebaldi (2008) and Buraschi et al. (2008). Buraschi et al. (2008) show that the Wishart model also enhances model flexibility by capturing various empirical regularities of yield curves, such as the predictability of excess bond returns, the persistence of conditional volatilities and correlations of yields, and the hump in the term structure of forward volatilities.

It can be seen from the literature that a flexible market price of risk and stochastic nature of factor correlation are important in matching features of the yield curve. However, evidence from the literature is one-dimensional. No direct test of the merits of each component can be made from research in a one-dimensional setting. In this paper we propose and analyse a term structure model that allows for both components, which we call in short the WTSM model. We will characterize the invariant transformation for our WTSM model, and provide three different sets of normalization conditions, so that under appropriate normalization, our WTSM model will either nest the Wishart model proposed by Buraschi et al. (2008) (hereafter the BCT model), or nest the multidimensional CIR model with extended affine market price of risk (hereafter MCIR model). The tradeoff and relative advantages/disadvantages of each approach can therefore be analyzed.

One would question why we take the BCT model and the MCIR model as the benchmark. Would other models be equally capable of beating those benchmarks? It is noted that Buraschi et al. (2008) compare a simple 3x3 Wishart model with various 3-factor (completely and essentially) affine models and find that it has better performance, whereas Cheridito et al. (2007) have made the case for the extended affine market price of risk versus completely and essentially affine. The comparison between these two benchmarks and the existing affine models are complete.

We find that the WTSM model provides a better overall match to the data as well as better short-term and long-term forecasts. Even though Wishart risk factors have resulted in much flexibility for modelling yield curves, there are still essential constraints for fitting empirical data when adopting a simple market price of risk (as in the BCT framework). Here we still find that a simple market price of risk forces a sacrifice of time series fitting in order to adjust cross-sectional fitting, consistent with the findings of Cheridito et al. (2007) and Duffee & Stanton (2008). More specifically, imposing a restricted market price of risk worsens the average fitting error (forecast error) by 160% (42%) whereas imposing a more restricted correlation structure on the factors only worsens the errors by 120% (6%).

In terms of pricing, we find that the correlation risk is priced significantly, and is the most significant pricing factor at the short end of the curve. The price of this risk factor is stable in the medium time to maturity range, then increases slightly, but not as significant as the level risk factor, at the long end of the curve.

In terms of portfolio hedging, we find that it is crucial to take into account the stochastic nature of correlation. On average, imposing a restriction on the market prices of risk only worsens the risk-adjusted return of the hedged portfolio by 5.8%, whereas imposing a restriction on the nature of the correlation between the underlying factors worsens the risk adjusted return by 21.2%. Allowing for both components ensures the best hedging performance, as well as the best match to empirical asset correlations.

Finally, we investigate the ability of the WTSM model in predicting (discrete) bond risk premia. We find that the factors implied by the Wishart model have very good predictive power. Once these factors are taken into account, other factors that have been considered in the literature are no longer significant. Those factors include the Cochrane & Piazzesi (2005) forward factor, the Ludvigson & Ng (2009) macroeconomic factors, the Cooper & Priestley (2009) output gap factor, the principal components and various measures of liquidity. Our bond risk premia factors are completely spanned by the yield curve, although a higher order factor of the yield curve is needed.

The remainder of the paper is organized as follows. Section 2 outlines the model as well as its properties in terms of conditional moments and stochastic correlation. It also discusses the model invariant transformation and normalization issues. Data, estimated models and estimation procedures are outlined in Section 3. Empirical evidence and analyses of bond fitting and forecasting are presented in Section 4, whereas portfolio hedging performance under stochastic correlation are analysed in Section 5. Discussion regarding bond risk premia, instantaneous as well as discrete, is given in Section 6. Section 7 concludes the paper and all technical details are placed in the Appendices.

2. Wishart Term Structure Models

If the instantaneous interest rate r_t at time t follows a *Wishart process*, the term structure model based upon it is called a *Wishart term structure model* (WTSM). In this section we will set up a new WTSM. The model is different from the set-up in Buraschi et al. (2008) by adopting a more general market price of risk specification. Our model is in a continuous time setting, and therefore quite different from the discrete-time set-up in Gourieroux et al. (2009).

2.1. **A Wishart Term Structure Model (WTSM).**

Definition 1. (The Wishart Process) Let X_t be a full-rank symmetric positive-definite $n \times n$-matrix diffusion process defined as

$$dX_t = \left(kQQ^\top + MX_t + X_tM^\top\right)dt + QdW_t\sqrt{X_t} + \sqrt{X_t}dW_t^\top Q^\top \ , \tag{1}$$

where $\top$ denotes the transpose, Q and M are $n \times n$ matrices, k is a constant satisfying

$$k \geq n + 1, \tag{2}$$

and W_t is an $n \times n$ standard Wiener process. The square root $\sqrt{\cdot}$ is in the matrix sense[1]. The matrix X_t is called a Wishart process with degree of freedom k. The matrix M is usually assumed to be negative definite so that the process X_t is stationary.

The condition $k \geq n + 1$ guarantees that the Wishart process is strictly positive definite, see Theorem $2''$ (p.745) in Bru (1991). We need to have the stronger requirement of positive definiteness than that of Buraschi et al. (2008) because later we will consider $\sqrt{X_t}^{-1}$ in our market price of risk specification.

Assumption 1. *The instantaneous rate r_t is a linear combination of the Wishart process X_t given by*

$$r_t = \alpha + \mathbf{tr}(\Psi X_t) = \alpha + \sum_{i,j=1}^{n} \Psi_{ij} X_{ij,t} \ , \tag{3}$$

where Ψ is an $n \times n$ matrix, and **tr** *is the trace operator. Without loss of generality Ψ is a symmetric matrix*[2]*. In order to guarantee the positivity of r_t, Ψ is required to be positive definite.*

There are market prices of risk, denoted by an $n \times n$ matrix Λ_t, associated with the $n \times n$ risk process W_t. The probability transformation from the empirical measure to

---

[1]For a positive definite matrix X which can be diagonalized as $X = P^\top \text{diag}(\lambda_1, \cdots, \lambda_n)P$ with P unitary, then $\sqrt{X} := P^\top \text{diag}(\sqrt{\lambda_1}, \cdots, \sqrt{\lambda_n})P$.

[2]This is because

$$\mathbf{tr}(\Psi X_t) = \mathbf{tr}\left(\frac{1}{2}(\Psi + \Psi^\top)X_t\right) \tag{4}$$

and the matrix $\frac{1}{2}(\Psi + \Psi^\top)$ is symmetric.

a risk-neutral measure is characterized by the transformed Wiener process

$$d\tilde{W}_t = dW_t + \Lambda_t dt \ , \tag{5}$$

where $\tilde{W}_t$ is an $n \times n$ standard Wiener process under the risk-neutral measure.

Assumption 2. *The market price of risk in the Wishart process is assumed to be of the form*

$$\Lambda_t = \Lambda_0 \sqrt{X_t}^{-1} + \Lambda_1 \sqrt{X_t} \ , \tag{6}$$

where Λ_0 and Λ_1 are $n \times n$ matrices and the inverse operator is the matrix inverse.

The form of the market prices of risk is in line with the extended affine term structure model in Cheridito et al. (2007). Under this assumption, the factor process under the risk-neutral measure is given by

$$dX_t = (\tilde{\Gamma} + \tilde{M}X + X\tilde{M}^\top)dt + Qd\tilde{W}_t\sqrt{X} + \sqrt{X_t}d\tilde{W}_t^\top Q^\top \ , \tag{7}$$

where

$$\tilde{\Gamma} \ := \ kQQ^\top - Q\Lambda_0 - \Lambda_0^\top Q^\top \ , \tag{8}$$

$$\tilde{M} \ := \ M - Q\Lambda_1 \ . \tag{9}$$

We observe that $\tilde{\Gamma}$ is a symmetric $n \times n$ matrix.

Remark 2.1. We require

$$\tilde{\Gamma} \geq^M (n+1)QQ^\top \tag{10}$$

(meaning that $\tilde{\Gamma} - (n+1)QQ^\top$ is a positive semi-definite matrix), in order that the Wishart process X_t be strictly positive under the risk-neutral measure; see Cuchiero et al. (2011)

Remark 2.2. Under the parameter restrictions (2) and (10) the boundary non-attainment conditions in Cheridito et al. (2007) are satisfied, so there exists an equivalent martingale measure and risk-neutral pricing is free of arbitrage. Without these restriction the change of measure cannot be guaranteed to be equivalent[3].

Remark 2.3. Our specification contrasts with that of Buraschi et al. (2008), whose market price of risk is of the simpler form $\Lambda_t = \sqrt{X_t}$, which can be derived in an elegant way from a general equilibrium argument. However, to derive it one must make the assumption of a log-utility function, which is restrictive.

In this paper we extend the model in Buraschi et al. (2008) by adopting a more general form of market price of risk (6) and also adding other parameters (for more detail see Section 3). Although the extension is quite simple as far as solving the no-arbitrage bond price is concerned, later in Section 4 we will see that this extension does greatly increase model flexibility in fitting empirical data.

Let $P(t,T)$ be the price at t of a bond maturing at T with payout of one unit of money. According to no-arbitrage pricing theory the bond price is equal to the expected value of the discounted future payoff with respect to a particular risk-neutral measure. Thus,

$$P(t,T) = \tilde{\mathbb{E}}_t[\exp\big(-\int_t^T r_s ds\big)] \tag{11}$$

where $\tilde{\mathbb{E}}_t$ is the expectation operator under the risk-neutral measure conditional on the information up to t. Based on the linear spot rate relation (3) and the factor dynamics (7), the bond price depends on the current state X_t in the form

$$P(t,T;X_t) = \exp\big(a(\tau) + \mathbf{tr}[C(\tau)X_t]\big)\ , \tag{12}$$

where $\tau = T - t$, $a(\tau)$ is a scalar function and $C(\tau)$ is a symmetric[4] $n \times n$ function, see Cuchiero et al. (2011).

[3]For example, the probability of the process hitting a boundary is nonzero under one measure but zero under another.

[4]This is for the same reason as given in footnote 2.

Proposition 1. *For the given well-defined factor dynamics (7) under the risk neutral measure and the instantaneous rate relation (3) the bond price given by (11) can be solved in the form (12) where the coefficients $a(\tau)$ and $C(\tau)$ solve the ordinary differential equations (ODEs)*

$$\begin{aligned} \frac{d}{d\tau}C(\tau) &= C(\tau)\tilde{M} + \tilde{M}^\top C(\tau) + 2C(\tau)QQ^\top C(\tau) - \Psi \ , && (13) \\ \frac{d}{d\tau}a(\tau) &= \mathbf{tr}[\tilde{\Gamma}C(\tau)] - \alpha && (14) \end{aligned}$$

with $\tilde{\Gamma}$ and $\tilde{M}$ defined in (8) and (9) and subject to the initial conditions $a(0) = 0$, $C(0) = 0$.

Proof. See Appendix A. □.

Proposition 2. *The solution of the $n \times n$ matrix valued function $C(\tau)$ satisfying the ODE (13) and the initial condition $C(0) = 0$ is given by*

$$C(\tau) = \Phi_{22}(\tau)^{-1}\Phi_{21}(\tau) \ , \tag{15}$$

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

$$\begin{pmatrix} \Phi_{11}(\tau) & \Phi_{12}(\tau) \\ \Phi_{21}(\tau) & \Phi_{22}(\tau) \end{pmatrix} := \exp\left[\tau \begin{pmatrix} \tilde{M} & -2QQ^\top \\ -\Psi & -\tilde{M}^\top \end{pmatrix}\right] .$$

The solution of $a(\tau)$ in (14) is given by

$$a(\tau) = -\mathbf{tr}\Big[\Big(\frac{(Q^\top Q)^{-1}\tilde{\Gamma}}{2}\Big)\Big(\ln \Phi_{22}(\tau) + \tau\tilde{M}^\top\Big)\Big] - \alpha\tau \ . \tag{16}$$

Proof. The solution for general solvable affine term structure models can be found in Grasselli & Tebaldi (2008). □.

Example 2.4. It is easy to see that a one-dimensional CIR process is a 1×1 Wishart process. In this case (1) becomes

$$dX_t = (kQ^2 + 2MX_t)dt + 2Q\sqrt{X_t}dW_t \ .$$

Note that in this case M is required to be negative in order that X_t be stationary.

Example 2.5. Consider an $n \times n$ process X_t in (1) with diagonal parameters Q, M and a diagonal Wiener process

$$dW_t = \begin{pmatrix} dW_{1t} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & dW_{nt} \end{pmatrix}.$$

Then it is easy to see that X_t is a diagonal process and each component on the diagonal follows a CIR process

$$dX_{it} = (kq_i^2 + 2m_i X_{it})dt + 2q_i\sqrt{X_i}dW_{it} \quad , i = 1, \cdots, n, \tag{17}$$

and the X_i processes are independent of each other. The terms m_i and q_i are the items on the diagonals of M and Q respectively, and m_i need to be negative in order to ensure stationary of x_{it} for all i.

We note that the multi-variate diagonal process (17) is more restrictive than a system of multiple independent one-dimensional CIR processes. The constant term kq_i^2 in the drift coefficient in (17) has a fixed linear relation to its variance q_i^2, while in a system of multiple independent one-dimensional CIR processes, each X_{it} can have a different k_i. We found that this kind of a proportional restriction on the risk-neutral dynamics (7) largely reduces the capability of the model to fit empirical data. Therefore we adopt the parametrization $\tilde{\Gamma}$ to relax this proportional restriction. In the model of Buraschi et al. (2008) $\tilde{\Gamma} = kQQ$ so it does not cover a system of multiple independent one-dimensional CIR processes.

2.2. **Properties of the Wishart Process.**

2.2.1. *Conditional Moments.*

Proposition 3. *The conditional first moment of the Wishart process (1) is given by*

$$E[X_{t+\tau}|X_t] = \Phi_\tau X_t \Phi_\tau^\top + kV_\tau \ , \tag{18}$$

where $\Phi_\tau := \exp(M\tau)$[5] *and*

$$V_\tau := \int_0^\tau \Phi_s QQ^\top \Phi_s^\top ds \ . \tag{19}$$

The conditional second moment is given by

$$\begin{aligned} &Var[vecX_{t+\tau}(vecX_{t+\tau})^\top|X_t] \\ &= (I_{n^2} + K_{n,n})\Big(\Phi_\tau X_t \Phi_\tau^\top \otimes V_\tau + k(V_\tau \otimes V_\tau) + V_\tau \otimes \Phi_\tau X_t \Phi_\tau^\top\Big) \ , \end{aligned} \tag{20}$$

where vec(X_t) *stacks all columns of* X_t *into an* $n^2 \times 1$ *vector,* I_{n^2} *is an* $n^2 \times n^2$ *unit matrix,* $K_{n,n}$ *is the commutative matrix defined by*

$$vec(H^\top) = K_{n,n} vec(H) \quad , \text{ for any } n \times n \text{ matrix } H \ . \tag{21}$$

Proof. See Buraschi et al (2008). □.

2.2.2. *Stochastic Correlation.*

Proposition 4. *For the* $n \times n$ *Wishart process* X_t *as defined in (1), the instantaneous correlation is given by*

$$Cov[dX_{ij}dX_{uv}] = \Big((QQ^\top)_{iu}X_{jv} + (QQ^\top)_{ju}X_{iv} + (QQ^\top)_{iv}X_{ju} + (QQ^\top)_{jv}X_{iu}\Big)dt \ . \tag{22}$$

Proof. See Appendix A. □.

---

[5]Recall M is negative definite so Φ_τ converges to zero for large τ.

For the special case of the covariance of the variables on the diagonal, we have

$$\text{Cov}[dX_{ii}\, dX_{jj}] = 4(QQ^\top)_{ji} X_{ij} dt \; . \tag{23}$$

Now we calculate the correlation

$$\text{Corr}[dX_{ii}\, dX_{jj}] = \frac{\text{Cov}(dX_{ii}\, dX_{jj})}{\sqrt{\text{Var}(dX_{ii})}\sqrt{\text{Var}(dX_{jj})}} = \eta_{ij} \frac{X_{ij}}{\sqrt{X_{ii}}\sqrt{X_{jj}}} \; , \tag{24}$$

where η_{ij} is a constant given by

$$\eta_{ij} = \frac{\sum_{u=1}^{n} Q_{iu} Q_{ju}}{\sqrt{\sum_{u=1}^{n} Q_{iu}^2}\sqrt{\sum_{u=1}^{n} Q_{ju}^2}} \; .$$

Proof. See Appendix A. □.

Proposition 5. *The stochastic covariance given in the last proposition can be also summarized in vector form:*

$$\begin{aligned}
& \mathit{Cov}[\mathit{vec}(dX)\mathit{vec}(dX)^\top] \\
= \; & (I_{n^2} + K_{n,n})\big(X \otimes QQ^\top\big)dt + (K_{n,n} + I_{n^2})\big(QQ^\top \otimes X\big)dt \qquad (25) \\
= \; & (X \otimes \mathbf{1}_{n\times n}).*(\mathbf{1}_{n\times n} \otimes QQ^\top)dt + (\mathbf{1}_{1\times n} \otimes X \otimes \mathbf{1}_{n\times 1}).*(\mathbf{1}_{n\times 1} \otimes QQ^\top \otimes \mathbf{1}_{1\times n})dt \\
& +(\mathbf{1}_{1\times n} \otimes QQ^\top \otimes \mathbf{1}_{n\times 1}).*(\mathbf{1}_{n\times 1} \otimes X \otimes \mathbf{1}_{1\times n})dt + (QQ^\top \otimes \mathbf{1}_{n\times n}).*(\mathbf{1}_{n\times n} \otimes X)dt \\
=: \; & S(X)dt \; , \qquad (26)
\end{aligned}$$

where I_{n^2} is the $n^2 \times n^2$ identity matrix, $K_{n,n}$ is defined in (21), $\mathbf{1}_{n\times n}$ is an $n \times n$ matrix with all elements equal to one and $.*$ represents element-wise multiplication.

Proof. See Appendix A. □.

2.3. **Invariant Transformations and Normalization of Parameters.** A Wishart term structure model is characterized by its model parameters $\Theta := \big(k, M, Q, \tilde{\Gamma}, \tilde{M}, \alpha, \Psi\big)$ given in the factor dynamics (1) under the real world measure, the factor dynamics (7) under the risk-neutral measure and the instantaneous rate (3). Dai and Singleton

(2000) pointed out that for the dynamic affine term structure model different parameter specifications can generate exactly the same model bond price. A straightforward example is to take any arbitrary $n \times n$ transformation $\mathcal{L}$ and apply it to

$$C^{\mathcal{L}}(\tau) = (\mathcal{L}^{\top})^{-1} C(\tau) \mathcal{L}^{-1}, \quad X^{\mathcal{L}} = \mathcal{L} X \mathcal{L}^{\top} ,$$

then the bond price (12) calculated from the pair $\big(C(\tau), X_t\big)$ is exactly the same as that calculated from the transformed pair $\big(C^{\mathcal{L}}(\tau), X_t^{\mathcal{L}}\big)$, in fact

$$P(t,T;X_t) = \exp\Big(a(\tau) + \mathbf{tr}[C(\tau)X_t]\Big) = \exp\Big(a(\tau) + \mathbf{tr}[C^{\mathcal{L}}(\tau)X_t^{\mathcal{L}}]\Big) , \quad \tau = T - t.$$

In the following discussion the parameter set Θ is added to the argument of the bond price $P(t,T;X_t,\Theta)$ in order to emphasize its role.

Definition 2. A transformation $\mathcal{L}$ is called *an invariant transformation* if

$$P(t,T;X_t,\Theta) = P(t,T;X_t^{\mathcal{L}},\Theta^{\mathcal{L}})$$

for all t, T and the whole process X_t.

The definition is the same as that in Dai and Singleton (2000). Here we stress that the pricing invariance holds not only for any one point $X_t = x$ but for the whole process $X_t, t \geq 0$.

The Proposition 6 gives the exact relation of an invariant transformation for our Wishart term structure model.

Proposition 6 (An Invariant Transformation)**.** *Consider the transformation*

$$X_t^{\mathcal{L}} := \mathcal{L} X_t \mathcal{L}^{\top} \text{ , and } \quad W_t^{\mathcal{O}} = \mathcal{O} W_t , \tag{27}$$

where $\mathcal{L}$ is an $n \times n$ matrix and $\mathcal{O}$ is an orthogonal matrix with $\mathcal{O}\mathcal{O}^{\top} = I_n$. The transformation is an invariant transformation if the parameters

$\Theta^{\mathcal{L}} := \left(k^{\mathcal{L}}, M^{\mathcal{L}}, Q^{\mathcal{L}\mathcal{O}}, \tilde{\Gamma}^{\mathcal{L}}, \tilde{M}^{\mathcal{L}}, \alpha^{\mathcal{L}}, \Psi^{\mathcal{L}}\right)$ *are transformed according to*

$$k^{\mathcal{L}} = k \,, \tag{28}$$

$$M^{\mathcal{L}} = \mathcal{L} M \mathcal{L}^{-1} \,, \tag{29}$$

$$Q^{\mathcal{L}\mathcal{O}} = \mathcal{L} Q \mathcal{O}^{\top} \,, \tag{30}$$

$$\tilde{\Gamma}^{\mathcal{L}} = \mathcal{L} \tilde{\Gamma} \mathcal{L}^{\top} \,, \tag{31}$$

$$\tilde{M}^{\mathcal{L}} = \mathcal{L} \tilde{M} \mathcal{L}^{-1} \,, \tag{32}$$

$$\alpha^{\mathcal{L}} = \alpha \,, \tag{33}$$

$$\Psi^{\mathcal{L}} = (\mathcal{L}^{\top})^{-1} \Psi \mathcal{L}^{-1} \,. \tag{34}$$

Technically, $a^{\mathcal{L}}(\tau)$ *and* $C^{\mathcal{L}}(\tau)$ *are the no-arbitrage bond pricing coefficients given in (15) and (16) calculated with the transformed parameter* $\Theta^{\mathcal{L}}$*. The transformation* $(X_t^{\mathcal{L}}, W_t^{\mathcal{O}}, \Theta^{\mathcal{L}})$ *is an invariant transformation if there hold the relations*

$$a^{\mathcal{L}}(\tau) = a(\tau) \quad \textit{and} \quad C^{\mathcal{L}}(\tau) = (\mathcal{L}^{\top})^{-1} C(\tau) \mathcal{L}^{-1} \,. \tag{35}$$

Proof. See Appendix A. □.

The sup-index $\mathcal{L}\mathcal{O}$ in $Q^{\mathcal{L}\mathcal{O}}$ indicates that the parameter Q is affected by the transformation $W_t^{\mathcal{O}} = \mathcal{O} W_t$, whereas the other parameters are not.

The invariant transformation above is characterized using the parametrization $(Q, \tilde{\Gamma}, \tilde{M})$ for the risk-neutral dynamics. Alternatively, we can use the parametrization $(k, M, Q, \Lambda_0, \Lambda_1)$ where Λ_0 and Λ_1 are given in equation (6). The market price of risk (5) remains the same under the transformation but will have a different drift adjustment under the measure change in the transformed system.

Proposition 7. *If we adopt the parametrization* $\Theta' = (k, M, Q, \Lambda_0, \Lambda_1, \alpha, \Psi)$*, the parameter relations (31) and (32) are replaced by*

$$\tilde{\Gamma}^{\mathcal{L}} = k Q^{\mathcal{L}\mathcal{O}} (Q^{\mathcal{L}\mathcal{O}})^{\top} - Q^{\mathcal{L}\mathcal{O}} \mathcal{O} \Lambda_0 \mathcal{L}^{\top} - (Q^{\mathcal{L}\mathcal{O}} \mathcal{O} \Lambda_0 \mathcal{L}^{\top})^{\top} \,, \tag{36}$$

$$\tilde{M}^{\mathcal{L}} \;=\; M^{\mathcal{L}} - Q^{\mathcal{L}\mathcal{O}}\mathcal{O}\Lambda_1\mathcal{L}^{-1}\,. \tag{37}$$

Proof see Appendix A. □.

Arbitrarily many parameters values can map to the same term structure using the invariant transformations, and Proposition 8 provides *normalization conditions* that allow us to exclude such invariant transformations, so that under the normalization conditions there is only one parameter specification mapping to one term structure. In order words, under the normalization conditions the only transformation $(\mathcal{L}, \mathcal{O})$ allowed is the identity transformation.

Proposition 8 (Normalization Conditions)**.** *We provide three sets of normalization conditions to facilitate comparison between different models. Assume M can be diagonalized. The three sets of normalization conditions are given by*

(S1) First set

- (a1) M *is diagonal.*
- (b1) Q *is lower triangular and the elements on the diagonal are all positive.*
- (c) *The elements on the diagonal of Ψ are equal to one. The elements in the first row of Ψ are nonnegative.*

(S2) Second set

- (a2) M *is lower triangular.*
- (b2) Q *is diagonal and the elements on the diagonal are all positive.*
- (c) *The elements on the diagonal of Ψ are equal to one. The elements in the first row of Ψ are nonnegative.*

(S3) Third set

- (a3) $\Lambda_1 \equiv I_n$ *(I_n is an n-dimensional unit matrix).*
- (b3) Q *is upper triangular.*
- (c3) *The elements on the diagonal and in the first row of Ψ are positive.*

Proof. See Appendix A. □.

Based on the model identification conditions provided by Proposition 8, the Wishart model in Buraschi et al. (2008) is a restricted version of our Wishart term structure model where the upper triangular part of M in (1), α in (3) and Λ_0 in (6) are all zero. Within the framework developed we can test these parameter restrictions later in the empirical investigation.

3. Empirical Procedures

3.1. **Summary of the Models.** We consider four models, (1) our Wishart term structure model (WTSM), (2) the multiple CIR (MCIR) model which is nested inside the WTSM model, (3) the Buraschi et al. (2008) (BCT) model, and (4) an extended Wishart (BCTEW) encompassing the BCT model, and equivalent to the WTSM model. Figure 2 summarizes the relationships between the different models. We note that

(i) The MCIR model is a restricted version of the WTSM model.
(ii) The WTSM and the BCTEW model are equivalent term structure models as discussed in Proposition 6, and
(iii) The BCT model is a restricted version of the BCTEW model.

[Figure 2 around here]

Regarding the question whether MCIR and BCT models are fair benchmarks, it is noted that Buraschi et al. (2008) compare a simple 3x3 Wishart model with various 3-factor (completely and essentially) affine models and find that it has better performance, whereas Cheridito et al. (2007) have made the case for the extended affine market price of risk versus completely and essentially affine. Therefore, both the MCIR and the BCT are superior models representing two different approaches to modelling the yield curve, namely to utilize a more flexible market price of risk specification and to model the stochastic nature of factor correlation. The equivalence transformation allows us to compare these two approaches and examine the relative contribution of each component to the success of yield curve modelling.

We consider the Wishart model with $n = 2$. Now we describe the models in detail. The first model, **WTSM**, is a Wishart term structure model with the normalization conditions S2 in Proposition 8. There are fifteen parameters to be estimated:

$$\Theta = \left(k, m_1, m_2, q_{11}, q_{21}, q_{22}, \tilde{\gamma}_{11}, \tilde{\gamma}_{12}, \tilde{\gamma}_{22}, \tilde{m}_{11}, \tilde{m}_{12}, \tilde{m}_{21}, \tilde{m}_{22}, \alpha, \psi\right),$$

which are in the form of

$$\begin{aligned} &M = \begin{pmatrix} m_{11} & 0 \\ 0 & m_{22} \end{pmatrix}, \; Q = \begin{pmatrix} q_{11} & 0 \\ q_{21} & q_{22} \end{pmatrix}, \; \tilde{\Gamma} = \begin{pmatrix} \tilde{\gamma}_{11} & \tilde{\gamma}_{12} \\ \tilde{\gamma}_{12} & \tilde{\gamma}_{22} \end{pmatrix}, \\ &\tilde{M} = \begin{pmatrix} \tilde{m}_{11} & \tilde{m}_{12} \\ \tilde{m}_{21} & \tilde{m}_{22} \end{pmatrix}, \; \Psi = \begin{pmatrix} 1 & \psi \\ \psi & 1 \end{pmatrix}, \; k \text{ and } \alpha. \end{aligned} \tag{38}$$

The second model, **MCIR**, is based on a two-dimensional CIR process and it can be considered as a restricted version of the first model (WTSM). There are ten parameters to be estimated:

$$\begin{aligned} &M = \begin{pmatrix} m_1 & 0 \\ 0 & m_2 \end{pmatrix}, \; Q = \begin{pmatrix} q_1 & 0 \\ 0 & q_2 \end{pmatrix}, \; \tilde{\Gamma} = \begin{pmatrix} \tilde{\gamma}_1 & 0 \\ 0 & \tilde{\gamma}_2 \end{pmatrix}, \\ &\tilde{M} = \begin{pmatrix} \tilde{m}_1 & 0 \\ 0 & \tilde{m}_2 \end{pmatrix}, \; \Psi = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \; k \text{ and } \alpha. \end{aligned} \tag{39}$$

The initial values $X_{0,21} = X_{0,12}$ are set to zero.

The third model **BCTEW** is an extended version of the Buraschi et al. (2008) Wishart model. The market price of risk is set to be $\Lambda_t \equiv \sqrt{X_t}$ therefore $\Lambda_1 = I$ and $\tilde{M} = M - Q$ in (9). This restriction is compensated for by the extra freedom in M and Ψ compared with the WTSM (38). The BCTEW adopts the normalization conditions S3 given in Proposition 8 and is statistically equivalent to the WTSM. It also has fifteen parameters:

$$M = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix}, \; Q = \begin{pmatrix} q_{11} & q_{12} \\ 0 & q_{22} \end{pmatrix}, \; \tilde{\Gamma} = \begin{pmatrix} \tilde{\gamma}_{11} & \tilde{\gamma}_{12} \\ \tilde{\gamma}_{12} & \tilde{\gamma}_{22} \end{pmatrix}, \\ \tilde{M} = M - Q, \;\; , \Psi = \begin{pmatrix} \psi_{11} & \psi_{12} \\ \psi_{12} & \psi_{22} \end{pmatrix}, \; k \text{ and } \alpha. \tag{40}$$

The fourth model is the **BCT** model. It has ten parameters:

$$M = \begin{pmatrix} m_{11} & 0 \\ m_{21} & m_{22} \end{pmatrix}, \; Q = \begin{pmatrix} q_{11} & q_{12} \\ 0 & q_{22} \end{pmatrix}, \; \tilde{M} = M - Q, \; \Psi = \begin{pmatrix} \psi_{11} & \psi_{12} \\ \psi_{12} & \psi_{22} \end{pmatrix}, \; k \; . \tag{41}$$

This model is a restricted BCTEW model with the restrictions $\alpha \equiv 0$ and $\tilde{\Gamma} \equiv kQQ^{\top}$.

3.2. **Data.** The observed variables are the US strip bond yields of fixed times to maturity from Bloomberg, shown in Figure 3. Yields are calculated based on a linear approximation method. There are 11 time series with times to maturity of 3 months, 6 months, 1-5, 7, 10, 20, 30 years. Bond yields are collected at the end of each month. The observation period is from 04/1991 to 07/2008, giving 208 data points.

In the literature, the level and the slope of the curve are usually considered as important determining factors for the term structure of interest rates. For our sample, the level of the 30 year long-term yield changes only moderately and has a slow downward tendency. The three month yield, which follows very closely the federal funds rate and is affected largely by US monetary policy, fluctuates more widely. This gives a volatile development of the yield spread.

[Figure 3 around here]

3.3. **Estimation method.** The Wishart risk factors are not directly observed but need to be inferred from observed bond yields. Buraschi et al. (2008) and Cheridito et al. (2007) assume that there are as many bond yields as the factors observed without measurement errors so that the factors can be obtained by an exact yield-factor correspondence. Based on this Buraschi et al. (2008) uses the GMM and Cheridito et al. (2007) employ an approximate likelihood method to estimate the parameters.

Different from the two approaches above, but in line with Duffie & Singleton (1997), we adopt the extended Kalman filter to filter out underlying factors from the observations with measurement/observation errors. We in fact assume all bonds are observed with errors and utilize the filter to derive the maximum likelihood estimator. There are two main reasons for this choice of estimation method.

First, filtering using all bonds assumed to be measured with errors allows us to estimate the model even when there are hidden factors in the yield curve. This is a term we borrow from Duffee (2011) to describe those factors about which a snapshot of the time t-yield curve conveys no information. Using filtering, Duffee (2011) finds the importance of hidden factors for his Gaussian term structure model. Our model, on the other hand, is not Gaussian. Other estimation method for hidden factors are available but they involve the use of additional data apart from yields, such as the use of macroeconomic variables in Joslin, Priebsch & Singleton (2011). Here we do not assume a factor is hidden but leave it to the data to tell us.

Second, we find that it's hard to monitor whether the underlying Wishart factors remain positive definite during the estimation procedure of the GMM method (used in Buraschi et al. (2008)). This is technically an important point because a Wishart process is not defined if it is not positive definite. Using methods without the control of positive definiteness, the parameters can still be obtained but there is no guarantee that the obtained Wishart process is well-defined. Filtering, on the other hand, allows us to look at each time step and monitor the behaviour of the process.

Details on the extended Kalman filter and positivity control can be found in Appendix B. The appendix also provides some notes on how we treat the initial values as parameters in the model and our optimization methods.

4. MODEL ESTIMATES, FIT AND FORECAST

4.1. **Estimation of MCIR and WTSM models.** Table 1 gives estimation results of the MCIR and the encompassing WTSM models.

[Table 1 around here]

For the model MCIR, under the risk-neutral measure the dynamics of X_t can be seen as the two independent CIR processes

$$dX_{iit} = (\tilde{\gamma}_{ii} - 2\tilde{m}_{ii}X_{it})dt + 2q_{ii}\sqrt{X_{iit}}dW_{iit}\ .$$

The degree of freedom k is estimated to be 20.17 which is greater than $n+1=3$ for the Wishart process X_t under the real-world measure so the Wishart process is a strictly positive-valued process. It is easy to see that the Feller condition $\tilde{\gamma}_{ii} > \frac{4q_{ii}^2}{2}$ is satisfied for both i so the process is also a strictly positive process under the risk-neutral measure. The boundary non-attainment condition is satisfied and so the martingale pricing formula (11) is free of arbitrage. The mean reversion parameters m_{ii} and $\tilde{m}_{ii}, i = 1, 2$ are all negative so the process is stationary under both the real-world and risk-neutral measures.

The WTSM model, on the other hand, does not assume independence between the two underlying factors. As suggested by Eq. (23) the third diagonal factor X_{12t} represents the covariance of the two off-diagonal factors.

The degree of freedom parameter k is estimated to be 7.28 $> 3 = n+1$. The condition $\tilde{\Gamma} \geq^M 3 * Q * Q^\top$ is satisfied. So the estimates of the WTSM satisfy the boundary non-attainment condition. The parameters M and $\tilde{M}$ are both negative definite so the Wishart process is a positive stationary process for both the real-world and risk-neutral measures.

The parameter α in (3) has the role of shifting the level of the factor X_t. A negative α helps to keep the X_t in the positive area. Duffee (1999) fixes a negative value for α

(on page 209), while we let the data determine α here. Both the MCIR model and the WTSM model give similar estimates of -12.33% and -11.21% respectively.

Figure 4 (a and b) plots the estimated factors of the two models, whereas Figure 5 shows the comparison between the estimated factors and the long term yield level and the yield spread (defined as the difference between 3-month and 30-year yields). There is a high correlation between the estimated factors and the two economic factors of level and spread. We therefore will name X_{11} the spread factor and X_{22} the level factor.

[Figure 4 around here]

[Figure 5 around here]

Statistically, the WTSM outperforms the MCIR model as is evident in Table 1. This is so since the estimated standard deviation of the measurement errors σ_ϵ is smaller. Also, the likelihood ratio test strictly rejects the restrictions of the MCIR model against the WTSM. In addition, the WTSM also has smaller fitting errors for the bond yield data both in bias and in mean-square error (MSE).[6]

4.2. **Estimation of the BCT and BCTEW models.** In the column "BCTEW" in Table 2 we present the parameter values which form an equivalent model to the WTSM results in Table 1. One can hardly recognize the equivalence from these numbers, nor from their factors as depicted in Panels (b) and (c) in Figure 4. This is because they are statistical factors and can be rotated. However, from the fact that they generate the same the risk premia (calculated by equation (45)) shown in Panels (b) and (c) in Figure 9 we can still recognize their equivalence. The trajectories of the BCTEW factors are totaly different from those of the WTSM and they have lost the correspondence to the long-term yield and the yield spread.

[Table 2 around here]

[6]For all models, bias is defined as the fitting errors of the model, which are the difference between the observed bond yields and the model bond yields calculated using the updated factor levels. The average is taken over all bonds.

In the BCT model the shift parameter α is set to zero. As a consequence one can observe that the BCT factors in Figure 4(d) are closer to zero. In this situation the positive definitiveness is more easily violated as can be seen from the fact that the X_{11t} factor crosses the zero line around the years 2000 and 2008.

The other parameter $\tilde{\Gamma}$ also accounts for flexibility in fitting bond yield data. The restriction $\tilde{\Gamma} = kQQ^\top$ requires that the constant drift term under the real world dynamics be equal to that under the risk-neutral measure. This restriction reduces the model flexibility to a large extent, as also pointed out in the extended affine term structure in Cheridito et al. (2007). The reduction of model capacity is evidenced by an increase of measurement errors σ_ϵ and yield fitting errors (average bias and average MSE), and a decrease of the likelihood. The likelihood ratio test strictly rejects the BCT model against the BCTEW model.

4.3. **Forecasting Power of the Models.** This section investigates the cross-sectional fitting and forecast performance. In Panel (a) in Figure 6 we give the average errors (bias, the left-hand figures) and the mean-square errors (MSE, the right-hand figures) of the bond yield fit. The three curves correspond to the measurement errors of the MCIR, the WTSM and the BCT models. We do not have a curve for the BCTEW model since it is coincident with the WTSM due to their equivalence. Error scales in this MSE figure correspond to the scale of the measurement errors σ_ϵ in Tables 1 and 2 and the “av. Bias” and “av. MSE” which are the averages over all bonds. The WTSM has the smallest MSE amongst the three models.

Panel (b) in Figure 6 gives the bias and the MSE of one-month ahead (ie. 1-step ahead) forecast errors of the all bonds. The WTSM has the best performance of the one-month ahead forecasts. Note that the one-month forecast errors are used for calculating the likelihood function. Therefore the superiority of the WTSM against the MCIR and the BCT models for the one-month forecast performance can be seen together with the clear results of the LR test in the pervious section. Compared to the random walk (hereafer RW) model, the WTSM model does better overall in terms of bias (the RW model consistently underestimates bond yields). However, in terms of MSE, the WTSM model only outperforms the RW model at 2-12 years maturity range. It performs slightly worse than RW model for short maturity (less than 2 years) and underperforms at the long end of the curve (more than 13 years maturity).

[Figure 6 around here]

Nevertheless, when it comes to anything longer than 1-step ahead forecasts, the WTSM model outperforms all of the other models (RW, MCIR, and BCT), which can be seen in Figure 7. The BCT model is the worst performing one, suggesting that the restrictions on the market price of risk greatly reduce the flexibility of the model to match observed yields. This restriction has much more negative impact on fitting and forecasting than the restriction on the correlation of factors. More specifically, imposing a restricted market price of risk worsens the average fitting error (forecast error) by 160% (42%) whereas imposing a more restricted correlation structure on the factors only worsens the errors by 120% (6%).

[Figure 7 around here]

5. Portfolio Strategies under Stochastic Correlation

This section explores the hedging performance of the three models under consideration. We adopt the minimal variance portfolio in Campbell et al. (1996) as the hedging strategy. Given that the MCIR has 2 different risk factors, whereas the WTSM and BCT models have 3 different risk factors, we are going to consider portfolios containing only 2 assets. This means that the WTSM and BCT models will not have complete hedging, and therefore their performance will not be optimal.

We consider monthly return of bonds with fixed time to maturity[7]. Based on the model, the asset covariance can be calculated for each time point. Using equation

[7]Concretely, an investor holds a bond with a given time to maturity τ_i for one period $[t, t+\Delta t]$ for some i. Next period, the investor sells the bond and reinvests in a new bond with the same time to maturity τ_i. For the calculation in each holding period $[t, t+\Delta t]$ the time to maturity is shortened from τ_i to $\tau_i - \Delta t$. We calculate the bond price of the shorter time to maturity $\tau_i - \Delta t$ using the estimated results of WTSM in Table 1 and then building the log return.

(65), (see Appendix A) the bond return (with fixed time to maturity) is calculated from

$$\frac{dP_{i,t}}{P_{i,t}} = \big(r_t + e_{i,t}\big)dt + \mathbf{tr}\big[dZ_t C_i\big] \ , \tag{42}$$

where

$$e_{i,t} = 2\mathbf{tr}\big[Q(\Lambda_0 + \Lambda_1 X_t)C_{i,t}\big] \ , \ \ C_i := C(\tau_i) \ , \ \ dZ_t := QdW_t\sqrt{X_t} + \sqrt{X_t}dW_t^\top Q^\top \ .$$

The instantaneous covariance at time t of the two assets is then given by

$$\sigma_{ij,t}dt = \mathrm{Cov}\Big(\mathbf{tr}[dZ_t C_{i,t}]\ \mathbf{tr}[dZ_t C_{j,t}]\Big) \ , \tag{43}$$

which changes over time and can be calculated based on the estimation results.

Heterogenous investors decide their hedging strategy based on their beliefs of the models. In other words they calculate the instantaneous variance $\sigma_{ij,t}$ in (43) based on the estimation results of the MCIR, the WTSM and the BCT models respectively. Given the variance structure, hedged portfolio positions can be obtained easily. The investment horizon is taken to be the same as the observation period (from April 1991 to July 2008), and portfolio position is normalized to one.

Table 3 summarises the performance of the hedged 2-asset portfolios based on the three models for different hedging pairs. For example “3Y10Y” means the hedging portfolio consisting of two bonds with time to maturity of three years and ten years, whereas the equally weighted portfolio considers a simple portfolio consisting of 50% of each bond.

[Table 3 around here]

From the table, it can be seen that hedged portfolios based on the estimates from any of the three models have quite low volatility. Volatility of the hedged positions is around 23%-50% of that of the equally weighted portfolio. Though WTSM is disadvantaged because we only use 2 assets to hedge 3 risk factors[8], it still delivers the lowest portfolio volatility (our objective function), as well as lowest downside risk.

[8]The BCT is similarly disadvantaged. The MCIR has only 2 risk factors which are hedged by 2 assets.

The risk-adjusted return[9] is also highest under the WTSM model, then BCT and finally MCIR. Note that Sharpe ratios obtained by the WTSM model are plausible, with the maximum less than 0.2, which does not suggest any overfitting.[10]

Given that correlation is the most important factor in determining portfolio diversification, we attribute the ranking of the performance of the three models on the ability of the WTSM and BCT to model stochastic correlation. It is not possible to calculate the stochastic correlation from empirical data directly, therefore to investigate correlation estimates, we look at the asset correlation implied by the three models, as seen in Figure 8. The posterior reference line (labelled "emp corr") is obtained using the realized sample correlation between the two assets used to form the portfolio. It can be observed that MCIR (BCT) estimated stochastic correlation is always higher (lower) than the empirically realized correlation, suggesting that the model overestimates (underestimates) the true stochastic correlation between assets. The WTSM estimated stochastic correlation, on the other hand, fluctuates around the realized sample correlation. The trajectory of the correlation is stable in general.

[Figure 8 here]

6. Bond Risk Premia

6.1. **Instantaneous Risk Premia.** The instantaneous risk premium of a bond is the excess instantaneous return of the bond over the risk-less instantaneous return.

Proposition 9. *The instantaneous risk premium of a bond under the Wishart model 1 is*

$$\textit{Risk Premium } = \mathbf{tr}\left[(kQQ^\top - \tilde{\Gamma})C(\tau)\right] + 2\mathbf{tr}\left[(M - \tilde{M})X_t C(\tau)\right] . \tag{45}$$

[9]Risk adjusted returns are measured by the Sharpe ratio and the Sortino ratio. The Sortino Ratio is defined by

$$\text{Sortino Ratio} = \frac{R-T}{DR} , \tag{44}$$

where R is the asset return, T is the target return (usually the risk-free rate), and the downside risk (DR) given by $\text{DR}^2 := \int_{-\infty}^{T}(T-x)^2 f(x)dx$ where $f(x)$ is the probability density of the return.

[10]Duffee (2010) examines properties of Sharpe ratio for GDTSMs, whereas Joslin, Priebsch & Singleton (2011) also investigates the Sharpe ratio under macro-GDTSMs.

Proof see Appendix A. □.

Figure 9 compares the development of the risk premia of the different maturity bonds for all the four models. Recall that Panels (b) and (c) provide the same picture of the risk premia since they are equivalent. The term structures of the risk premia of the four models share some similarity. All of them are positive most of the time. Furthermore the risk premia of long-term bonds fluctuate more than those of short term bonds.

[Figure 9 around here]

Consider the risk premium (45) more in detail. The risk premium is a linear combination between the underlying factors, adjusted by a constant term. Figure 10 illustrates the absolute contributions of each factor in the risk premia. For all of the models, a positive shock in the spread (ie. short term minus long term yield) decreases the risk premium, and the longer the maturity, the more the bond is affected. This is an expected behaviour. An increase in the spread means short term yield increases (decreases) more (less) than the long term yield. If the price of a unit of risk does not change, the excess return of bonds should decrease to reflect the expected lower return in the future.

[Figure 10 around here]

Regarding the second risk factor, the level of long term yields, the MCIR and the WTSM models predict that a positive shock in this factor leads to an increase in bond excess return. The BCT model, on the other hand, gives an unexpected prediction of an opposite movement. We conjecture that the flexible market price of risk specification of the MCIR and the WTSM model allows a better match between the estimated risk premia and the expected behaviours of the yield curve.

The BCT and the WTSM models allow for stochastic correlation between factors, and both predict an increase in risk premia due to a positive shock in the correlation.

This implies that financial market participants do price the correlation risk. This finding is consistent with other markets, such as the options market. A study by Driessen et al. (2009) in the options market shows that the correlation risk is priced, and the implied volatility smile from individual stock options is flatter than that from index options.

6.2. **Discrete Bond Risk Premia.** In the previous section we focused on the instantaneous excess return of a bond over the risk free rate. In this section we will look at the bond risk premia commonly defined in the literature as the excess holding period return of a long term bond over the return of a shorter term bond. We name it "discrete bond risk premia" to distinguish it from the instantaneous risk premia.

The excess return of an m-period bond over j-periods, expressed in bond prices, is

$$xr_{t,t+j}^{(m)} = -\ln P(t, t+m) + \ln P(t+j, t+m) + \ln P(t, t+j).$$

Using (12), this in turn can be written in terms of the state factors as

$$xr_{t,t+j}^{(m)} = -a(m) + a(m-j) + a(j) - \mathbf{tr}[(C(m) - C(j))X_t] + \mathbf{tr}[C(m-j)X_{t+j}].$$

The predictability of excess bond returns has long been of interest. Campbell & Shiller (1991) find that excess bond returns are forecastable by Treasury yield spreads. Cochrane & Piazzesi (2005) find that a linear combination of forward rates has a very good predictive power, and the combination is now called the CP factor. Ludvigson & Ng (2009) use 8 factors extracted from 132 different macroeconomic variables to explain the risk premia. Cooper & Priestley (2009) report the predictive power of the output gap. Cieslak & Povala (2011) explain the premia using cycles and a persistent factor proxied by the discounted moving average of core inflation data. Given that our model provides bond prices, a question of interest is about the ability of our model to predict excess bond returns, and how it compares to other existing predictive factors in the literature.

In order to make it easier to compare to existing literature, we will examine a one year holding period (ie. j is one year) and long term bonds of 2-5 years (ie. $m = 2, \ldots, 5$ years). We will cut down the sample period to 2007:12, so that we can compare with the Ludvigson & Ng (2009) results.[11]

Table 4 reports a simple regression of the excess log bond return on our factors extracted from the Wishart, the BCT and the MCIR models. The t-statistics are calculated using Newey-West adjustments with 18 lags. Our Wishart factors can explain 58%-62% of the variations of the one year excess bond return. Factors from the BCT model can explain slightly smaller portion, namely from 56%-60%, however, none of the factors is statistically significant. Regarding the MCIR model, only X_1 is significant and the model has the least predictive power of 23%-32%.

[Table 4 around here]

Table 5 and Table 6 look at other factors that have been used to explain the bond risk premia. In Table 5, the explanatory variables are various measures of liquidity and the measure of output gap proposed by Cooper & Priestley (2009)[12]. None of the factors is significant, apart from the Moody Baa-Aaa spread. In Table 6, the explanatory variables are the Cochrane & Piazzesi (2005) forward factors, the 8 macro-economic factors extracted by Ludvigson & Ng (2009) and the first three principal components.[13] As reported in the various papers, all of the factors have high predictive power for bond risk premia. The question would be whether these predictive factors still have predictive power once controlling for our model factors.

[Table 5 around here]

[Table 6 around here]

It is noted that the above regressions treat the dependence of the bond risk premia on the Wishart factors as constant over time. However, we do have a model that predicts

[11] We would like to thank these authors for making their macroeconomic data publicly available.
[12] The gap is constructed using unrevised data on industrial production, applying a quadratic time trend. Definition of liquidity measures can be found in Appendix C.
[13] We do not have access to Cieslak & Povala (2011) results to report their factors' predictive power.

the changing dependency of the risk premia and the factors. Table 7 looks at the errors from this prediction, and regresses the errors on other predictive factors (we only use those factors that prove to have predictive power in the previous regressions). We find that almost all of the other factors become insignificant, including the principal components. The only notable exception is $\hat{F}_4$, the inflation factor.

[Table 7 around here]

Our finding indicates that information on the yield curve alone is still sufficient in explaining the bond risk premia. However, unlike the current literature that looks at only linear combination of information from the yield curve (be it the yields, the forward rates, or the bond returns), we also use higher order information, namely the correlation of two factors. This higher order information falls outside any possible rotation of the linear combination of first order yields, the general framework of which is discussed in Joslin, Singleton & Zhu (2011), Joslin, Le & Singleton (2011) and Hamilton & Wu (2011). This extra information that has not been used before turns out to provide essential information about the future yields and returns.

7. Conclusion

In this paper we investigate the significance of allowing for a flexible specification of the market price of risk and allowing flexible and stochastically correlated factors in modelling the term structure of interest rates. We propose a Wishart model that is an extension of the Wishart model in Buraschi et al. (2008) (BCT) and incorporating the more flexible market price of risk given in Cheridito et al. (2007). The advantage of this approach is that by using appropriate invariant transformations between different parametrizations we can nest models with different restrictions and therefore determine the roles of each component.

The empirical analysis shows that relaxing both restrictions plays a crucial role in improving the fit, as well as the forecasting, of bond yields. Imposing a restricted market price of risk worsens the average fitting error (forecast error) by 160% (42%)

whereas imposing a more restricted correlation structure on the factors only worsens the fitting errors by 120% (6%). However, given that the underlying factors can be interpreted as the long yields and yield spread (correlation between our statistical factors and the economic factors is around 98%), modelling stochastic factor correlation explicitly allows a better understanding of how changes in those factors affect the bond yields over time. In addition, the explicit modelling of the stochastic correlation reveals that market participants price correlation risk significantly. Its price is more important than the prices of the level and slope factors at the short and medium sections of the yield curve, and only be slightly overdone by the price of the level factor at the long end of the curve.

It should also be noted that though the restricted models are not as good as the general model in fitting and forecasting bond yields, their absolute model performance (in terms of fit and forecasting) is still good. The model restrictions have much larger impact on the implied behaviour of the market price of risk, risk premia and therefore bond portfolio construction. We find that hedged portfolios built under the more general Wishart model, which allows for an extended affine market price of risk, outperform those built by more restricted models by a considerable margin. However, in contrast to the fitting and forecasting performance, taking into account stochastic correlation (with a simple market price of risk specification) improves the hedging performance significantly compared to the model that only allows for a flexible market price of risk. On average, imposing a restriction on the market prices of risk only worsens the risk-adjusted return of the hedged portfolio by 5.8%, whereas imposing a restriction on the nature of the correlation between the underlying factors worsens the risk adjusted return by 21.2%.

In brief, we find that both a flexible specification for market prices of risk and a specification that allows for stochastic correlation between underlying factors are important in term structure models. However, the first specification is more important for bond fitting and forecasting, whereas the latter specification is more important for portfolio hedging purposes. The combination of flexible market prices of risk and stochastic correlation also results in good predictive power for (discrete) bond risk premia. After taking into account the prediction by the model, almost all of other predictive factors considered in the literature become insignificant. Those factors include the Cochrane & Piazzesi (2005) forward factors, the Ludvigson & Ng (2009) macroeconomic factors, the Cooper & Priestley (2009) output gap factor, the principal components and various measures of liquidity. Our bond risk premia factors are completely spanned by the yield curve, although higher order information of the yield curve is needed.

APPENDIX A. PROOFS OF ALL PROPOSITIONS

Proof of Proposition 1

Apply Itô's Lemma to (12) we obtain the bond return process

$$\begin{aligned}\frac{dP}{P} &= \big(-\frac{d}{d\tau}a(\tau)-\textbf{tr}[\frac{d}{d\tau}C(\tau)X_t]\big)dt+\textbf{tr}[C(\tau)dX_t] \\ &\quad +\frac{1}{2}\sum_{i,j=1}^{n}\sum_{u,v=1}^{n}C_{ij}C_{uv}dX_{ij,t}dX_{uv,t}\ .\end{aligned} \tag{46}$$

We replace the cross term in the last term by (22) so we obtain

$$\begin{aligned}&\sum_{i,j=1}^{n}\sum_{u,v=1}^{n}C_{ij}C_{uv}dX_{ij,t}dX_{uv,t} \\ =\ &\sum_{i,j=1}^{n}\sum_{u,v=1}^{n}C_{ij}C_{uv}\Big((QQ^\top)_{iu}X_{jv}+(QQ^\top)_{ju}X_{iv}+(QQ^\top)_{iv}X_{ju}+(QQ^\top)_{jv}X_{iu}\Big)dt\end{aligned} \tag{47}$$

Now calculate each term in (47) we start with the first term

$$\begin{aligned}\sum_{i,j=1}^{n}\sum_{u,v=1}^{n}C_{ij}C_{uv}(QQ^\top)_{iu}X_{jv} &= \sum_{j,u=1}^{n}\big(\sum_i C_{ij}(QQ^\top)_{iu}\big)\big(\sum_v C_{uv}X_{jv}\big) \\ = \sum_{j,u=1}^{n}(CQQ^\top)_{ju}(CX^\top)_{uj} &= \textbf{tr}[CQQ^\top CX^\top]=\textbf{tr}[CQQ^\top CX]\ .\end{aligned}$$

For the second term in (47) we follow the same calculation but with interchange of i and j. Similarly for the third and term we obtain the same result with interchange of u and v. And the last term again runs with pair interchange $i\leftrightarrow j$ and $u\leftrightarrow v$. These calculations lead to rewrite the last term in (46) by

$$\frac{1}{2}\sum_{i,j=1}^{n}\sum_{u,v=1}^{n}C_{ij}C_{uv}dX_{ij,t}dX_{uv,t}=2\textbf{tr}[CQQ^\top CX]\ .$$

Use this and the dynamics (7) we then rewrite the bond return (46) into

$$\begin{aligned}\frac{dP}{P} &= \Big(-a'+\textbf{tr}[CX]+\textbf{tr}\Big[\big(-C'+C\tilde{M}+\tilde{M}^\top C+2CQQ^\top C\big)X\Big]\Big)dt \\ &\quad +\textbf{tr}\Big[C\big(Qd\tilde{W}\sqrt{X}+\sqrt{X}d\tilde{W}^\top Q^\top\big)\Big]\ .\end{aligned}$$

According to the no-arbitrage principle, the instantaneous return under the risk neutral measure is equal to the sport rate r_t. The ODEs (13) and (14) are obtained by comparing the coefficients with the r_t given in (3).

□

Proof of Proposition 4

In order to prove this proposition we need the equation

$$\mathrm{E}[\mathbf{tr}(HdW_t)\mathbf{tr}(GdW_t)] = \mathbf{tr}(HG^\top)dt \ , \tag{48}$$

where H and G are $n \times n$ constant matrices and W_t is an $n \times n$ standard Wiener process. This is because

$$\mathrm{E}[\mathbf{tr}(HdW_t)\mathbf{tr}(GdW_t)] = \mathrm{E}[(\sum_{i,j=1}^{n} H_{ij}dW_{ij,t})(\sum_{i,j=1}^{n} G_{ij}dW_{ij,t})] = \mathrm{E}[\sum_{i,j=1}^{n} H_{ij}G_{ij}]dt = \mathbf{tr}(HG^\top)dt \ .$$

Calculating the covariance $\mathrm{Cov}[dX_{ij}\ dX_{uv}]$ using the definition (1) we have

$$\mathrm{Cov}[dX_{ij}\ dX_{uv}] = \mathrm{E}\Big[\big(Q_{i\cdot}dW\sqrt{X}_{\cdot j} + Q_{j\cdot}dW\sqrt{X}_{\cdot i}\big)\big(Q_{u\cdot}dW\sqrt{X}_{\cdot v} + Q_{v\cdot}dW\sqrt{X}_{\cdot u}\big)\Big] \ , \tag{49}$$

where $Q_{i\cdot}$ is the i-th row of the Q-matrix and $\sqrt{X}_{\cdot j}$ is the j-th column.

Rewrite each single term as

$$\begin{aligned} Q_{i\cdot}dW\sqrt{X}_{\cdot j} &= Q_{i\cdot}\begin{pmatrix} \sum_{u=1}^{n} W_{1u}\sqrt{X}_{uj} \\ \vdots \\ \sum_{u=1}^{n} W_{nu}\sqrt{X}_{uj} \end{pmatrix} = Q_{i1}\sum_{u=1}^{n} W_{1u}\sqrt{X}_{uj} + \cdots + Q_{in}\sum_{u=1}^{n} W_{nu}\sqrt{X}_{uj} \\ &= \sum_{v=1}^{n}\sum_{u=1}^{n} Q_{iv}W_{vu}\sqrt{X}_{uj} = \mathbf{tr}[(Q_{i\cdot})^\top(\sqrt{X}_{\cdot j})^\top dW] \ . \end{aligned}$$

Apply it to $\mathrm{E}[(Q_{i\cdot}dW\sqrt{X}_{\cdot j})(Q_{u\cdot}dW\sqrt{X}_{\cdot v})]$ then we have

$$\begin{aligned} &\mathrm{E}[(Q_{i\cdot}dW\sqrt{X}_{\cdot j})(Q_{u\cdot}dW\sqrt{X}_{\cdot v})] = \mathrm{E}[\mathbf{tr}\Big((Q_{i\cdot})^\top(\sqrt{X}_{\cdot j})^\top dW\Big)\mathbf{tr}\Big((Q_{u\cdot})^\top(\sqrt{X}_{\cdot v})^\top dW\Big)] \\ = \ &\mathbf{tr}[(Q_{i\cdot})^\top(\sqrt{X_{\cdot j}})^\top\sqrt{X}_{\cdot v}Q_{u\cdot}]dt = \mathbf{tr}(Q_{i\cdot})^\top X_{jv}Q_{u\cdot} = \mathbf{tr}[X_{jv}Q_{u\cdot}(Q_{i\cdot})^\top]dt \\ = \ &X_{jv}(QQ^\top)_{ui}dt = (QQ^\top)_{iu}X_{jv}dt \ . \end{aligned}$$

The second equality is according to the precalculation (48). Now calculate each cross term in (49) we will obtain (22).

□

Proof of Proposition 5

Recall vec(dX) stack the columns of dX into an $n^2 \times 1$ matrix so Cov[vec(dX)vec$(dX)^\top$] contains the items Cov$[dX_{ij}dX_{uv}], i, j, v, v = 1, \cdots, n$ locating in the matrix

$$\text{Cov}[\text{vec}(dX)\text{vec}(dX)^\top] = \begin{pmatrix} \boxed{\begin{array}{c} \text{Cov}[dX_{ij}dX_{uv}]\ , \\ j = 1, i = 1 \cdots, n, \\ v = 1, u = 1 \cdots, n \end{array}} & \cdots & \boxed{\begin{array}{c} \text{Cov}[dX_{ij}dX_{uv}]\ , \\ j = 1, i = 1 \cdots, n, \\ v = n, u = 1 \cdots, n \end{array}} \\ \vdots & \ddots & \vdots \\ \boxed{\begin{array}{c} \text{Cov}[dX_{ij}dX_{uv}]\ , \\ j = n, i = 1 \cdots, n, \\ v = 1, u = 1 \cdots, n \end{array}} & \cdots & \boxed{\begin{array}{c} \text{Cov}[dX_{ij}dX_{uv}]\ , \\ j = n, i = 1 \cdots, n, \\ v = n, u = 1 \cdots, n \end{array}} \end{pmatrix} . \quad (50)$$

Using this location plan we put the terms in (22) into the matrix form. We start with the first term $(QQ^\top)_{iu}X_{jv}$ and give its matrix expression by

$$\begin{aligned} & \left(\begin{array}{ccccccc} (QQ^\top)_{11}X_{11} & \cdots & (QQ^\top)_{1n}X_{11} & & (QQ^\top)_{11}X_{1n} & \cdots & (QQ^\top)_{1n}X_{1n} \\ \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\ (QQ^\top)_{n1}X_{1}1 & \cdots & (QQ^\top)_{nn}X_{11} & & (QQ^\top)_{n1}X_{1}n & \cdots & (QQ^\top)_{nn}X_{1n} \\ & \vdots & & \ddots & & \vdots & \\ (QQ^\top)_{11}X_{n1} & \cdots & (QQ^\top)_{1n}X_{n1} & & (QQ^\top)_{11}X_{nn} & \cdots & (QQ^\top)_{1n}X_{nn} \\ \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\ (QQ^\top)_{n1}X_{n}1 & \cdots & (QQ^\top)_{nn}X_{n1} & & (QQ^\top)_{n1}X_{n}n & \cdots & (QQ^\top)_{nn}X_{nn} \end{array}\right) \\ = \quad & X \otimes (QQ^\top) = (X \otimes \mathbf{1}_{n\times n}).*\left(\mathbf{1}_{n\times n} \otimes (QQ^\top)\right) . \end{aligned}$$

Put the other terms in (22) into the matrix expression then we can obtain (25).

□

Proof of Proposition 6

<u>Step 1</u>

In order to prove (35) we prove first the factor dynamics (1) under the real world measure, the factor dynamics (7) under the risk-neutral measure and the instantaneous rate relation (3) are satisfied in the transformed system.

For (3) it is easy to check

$$r_t = \alpha + \mathbf{tr}(\Psi X) = \alpha^{\mathcal{L}} + \mathbf{tr}(\Psi^{\mathcal{L}} X^{\mathcal{L}}) .$$

For factor dynamics (1) we want to show $dX_t^{\mathcal{L}}$ satisfies

$$dX_t^{\mathcal{L}} = \Big(kQ^{\mathcal{L}\mathcal{O}}(Q^{\mathcal{L}\mathcal{O}})^\top + M^{\mathcal{L}}X_t^{\mathcal{L}} + X_t^{\mathcal{L}}(M^{\mathcal{L}})^\top\Big)dt + Q^{\mathcal{L}}d\check{W}_t\sqrt{X_t^{\mathcal{L}}} + \sqrt{X_t^{\mathcal{L}}}(Q^{\mathcal{L}\mathcal{O}})^\top d(\check{W}_t)^\top \ , \tag{51}$$

for some $\check{W}_t$ standard independent $n \times n$ Wiener process.

We first calculate

$$dX_t^{\mathcal{L}} = \mathcal{L}dX_t\mathcal{L}^\top \tag{52}$$

$$= \Big(kQ^{\mathcal{L}\mathcal{O}}(Q^{\mathcal{L}\mathcal{O}})^\top + M^{\mathcal{L}}X_t^{\mathcal{L}} + X_t^{\mathcal{L}}(M^{\mathcal{L}})^\top\Big)dt + Q^{\mathcal{L}\mathcal{O}}dW_t^{\mathcal{O}}\sqrt{X_t}\mathcal{L}^\top + \mathcal{L}\sqrt{X_t}d(W_t^{\mathcal{O}})^\top Q^{\mathcal{L}\mathcal{O}} \ , \tag{53}$$

with $M^{\mathcal{L}}$ and $Q^{\mathcal{L}\mathcal{O}}$ specified in (29) and (30).

Comparing the last term in (53) with that in (51) we need the equality held in distribution sense

$$\mathcal{L}\sqrt{X}d(W^{\mathcal{O}})^\top \overset{\text{dist.}}{\simeq} \sqrt{X^{\mathcal{L}}}d\check{W}^\top \ . \tag{54}$$

First both of them are $n \times n$ normal distribution with mean zero. The variance of the term for the left hand side is calculated as given

$$\begin{aligned}
&\mathrm{Var}[\mathrm{vec}(\mathcal{L}\sqrt{X_t}d(W_t^{\mathcal{O}})^\top)] \\
=& \underbrace{\begin{pmatrix} \mathcal{L}\sqrt{X_t} & 0 & \cdots & 0 \\ 0 & \mathcal{L}\sqrt{X_t} & \cdots & 0 \\ \vdots & & \vdots & \\ 0 & 0 & \cdots & \mathcal{L}\sqrt{X_t} \end{pmatrix}}_{(n^2 \times n^2)} \mathrm{E}\Big[\underset{(n^2\times 1)}{\mathrm{vec}(dW_t^{\mathcal{O}\top})}\underset{(1\times n^2)}{\mathrm{vec}(dW_t^{\mathcal{O}\top})^\top}\Big] \underbrace{\begin{pmatrix} \sqrt{X_t}\mathcal{L}^\top & 0 & \cdots & 0 \\ 0 & \sqrt{X_t}\mathcal{L}^\top & \cdots & 0 \\ \vdots & & \vdots & \\ 0 & 0 & \cdots & \sqrt{X_t}\mathcal{L}^\top \end{pmatrix}}_{(n^2 \times n^2)} \\
=& \begin{pmatrix} \mathcal{L}X_t\mathcal{L}^\top & 0 & \cdots & 0 \\ 0 & \mathcal{L}X_t\mathcal{L}^\top & \cdots & 0 \\ \vdots & & \vdots & \\ 0 & 0 & \cdots & \mathcal{L}X_t\mathcal{L}^\top \end{pmatrix} dt \ .
\end{aligned}$$

And the variance of term on the right hand side is given by

$$\mathrm{Var}[\mathrm{vec}(\sqrt{X_t^{\mathcal{L}}}d\check{W}_t^\top)]$$

$$= \underbrace{\begin{pmatrix} \sqrt{X_t^{\mathcal{L}}} & 0 & \cdots & 0 \\ 0 & \sqrt{X_t^{\mathcal{L}}} & \cdots & 0 \\ \vdots & & \vdots & \\ 0 & 0 & \cdots & \sqrt{X_t^{\mathcal{L}}} \end{pmatrix}}_{(n^2\times n^2)} \mathrm{E}\Big[\underbrace{\mathrm{vec}(d\check{W}_t^\top)}_{(n^2\times 1)}\underbrace{\mathrm{vec}(d\check{W}_t^\top)^\top}_{(1\times n^2)}\Big] \underbrace{\begin{pmatrix} \sqrt{X_t^{\mathcal{L}}} & 0 & \cdots & 0 \\ 0 & \sqrt{X_t^{\mathcal{L}}} & \cdots & 0 \\ \vdots & & \vdots & \\ 0 & 0 & \cdots & \sqrt{X_t^{\mathcal{L}}} \end{pmatrix}}_{(n^2\times n^2)}$$

$$= \begin{pmatrix} X_t^{\mathcal{L}} & 0 & \cdots & 0 \\ 0 & X_t^{\mathcal{L}} & \cdots & 0 \\ \vdots & & \vdots & \\ 0 & 0 & \cdots & X_t^{\mathcal{L}} \end{pmatrix} dt \ .$$

The variance of both sides are equal so the distribution equivalence (54) is proved.

Then apply the distribution equivalent (54) into (53) then we can get (51).

For the risk-neutral dynamics (7) we follow the same argument of the real world dynamics to prove

$$dX_t = \big(\tilde{\Gamma}^{\mathcal{L}} + \tilde{M}^{\mathcal{L}}X^{\mathcal{L}} + X^{\mathcal{L}}(\tilde{M}^{\mathcal{L}})^\top\big)dt + Q^{\mathcal{L}}d\hat{W}_t\sqrt{X^{\mathcal{L}}} + \big(Q^{\mathcal{L}}d\hat{W}_t\sqrt{X^{\mathcal{L}}}\big)^\top , \tag{55}$$

$\tilde{\Gamma}^{\mathcal{L}}$ and $\tilde{M}^{\mathcal{L}}$ are given by (31) and (32), and $\hat{W}_t$ is an $n \times n$ standard Wiener process under the risk neutral measure satisfying the distribution equality.

$$\mathcal{L}\sqrt{X_t}d\tilde{W}_t^\top \overset{\text{dist.}}{\simeq} \sqrt{X_t^{\mathcal{L}}}d\hat{W}_t^\top \ . \tag{56}$$

<u>Step 2</u>

Based on Proposition 1, the no-arbitrage bond price is solved in form of

$$P(t,T) = \exp\left(a^{\mathcal{L}}(\tau) + \mathbf{tr}(C^{\mathcal{L}}(\tau)X)\right) , \tag{57}$$

where $a^{\mathcal{L}}(\tau)$ and $C^{\mathcal{L}}(\tau)$ solve the ODEs

$$\frac{d}{d\tau}C^{\mathcal{L}}(\tau) = C^{\mathcal{L}}(\tau)\tilde{M}^{\mathcal{L}} + (\tilde{M}^{\mathcal{L}})^\top C^{\mathcal{L}}(\tau) + 2C^{\mathcal{L}}(\tau)Q^{\mathcal{LO}}(Q^{\mathcal{LO}})^\top C^{\mathcal{L}}(\tau) - \Psi^{\mathcal{L}} , \tag{58}$$

$$\frac{d}{d\tau}a^{\mathcal{L}}(\tau) = \mathbf{tr}[Q^{\mathcal{LO}}(Q^{\mathcal{LO}})^\top C^{\mathcal{L}}(\tau)] , \tag{59}$$

with initial conditions $C^{\mathcal{L}}(0) = 0$ and $a^{\mathcal{L}}(0) = 0$.

We can check easily that the solution $C^{\mathcal{L}}(\tau) = (\mathcal{L}^\top)^{-1}C(\tau)\mathcal{L}^{-1}$ satisfies (58) and $a^{\mathcal{L}}(\tau) = a(\tau)$ satisfies (59) respectively. Since the solution is unique for each equation we prove (35).

□

Proof of Proposition 7

We replace $dW_t^{\mathcal{O}} = \mathcal{O}dW_t = \mathcal{O}\big(d\tilde{W}_t - \Lambda_0\sqrt{X_t}^{-1} - \Lambda_1\sqrt{X_t}\big)$ in (53) and obtain

$$\begin{aligned}
dX_t^{\mathcal{L}} \;&=\; \Big(kQ^{\mathcal{LO}}(Q^{\mathcal{LO}})^\top + M^{\mathcal{L}}X_t^{\mathcal{L}} + X_t^{\mathcal{L}}(M^{\mathcal{L}})^\top\Big)dt \\
&\quad + Q^{\mathcal{LO}}\mathcal{O}\big(d\tilde{W}_t - \Lambda_0\sqrt{X_t}^{-1} - \Lambda_1\sqrt{X_t}\big)\sqrt{X_t}\mathcal{L}^\top \\
&\quad + \mathcal{L}\sqrt{X_t}\big(d\tilde{W}_t - \Lambda_0\sqrt{X_t}^{-1} - \Lambda_1\sqrt{X_t}\big)^\top\mathcal{O}^\top Q^{\mathcal{LO}} \\
&= \Big(kQ^{\mathcal{LO}}(Q^{\mathcal{LO}})^\top - Q^{\mathcal{LO}}\mathcal{O}\Lambda_0\mathcal{L}^\top - (Q^{\mathcal{LO}}\mathcal{O}\Lambda_0\mathcal{L}^\top)^\top\Big)dt \\
&\quad + \big(M^{\mathcal{L}} - Q^{\mathcal{LO}}\mathcal{O}\Lambda_1\mathcal{L}^{-1}\big)X_t^{\mathcal{L}}dt + (X_t^{\mathcal{L}})^\top\big(M^{\mathcal{L}} - Q^{\mathcal{LO}}\mathcal{O}\Lambda_1\mathcal{L}^{-1}\big)^\top \\
&\quad + Q^{\mathcal{LO}}d\tilde{W}_t\sqrt{X_t}\mathcal{L}^\top + \mathcal{L}\sqrt{X_t}d\tilde{W}_t^\top(Q^{\mathcal{LO}})^\top\ .
\end{aligned}$$

Comparing the equation above with (7) we obtain the relations (31) and (32).

□

Proof of Proposition 8

Two things need to be shown for a proof of normalization conditions. First we show these the conditions do not restrict the WTSM so that for arbitrary given M, Q and Ψ (M can be diagonalized) we can find a transformation $\mathcal{L}$ so that the transformed parameters satisfy the conditions S1, S2 or S3. Then we show this transformation is unique.

We start with the first set S1. To a square matrix M, which can be diagonalized in (1) we apply *eigen decomposition* on M so that $M = \mathcal{L}_d\mathcal{D}\mathcal{L}_d^{-1}$ where $\mathcal{D}$ is a diagonal matrix and $\mathcal{L}_d$ consists of the eigenvectors as column vectors. Replace the M and consider the transformed factor $X_t^d := \mathcal{L}_d^{-1}X_t(\mathcal{L}_d^{-1})^\top$, the dynamics in (1) becomes

$$dX_t^d = \big(kQ_dQ_d^\top + \mathcal{D}X_t^d + X_t^d\mathcal{D}\big)dt + Q_d dW_t\sqrt{X_t}(\mathcal{L}_d^{-1})^\top + \mathcal{L}_d^{-1}\sqrt{X_t}dW_t^\top Q_d^\top\ , \tag{60}$$

with $Q_d = \mathcal{L}_d^{-1}Q$.

Use the distribution equivalence (56) we have

$$\mathcal{L}_d^{-1}\sqrt{X_t}dW_t^\top \overset{\text{dist.}}{\simeq} \sqrt{X_t^d}d\breve{W}_t^\top\ , \quad dW_t\sqrt{X_t}(\mathcal{L}_d^{-1})^\top \overset{\text{dist.}}{\simeq} d\breve{W}_t\sqrt{X_t^d}\ .$$

Now we apply the *QR decomposition*[14] to $Q_d^\top$ so that $Q_d = \mathcal{R}^\top \mathcal{O}^\top$ where $\mathcal{R}$ is an upper triangular matrix and $\mathcal{O}$ is an orthogonal matrix. Use it to rewrite the term

$$Q_d dW_t \sqrt{X_t} (\mathcal{L}_d^{-1})^\top \overset{\text{dist.}}{\simeq} \mathcal{R}^\top \mathcal{O}^\top d\check{W}_t \sqrt{X_t^d} \overset{\text{dist.}}{\simeq} \mathcal{R}^\top d\check{W}_t \sqrt{X_t^d} \ ,$$

where $\check{W}_t := \mathcal{O}^\top \check{W}_t$ is a new $n \times n$ standard Wiener process. Summarize the transformation above the dynamics (60) becomes

$$dX_t^d = \big(kR^\top R + \mathcal{D}X_t^d + X_t^d \mathcal{D}\big)dt + \mathcal{R}^\top d\check{W}_t \sqrt{X_t^d} + \sqrt{X_t^d} d\check{W}_t^\top \mathcal{R} \ , \tag{61}$$

where $\mathcal{D}$ is diagonal and $\mathcal{R}^\top$ is lower triangular. Note that the sign of the i-th element on the diagonal of $\mathcal{R}$ can be changed by changing the sign of the i-th row in $\check{W}_t$.

In order to fit (c) we preform a re-scaling. From Proposition 6 we know Ψ is transformed to $\Psi_d := \mathcal{L}_d^\top \Psi \mathcal{L}_d$. Since Ψ is assumed to be strictly positive definite the diagonal elements in Ψ are positive So we can define the re-scaling $\mathcal{C} = \text{diag}\big(\sqrt{\Psi_{d,11}}, \cdots, \sqrt{\Psi_{d,nn}}\big)$ where $\Psi_{d,ii}$ is the i-th element on the diagonal of Ψ_d. Consider the transformation

$$X_t^c := \mathcal{C} X_t^d \mathcal{C} = \mathcal{C} \mathcal{L}_d^{-1} X_t (\mathcal{L}_d^{-1})^\top \mathcal{C}.$$

We still have freedom for ”sign transformation” which is

$$\mathcal{S} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \delta_2 & 0 & 0 \\ 0 & 0 & \ddots & \\ 0 & 0 & 0 & \delta_n \end{pmatrix} .$$

This transformation $(\mathcal{S}^\top)^{-1} \Psi \mathcal{S}^{-1}$ gives a new Ψ^S that

$$\Psi^S = \Big(\delta_i \delta_j \Psi_{ij}\Big) .$$

So this transformation keeps the sign of the diagonal of Ψ and the requirement that the first row of Ψ has positive elements can fix $\mathcal{S}$.

[14]It means a matrix can be decomposed into an orthogonal matrix times a upper triangular matrix.

Through direct calculation we know the new M, Q and Ψ obtained from (29), (30) and (34) satisfy the conditions (a1), (b1) and (c). The transformation is unique since $(\mathcal{L}_d, \mathcal{R}, \mathcal{C}, \mathcal{S})$ are uniquely defined by (M, Q, Ψ).

The proof for the conditions S2 is similar with that above. To any given square matrix M in (1) we apply *Schur decomposition*[15] on M so that $M = \mathcal{L}_b^\top \mathcal{B} \mathcal{L}_b$ where $\mathcal{L}_b$ is a unitary matrix and $\mathcal{B}$ is a lower triangular matrix[16] Consider the transformed factor $X_t^b := \mathcal{L}_b X_t \mathcal{L}_b^\top$ and the distribution equivalence $\mathcal{L}_b \sqrt{X_t} dW_t = \sqrt{X_t^b} d\breve{W}_t^\top$ using (56) we can obtain

$$dX_t^b = \left(kQ_bQ_b^\top + \mathcal{B}X_t^b + X_t^b\mathcal{B}^\top\right)dt + Q_b d\breve{W}_t \sqrt{X_t^b} + \sqrt{X_t^b} d\breve{W}_t^\top Q_b^\top \ , \tag{62}$$

with $Q_b := \mathcal{L}_b Q$. Apply again the QR decomposition $Q_b^\top$ so that $Q_b = \mathcal{R}^\top \mathcal{O}^\top$ where $\mathcal{R}$ is an upper triangular matrix and $\mathcal{O}$ is an orthogonal matrix. Replace Q_b and consider the dynamics of the transformed factor $X_t^r := (\mathcal{R}^\top)^{-1} X_t^b \mathcal{R}^{-1}$ we can have

$$dX_t^r = \left(k + \mathcal{B}_r X_t^r + X_t^r \mathcal{B}_r^\top\right)dt + \mathcal{O}^\top d\breve{W}_t \sqrt{X_t^b} \mathcal{R}^{-1} + (\mathcal{R}^\top)^{-1} \sqrt{X_t^b} d\breve{W}_t^\top \mathcal{O} \ , \tag{63}$$

where $\mathcal{B}_r := (\mathcal{R}^\top)^{-1} \mathcal{B} \mathcal{R}^\top$ is still lower triangular.

Define a new $n \times n$ Wiener process $\breve{W}_t := \mathcal{O}^\top \breve{W}_t$ and note that $d\breve{W}_t \sqrt{X_t^b} \mathcal{R}^{-1} \overset{\text{dist.}}{\simeq} d\breve{W}_t \sqrt{X_t^r}$ due to the distribution equivalence (56). The conditions (a2) and (b2) are satisfied. Condition (c) is obtained again through re-scaling.

Regarding the normalization conditions S3, we observe first that $\Lambda_1 Q$ is invariant which means $\Lambda_1^{\mathcal{L}} = Q^{\mathcal{L}}$. Adopt the Schur decomposition that $\Lambda_1 Q = \mathcal{O}^\top U \mathcal{O}$ where $\mathcal{O}$ is unitary orthogonal and U is upper triangular.

Let $\mathcal{L} = \mathcal{O}\Lambda_1$. Then based on (30) we obtain

$$Q^{\mathcal{L}\mathcal{O}} = \mathcal{L} Q \mathcal{O}^\top = \mathcal{O} \Lambda Q \mathcal{O}^\top = U \ \text{ (upper triangular)} \ ,$$

$$\Lambda_1^{\mathcal{L}\mathcal{O}} = \mathcal{O}\Lambda_1 \mathcal{L}^{-1} = \mathcal{O}\Lambda_1 (\mathcal{O}\Lambda_1)^{-1} = I \ .$$

□

Proof of Proposition 9

[15]Schur Decomposition is stated for a square complex matrix and the transpose operator $\top$ is actually conjugate transpose.
[16]For a unitary matrix $\mathcal{L}_b$, $\mathcal{L}_b^\top = \mathcal{L}_b^{-1}$.

The bond return under the risk neutral measure is given by

$$\frac{dP(t,T;X_t)}{P(t,T;X_t)} = r_t dt + \mathbf{tr}\Big[(Qd\tilde{W}_t\sqrt{X_t} + \sqrt{X_t}d\tilde{W}_t^\top Q^\top)C(\tau)\Big] \, , \tag{64}$$

with the bond having the risk neutral return r_t. Using the change of measure (5) and the assumption of the market price of risk (6) we obtain the risk premia under the real world measure, so that

$$\begin{aligned}\frac{dP(t,T;X_t)}{P(t,T;X_t)} &= r_t dt + \mathbf{tr}\Big[\Big(Q(\Lambda_0 + \Lambda_1 X_t)\sqrt{X_t} + \sqrt{X_t}(\Lambda_0^\top + X_t^\top \Lambda_1^\top)Q^\top\Big)C(\tau)\Big]dt \\ &\quad + \mathbf{tr}\Big[(QdW_t\sqrt{X_t} + \sqrt{X_t}dW_t^\top Q^\top)C(\tau)\Big] \\ &= r_t dt + 2\mathbf{tr}\big[Q(\Lambda_0 + \Lambda_1 X_t)C(\tau)\big]dt + \mathbf{tr}\Big[(QdW_t\sqrt{X_t} + \sqrt{X_t}dW_t^\top Q^\top)C(\tau)\Big] \qquad (65) \\ &= r_t dt + \mathbf{tr}\big[(kQQ^\top - \tilde{\Gamma})C(\tau)\big]dt + 2\mathbf{tr}\big[(M - \tilde{M})X_t C(\tau)\big]dt \qquad (66) \\ &\quad + \mathbf{tr}\Big[(QdW_t\sqrt{X_t} + \sqrt{X_t}dW_t^\top Q^\top)C(\tau)\Big] \, .\end{aligned}$$

The excess return above the instantaneous rate is called the (instantaneous) risk premia, so we write

$$\begin{aligned}\text{Risk Premium} = e_t &= 2\mathbf{tr}\big[Q(\Lambda_0 + \Lambda_1 X_t)C(\tau)\big] \\ &= \mathbf{tr}\big[(kQQ^\top - \tilde{\Gamma})C(\tau)\big] + 2\mathbf{tr}\big[(M - \tilde{M})X_t C(\tau)\big] \, .\end{aligned}$$

Appendix B. Estimation using extended Kalman filter

Recall that the Wishart process (1) is:

$$dX_t = \big(kQQ^\top + MX_t + X_t M^\top\big)dt + QdW_t\sqrt{X_t} + \sqrt{X_t}dW_t^\top Q^\top \, . \tag{67}$$

The Eular-Maruyama method is used to discretize this process according to

$$X_{t+\Delta} = X_t + \big(kQQ^\top + MX_t + X_t M^\top\big)\Delta + Q\Delta W_t\sqrt{X_t} + \sqrt{X_t}\Delta W_t Q^\top \, . \tag{68}$$

A positive control given later in equation (73) is imposed on X_t in order to retain the positivity of the simulated X_t. It corresponds to a *full truncation scheme* which is the best approximation among several schemes as shown in Lord et al. (2010). Let $\vec{X}_t$ denote the column obtained by stacking the columns of the matrix X_t. The discretized dynamics can then be represented by

$$\vec{X}_{t+\Delta} = f + F\vec{X}_t + U_{t+\Delta} \ , \tag{69}$$

where

$$f = \mathrm{vec}(kQQ^\top)\Delta \ , \quad F = I_{n^2} + \big(I_n \otimes M + M \otimes I_n\big)\Delta$$

and

$$\mathrm{Cov}[U_{t+\Delta}] = (I_{n^2} + K_{n,n})\big(X \otimes QQ^\top + QQ^\top \otimes X\big)\Delta$$

according to Proposition 5.

The bond yields are modelled based on the bond price formula (12) but the observation of the bond yields is contaminated with measurement errors. Let a $d \times 1$ vector y_t represent the bond yields of the d different times to maturity $\tau_1, \cdots, \tau_d$ observed at time t. The bond yield equation is given by ϵ_t so that the bond yield y_t is given by

$$y_t := \begin{pmatrix} -\frac{1}{\tau_1}\ln P(t, t+\tau_1; X_t) \\ \vdots \\ -\frac{1}{\tau_d}\ln P(t, t+\tau_d; X_t) \end{pmatrix} = j + J\vec{X}_t + \epsilon_t \ , \tag{70}$$

where

$$j = -\begin{pmatrix} a(\tau_1)/\tau_1 \\ \vdots \\ a(\tau_d)/\tau_d \end{pmatrix}_{d\times 1} ,$$

$$J = -\begin{pmatrix} C_{11}(\tau_1)/\tau_1 & C_{21}(\tau_1)/\tau_1 & C_{12}(\tau_1)/\tau_1 & C_{22}(\tau_1)/\tau_1 \\ \vdots & \vdots & \vdots & \vdots \\ C_{11}(\tau_d)/\tau_d & C_{21}(\tau_d)/\tau_d & C_{12}(\tau_d)/\tau_d & C_{22}(\tau_d)/\tau_d \end{pmatrix}_{d\times 4} ,$$

$\vec{X} = (X_{11}, X_{21}, X_{12}, X_{22})^\top$, and the measurement error ϵ_t is a $d \times 1$ zero mean random variable with the distribution $\mathcal{N}(0, \sigma_\epsilon^2 I_d)$, i.i.d. across all time points t.

We summarize the algorithm of the *extended Kalman filter* as follows[17]. Let Y_t denote all the bond yields observed until t. Let $\hat{X}_{t|s} = \mathrm{E}[\vec{X}_t|Y_s]$ and $P_{t|s} = \mathrm{Cov}[\vec{X}_t|Y_s] = \mathrm{E}[(\vec{X}_t -$

[17]For details see Harvey (1989).

$\hat{X}_{t|s})(\vec{X}_t - \hat{X}_{t|s})^\top|Y_s]$. We start with the initial state $\hat{X}_{0|0}$ and the covariance $P_{0|0}$. The algorithm runs iteratively and each iteration consists of two steps.

The first step, the *prediction* step, predicts the states based on the last time $t - \Delta$:

$$
\begin{aligned}
\hat{X}_{t|t-\Delta} &= f + F\hat{X}_{t-\Delta|t-\Delta} \ , \\
P_{t|t-\Delta} &= FP_{t-\Delta|t-\Delta}F^\top + \text{Cov}[U_t] \ .
\end{aligned}
\tag{71}
$$

The second step, the *updating* step, updates the states as new information y_t comes in:

$$
\begin{aligned}
\hat{X}_{t|t} &= \hat{X}_{t|t-\Delta} + K_t\big(y_t - JX_{t|t-\Delta} - j\big) \ , \\
P_{t|t} &= P_{t|t-\Delta} - K_tJP_{t|t-\Delta} \ ,
\end{aligned}
\tag{72}
$$

where the gain matrix K_t is given by

$$K_t := P_{t|t-\Delta}J^\top\Sigma_{t|t-\Delta}^{-1}$$

and the observation covariance $\Sigma_{t|t-\Delta}$

$$\Sigma_{t|t-\Delta} := JP_{t|t-\Delta}J^\top + \sigma_\epsilon^2 I_d.$$

Let $\hat{y}_{t|t-\Delta}$ be the prediction of the observation $\hat{y}_t = j + J\hat{X}_{t|t-\Delta}$ and $v_{t|t-\Delta}$ be the prediction error $v_t = y_t - \hat{y}_{t|t-\Delta}$. Note the $\text{Cov}[v_{t|t-\Delta}|Y_{t-\Delta}] = \Sigma_{t|t-\Delta}$. The likelihood function $L(Y_T, \Theta)$ of the observations Y_T and the parameter Θ is given by

$$L(Y_T, \Theta) = \prod_{i=1}^{N} l(y_{i\Delta}|Y_{(i-1)^\Delta}, \Theta) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}^d\sqrt{\det\Sigma_i}} \exp\big(-\frac{1}{2}v_i^\top\Sigma_i^{-1}v_i\big) \ ,$$

where l represents the conditional likelihood function, $\Sigma_i := \Sigma_{i\Delta|(i-1)\Delta}$ and $v_i = v_{i\Delta|(i-1)\Delta}$. The *maximum likelihood estimator* maximizes the likelihood function based on the extended Kalman filter method, so that

$$\hat{\Theta}_{ml} = \max_{\Theta} L(Y_T, \Theta) \ .$$

We impose **positivity control** in the estimation of the Wishart model because by construction the Wishart process (1) at each t is a positive definite matrix almost surely so that the square root of X_t is well-defined. In the estimation the positivity cannot be guaranteed due to errors caused by the discretization (68). Also the updating step (72) in the extended Kalman filter cannot guarantee that the updated factor $\hat{X}_{t|t}$ is positive definite. The positive control for a 2×2 Wishart model is given by

$$X_{11,t} > 0 \ , \quad X_{22,t} > 0 \ , \quad X_{11,t}X_{22,t} > X_{12,t}X_{21,t} \ , \tag{73}$$

for all observation times t and all samples. All the inequalities are required to hold strictly. The positivity control is imposed for every prediction step (71) and every updating step (72).

For the choice of the initial state level and state covariance we do not take the unconditional expectation of the level and covariance as in Duan and Simonato (1999) but rather we treat them as unknown parameters to be estimated. This is necessary for two reasons: (1) the estimation results later show that the initial state level is not its mean level, (2) the unconditional expectation is hard to pindown if the reversion speed is very low as in our case.

We adopt an iterative maximization procedure: first we maximize the likelihood function with respect to just the model parameters for a fixed initial state level and covariance and then adjust the initial state and the covariance through the smoothing algorithm. We keep iterating the process until it converges. We found that this is a more efficient way to obtain the maximum because the sensitivity of the likelihood function to the initial state level and covariance is much lower than to the parameters, so it is difficult to attain the minimum for both simultaneously.

In terms of optimization, three methods are used: the simplex method `fminsearch`[18], the gradient method `fminunc`[19] and the simulated annealing method[20] provided by `Matlab`. The simplex method is a derivative-free method and is used for the initial search of parameters. It sometimes has problems in converging so we use the other two methods to find the local minimum after a global search. Note that the gradient method experiences the difficulty that even though the method has converged according to the stopping criterion, the minimum has not in fact been attained. The simulated annealing method is the most reliable method among the three for attaining the local minimum.

---

[18] See `http://www.mathworks.com/access/helpdesk/help/toolbox/optim/ug/fminsearch.html`.

[19] `fminunc` uses the BFGS Quasi-Newton method with a cubic line search procedure to search for the minimum, see `http://www.mathworks.com/access/helpdesk/help/toolbox/optim/ug/fminunc.html`.

[20] Obtained from Matlab Central.

Appendix C. Liquidity measures

Commercial paper spread is defined as the difference between the yield on a 3-month commercial paper and that of a 3-month T-bill.

TED spread is defined as the difference between 3-month LIBOR rate and the yield on a 3-month T-bill.

10 year swap spread is defined as the difference between 10-year swap rate and 10-year strip bond yield.

3 month T-bill spread is defined as the difference between the yield on 3-month T-bill and the Fed funds rate.

Moody Baa Aaa spread is defined as the difference between Moody’s Baa corporate bond yield and Moody’s Aaa corporate bond yield.

Data are downloaded from the Federal Reserve website and Bloomberg.

References

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

Buraschi, A., Cieslak, A. & Trojani, F. (2008), Correlation Risk and the Term Structure of Interest Rates, Working paper, Imperial College London.

Buraschi, A., Porchia, P. & Trojani, F. (2010), 'Correlation risk and the term structure of interest rates', *Journal of Finance* **65**(1).

Campbell, J. Y., Lo, A. W. & MacKinlay, A. G. (1996), *The Econometrics of Financial Markets*, Princeton University Press.

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

Cheridito, P., Filipović, D. & Kimmel, R. L. (2007), 'Market Price of Risk Specifications for Affine Models: Theory and evidence', *Journal of Financial Economics* **83**(1), 123–170.

Cieslak, A. & Povala, P. (2011), Undertanding Bond Risk Premia, Working paper, University of Lugano, Swiss Finance Institute.

Cochrane, J. H. & Piazzesi, M. (2005), 'Bond Risk Premia', *The American Economic Review* **95**(1), 138–160.

Cooper, I. & Priestley, R. (2009), 'Time-Varying Risk Premiums and the Output Gap', *Review of Financial Studies* **22**(7), 2801–2833.

Cox, J. C., Ingersoll, J. E. & Ross, S. A. (1985b), 'A Theory of the Term Structure of Interest Rates', *Econometrica* **53**(2), 385–407.

Cuchiero, C., Filipović, D., Mayerhofer, E. & Teichmann, J. (2011), 'Affine Processes on Positive Semidefinite Matrices', *The Annals of Applied Probability* . forthcoming.

Da Fonseca, J., Grasselli, M. & Tebaldi, C. (2007), 'Option Pricing When Correlation Are Stochastic: An Analytical Framework', *Review of Derivatives Research* **10**(2), 151–180.

Dai, Q. & Singleton, K. J. (2000), ‘Specification Analysis of Affine Term Structure Models’, *Journal of Finance* **55**(5), 1943–1978.

Driessen, J., Maenhout, P. J. & Vilkov, G. (2009), ‘The Price of Correlation Risk: Evidence from Equity Options’, *Journal of Finance* **64**(3), 1377–1406.

Duan, J.-C. & Simonato, J.-G. (1999), ‘Estimating and Testing Exponential-Affine Term Structure Models by Kalman Filter’, *Review of Quantitative Finance and Accounting* **13**(2), 111–135.

Duarte, J. (2004), ‘Evaluating an Alternative Risk Preference in Affine Term Structure Models’, *The Review of Financial Studies* **17**(2), 379–404.

Duffee, G. R. (1999), ‘Estimating the Price of Default Risk’, *The Review of Financial Studies* **12**(1), 197–226.

Duffee, G. R. (2002), ‘Term Premia and Interest Rate Forecasts in Affine Models’, *Journal of Finance* **57**(1), 405–443.

Duffee, G. R. (2010), Sharpe Ratios in Term Structure Models, Working paper, Johns Hopkins University.

Duffee, G. R. (2011), ‘Information in (and not in) the Term Structure’, *Review of Financial Studies* .

Duffee, G. R. & Stanton, R. H. (2008), ‘Evidence on Simulation Inference for Near Unit-Root Processes with Implications for Term Structure Estimation’, *Journal of Financial Econometrics* **6**, 108–142.

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

Duffie, D. & Singleton, K. J. (1997), ‘An Econometric Model of the Term Structure of Interest Rate Swap Yields’, *Journal of Finance* **52**, 1287–1321.

Gourieroux, C. (2006), ‘Continuous Time Wishart Process for Stochastic Risk’, *Econometric Reviews* **25**(2-3), 177–217.

Gourieroux, C., Jasiak, J. & Sufana, R. (2009), ‘The Wishart Autoregressive Process of Multivariate Stochastic Volatility’, *Journal of Econometrics* **150**, 187–181.

Gourieroux, C. & Sufana, R. (2006), ‘A Classification of Two-Factor Affine Diffusion Term Structure Models’, *Journal of Financial Econometrics* **4**, 31–52.

Grasselli, M. & Tebaldi, C. (2008), ‘Solvable Affine Term Structure Models’, *Mathematical Finance* **18**(1), 135–153.

Hamilton, J. D. & Wu, J. (2011), Identification and Estimation of Gaussian Affine Term Structure Models, Working paper, University of California, San Diego.

Joslin, S., Le, A. & Singleton, K. J. (2011), Why Gaussian Macro-Finance Term Structure Models Are (Nearly) Unconstrained Factor-VARs, Working paper, MIT Sloan School of Management.

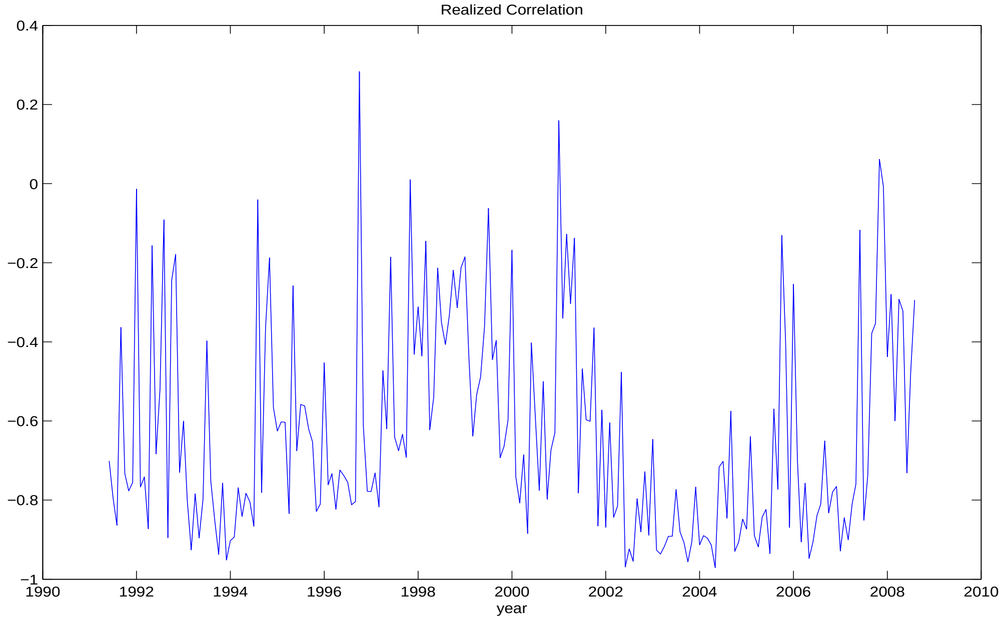
Joslin, S., Priebsch, M. & Singleton, K. J. (2011), Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks, Working paper, MIT Sloan School of Management.

Joslin, S., Singleton, K. J. & Zhu, H. (2011), ‘A New Perspective on Gaussian Dynamic Term Structure Models’, *Review of Financial Studies* **24**(3), 926–970.

Lord, R., Koekkoek, R. & van Dijk, D. (2010), ‘A comparison of biased simulation schemes for stochastic volatility’, *Quantitative Finance* **10**(2), 177–194.

Ludvigson, S. C. & Ng, S. (2009), ‘Macro Factors in Bond Risk Premia’, *The Review of Financial Studies* **22**(12), 5027–5067.

Vasicek, O. (1977), ‘An Equilibrium Characterization of the Term Structure’, *Journal of Financial Economics* **5**(2), 177–188.



The monthly realized correlation is between a 30-year yield and the yield spread (3-Month v.s. 30-Year yields) of US STRIPS from May 1991 to July 2008. It is a monthly realized correlation calculated from daily data collected from *Bloomberg*.

FIGURE 1. Realized correlation

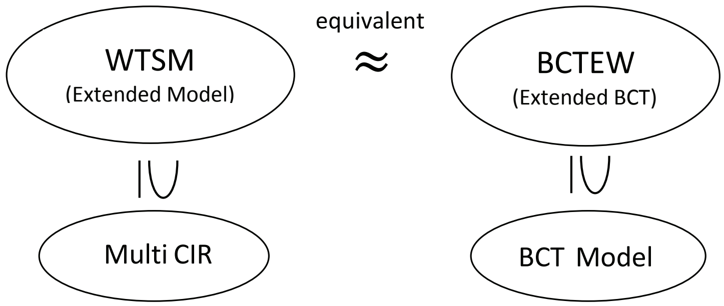


FIGURE 2. Relations of the models

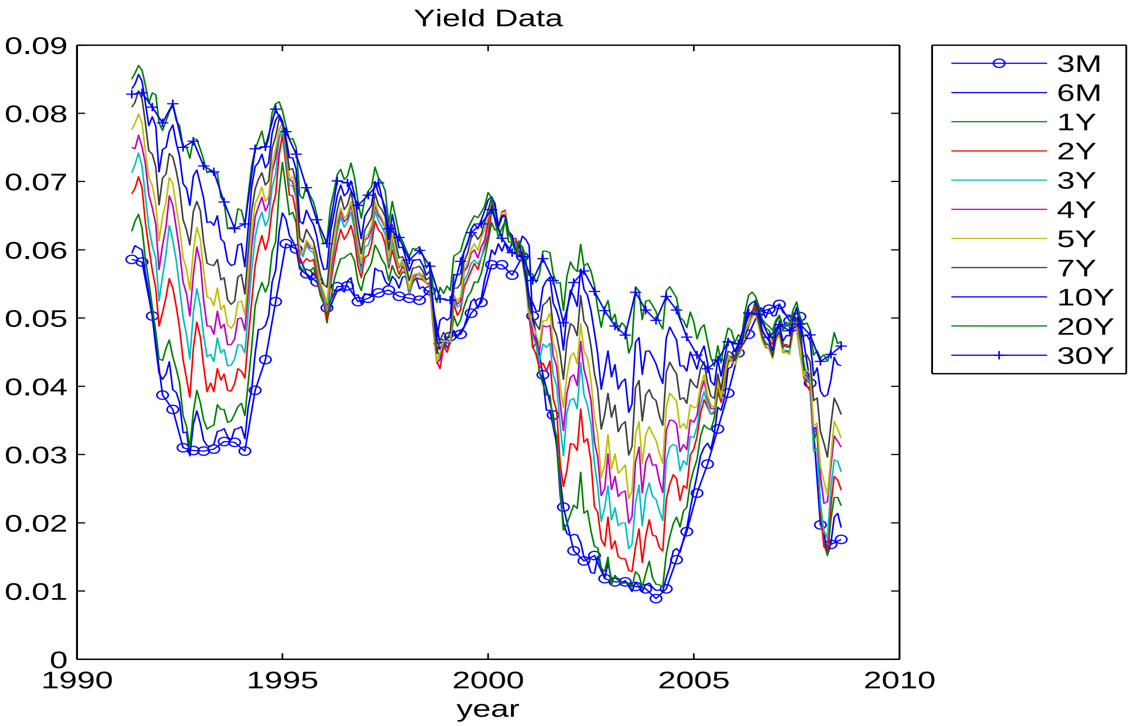
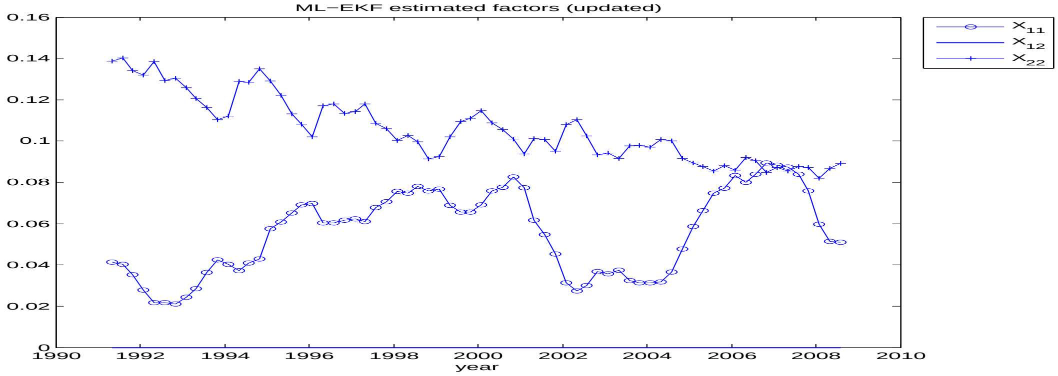
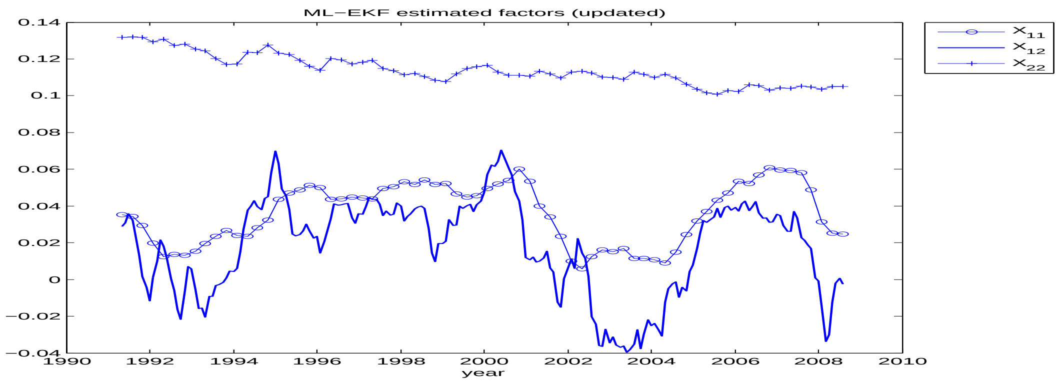


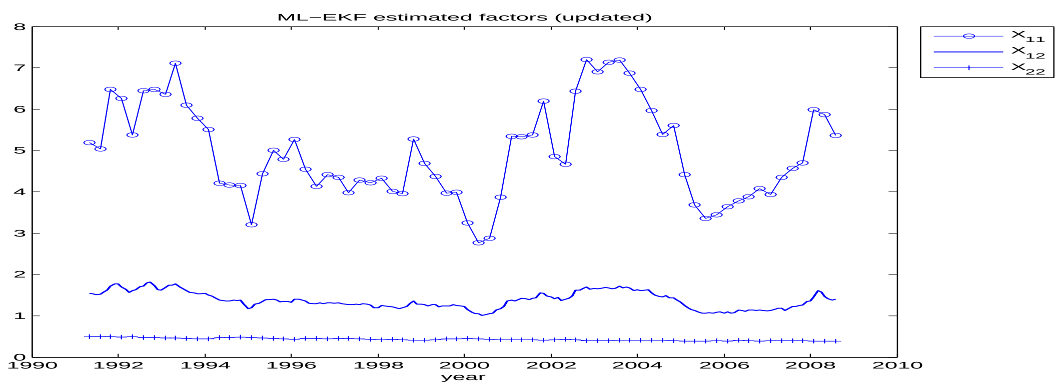
FIGURE 3. Monthly US yield data



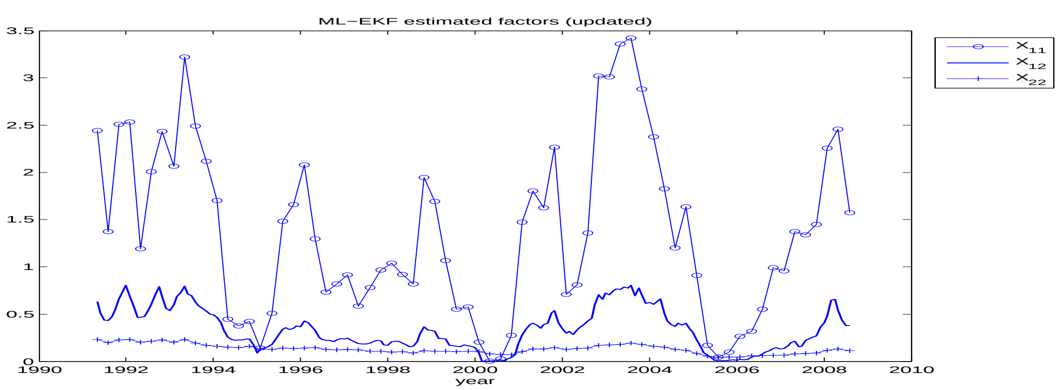
(a) CIR



(b) WTSM

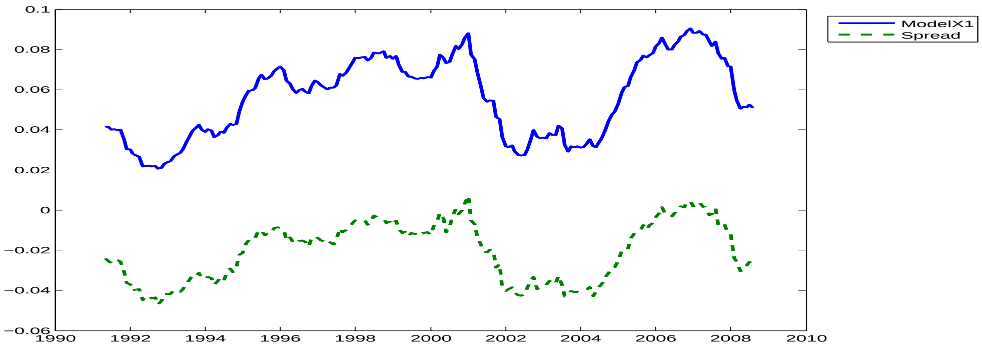


(c) BCTEW

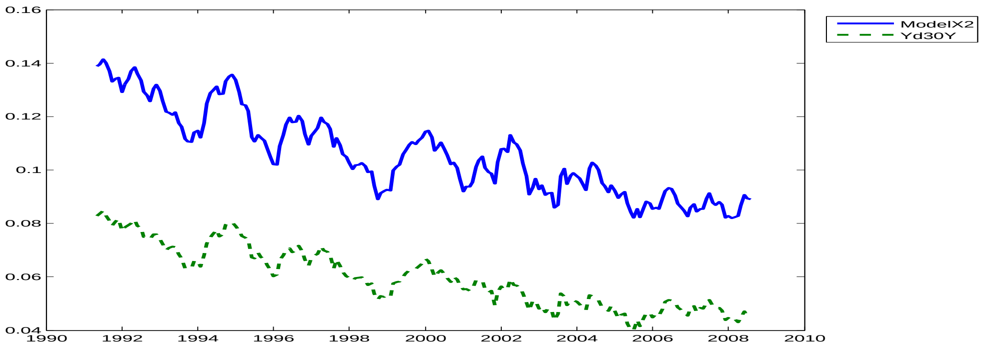


(d) BCT

FIGURE 4. Filtered factors from the four models

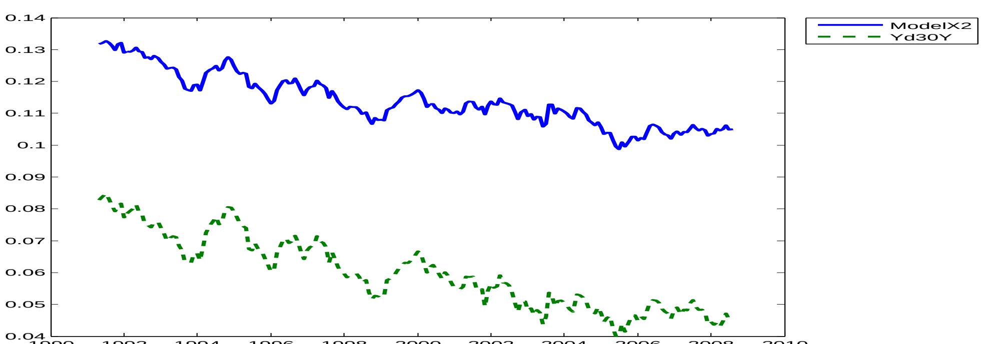


(a) MCIR X_1 factor and Spread (Corr = 98.74%)



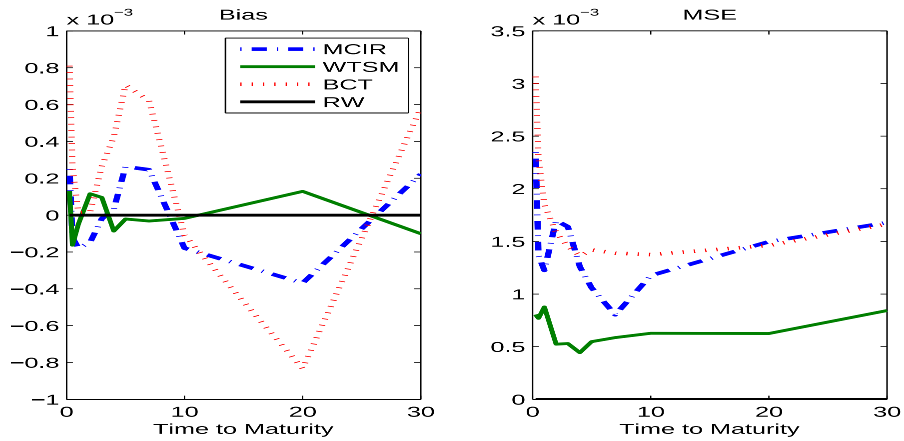
(b) MCIR X_2 factor and Long yield (Corr = 97.87%)

(c) WTSM X_{11} factor and Spread (Corr =98.26%)

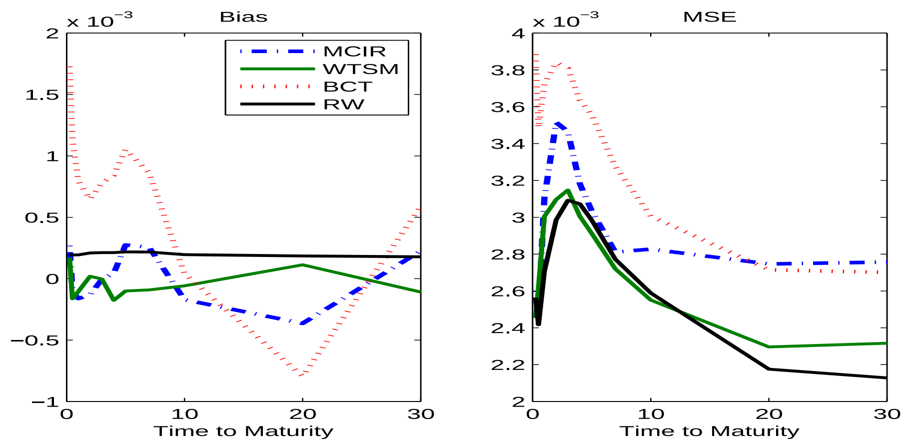


(d) WTSM X_{22} factor and Long yield (Corr =97.31%)

FIGURE 5. Comparison of the estimated factors and the economic factors



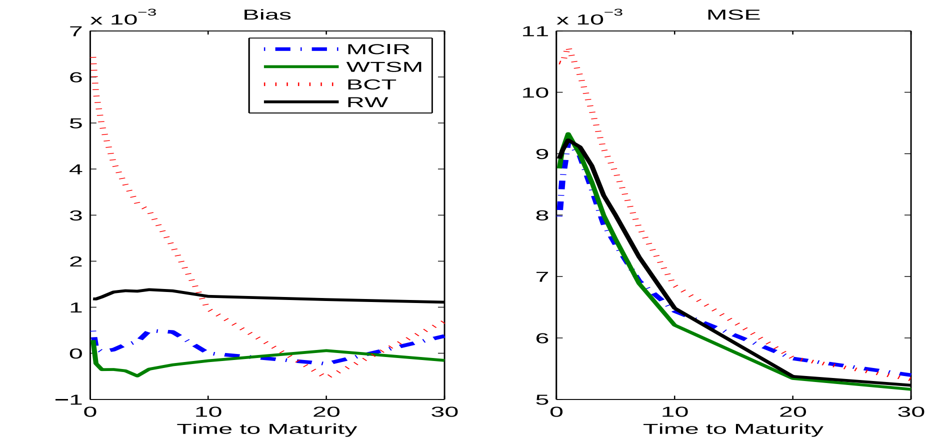
(a) Fitting Errors



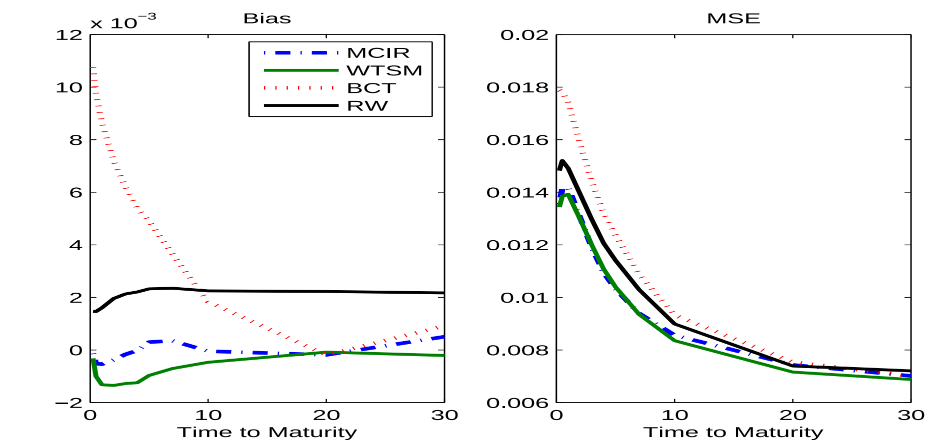
(b) One Month Ahead Forecast

The left hand panel in (a) gives average fitting errors of the bonds against the time to maturity for all three models: MCIR, WTSM and BCT. The right hand panel in (a) depicts the mean square errors (MSE) of the fitting errors. The panels in (b) illustrate these two error measures for one month ahead forecasts (ie. 1-step ahead forecasts). All forecasts are in-sample.

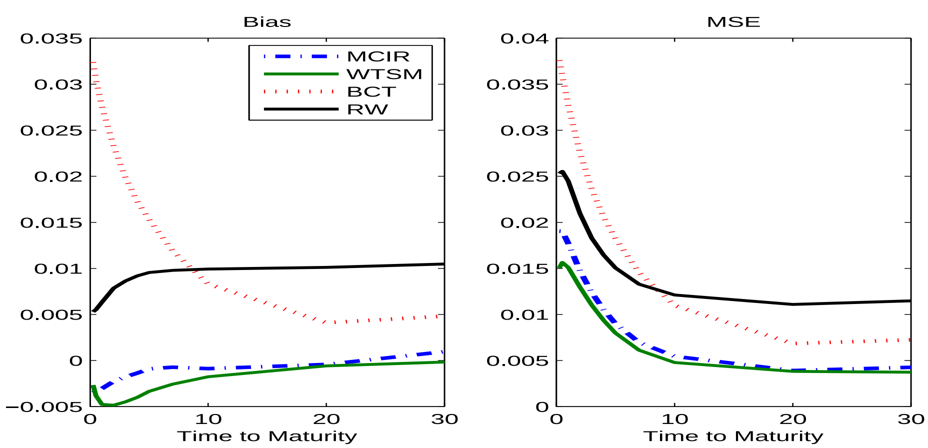
FIGURE 6. Fitting Errors and 1-step ahead forecasts of all the Bonds



(a) 6-month Ahead Forecast



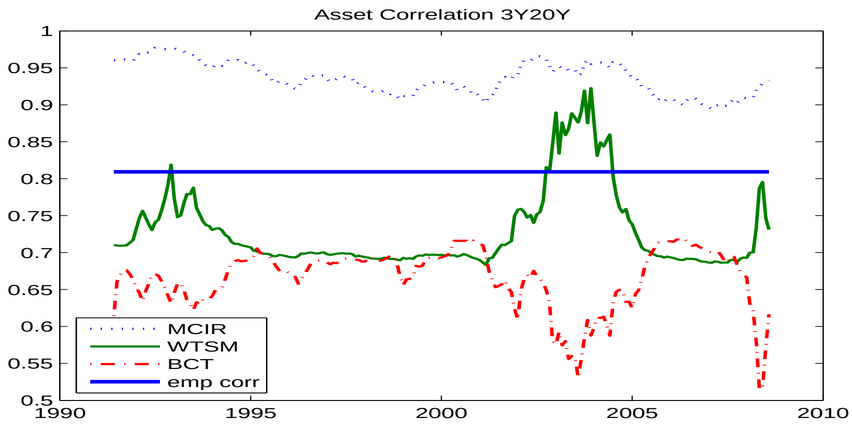
(b) One Year Ahead Forecast



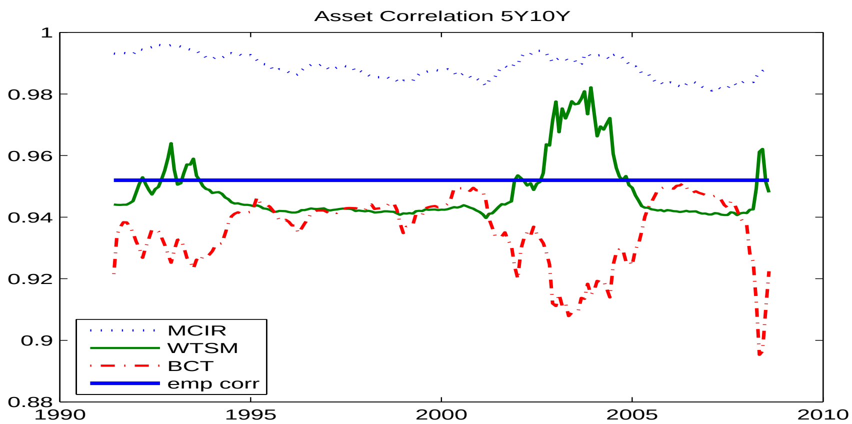
(c) Five Year Ahead Forecast

The left hand panels gives average forecast errors of the bonds against the time to maturity for all three models: MCIR, WTSM and BCT. The right hand panels depict the mean square errors (MSE) of the forecast errors. The forecast horizons are 6-month, one-year ahead and two-year ahead in each panel respectively. All forecasts are in-sample.

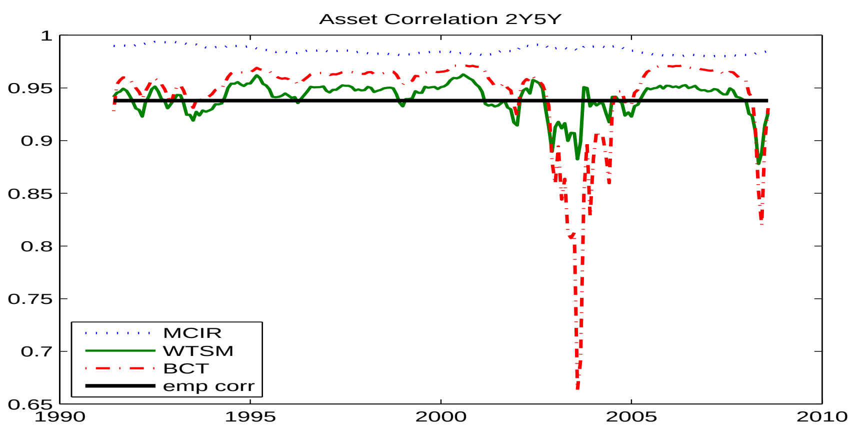
FIGURE 7. Medium-Term Forecast of all the Bonds



(a) Large difference in time to maturity: 3- and 20- year bonds

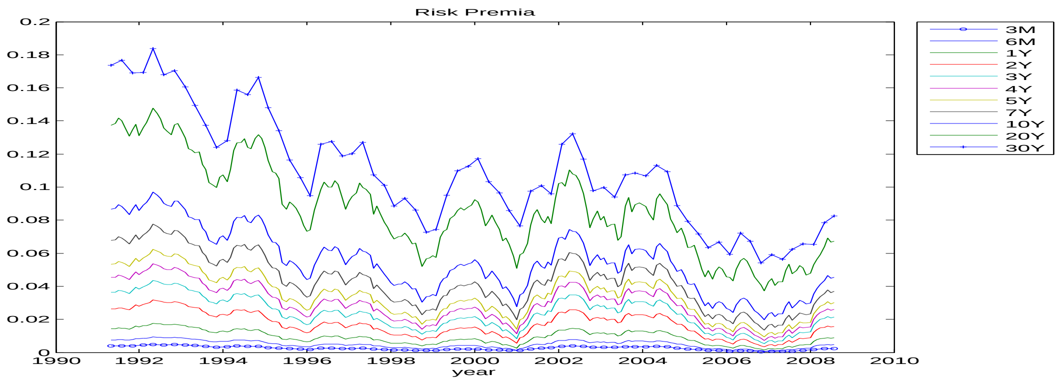


(b) Medium difference in time to maturity: 5- and 10- year bonds

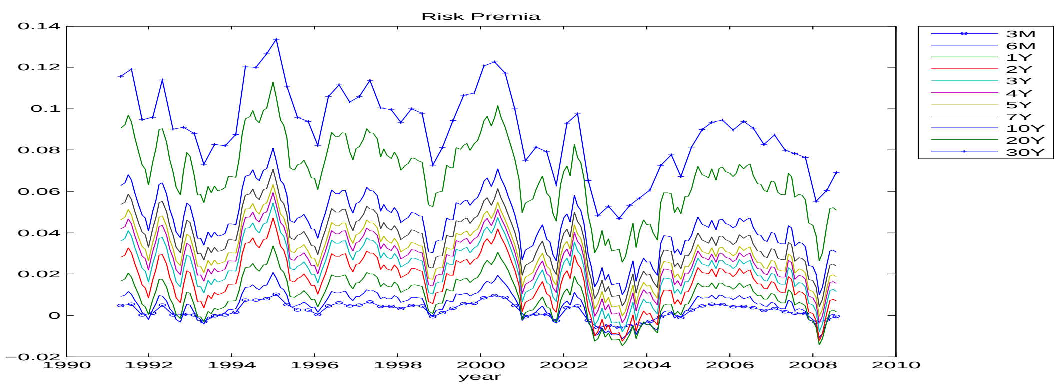


(c) Small difference in time to maturity: 2- and 5- year bonds

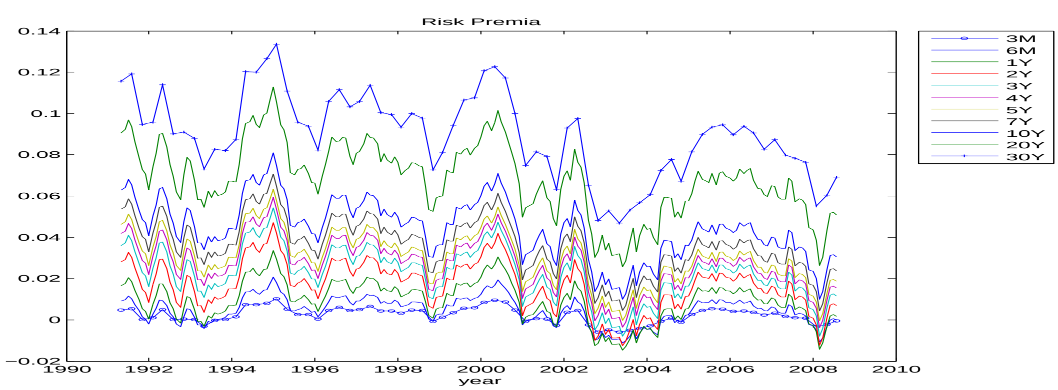
FIGURE 8. Asset Correlation



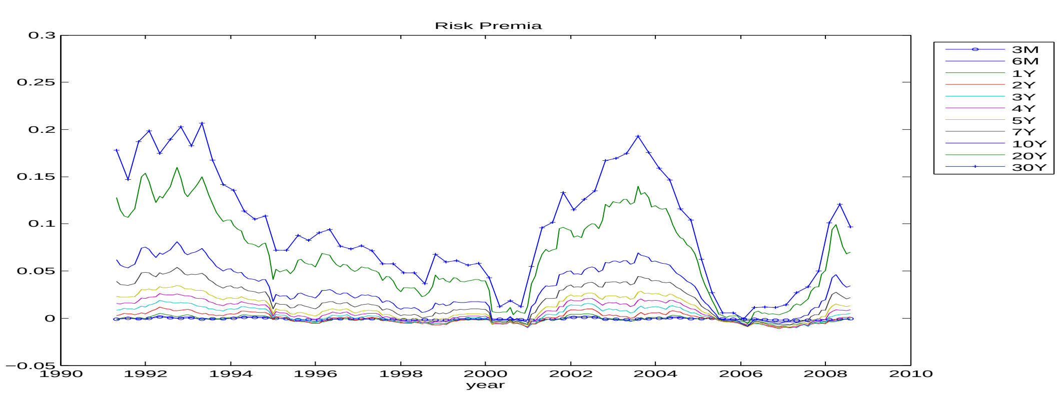
(a) MCIR



(b) WTSM

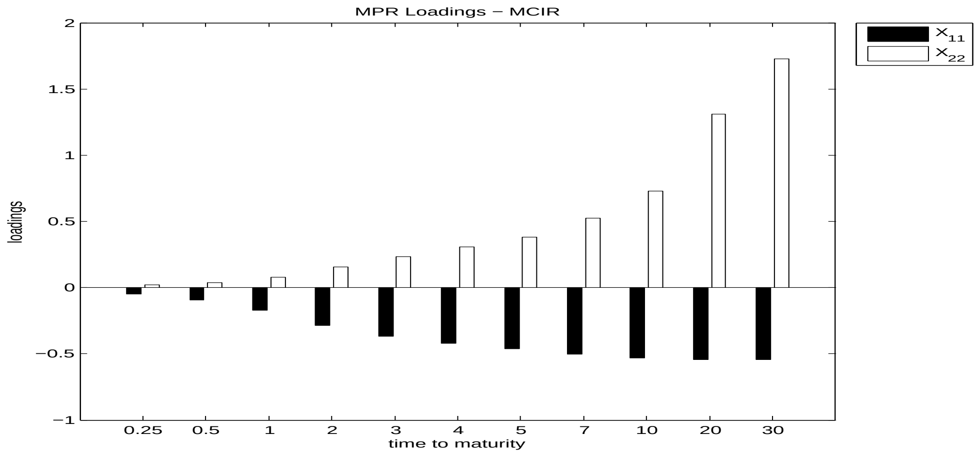


(c) BCTEW

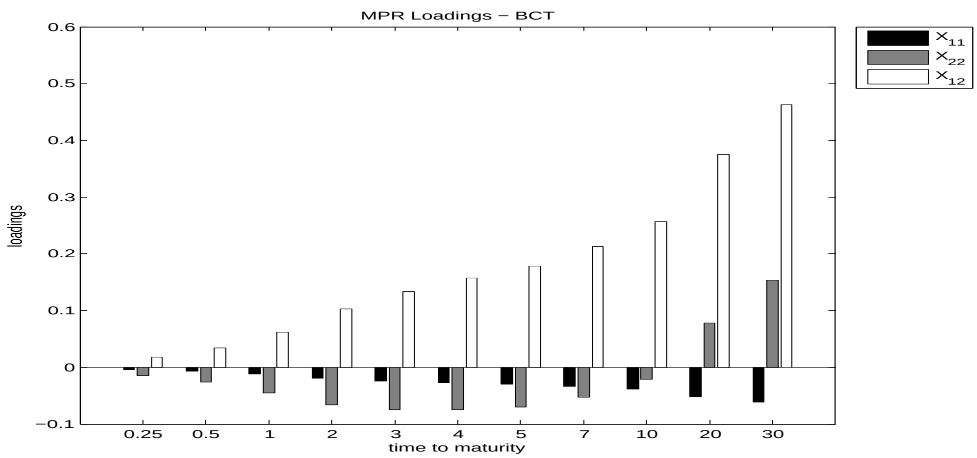


(d) BCT

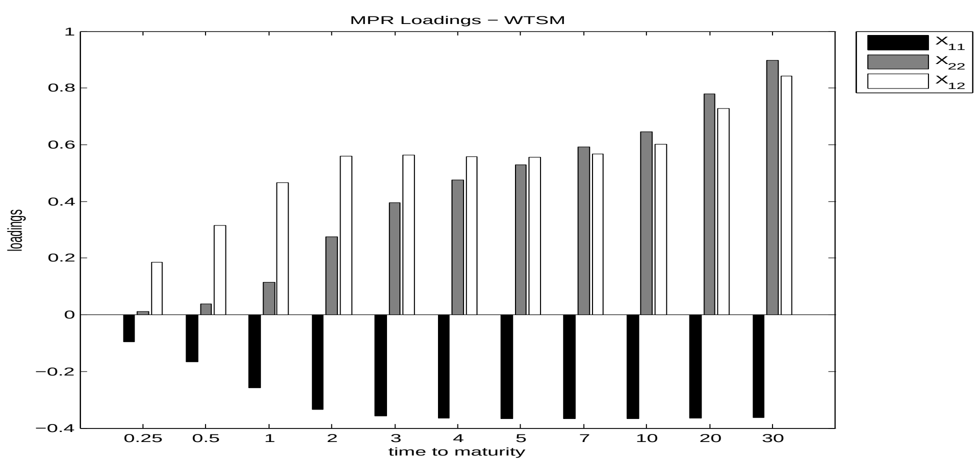
FIGURE 9. Term Structures of Instantaneous Risk Premia



(a) MCIR



(b) BCT



(c) WTSM

The figures illustrate the absolute factor contributions in the instantaneous risk premia, ignoring the constant term. X_{11} = slope, X_{22} = level, X_{12} = correlation.

FIGURE 10. Factor contributions to the Risk Premia

Parameter	$\hat{\Theta}_{ml}$(MCIR)	t-Stat	$\hat{\Theta}_{ml}$(WTSM)	t-Stat
k	20.17700	4.95	7.16420	7.39
m_{11}	-0.08463	-3.69	-0.36870	-4.76
m_{22}	-0.04492	-3.31	-0.00743	-1.58
q_{11}	0.02261	18.72	0.06153	204.65
q_{21}	≡0	–	-0.00075	-103.55
q_{22}	0.01838	185.47	0.00663	240.02
$\tilde{m}_{11}$	-0.18655	-165.61	-0.58970	-364.05
$\tilde{m}_{12}$	≡0	–	0.43503	262.12
$\tilde{m}_{21}$	≡0	–	0.00732	155.11
$\tilde{m}_{22}$	-0.00550	-73.12	-0.00802	-308.68
α	-0.12332	-211.50	-0.11203	-129.94
ψ	≡0	–	0.00909	1.48
$\tilde{\gamma}_{11}$	0.02779	129.62	0.03544	36.13
$\tilde{\gamma}_{22}$	0.00270	187.41	0.00164	295.00
$\tilde{\gamma}_{21}$	≡0	–	-0.03431	-161.97
σ_ϵ (bp)	16.05	51.88	7.71	40.82
Loglik	10393		12079	
LR stat	3372			
$\chi^2(5, 0.95)$	11.07			
av. Bias (bp)	-0.00311		-0.00086	
av. MSE (bp)	14.32		6.53	

The column $\hat{\Theta}_{ml}(\cdot)$ contains the estimates of the parameters in the corresponding models using maximum likelihood methods based on the extended Kalman filter. The columns "t-Stat" gives the t-statistics calculated by element-wise standard deviation. The likelihood ratio test tests the restrictions $q_{12} = \tilde{m}_{12} = \tilde{m}_{21} = \psi = \tilde{\gamma}_{12} = 0$ whose statistic is given by $2\big(Loglik(\hat{\Theta}_{ml,WTSM}) - Loglik(\hat{\Theta}_{ml,MCIR})$. "$\chi^2(5, 0.95)$" is the 95% cutoff value for the χ^2-distribution of degree 5. The "av. Bias" and "av. MSE" indicate the average fitting bias and the average mean square errors of the all eleven bond yields.

TABLE 1. Estimates for WTSM and MCIR models

Parameters	$\hat{\Theta}_{ml}$(BCTEW)	t-Stat	$\hat{\Theta}_{ml}$(BCT)	t-Stat
k	7.28110	7.43	6.04530	102.49
m_{11}	-0.34811	-380.00	-0.35497	-147.80
m_{12}	1.24300	407.66	$\equiv$0	–
m_{21}	0.00543	503.09	-0.04805	-192.42
m_{22}	-0.02701	-1034.60	-0.01919	-109.61
q_{11}	0.23743	456.27	0.13318	151.90
q_{12}	-0.00588	-20.62	-0.28087	-227.75
q_{22}	-0.01285	-569.90	-0.03857	-386.69
ψ_{11}	0.06508	474.46	0.06031	193.86
ψ_{12}	-0.23320	-530.16	-0.22434	-268.67
ψ_{22}	1.10970	310.99	0.84357	161.82
α	-0.11248	-129.37	$\equiv$0	–
$\tilde{\gamma}_{11}$	2.33110	156.90	$\equiv$0.58413	–
$\tilde{\gamma}_{12}$	0.24304	148.53	$\equiv$0.06549	–
$\tilde{\gamma}_{22}$	-0.00162	-77.28	$\equiv$0.00899	–
σ_ϵ (bp)	7.71	40.83	12.36	49.09
Loglik	12079		11262	
LR stat	1634			
$\chi^2(5, 0.95)$	11.07			
av. Bias (bp)	-0.00086		2.5087	
av. MSE (bp)	6.53		17.19	

The column $\hat{\Theta}_{ml}(\cdot)$ contains the estimates of the parameters in the corresponding models using maximum likelihood methods based on the extended Kalman filter. The columns "t-Stat" gives the t-statistics calculated by element-wise standard deviation. The likelihood ratio test tests the restrictions $m_{12} = \alpha = 0$ and $\tilde{\Gamma} = kQQ^\top$ whose statistic is given by $2\big(Loglik(\hat{\Theta}_{ml,WTSM}) - Loglik(\hat{\Theta}_{ml,MCIR}\big)$. "$\chi^2(5, 0.95)$" is the 95% cutoff value for the χ^2-distribution of degree 5. The "av. Bias" and "av. MSE" indicate the average fitting bias and the average mean square errors of the all eleven bond yields.

TABLE 2. Estimates for BCTEW and BCT models

Panel 1. Assets: 3 year bonds and 10 years bond.

3Y10Y	MCIR	WTSM	BCT	Eq Weight
Gain	0.14734	0.16473	0.14811	0.67774
Volatility	0.0058739	0.0056947	0.0059831	0.017184
Sharpe Ratio	0.12117	0.13974	0.11959	0.19054
Downturn Risk	0.0034897	0.0034948	0.0036371	0.010491
Sortino Ratio	0.20396	0.22771	0.19673	0.31209

Panel 2. Assets: 3 year bonds and 20 years bond.

3Y20Y	MCIR	WTSM	BCT	Eq Weight
Gain	0.19696	0.24418	0.24185	0.88733
Volatility	0.0065955	0.0064595	0.0067319	0.025559
Sharpe Ratio	0.14427	0.18262	0.17356	0.16772
Downturn Risk	0.0038438	0.0037595	0.0039921	0.015954
Sortino Ratio	0.24755	0.31377	0.29267	0.26869

Panel 3. Assets: 5 year bonds and 10 years bond.

5Y10Y	MCIR	WTSM	BCT	Eq Weight
Gain	0.20598	0.25946	0.25944	0.77222
Volatility	0.0096978	0.0093802	0.0096029	0.020153
Sharpe Ratio	0.10261	0.13363	0.13052	0.18511
Downturn Risk	0.005825	0.0054615	0.0055394	0.012413
Sortino Ratio	0.17083	0.2295	0.22626	0.30055

Panel 4. Assets: 5 year bonds and 20 years bond.

5Y20Y	MCIR	WTSM	BCT	Eq Weight
Gain	0.27738	0.35218	0.35352	0.98181
Volatility	0.010471	0.010299	0.010601	0.028412
Sharpe Ratio	0.12797	0.16519	0.1611	0.16694
Downturn Risk	0.0062027	0.0058632	0.0061333	0.017721
Sortino Ratio	0.21604	0.29017	0.27846	0.26766

Panel 5. Assets: 10 year bonds and 20 years bond.

10Y20Y	MCIR	WTSM	BCT	Eq Weight
Gain	0.41967	0.5302	0.51288	1.1579
Volatility	0.018574	0.01737	0.01761	0.034245
Sharpe Ratio	0.10915	0.14746	0.1407	0.16335
Downturn Risk	0.013087	0.011334	0.011544	0.02179
Sortino Ratio	0.15491	0.226	0.21463	0.25672

The columns “CIR”, “WTSM”, and “BCT” report the statistics for minimum variance portfolio built using each model estimate. The “Equally Weighted” column is based on a naive portfolio consisting of 50% of each bond. The “Portfolio Gain” is the excess return over the whole period from 1991.04 - 2008.07. The “Sharpe Ratio” is the mean of monthly excess return over its volatility. The“Sortino Ratio” is a risk measure which is defined by the excess return over the downside risk (DR) given in Eq. (44). The downturn risk is calculated by $\mathrm{DR}^2 = \frac{1}{N^-}\sum_{t,\pi_t<0}\pi_t^2$, where π_t is the empirical monthly excess return over R_t the instantaneous rate R_t. R_t is obtained from the WTSM and treated as a known time series. N^- is the number of negative excess monthly returns.

TABLE 3. Minimum variance portfolios - 2 assets

Panel A. WTSM model

Regressor	$xr^{(2yr)}$	$xr^{(3yr)}$	$xr^{(4yr)}$	$xr^{(5yr)}$
Const	**-0.000**	**-0.000**	**-0.000**	**-0.000**
	(-0.000)	(-0.000)	(-0.000)	(-0.000)
X_1	**0.290**	**0.463**	**0.529**	**0.556**
	(1.103)	(1.950)	(2.479)	(2.691)
X_{12}	**-0.934**	**-1.087**	**-1.125**	**-1.125**
	(-5.028)	(-6.449)	(-6.683)	(-6.572)
X_2	**0.344**	**0.304**	**0.297**	**0.267**
	(2.568)	(2.578)	(2.650)	(2.354)
Adj Rsqr	58.01%	62.65%	62.59%	60.21%

Panel B. BCT model

Regressor	$xr^{(2yr)}$	$xr^{(3yr)}$	$xr^{(4yr)}$	$xr^{(5yr)}$
Const	**0.000**	**0.000**	**0.000**	**0.000**
	(0.000)	(0.000)	(0.000)	(0.000)
X_1	**0.352**	**0.736**	**0.913**	**0.953**
	(0.452)	(1.066)	(1.494)	(1.622)
X_{12}	**0.332**	**0.016**	**-0.188**	**-0.217**
	(0.289)	(0.015)	(-0.208)	(-0.250)
X_2	**0.096**	**0.039**	**0.063**	**0.016**
	(0.206)	(0.095)	(0.169)	(0.043)
Adj Rsqr	56.34%	60.13%	60.10%	57.09%

Panel C. MCIR model

Regressor	$xr^{(2yr)}$	$xr^{(3yr)}$	$xr^{(4yr)}$	$xr^{(5yr)}$
Const	**0.000**	**0.000**	**0.000**	**-0.000**
	(0.000)	(0.000)	(0.000)	(-0.000)
X_1	**-0.660**	**-0.626**	**-0.593**	**-0.562**
	(-3.799)	(-3.584)	(-3.339)	(-3.097)
X_2	**-0.268**	**-0.367**	**-0.383**	**-0.403**
	(-1.161)	(-1.631)	(-1.766)	(-1.937)
Adj Rsqr	32.01%	28.48%	25.95%	23.97%

This table reports point estimates (in bold font) and t-statistics (in brackets) from a simple regression of excess log bond returns on predictive factors. The dependent variable is xr^j, which is the 1-year holding period return of a $j-$year bond over the return of a 1-year bond. X_1, X_2 and X_{12} comes from the Wishart/BCT/MCIR models. All variables are standardized, and t-statistics are calculated based on Newey-West adjustment with 18 lags.

TABLE 4. Bond risk premia - simple regression on predicted factors

Panel A. Output gap

Regressor	$xr^{(2yr)}$	$xr^{(3yr)}$	$xr^{(4yr)}$	$xr^{(5yr)}$
Const	**-0.000**	**0.000**	**-0.000**	**-0.000**
	(-0.000)	(0.000)	(-0.000)	(-0.000)
output gap	**-0.243**	**-0.216**	**-0.194**	**-0.183**
	(-1.668)	(-1.468)	(-1.358)	(-1.299)
Adj Rsqr	5.42%	4.14%	3.23%	2.84%

Panel B. Liquidity factors

Regressor	$xr^{(2yr)}$	$xr^{(3yr)}$	$xr^{(4yr)}$	$xr^{(5yr)}$
Const	**-0.000**	**-0.000**	**-0.000**	**-0.000**
	(-0.000)	(-0.000)	(-0.000)	(-0.000)
Commercial paper spread	**-0.235**	**-0.209**	**-0.200**	**-0.199**
	(-1.630)	(-1.442)	(-1.453)	(-1.491)
Adj Rsqr	5.00%	3.85%	3.47%	3.46%
Const	**-0.000**	**0.000**	**-0.000**	**-0.000**
	(-0.000)	(0.000)	(-0.000)	(-0.000)
TED spread	**-0.280**	**-0.259**	**-0.248**	**-0.248**
	(-1.579)	(-1.446)	(-1.454)	(-1.503)
Adj Rsqr	7.32%	6.19%	5.65%	5.63%
Const	**-0.000**	**-0.000**	**-0.000**	**-0.000**
	(-0.000)	(-0.000)	(-0.000)	(-0.000)
10yr swap spread	**-0.350**	**-0.339**	**-0.349**	**-0.343**
	(-2.007)	(-1.945)	(-1.925)	(-1.877)
Adj Rsqr	11.77%	11.01%	11.73%	11.28%
Const	**-0.000**	**-0.000**	**-0.000**	**-0.000**
	(-0.000)	(-0.000)	(-0.000)	(-0.000)
3m T-bill spread	**-0.153**	**-0.207**	**-0.205**	**-0.208**
	(-1.090)	(-1.471)	(-1.443)	(-1.447)
Adj Rsqr	1.83%	3.76%	3.70%	3.82%
Const	**-0.000**	**-0.000**	**-0.000**	**-0.000**
	(-0.000)	(-0.000)	(-0.000)	(-0.000)
Moody Baa Aaa spread	**0.325**	**0.321**	**0.284**	**0.261**
	(2.600)	(2.581)	(2.304)	(2.123)
Adj Rsqr	10.06%	9.83%	7.55%	6.29%

This table reports point estimates (in bold font) and t-statistics (in bracket) from a simple regression of excess log bond returns on the output gap proposed by Cooper & Priestley (2009) and liquidity factors. The dependent variable is xr^j, which is the 1-year holding period return of a $j-$year bond over the return of a 1-year bond. All variables are standardized, and t-statistics are calculated based on Newey-West adjustment with 18 lags.

TABLE 5. Bond risk premia - simple regression on output gap and liquidity factors

Regressor	$xr^{(2yr)}$	$xr^{(3yr)}$	$xr^{(4yr)}$	$xr^{(5yr)}$
Const	**-0.000**	**0.000**	**-0.000**	**-0.000**
	(-0.000)	(0.000)	(-0.000)	(-0.000)
CP	**0.541**	**0.545**	**0.559**	**0.561**
	(4.881)	(5.569)	(5.913)	(6.363)
Adj Rsqr	28.88%	29.30%	30.93%	31.14%

Const	**-0.000**	**0.000**	**-0.000**	**-0.000**
	(-0.000)	(0.000)	(-0.000)	(-0.000)
$\hat{F}_1$ (real)	**-0.690**	**-0.701**	**-0.674**	**-0.657**
	(-8.576)	(-6.889)	(-5.862)	(-5.360)
$\hat{F}_2$ (fin spread)	**-0.286**	**-0.240**	**-0.237**	**-0.214**
	(-4.275)	(-3.012)	(-2.559)	(-2.135)
$\hat{F}_3$ (inflation)	**-0.095**	**-0.080**	**-0.080**	**-0.067**
	(-4.096)	(-3.063)	(-2.448)	(-1.886)
$\hat{F}_4$ (inflation)	**-0.464**	**-0.361**	**-0.306**	**-0.260**
	(-5.778)	(-3.803)	(-2.790)	(-2.237)
$\hat{F}_5$	**-0.119**	**-0.127**	**-0.127**	**-0.127**
	(-2.865)	(-2.533)	(-2.217)	(-2.124)
$\hat{F}_6$ (monetary)	**0.119**	**0.126**	**0.114**	**0.110**
	(2.155)	(2.001)	(1.667)	(1.578)
$\hat{F}_7$ (monetary)	**-0.095**	**-0.125**	**-0.133**	**-0.143**
	(-2.029)	(-2.471)	(-2.423)	(-2.581)
$\hat{F}_8$ (stock mkt)	**0.200**	**0.201**	**0.193**	**0.196**
	(3.809)	(3.722)	(3.199)	(3.092)
Adj Rsqr	74.05%	68.05%	60.88%	56.22%

Const	**-0.000**	**0.000**	**-0.000**	**-0.000**
	(-0.000)	(0.000)	(-0.000)	(-0.000)
PC1	**-0.325**	**-0.383**	**-0.385**	**-0.392**
	(-2.141)	(-2.782)	(-2.943)	(-3.046)
PC2	**0.456**	**0.347**	**0.303**	**0.256**
	(3.970)	(3.175)	(2.708)	(2.185)
PC3	**0.499**	**0.577**	**0.600**	**0.598**
	(4.461)	(5.782)	(6.315)	(6.272)
Adj Rsqr	57.53%	62.09%	61.96%	59.69%

This table reports point estimates (in bold font) and t-statistics (in bracket) from a simple regression of excess log bond returns on predictive factors. The dependent variable is xr^j, which is the 1-year holding period return of a $j-$year bond over the return of a 1-year bond. The predictive factors include the Cochrane & Piazzesi (2005) factor, the 8 factors extracted from 132 macroeconomic variables as in Ludvigson & Ng (2009), and the first 3 Principal Components. All variables are standardized, and t-statistics are calculated based on Newey-West adjustment with 18 lags.

TABLE 6. Bond risk premia - simple regression on known predictive factors

Regressor	$xr^{(2yr)}$	$xr^{(3yr)}$	$xr^{(4yr)}$	$xr^{(5yr)}$
Const	**0.000**	**0.000**	**-0.000**	**0.000**
	(0.000)	(0.000)	(-0.000)	(0.000)
CP	**0.127**	**0.078**	**0.268**	**0.283**
	(1.046)	(0.602)	(1.452)	(1.631)
Adj Rsqr	1.08%	0.08%	6.71%	7.52%

Const	**-0.000**	**0.000**	**-0.000**	**-0.000**
	(-0.000)	(0.000)	(-0.000)	(-0.000)
$\hat{F}_1$ (real)	**-0.024**	**-0.120**	**-0.061**	**-0.230**
	(-0.308)	(-0.758)	(-0.469)	(-2.047)
$\hat{F}_2$ (fin spread)	**-0.099**	**0.099**	**-0.168**	**-0.060**
	(-1.239)	(0.821)	(-1.575)	(-0.476)
$\hat{F}_3$ (inflation)	**0.021**	**0.058**	**-0.067**	**0.031**
	(0.542)	(0.852)	(-1.593)	(0.810)
$\hat{F}_4$ (inflation)	**-0.418**	**-0.121**	**-0.197**	**-0.226**
	(-4.783)	(-0.944)	(-2.057)	(-2.332)
$\hat{F}_5$	**-0.089**	**-0.043**	**0.041**	**0.071**
	(-1.656)	(-0.719)	(0.621)	(1.112)
$\hat{F}_6$ (monetary)	**-0.012**	**0.121**	**0.003**	**0.008**
	(-0.280)	(1.899)	(0.053)	(0.153)
$\hat{F}_7$ (monetary)	**-0.004**	**-0.081**	**-0.096**	**-0.216**
	(-0.053)	(-1.242)	(-1.119)	(-2.768)
$\hat{F}_8$ (stock mkt)	**0.107**	**0.023**	**-0.135**	**-0.100**
	(1.341)	(0.315)	(-1.871)	(-1.696)
Adj Rsqr	15.94%	3.02%	4.74%	11.94%

Const	**0.000**	**0.000**	**-0.000**	**0.000**
	(0.000)	(0.000)	(-0.000)	(0.000)
PC1	**-0.075**	**-0.245**	**-0.139**	**-0.282**
	(-0.649)	(-1.389)	(-1.189)	(-1.953)
PC2	**0.404**	**0.011**	**0.196**	**0.162**
	(5.495)	(0.099)	(1.693)	(1.284)
PC3	**-0.049**	**0.024**	**0.121**	**0.156**
	(-0.593)	(0.182)	(0.819)	(1.318)
Adj Rsqr	16.95%	4.70%	6.28%	12.93%

Const	**-0.000**	**-0.000**	**-0.000**	**-0.000**
	(-0.000)	(-0.000)	(-0.000)	(-0.000)
Moody Baa Aaa spread	**0.181**	**0.348**	**0.186**	**0.256**
	(1.166)	(2.135)	(1.611)	(1.883)
Adj Rsqr	2.76%	11.61%	2.95%	6.04%

Similar to the regression in Table 6, but the independent variables are the errors of the predicted Bond Risk Premia after taking into account the estimated Wishart model.

TABLE 7. Bond risk premia - Regression using Wishart model prediction errors