QUANTITATIVE FINANCE
RESEARCH CENTRE

UNIVERSITY OF
TECHNOLOGY SYDNEY

QUANTITATIVE FINANCE
RESEARCH CENTRE

UTS

THINK.CHANGE.DO

QUANTITATIVE FINANCE RESEARCH CENTRE

Research Paper 298 December 2011

Stochastic Correlation and Risk Premia in Term Structure Models

Carl Chiarella, Chih-Ying Hsiao and Thuy-Duong To

ISSN 1441-8010 www.qfrc.uts.edu.au

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 is obtained by both allowing the market price of risk to have an extended affine form, and allowing the correlation between underlying factors to be stochastic as well as of variable sign. The overall model fit is more negatively impacted by the restriction on the market price of risk than the restriction of correlated factors. However, the stochastic correlation is priced significantly by market participants, though its impact on the risk premia reduces gradually as time to maturity increases. In addition, stochastic correlation is vital in obtaining good hedged portfolio positions. Certainly, the best hedged portfolio is the one that is built based on the model that takes into account both stochastic correlation and extended market price of risk.

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

JEL classifications: E43, C51

Date: February 16, 2011.

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 squared 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 restriction they impose on the correlation structure of the state variables. Take the two fundamental factors long term yield and yield spread as an example. These two factors are usually found to be the important factors influencing the term structure of interest rate (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 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 this stochastic nature of the correlation. In addition, it is expected that correlation risk is priced. Though we are not aware of empirical research in the interest rate market to test this hypothesis, evidence has been found in the equity option market by Driessen et al. (2009). It is therefore of special importance to model the stochastic correlation between the underlying factors of the term structure of interest.

Recently, financial market researchers have explored the use of the Wishart distribution to model dynamic correlation structure of the state variables. The risk factors are assumed to follow a continuous time affine process of positive definite matrices, whose transition probability is a Wishart distribution. Some of the Wishart term structure analysis work can be found in Gourieroux & Sufana (2003), Gourieroux (2006), Gourieroux et al. (2009), Da Fonseca et al. (2007), Da Fonseca et al. (2008), Buraschi et al. (2008) and Cuchiero et al. (2009). In their comprehensive paper, Buraschi et al. (2008) show that the Wishart model also enhances model flexibility to capture 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 term structure of forward volatilities.

Similar to the Cox, Ingersoll & Ross (1985b) (hereafter CIR) model, the Wishart model of Buraschi et al. (2008) is an equilibrium model. It is argued that though the market price of risk specification is simple under the equilibrium setting (square root function without a constant term), the model is still capable of matching various features of the yield curve. The theoretical advantage of Wishart term structure models over the affine term structure class is that they simultaneously allow for stochastically and negatively correlated factors, as well as allow for a variable sign of the market price of risk and excess bond returns. 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.

In this paper we propose and analyse a term structure model that allows for both stochastic correlation between underlying factors and a sophisticated market price of risk specification for each factor. This allows us to compare the significance of each component in matching the features of the yield curve. It also allows us to examine how investors value each component of risk, whether they give a higher price for the factor risk, or for the correlation risk. Finally, we believe that modelling both risk premia and correlation structure dynamically will result in a better portfolio performance.

It is noted again that financial market researchers have provided one-dimensional empirical evidence of the significance of the above two components, namely stochastic correlation and market price of risk. Buraschi et al. (2008) has provided extended analysis of the Wishart versus completely affine model, and similarly Cheridito et al. (2007) have made the case for the extended affine market price of risk versus completely and essentially affine. However, direct comparison between the two set-ups is not possible. Our model (hereafter our WTSM model) allows this comparison, though not directly. 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 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.

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 them 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. Filtering techniques are advantageous for estimating Wishart factors because we can control positive definiteness easily for each time step. This is technically an important point because a Wishart process is not defined if it is not positive definite. As far as we are aware, this issue has not been considered, or considered but not discussed, in the current literature. Using methods without the control of positive definiteness, the parameters can be still obtained but there is no guarantee that the obtained Wishart process is well-defined.

Confirming the findings of Duffee (1999) and Duan & Simonato (1999), we also find that the underlying risk factors in our MCIR model can be interpreted as the long-term yield and the term spread. Moreover in our framework we can estimate the time-varying correlation of these two factors. The correlations need not be constrained to be positive as in Dai & Singleton (2000) but are allowed to change signs.

The estimation results show that the extension of the WTSM is statistically significant against the BCT and the MCIR model. The WTSM 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. 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 Duffee & Stanton (2004) and Cheridito et al. (2007).

In addition, we find that the risk factors are priced very differently under different models. If we do not allow for stochastic correlation, the level-factor risk totally dominates the spread-factor risk in the risk premia. On the other hand, if we only allow a simple market price of risk specification and stochastic correlation, the correlation risk is priced significantly, and is even more significant than the level-factor risk at the long end of the curve. If we allow for both flexible market price of risk and stochastic correlation, we find that all risk factors play an important role in determining the risk premia. However, the correlation risk and the spread-factor risk are priced more significantly at the shorter end of the curve, then the significance reduces gradually as the time to maturity increases. The level-factor risk, on the other hand, has a more permanent presence.

We further analyze the performance of different models based on the ability to obtain the best hedged position. WTSM, which takes into account both stochastic correlation and flexible market price of risk, indeed delivers the best outcome. However, in contrast to the fitting performance, the restrictions on the market price of risk impact less negatively on the portfolio hedging performance than the restriction on correlated factors.

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. Estimation procedures are given in Section 3. Empirical evidence and analyses of bond fitting and forecasting are presented in Section 4, whereas risk premia and portfolio hedging performance under stochastic correlation are analysed in Section 5. Section 6 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

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

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 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 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.

---

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

[2]This is because

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

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

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 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

$$\begin{aligned} \tilde{\Gamma} \quad &:= \quad kQQ^\top - Q\Lambda_0 - \Lambda_0^\top Q^\top \ , &(8)\\ \tilde{M} \quad &:= \quad M - Q\Lambda_1 \ . &(9)\end{aligned}$$

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), so that the Wishart process X_t is strictly positive under the risk-neutral measure; see Cuchiero et al. (2009)

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

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

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“cheap” 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 for 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 the 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 conditioning 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. (2009).

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 equations*

$$\frac{d}{d\tau}C(\tau) \;=\; C(\tau)\tilde{M} + \tilde{M}^\top C(\tau) + 2C(\tau)QQ^\top C(\tau) - \Psi \,, \tag{13}$$

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

$$\frac{d}{d\tau}a(\tau) \quad = \quad \mathbf{tr}[\tilde{\Gamma}C(\tau)] - \alpha \tag{14}$$

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

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. 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 an 1×1 Wishart process. In this case (1) becomes

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

Note that 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 are 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 for the risk-neutral dynamics (7) largely reduces the model capability for fitting empirical data. Therefore we adopt the parametrization $\tilde{\Gamma}$ in order 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*

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

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

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*

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

Proof see Buraschi et al (2008). □.

2.2.2. *Stochastic Correlation*.

Proposition 4. *For the* $n \times n$ *Wishart process* X_t *is 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. □.

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

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

where η is a constant given by

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

The stochastic covariance given in the last proposition can be also summarized in vector form:

Proposition 5.

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

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.

2.2.3. *Risk Premia*. The bond return under the risk neutral measure is given by

$$
\frac{dP(t,T;X_t)}{P(t,T;X_t)} = r_t dt + \mathbf{tr}\Big[(Qd\tilde{W}_t\sqrt{X_t} + \sqrt{X_t}d\tilde{W}_t^\top Q^\top)C(\tau)\Big]\ , \tag{27}
$$

with 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

$$
\begin{aligned}
\frac{dP(t,T;X_t)}{P(t,T;X_t)} &= r_t dt + \mathbf{tr}\Big[\Big(Q(\Lambda_0 + \Lambda_1 X_t)\sqrt{X_t} + \sqrt{X_t}(\Lambda_0^\top + X_t^\top \Lambda_1^\top)Q^\top\Big)C(\tau)\Big]dt \\
&\quad + \mathbf{tr}\Big[(QdW_t\sqrt{X_t} + \sqrt{X_t}dW_t^\top Q^\top)C(\tau)\Big]
\end{aligned}
$$

$$
= r_t dt + 2\mathbf{tr}\big[Q(\Lambda_0 + \Lambda_1 X_t)C(\tau)\big]dt + \mathbf{tr}\Big[(QdW_t\sqrt{X_t} + \sqrt{X_t}dW_t^\top Q^\top)C(\tau)\Big] \tag{28}
$$

$$
\begin{aligned}
&= r_t dt + \mathbf{tr}\big[(kQQ^\top - \tilde{\Gamma})C(\tau)\big]dt + 2\mathbf{tr}\big[(M - \tilde{M})X_t C(\tau)\big]dt && (29)\\
&\quad + \mathbf{tr}\Big[(QdW_t\sqrt{X_t} + \sqrt{X_t}dW_t^\top Q^\top)C(\tau)\Big]\ .
\end{aligned}
$$

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

$$\begin{aligned}\text{Risk Premia} = e_t &= 2\mathbf{tr}\big[Q(\Lambda_0 + \Lambda_1 X_t)C(\tau)\big] \\ &= \mathbf{tr}\big[(kQQ^\top - \tilde{\Gamma})C(\tau)\big] + 2\mathbf{tr}\big[(M - \tilde{M})X_t C(\tau)\big]\ .\end{aligned} \tag{30}$$

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 of 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 out Wishart term structure model.

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

$$X_t^{\mathcal{L}} := \mathcal{L} X_t \mathcal{L}^\top \text{ , and } \quad W_t^{\mathcal{O}} = \mathcal{O} W_t \ , \tag{31}$$

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*

$$k^{\mathcal{L}} = k \ , \tag{32}$$

$$M^{\mathcal{L}} = \mathcal{L} M \mathcal{L}^{-1} \ , \tag{33}$$

$$Q^{\mathcal{L}\mathcal{O}} = \mathcal{L} Q \mathcal{O}^\top \ , \tag{34}$$

$$\tilde{\Gamma}^{\mathcal{L}} = \mathcal{L} \tilde{\Gamma} \mathcal{L}^\top \ , \tag{35}$$

$$\tilde{M}^{\mathcal{L}} = \mathcal{L} \tilde{M} \mathcal{L}^{-1} \ , \tag{36}$$

$$\alpha^{\mathcal{L}} = \alpha \ , \tag{37}$$

$$\Psi^{\mathcal{L}} = (\mathcal{L}^\top)^{-1} \Psi \mathcal{L}^{-1} \ . \tag{38}$$

Technically, let $a^{\mathcal{L}}(\tau)$ 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 relations

$$a^{\mathcal{L}}(\tau) = a(\tau) \quad \text{and} \quad C^{\mathcal{L}}(\tau) = (\mathcal{L}^\top)^{-1} C(\tau) \mathcal{L}^{-1} \ . \tag{39}$$

Proof see Appendix. □.

The sup-index $\mathcal{L}\mathcal{O}$ in $Q^{\mathcal{L}\mathcal{O}}$ indicates that the parameter Q is affected by the transformation $W_t^{\mathcal{O}} = \mathcal{O} W_t$. 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) is kept the same under the transformation but we 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 (35) and (36) are replaced by*

$$\tilde{\Gamma}^{\mathcal{L}} = kQ^{\mathcal{L}\mathcal{O}}(Q^{\mathcal{L}\mathcal{O}})^\top - Q^{\mathcal{L}\mathcal{O}}\mathcal{O}\Lambda_0\mathcal{L}^\top - (Q^{\mathcal{L}\mathcal{O}}\mathcal{O}\Lambda_0\mathcal{L}^\top)^\top , \tag{40}$$

$$\tilde{M}^{\mathcal{L}} = M^{\mathcal{L}} - Q^{\mathcal{L}\mathcal{O}}\mathcal{O}\Lambda_1\mathcal{L}^{-1} . \tag{41}$$

Proof see Appendix. □.

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. $\square$.

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 zero. Within the framework we can test these parameter restrictions later in an empirical investigation.

3. Empirical Procedures

In this section we describe our empirical procedure for estimation of the Wishart term structure model.

3.1. **Summary of the Models.**

We consider four models, (1) our Wishart term structure model (WTSM), (2) the multiple CIR (MCIR) model, (3) an extended Wishart encompassing Buraschi et al. (2008) model (hereafter BCTEW) and the the Buraschi et al. (2008) (BCT) model. Figure 2 summarizes the relationships between the different models. We note that

(i) The MCIR model is a restricted version of the WTSM.

(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.

Due to the link of the model equivalence, the MCIR and the BCT models can now be only compared with each other.

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:

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

which are in the form of

$$\begin{aligned} M &= \begin{pmatrix} m_{11} & 0 \\ 0 & m_{22} \end{pmatrix}, \; Q = \begin{pmatrix} q_{11} & 0 \\ q_{21} & q_{22} \end{pmatrix}, \; \tilde{\Gamma} = \begin{pmatrix} \tilde{\gamma}_{11} & \tilde{\gamma}_{12} \\ \tilde{\gamma}_{12} & \tilde{\gamma}_{22} \end{pmatrix}, \\ \tilde{M} &= \begin{pmatrix} \tilde{m}_{11} & \tilde{m}_{12} \\ \tilde{m}_{21} & \tilde{m}_{22} \end{pmatrix}, \; \Psi = \begin{pmatrix} 1 & \psi \\ \psi & 1 \end{pmatrix}, \; k \text{ and } \alpha. \end{aligned} \tag{42}$$

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:

$$\begin{aligned} M &= \begin{pmatrix} m_1 & 0 \\ 0 & m_2 \end{pmatrix}, \; Q = \begin{pmatrix} q_1 & 0 \\ 0 & q_2 \end{pmatrix}, \; \tilde{\Gamma} = \begin{pmatrix} \tilde{\gamma}_1 & 0 \\ 0 & \tilde{\gamma}_2 \end{pmatrix}, \\ \tilde{M} &= \begin{pmatrix} \tilde{m}_1 & 0 \\ 0 & \tilde{m}_2 \end{pmatrix}, \; \Psi = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \; k \text{ and } \alpha. \end{aligned} \tag{43}$$

The initial value $X_{0,21} = X_{0,12}$ are set to be 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 by the extra freedom in M and Ψ compared with the WTSM (42). The BCTEW adopts normalization conditions S3 given in Proposition 8 and is statistically equivalent to the WTSM. It also has fifteen parameters:

$$
\begin{aligned}
&M = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix}, \ Q = \begin{pmatrix} q_{11} & q_{12} \\ 0 & q_{22} \end{pmatrix}, \ \tilde{\Gamma} = \begin{pmatrix} \tilde{\gamma}_{11} & \tilde{\gamma}_{12} \\ \tilde{\gamma}_{12} & \tilde{\gamma}_{22} \end{pmatrix}, \\
&\tilde{M} = M - Q, \ \ , \Psi = \begin{pmatrix} \psi_{11} & \psi_{12} \\ \psi_{12} & \psi_{22} \end{pmatrix}, \ k \text{ and } \alpha.
\end{aligned} \tag{44}
$$

The fourth model is the Buraschi, Cieslak & Trojani (2008) model (hereafter **BCT** model). It has ten parameters:

$$
M = \begin{pmatrix} m_{11} & 0 \\ m_{21} & m_{22} \end{pmatrix}, \ Q = \begin{pmatrix} q_{11} & q_{12} \\ 0 & q_{22} \end{pmatrix}, \ \tilde{M} = M - Q, \ \Psi = \begin{pmatrix} \psi_{11} & \psi_{12} \\ \psi_{12} & \psi_{22} \end{pmatrix}, \ k \ . \tag{45}
$$

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

3.2. **Maximum Likelihood Estimation based on the Extended Kalman Filter.** In order to carry out the estimation task we adopt maximum likelihood estimation (MLE). The *extended Kalman filter* is imbedded in the likelihood function in order to filter the unobservable factor X_t from the bond yield observations y_t^τ. We do not adopt the generalized method of moments (GMM) as in Buraschi et al. (2008) because we have found that it is difficult to monitor whether the underlying Wishart factors remain positive definite during the estimation procedure.

The Eular-Maruyama method is used to discretize the Wishart process (1) according to

$$
X_{t+\Delta} = X_t + \big(kQQ^\top + MX_t + X_tM^\top\big)\Delta + Q\Delta W_t\sqrt{X_t} + \sqrt{X_t}\Delta W_t Q^\top \ . \tag{46}
$$

A positive control given later in Eq. (51) 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. (2008). Let $\vec{X}_t$ denote the column obtained by staking the columns of the matrix X_t. The discretized dynamics can then be represented by

$$\vec{X}_{t+\Delta} = f + F\vec{X}_t + U_{t+\Delta} \ , \tag{47}$$

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^τ is given by

$$y_t := \begin{pmatrix} -\frac{1}{\tau_1} \ln P(t, t+\tau_1; X_t) \\ \vdots \\ -\frac{1}{\tau_d} \ln P(t, t+\tau_d; X_t) \end{pmatrix} = j + J\vec{X}_t + \epsilon_t \ , \tag{48}$$

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[6]. Let Y_t denote all the bond yields observed until t. Let $\hat{X}_{t|s} = \mathrm{E}[\vec{X}_t|Y_s]$ and $P_{t|s} = \mathrm{Cov}[\vec{X}_t|Y_s] = \mathrm{E}[(\vec{X}_t - \hat{X}_{t|s})(\vec{X}_t - \hat{X}_{t|s})^\top|Y_s]$. We start with the initial state $\hat{X}_{0|0}$ and the covariance $P_{0|0}$. The algorithm runs iteratively and every iteration consists of two steps.

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

$$
\begin{aligned}
\hat{X}_{t|t-\Delta} &= f + F\hat{X}_{t-\Delta|t-\Delta} , \\
P_{t|t-\Delta} &= FP_{t-\Delta|t-\Delta}F^\top + \mathrm{Cov}[U_t] .
\end{aligned} \tag{49}
$$

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

$$
\begin{aligned}
\hat{X}_{t|t} &= \hat{X}_{t|t-\Delta} + K_t\big(y_t - JX_{t|t-\Delta} - j\big) , \\
P_{t|t} &= P_{t|t-\Delta} - K_tJP_{t|t-\Delta} ,
\end{aligned} \tag{50}
$$

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 $\mathrm{Cov}[v_{t|t-\Delta}|Y_{t-\Delta}] = \Sigma_{t|t-\Delta}$. 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

[6]For details see Harvey (1989).

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 (46). Also the updating step (50) 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

$$X_{11,t} > 0 \ , \quad X_{22,t} > 0 \ , \quad X_{11,t}X_{22,t} > X_{12,t}X_{21,t} \ , \tag{51}$$

for all observation times t and all samples. All the inequalities are required to hold strictly. The positive control is imposed for every prediction step (49) and every updating step (50).

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.

4. Empirical Results

4.1. **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, 2, 3, 4, 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. During the observation period the level of the 30 year long-term yield changes only moderately and has a slow downward development. The three month yield, which follows very closely the federal funds rate and is affected largely by the US monetary policy, fluctuates more widely. This gives a volatile development of the yield spread as illustrated in Figure 4. In the literature, the long-term yield and the yield spread are often considered as important determining factors for the term structure of interest rates; see for example, Duffee (1999).

[Figure 3 here]

[Figure 4 here]

4.2. **Estimation.** In order to obtain parameter estimates we utilize three optimization methods: the simplex method `fminsearch`[7], the gradient method `fminunc`[8] and the simulated annealing method[9] 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.

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

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

[9]Obtained from Matlab Central.

4.2.1. *Estimation of MCIR*. Table (1) gives estimation results of the MCIR and WTSM models. 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. Under the risk-neutral measure the dynamics of X_t can be seen as the two independent CIR process

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

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 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 here we are able to estimate it and find that the estimated value is negative.

Figure 5(a) plots the estimated factors of the MCIR model. As we compare the estimated factors with the long-term yield level and the yield spread, we find that the estimated factors in the MCIR model are incredibly highly correlated with the two economic factors. Figure 6 shows the comparison of these time series. The correlation coefficients of both pairs are 98.74% and 97.87% respectively. Duffee (1999) has also found a high correlation of 97% in his two-dimensional CIR model. This finding suggests that the estimated factors in the MCIR model have a correspondence with long-term yield and yield spread.

4.2.2. *Estimation of WTSM*. One novel feature of the WTSM is its capacity to model time-varying correlation between positive factors. Here we intend to utilize this fact to investigate the correlation between the two factors incorporated with economic meaning such as yield spread and long term yield. We consider these two factors to be on

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 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

the diagonal of the 2×2 Wishart process. As suggested by Eq. (23) X_{12t} represents the covariance of the two 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 positive stationary process for both the real-world and risk-neutral measures. The "shift" parameter α is -11.25$\%$ similar with the $\alpha = -12.33\%$ estimated in the MCIR model.

Statistically, the WTSM outperforms the MCIR model as is evident in Table 1. The estimated standard deviation of the measurement errors σ_ϵ is smaller. The likelihood ratio test strictly rejects the restrictions of the MCIR model against the WTSM since since the LR statistic (3372) is far higher than the 95% cutoff value of the asymptotic distribution of $\chi^2(5) = 11.07$. The WTSM also has smaller fitting errors for the bond yield data both in bias and in mean-square error (MSE).

Figure 5(b) plots the estimated factors of the WTSM and Figure 6(c-d) compares the estimated factors $\hat{X}_{11t}$ $\hat{X}_{22t}$ with the long-term yield and the yield spread. The factors are also highly correlated with the long-term yield (with correlation 97.31%) and the yield spread (with correlation 98.26%).

4.2.3. *Estimation of BCTEW and BCT 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 5. However, from that fact that they generate the same the risk premia (calculated by equation (30) 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.

In the BCT model the shift parameter α is set to zero. As a consequence one can observe that the BCT factors in Figure 5(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 for fitting bond yields data. The restriction $\tilde{\Gamma} = kQQ^\top$ requires that the constant drift term under the real world dynamics is 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 (av. Bias and av. MSE), and

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 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

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 7 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. Fitting errors are the difference between the observed bond yields and the model bond yields calculated using the updated factor levels. The three curves correspond to the error measurements of the MCIR, the WTSM and the BCT models. We do not have a curve of 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. It is obvious that the WTSM has the smallest MSE amongst the three models.

Panel (b) in Figure 7 gives the bias and the MSE of one-month 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.

The longer term predictions are given in Panel (c) in Figure 7 and all panels in Figure 8. Out-performance of the WTSM remains for the all cases though the difference between the MCIR and the WTSM reduces as the time to maturity increases. Both give clearly better forecasts than the BCT model. This provides a clear evidence that the restrictions of the BCT model greatly reduce the flexibility of the model.

5. Risk Premia and Portfolio Strategies under Stochastic Correlation

In this section we investigate how the stochastic correlation affects the risk premia and the portfolio strategies.

5.1. **Risk Premia.**

[Figure 9 here]

The risk premium of a bond is the excess return of the bond over the risk-less instantaneous return. The model risk premia are calculated based on Equation (30). Figure 9 compares the development of the risk premia of the different maturity bonds for all the four models. Recall 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.

Consider the risk premia (30) more in detail. Figure 10 illustrates the contributions of each factor in the risk premia. The patterns vary across the three models. For the MCIR model, the factor X_{22}, corresponding the 30-year yield level, dominates the slope factor over the whole yield curve. Its contribution reaches almost 100% for longer term bonds. The BCT model, on the other hand, shows a significant contribution of the correlation factor at the 5-10 year segment of the curve. In the WTSM model, which allows for both stochastic correlation and sophisticated market price of risk, all risk factors play an important role in determining the risk premia. The correlation risk and the spread-factor risk is priced more significantly at the shorter end of the curve, then reduces gradually as the time to maturity increases, whereas the level-factor risk has a more permanent presence.

[Figure 10 here]

5.2. **Hedging performance.** 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 WTSM will not have a complete hedging, and therefore its performance will not be as optimal.

We consider monthly return of bonds with fixed time to maturity[10]. Based on the model, the asset covariance can be calculated for each time point. Using Eq. (28) bond return (with fixed time to maturity) is calculated by

$$\frac{dP_{i,t}}{P_{i,t}} = \big(r_t + e_{i,t}\big)dt + \mathbf{tr}\big[dZ_tC_i\big] \ , \tag{52}$$

[10]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 reinvest 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 shorten 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 build the log return.

where

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

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

$$\sigma_{ij,t}dt = \mathrm{Cov}\Big(\mathbf{tr}[dZ_tC_{i,t}]\ \mathbf{tr}[dZ_tC_{j,t}]\Big) \ , \tag{53}$$

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 (55) based on the estimation results of the MCIR, the WTSM and the BCT models respectively. Given the variance structure, hedged portfolio position can be obtained easily. The investment horizon is taken to be the same as the observation (from April 1991 to July 2008), and portfolio position is normalized to one.

Table 3 summaries 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.

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, it still delivers the lowest portfolio volatility (our objective function), as well as lowest downturn risk. The risk-adjusted return[11] is also highest under the WTSM model, then BCT and finally MCIR.

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

$$\text{Sortino Ratio} = \frac{R-T}{DR} \ , \tag{54}$$

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

Figures 11 plot the asset correlation based on the three models. The posterior reference line 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 empirical 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.

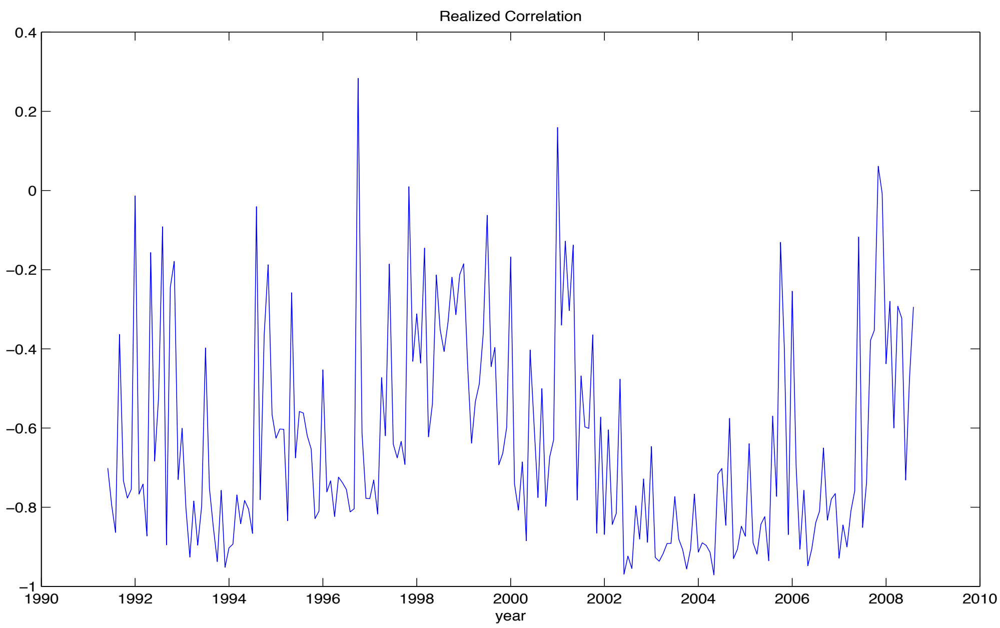
[Figure 11 here]

6. 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 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 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 price of the underlying factors at the short end of the curve. The significance gradually reduces as the time to maturity increases.

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 extended affine market price of risk, outperforms those built by more restricted models by a considerable margin. However, in contrast 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 flexible market price of risk.



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

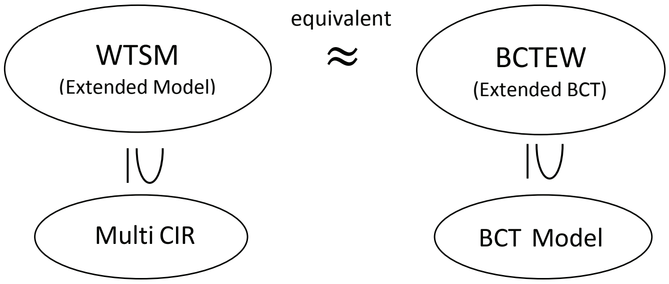


FIGURE 2. Relations of the models

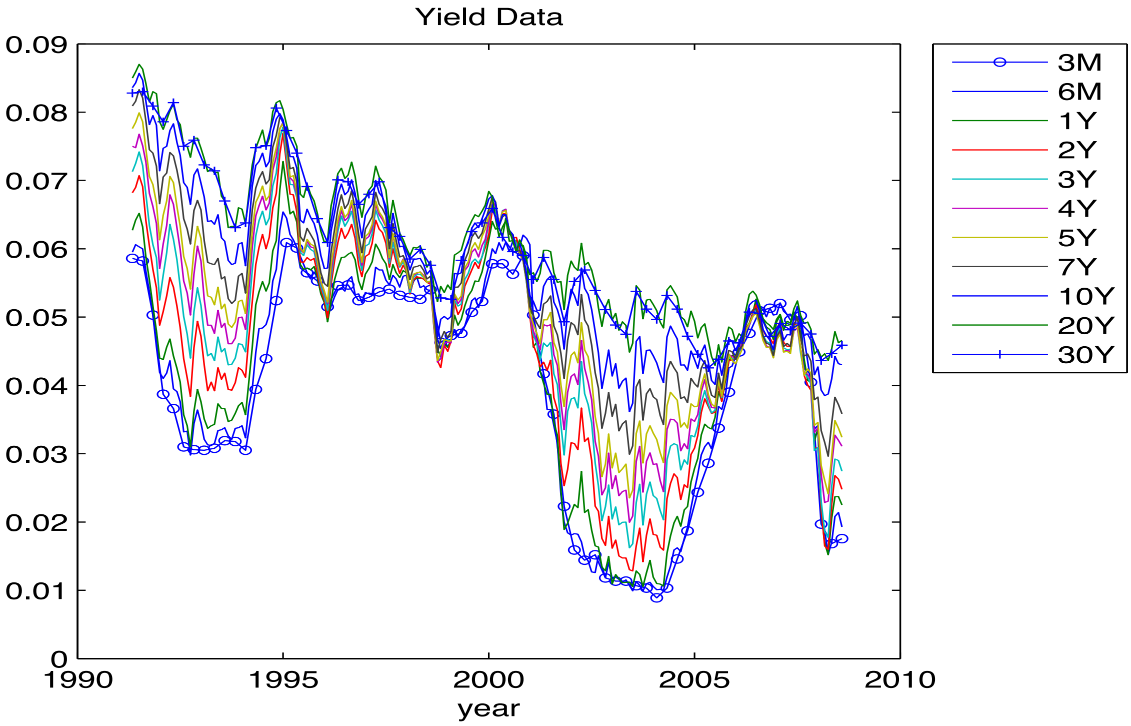
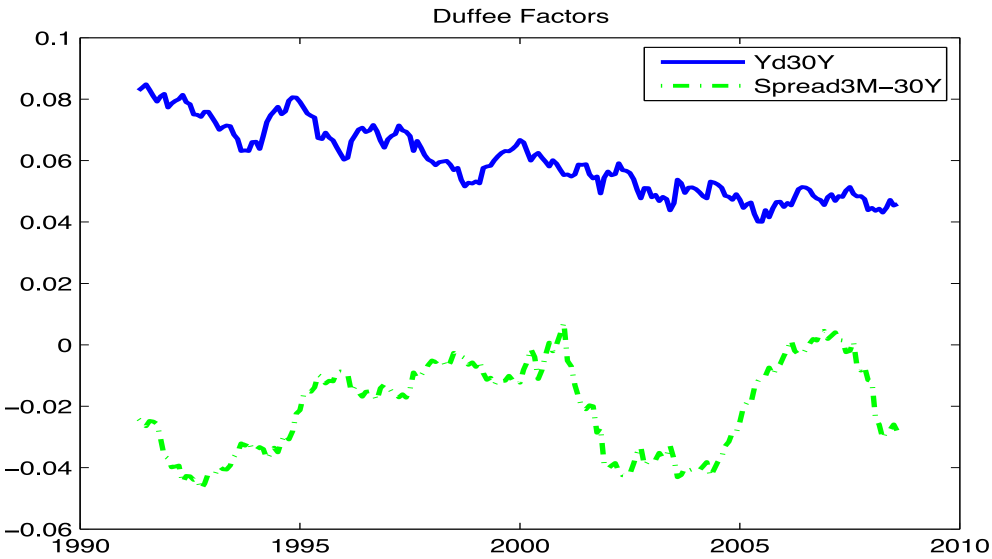
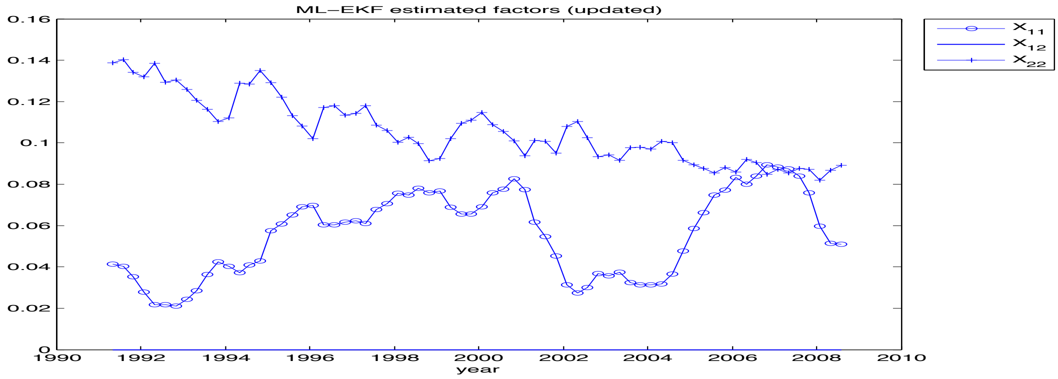


FIGURE 3. Monthly US yield data

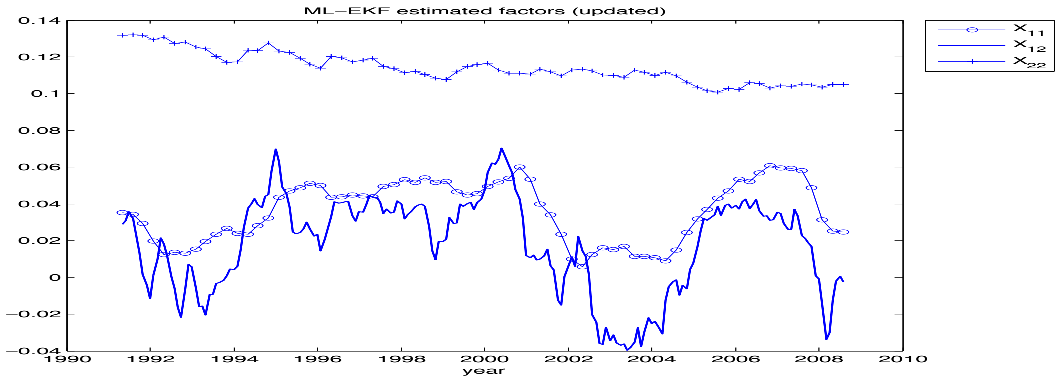


The long-term yield is the 30 year bond yield and the yield spread is the difference between of the 3 month and 30 year yields.

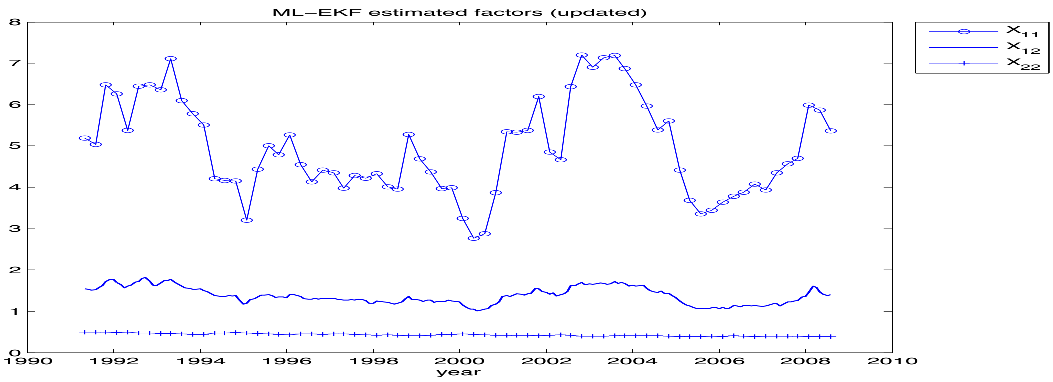
FIGURE 4. Long-term yield and yield spread



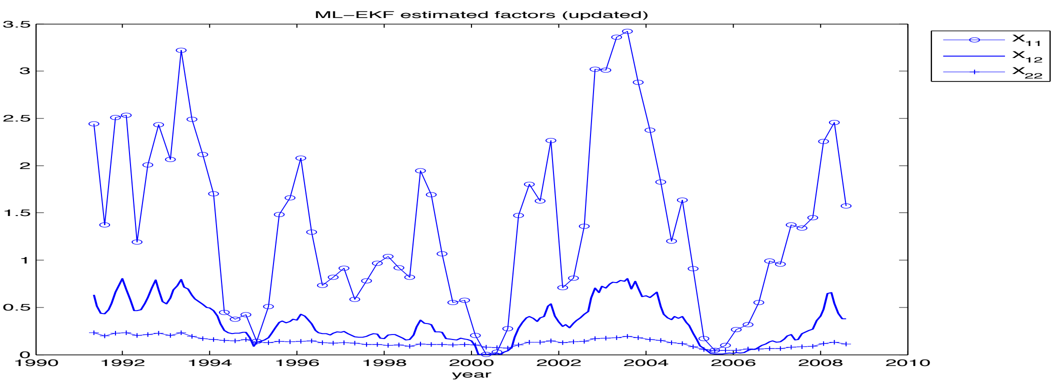
(a) CIR



(b) WTSM

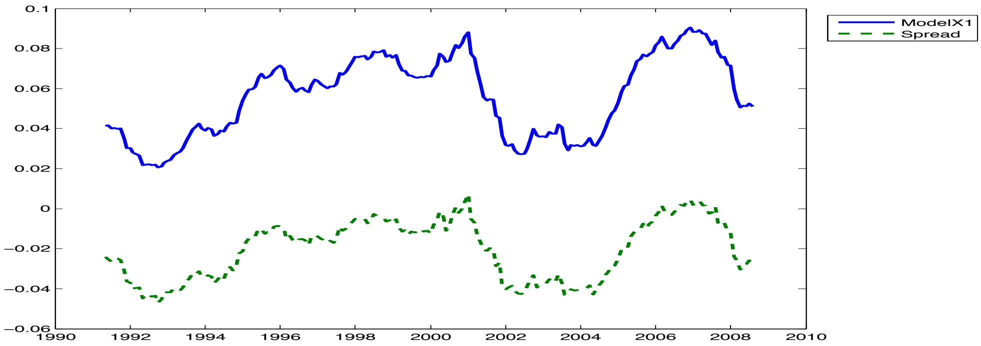


(c) BCTEW

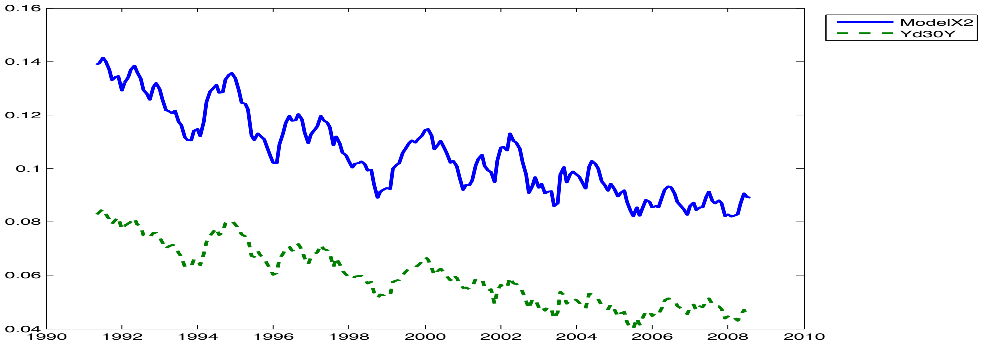


(d) BCT

FIGURE 5. Filtered factors from the four models

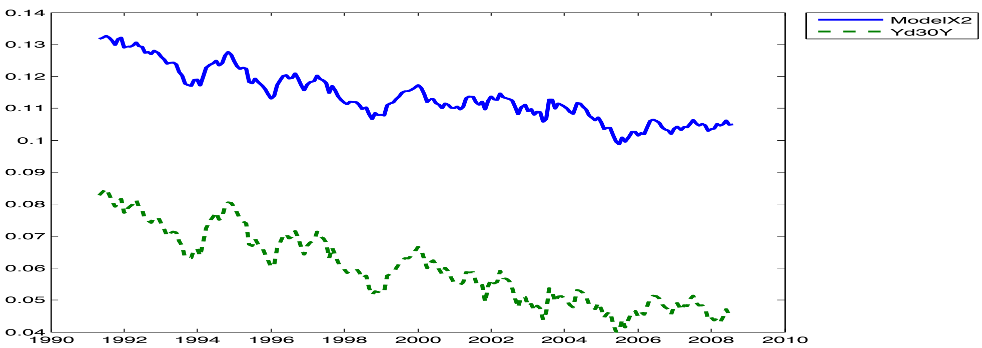


(a) One MCIR factor and Spread (Corr = 98.74%)



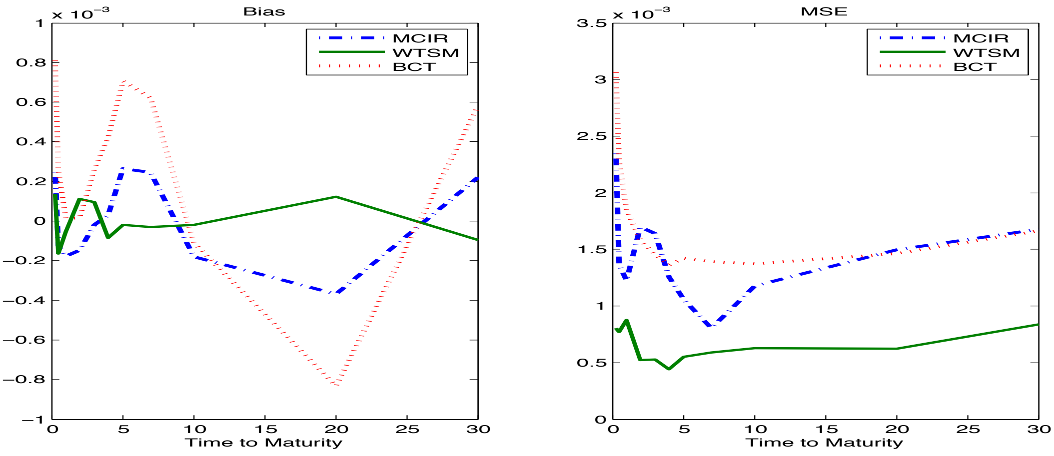
(b) One MCIR factor and Long yield (Corr = 97.87%)

(c) One WTSM factor and Spread (Corr =98.26%)

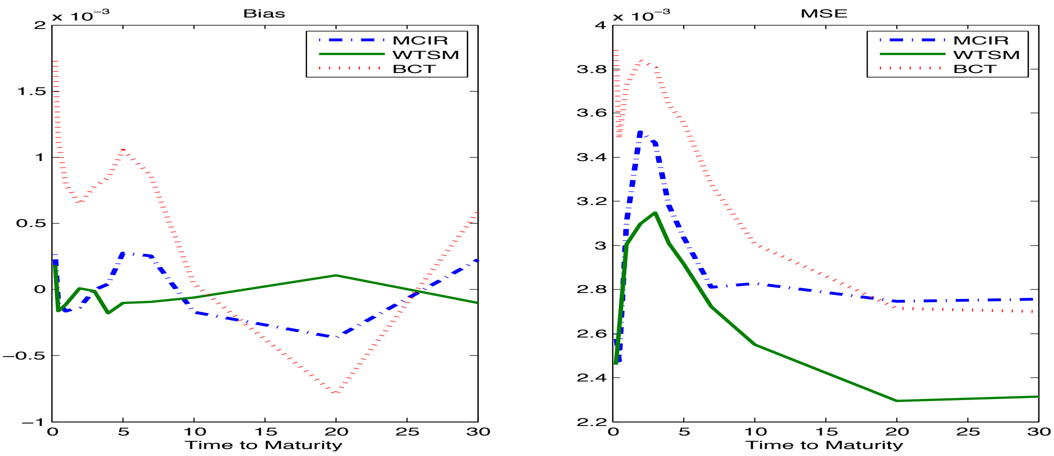


(d) One WTSM factor and Long yield (Corr =97.31%)

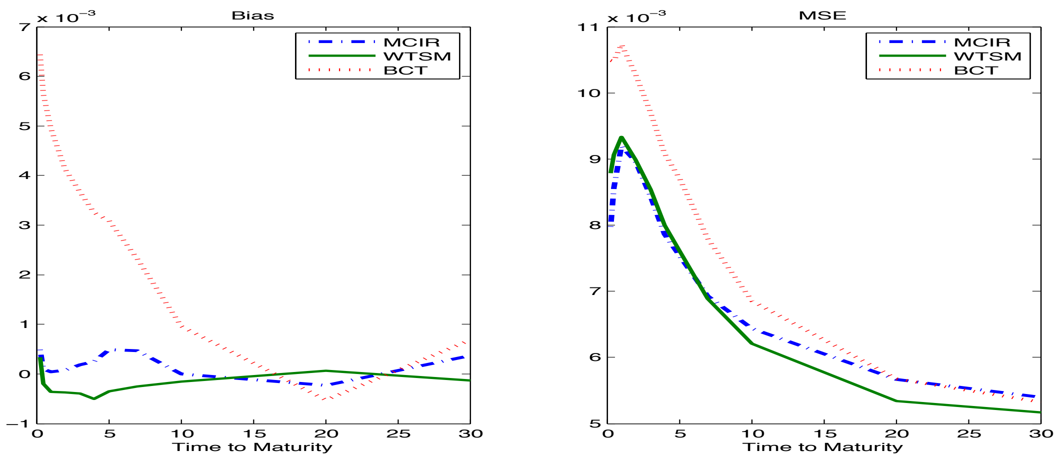
FIGURE 6. Comparison of the estimated factors and the economic factors



(a) Fitting Errors



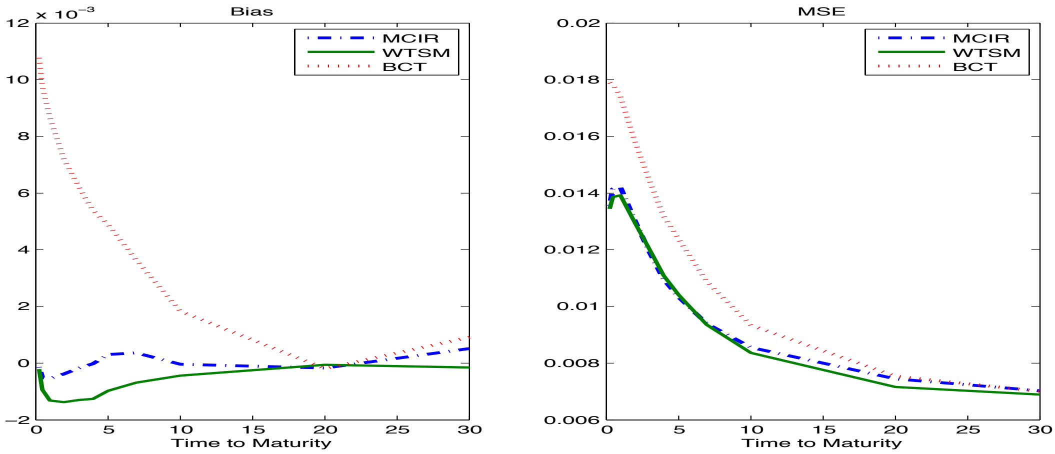
(b) One Month Ahead Forecast



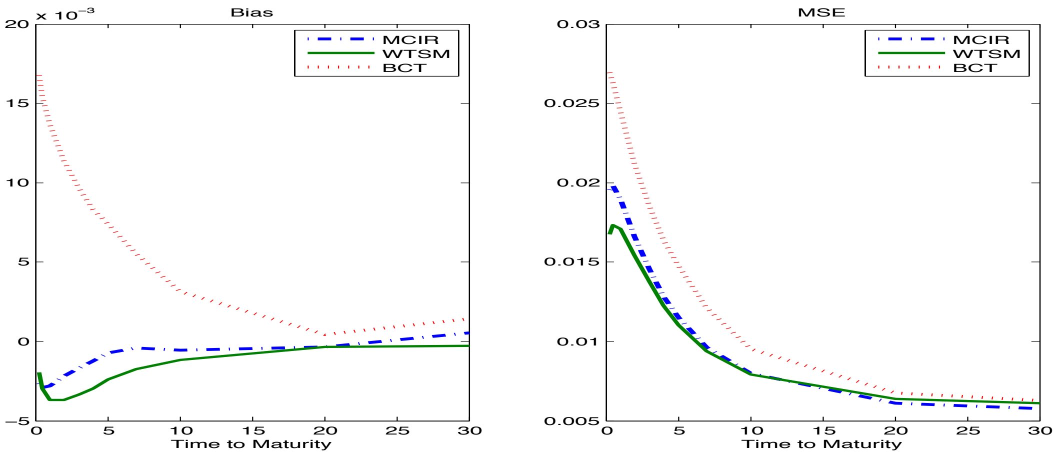
(c) Six Months ahead Forecast

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. Panels in (b) illustrate these two error measures for one month ahead forecast and Panels in (c) give for a six month ahead forecast. All forecasts are in-sample forecast.

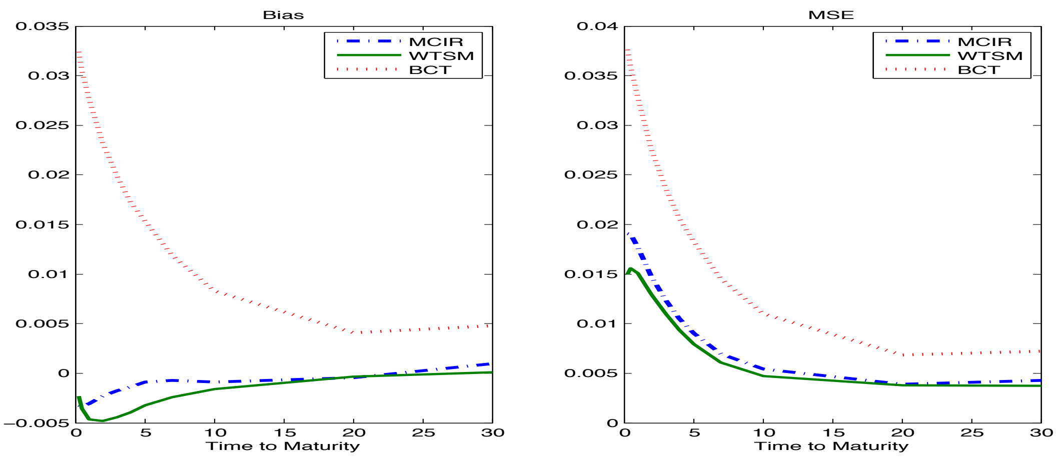
FIGURE 7. Fitting Errors and Short-Term Forecasts of all the Bonds



(a) One Year Ahead Forecast



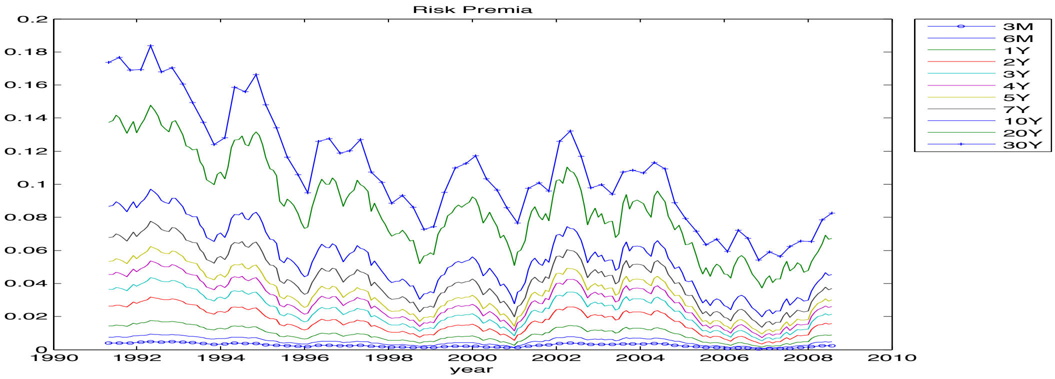
(b) Two Year Ahead Forecast



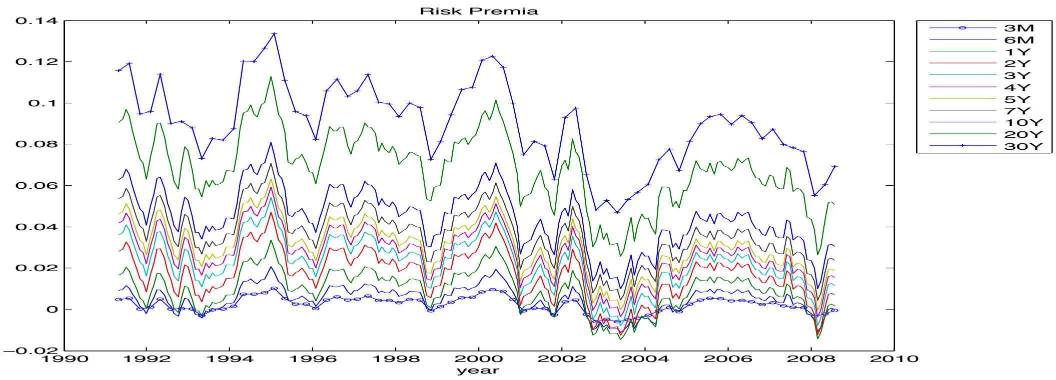
(c) Five Year Ahead Forecast

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 one year ahead, two year ahead and five year ahead in the each panel respectively. All forecasts are in-sample forecast.

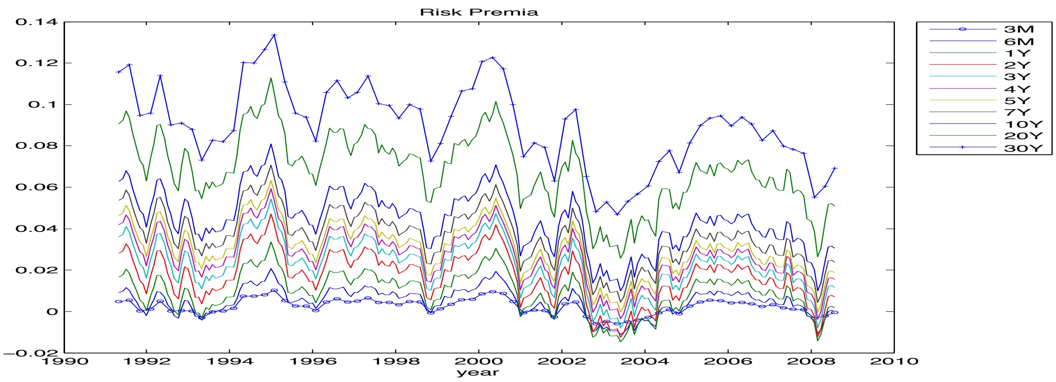
FIGURE 8. Medium-Term Forecast of all the Bonds



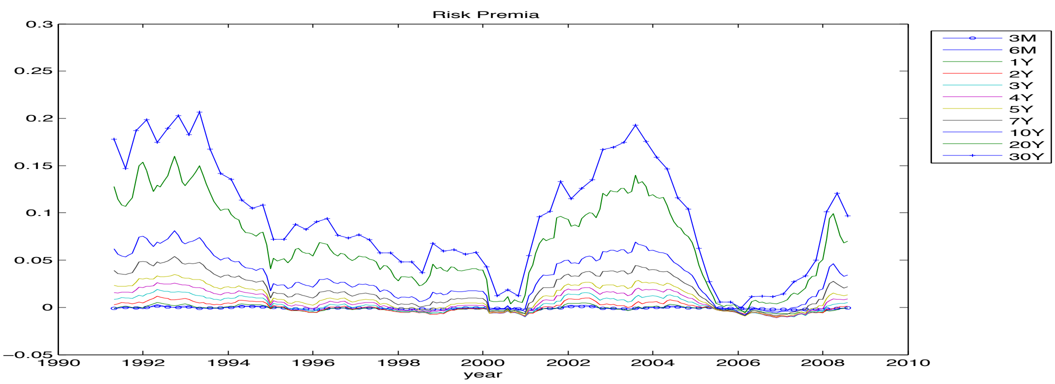
(a) MCIR



(b) WTSM

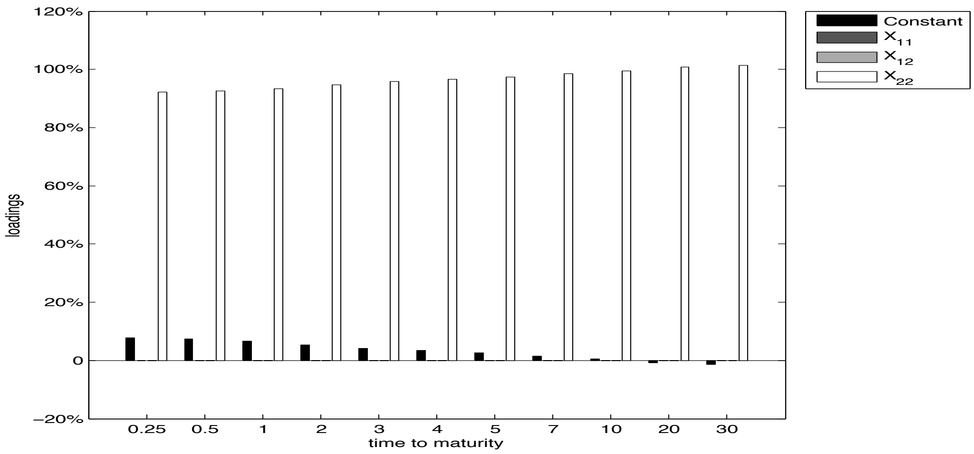


(c) BCTEW

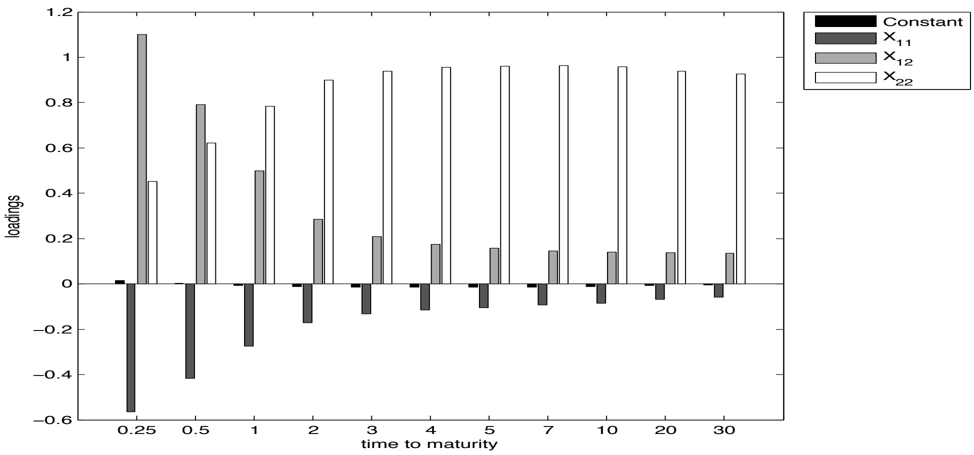


(d) BCT

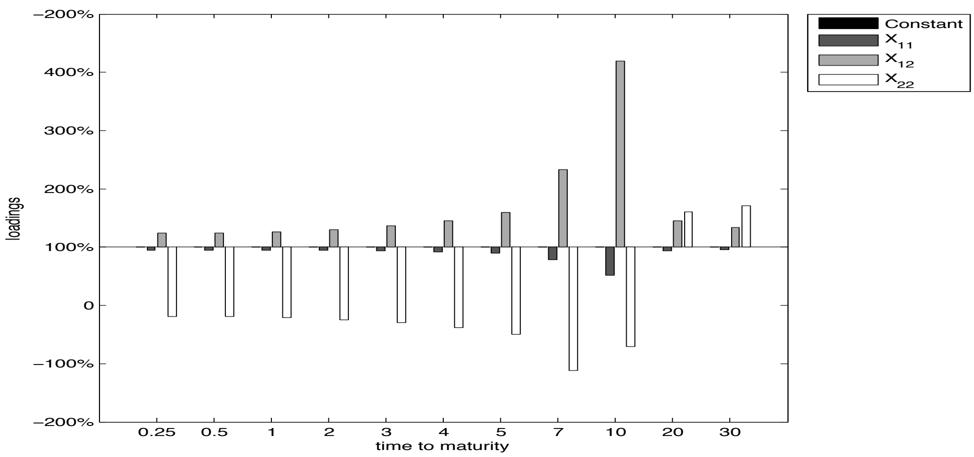
FIGURE 9. Term Structures of Risk Premia of the four Models



(a) MCIR



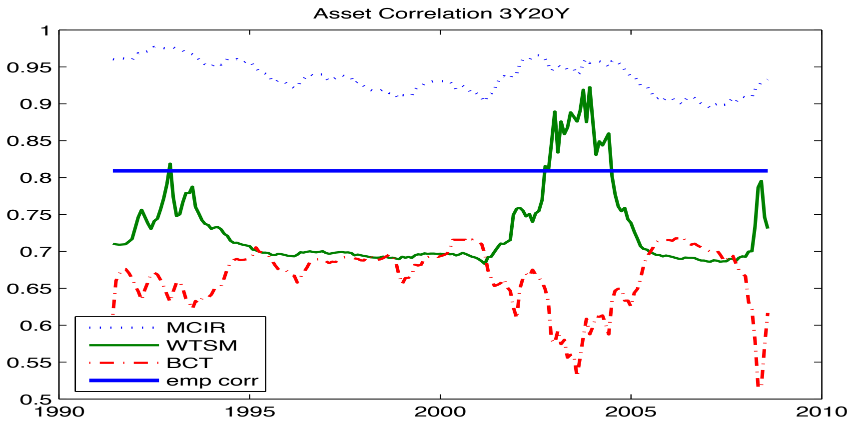
(b) WTSM



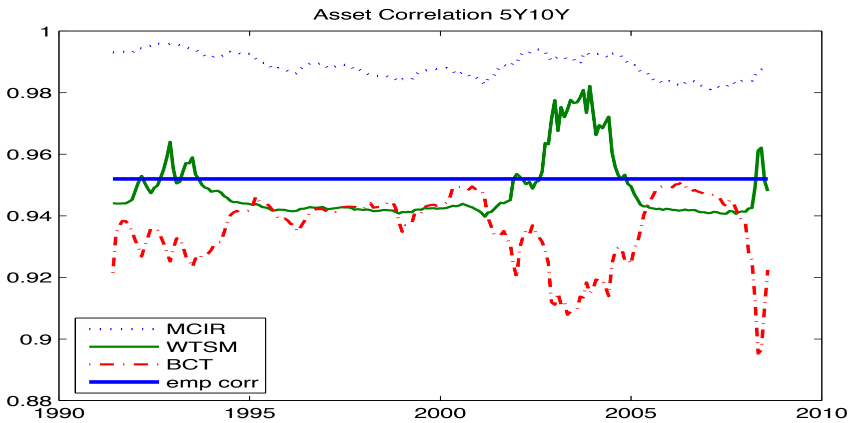
(c) BCT

The figures illustrate the factor contributions in the risk premia in percentage, calculated at the average level of the factors.

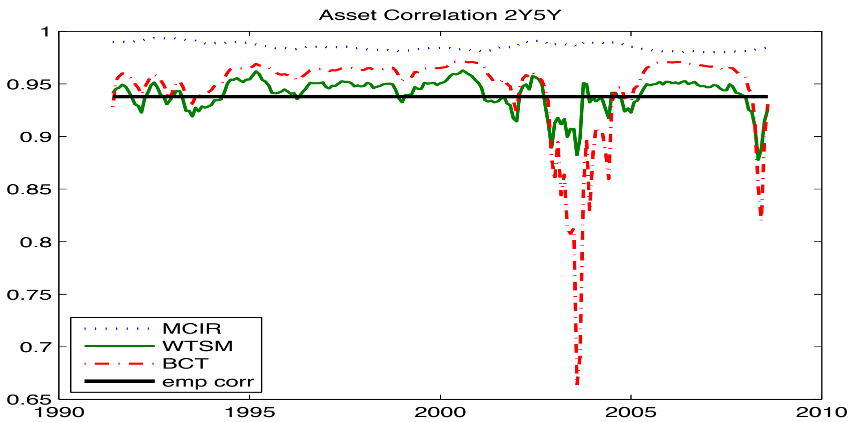
FIGURE 10. Factor contributions in the Risk Premia



(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 11. Asset Correlation

APPENDIX A. PROOFS OF ALL PROPOSITIONS

Proof of Proposition 1

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

$$\begin{aligned} \frac{dP}{P} &= \big(-\frac{d}{d\tau}a(\tau) - \mathbf{tr}[\frac{d}{d\tau}C(\tau)X_t]\big)dt + \mathbf{tr}[C(\tau)dX_t] \\ &+\frac{1}{2}\sum_{i,j=1}^{n}\sum_{u,v=1}^{n} C_{ij}C_{uv}dX_{ij,t}dX_{uv,t} \ . \end{aligned} \tag{55}$$

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

$$\begin{aligned} &\sum_{i,j=1}^{n}\sum_{u,v=1}^{n} C_{ij}C_{uv}dX_{ij,t}dX_{uv,t} \\ =& \sum_{i,j=1}^{n}\sum_{u,v=1}^{n} C_{ij}C_{uv}\Big((QQ^\top)_{iu}X_{jv} + (QQ^\top)_{ju}X_{iv} + (QQ^\top)_{iv}X_{ju} + (QQ^\top)_{jv}X_{iu}\Big)dt \end{aligned} \tag{56}$$

Now calculate each term in (58) we start with the first term

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

For the second term in (58) 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 (57) by

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

Use this and the dynamics (7) we then rewrite the bond return (57) into

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

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

$$\mathrm{E}[\mathbf{tr}(HdW_t)\mathbf{tr}(GdW_t)] = \mathbf{tr}(HG^\top)dt \ , \tag{57}$$

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

$$\mathrm{Cov}[dX_{ij}\ dX_{uv}] = \mathrm{E}\Big[\big(Q_{i\cdot}dW\sqrt{X}_{\cdot j} + Q_{j\cdot}dW\sqrt{X}_{\cdot i}\big)\big(Q_{u\cdot}dW\sqrt{X}_{\cdot v} + Q_{v\cdot}dW\sqrt{X}_{\cdot u}\big)\Big] \ , \tag{58}$$

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

$$Q_{i\cdot}dW\sqrt{X}_{\cdot j} = Q_{i\cdot}\begin{pmatrix}\sum_{u=1}^{n} W_{1u}\sqrt{X}_{uj} \\ \vdots \\ \sum_{u=1}^{n} W_{nu}\sqrt{X}_{uj}\end{pmatrix} = Q_{i1}\sum_{u=1}^{n} W_{1u}\sqrt{X}_{uj} + \cdots + Q_{in}\sum_{u=1}^{n} W_{nu}\sqrt{X}_{uj}$$

$$= \sum_{v=1}^{n}\sum_{u=1}^{n} Q_{iv}W_{vu}\sqrt{X}_{uj} = \mathbf{tr}[(Q_{i\cdot})^\top(\sqrt{X}_{\cdot j})^\top dW] \ .$$

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 (59). Now calculate each cross term in (60) 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

$$\text{Cov}[\text{vec}(dX)\text{vec}(dX)^\top] = \begin{pmatrix} \boxed{\begin{array}{c} \text{Cov}[dX_{ij}dX_{uv}] \ , \\ j = 1, i = 1 \cdots, n, \\ v = 1, u = 1 \cdots, n \end{array}} & \cdots & \boxed{\begin{array}{c} \text{Cov}[dX_{ij}dX_{uv}] \ , \\ j = 1, i = 1 \cdots, n, \\ v = n, u = 1 \cdots, n \end{array}} \\ \vdots & \ddots & \vdots \\ \boxed{\begin{array}{c} \text{Cov}[dX_{ij}dX_{uv}] \ , \\ j = n, i = 1 \cdots, n, \\ v = 1, u = 1 \cdots, n \end{array}} & \cdots & \boxed{\begin{array}{c} \text{Cov}[dX_{ij}dX_{uv}] \ , \\ j = n, i = 1 \cdots, n, \\ v = n, u = 1 \cdots, n \end{array}} \end{pmatrix} . \quad (59)$$

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

$$\begin{aligned} &\begin{pmatrix} \begin{matrix} (QQ^\top)_{11}X_{11} & \cdots & (QQ^\top)_{1n}X_{11} \\ \vdots & \ddots & \vdots \\ (QQ^\top)_{n1}X_1 1 & \cdots & (QQ^\top)_{nn}X_{11} \end{matrix} & \cdots & \begin{matrix} (QQ^\top)_{11}X_{1n} & \cdots & (QQ^\top)_{1n}X_{1n} \\ \vdots & \ddots & \vdots \\ (QQ^\top)_{n1}X_1 n & \cdots & (QQ^\top)_{nn}X_{1n} \end{matrix} \\ \vdots & \ddots & \vdots \\ \begin{matrix} (QQ^\top)_{11}X_{n1} & \cdots & (QQ^\top)_{1n}X_{n1} \\ \vdots & \ddots & \vdots \\ (QQ^\top)_{n1}X_n 1 & \cdots & (QQ^\top)_{nn}X_{n1} \end{matrix} & \cdots & \begin{matrix} (QQ^\top)_{11}X_{nn} & \cdots & (QQ^\top)_{1n}X_{nn} \\ \vdots & \ddots & \vdots \\ (QQ^\top)_{n1}X_n n & \cdots & (QQ^\top)_{nn}X_{nn} \end{matrix} \end{pmatrix} \\ = \quad & X \otimes (QQ^\top) = (X \otimes \mathbf{1}_{n\times n}).*\left(\mathbf{1}_{n\times n} \otimes (QQ^\top)\right) . \end{aligned}$$

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 (39) 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

$$dX_t^{\mathcal{L}} = \Big(kQ^{\mathcal{L}\mathcal{O}}(Q^{\mathcal{L}\mathcal{O}})^\top + M^{\mathcal{L}}X_t^{\mathcal{L}} + X_t^{\mathcal{L}}(M^{\mathcal{L}})^\top\Big)dt + Q^{\mathcal{L}}d\check{W}_t\sqrt{X_t^{\mathcal{L}}} + \sqrt{X_t^{\mathcal{L}}}(Q^{\mathcal{L}\mathcal{O}})^\top d(\check{W}_t)^\top \ , \tag{60}$$

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

We first calculate

$$dX_t^{\mathcal{L}} = \mathcal{L}dX_t\mathcal{L}^\top \tag{61}$$

$$= \Big(kQ^{\mathcal{L}\mathcal{O}}(Q^{\mathcal{L}\mathcal{O}})^\top + M^{\mathcal{L}}X_t^{\mathcal{L}} + X_t^{\mathcal{L}}(M^{\mathcal{L}})^\top\Big)dt + Q^{\mathcal{L}\mathcal{O}}dW_t^{\mathcal{O}}\sqrt{X_t}\mathcal{L}^\top + \mathcal{L}\sqrt{X_t}d(W_t^{\mathcal{O}})^\top Q^{\mathcal{L}\mathcal{O}} \ , \tag{62}$$

with $M^{\mathcal{L}}$ and $Q^{\mathcal{L}\mathcal{O}}$ specified in (33) and (34).

Comparing the last term in (64) with that in (62) we need the equality held in distribution sense

$$\mathcal{L}\sqrt{X}d(W^{\mathcal{O}})^\top \overset{\text{dist.}}{\simeq} \sqrt{X^{\mathcal{L}}}d\check{W}^\top \ . \tag{63}$$

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

$$\text{Var}[\text{vec}(\mathcal{L}\sqrt{X_t}d(W_t^{\mathcal{O}})^\top)]$$

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

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

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

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

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

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

The variance of both sides are equal so the distribution equivalence (65) is proved.

Then apply the distribution equivalent (65) into (64) then we can get (62).

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

$$dX_t = \big(\tilde{\Gamma}^{\mathcal{L}} + \tilde{M}^{\mathcal{L}}X^{\mathcal{L}} + X^{\mathcal{L}}(\tilde{M}^{\mathcal{L}})^\top\big)dt + Q^{\mathcal{L}}d\hat{W}_t\sqrt{X^{\mathcal{L}}} + \big(Q^{\mathcal{L}}d\hat{W}_t\sqrt{X^{\mathcal{L}}}\big)^\top\ , \quad (64)$$

$\tilde{\Gamma}^{\mathcal{L}}$ and $\tilde{M}^{\mathcal{L}}$ are given by (35) and (36), and $\hat{W}_t$ is an $n \times n$ standard Wiener process under the risk neutral measure satisfying the distribution equality.

$$\mathcal{L}\sqrt{X_t}d\tilde{W}_t^\top \overset{\text{dist.}}{\simeq} \sqrt{X_t^{\mathcal{L}}}d\hat{W}_t^\top \ . \tag{65}$$

Step 2

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

$$P(t,T) = \exp\Big(a^{\mathcal{L}}(\tau) + \textbf{tr}(C^{\mathcal{L}}(\tau)X)\Big) \ , \tag{66}$$

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

$$\frac{d}{d\tau}C^{\mathcal{L}}(\tau) = C^{\mathcal{L}}(\tau)\tilde{M}^{\mathcal{L}} + (\tilde{M}^{\mathcal{L}})^\top C^{\mathcal{L}}(\tau) + 2C^{\mathcal{L}}(\tau)Q^{\mathcal{L}\mathcal{O}}(Q^{\mathcal{L}\mathcal{O}})^\top C^{\mathcal{L}}(\tau) - \Psi^{\mathcal{L}}, \tag{67}$$

$$\frac{d}{d\tau}a^{\mathcal{L}}(\tau) = \textbf{tr}[Q^{\mathcal{L}\mathcal{O}}(Q^{\mathcal{L}\mathcal{O}})^\top C^{\mathcal{L}}(\tau)] \ , \tag{68}$$

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 (69) and $a^{\mathcal{L}}(\tau) = a(\tau)$ satisfies (70) respectively. Since the solution is unique for each equation we prove (39).

□

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 (64) and obtain

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

Comparing the equation above with (7) we obtain the relations (35) and (36).

□

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

$$dX_t^d = \big(kQ_dQ_d^\top + \mathcal{D}X_t^d + X_t^d\mathcal{D}\big)dt + Q_d dW_t \sqrt{X_t}(\mathcal{L}_d^{-1})^\top + \mathcal{L}_d^{-1}\sqrt{X_t}dW_t^\top Q_d^\top \ , \tag{69}$$

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

Use the distribution equivalence (67) 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*[12] 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

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

where $\breve{W}_t := \mathcal{O}^\top \breve{W}_t$ is a new $n \times n$ standard Wiener process. Summarize the transformation above the dynamics (71) becomes

$$dX_t^d = \big(kR^\top R + \mathcal{D}X_t^d + X_t^d\mathcal{D}\big)dt + \mathcal{R}^\top d\breve{W}_t\sqrt{X_t^d} + \sqrt{X_t^d}d\breve{W}_t^\top\mathcal{R} \ , \tag{70}$$

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

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 $\mathcal{C} = \operatorname{diag}\big(\sqrt{\Psi_{d,11}}, \cdots, \sqrt{\Psi_{d,nn}}\big)$ 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 $\mathcal{S}$.

Through direct calculation we know the new M, Q and Ψ obtained from (33), (34) and (38) 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*[13] 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[14] Consider the transformed

[13]Schur Decomposition is stated for a square complex matrix and the transpose operator $\top$ is actually conjugate transpose.

[14]For a unitary matrix $\mathcal{L}_b$, $\mathcal{L}_b^\top = \mathcal{L}_b^{-1}$.

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 (67) we can obtain

$$dX_t^b = \left(kQ_bQ_b^\top + \mathcal{B}X_t^b + X_t^b\mathcal{B}^\top\right)dt + Q_b d\breve{W}_t \sqrt{X_t^b} + \sqrt{X_t^b} d\breve{W}_t^\top Q_b^\top \ , \qquad (71)$$

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

$$dX_t^r = \left(k + \mathcal{B}_r X_t^r + X_t^r \mathcal{B}_r^\top\right)dt + \mathcal{O}^\top d\breve{W}_t \sqrt{X_t^b} \mathcal{R}^{-1} + (\mathcal{R}^\top)^{-1} \sqrt{X_t^b} d\breve{W}_t^\top \mathcal{O} \ , \qquad (72)$$

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 (67). 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 (34) we obtain

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

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

□

APPENDIX B. NOTE FOR $\psi = 0$ IN (42)

In estimation WTSM we set ψ in (42) equal to zero. This intends to prevent a "counteracting" of factors as explained in the following.

In the WTSM we introduce the off-diagonal factor $X_{12,t}$ in order to model stochastic covariance between $X_{11,t}$ and $X_{22,t}$. Filtering techniques determine the whole factor trajectories of the unobserved factors X_t for fitting the observed bond yields in (48). In the estimation the factors are allowed to move quite freely. In filtering multiple factors in the maximum likelihood estimation, sometimes we can observe that a pair of factors has exaggerated and similar trajectories. The factors counteract each other for the most of time and this pattern can achieve high likelihood.

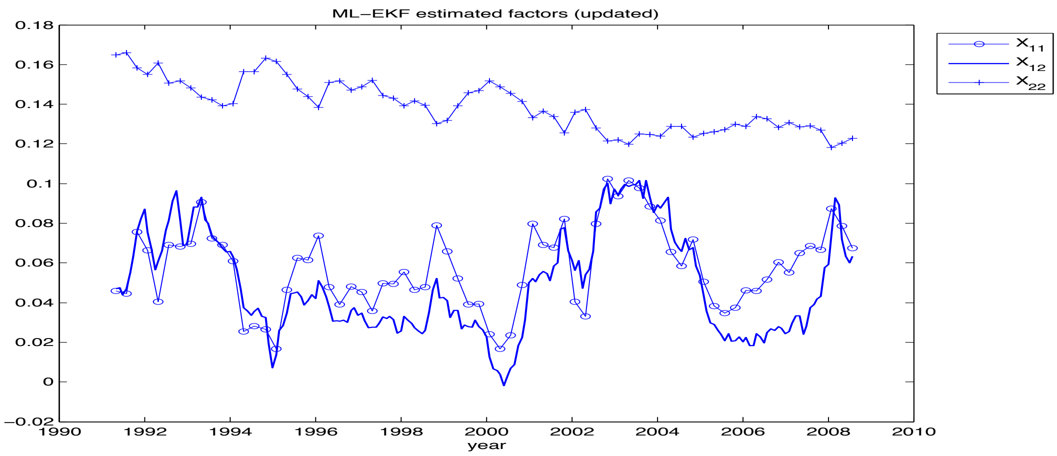


FIGURE 12. Estimated factors for $\psi \neq 0$

Figure 12 shows the filtered factors when ψ is freely estimated in the WTSM (42). The trajectory of $X_{12,t}$ runs close to $X_{11,t}$ most of them. The estimated ψ is -0.5127 so that the short rate (3) is modelled by

$$r_t = X_{11,t} - 1.0254X_{12,t} + X_{22,t} \ .$$

The five year bond yield is estimated by

$$y_t = j(5) + 0.1604X_{11,t} - 0.2073X_{12,t} + 1.0289X_{22,t} \ .$$

We see that the off-diagonal factor $X_{12,t}$ counteracts $X_{11,t}$ for their coefficients being close and of opposite signs.

This "counteracting" effect deteriorates the estimates and gives rise to some undesired results. Table 4 gives the estimates for setting $\psi = 0$ comparing with estimating ψ freely. We see that for the freely estimation case (the second column) $m_{22} > 0$. It indicates unstable real-world dynamics of X_t and it leads to large biases for longer term forecast. Furthermore, the $\tilde{\gamma}_{22}$ is negative which gives a negative definite $\tilde{\Gamma}$ in the risk-neutral dynamics (7). In this case $X_{22,t}$ will hit zero under risk-neutral measure while it wont (means with probability zero) under the real world measure (because $k > n+1$). So the measure transformation is not equivalent and the whole no-arbitrage argument breaks down. We also note that the estimation for $\psi \equiv 0$ achieves higher likelihood value. The estimation with free ψ encounters difficulty for converging.

If the our main intention of introducing the off-diagonal factor $X_{12,t}$ is to model the stochastic covariance of $X_{11,t}$ and $X_{22,t}$ rather than to given extra freedom for fitting the yields, we need to exclude this freedom so we set $\psi = 0$ in estimation.

In general setting $\psi = 0$ imposes restriction on model fitting capacity. In this paper we focus first the role of $X_{12,t}$ in modelling the stochastic covariance and compare with the MCIR model so we keep this restriction. It is our future work to maintain model fitting capacity properly but avoid the counteracting effect.

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.

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

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.

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. (2009), Affine Processes on Positive Semidefinite Matrices. Working Paper.

Da Fonseca, J., Grasselli, M. & Ielpo, F. (2008), 'Estimating the Wishart affine stochastic correlation model using the empirical characteristic function', *ESILV Working paper* (RR-35).

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. & Stanton, R. H. (2004), Estimation of Dynamic Term Structure Models, Working paper, Haas School of Business, U.C. Berkeley.

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), 177–271.

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. (2003), 'Wishart Term Structure Models', *CREF of HEC Montreal Working Paper* (No. 03-10).

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

Lord, R., Koekkoek, R. & van Dijk, D. (2008), 'A comparison of biased simulation schemes for stochastic volatility', *SSRN-working paper 903116* **http://papers.ssrn.com/**.

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

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 estimates. The “Equally Weighted” column is based on a naive portfolio consisting of 50% of each bonds. 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. (56). The downturn risk is calculated by $\text{DR}^2 = \frac{1}{N^-}\sum_{t,\pi_t<0}\pi_t^2$, 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

Parameter	WTSM ($\psi \neq 0$)	WTSM ($\psi = 0$)
k	15.89400	7.28110
m_{11}	-0.46350	-0.36789
m_{22}	0.00974	-0.00723
q_{11}	0.06002	0.06144
q_{21}	-0.00572	-0.00075
q_{22}	0.00847	0.00663
$\tilde{m}_{11}$	-0.57849	-0.59168
$\tilde{m}_{12}$	-0.03012	0.44139
$\tilde{m}_{21}$	0.05820	0.00731
$\tilde{m}_{22}$	0.00112	-0.00802
α	-0.10761	-0.11248
ψ	-0.51267	$\equiv 0$
$\tilde{\gamma}_{11}$	0.07021	0.03491
$\tilde{\gamma}_{22}$	-0.00448	0.00164
$\tilde{\gamma}_{21}$	0.02232	-0.03433
σ_ϵ (bp)	7.65	7.71
Loglik	12038	12079

TABLE 4. Estimates Comparison for free ψ and setting $\psi = 0$ for WTSM.